BMC Genomics

RESEARCH

Open Access

Genomic analyses of *Purpureocillium lilacinum* reveal unique DNA regions for developing isolate-specific molecular markers

Zhi-Yu Yeh[1†], Kusum Mushyakhwo[2†], Nian-Tong Ni[1], Pei-Hsin Lo[3], Chuen-Fu Lin[4*] and Yu-Shin Nai[1*]

Abstract

Background The *Purpureocillium lilacinum* NCHU-NPUST-175 (PI-NCHU-NPUST-175) strain has been previously reported as an entomopathogenic fungus (EPF) with potential for controlling *Forcipomyia taiwana* (little black mosquito). In this study, genome sequencing of PI-NCHU-NPUST-175 was performed via long-read sequencing (Oxford Nanopore Technologies, ONT). The comparative genomic analysis was conducted to identify isolate- and species-specific DNA regions for developing markers for rapid molecular identification, in order to detect the fungus in an environment and to confirm species identity before practical field applications.

Results The genome assembly of PI-NCHU-NPUST-175 is 36.55 Mb, consisting of 13 contigs with a GC content of 58.64%. The largest contig of the PI-NCHU-NPUST-175 genome was 6.79 Mb, and the N50 value was 4.08 Mb. The genome contains 14,069 putative protein-coding genes, with a gene density of 384.90 genes per Mb and a median number of three exons per gene. Additionally, one mitogenome, PI-NCHU-NPUST-175, was assembled with a size of 0.23 Mb and contained 16 protein-coding genes. In the PI-NCHU-NPUST-175 genome, repeat regions accounted for 3.70% of the genome. Most repetitive sequences exhibit a nucleotide sequence divergence of less than 30%. Phylogenetic analysis revealed that PI-NCHU-NPUST-175 is most closely related to other *P. lilacinum* strains. To identify distinctive genomic regions (DGRs), a comparative genomics analysis was conducted between PI-NCHU-NPUST-175, *P. takamizusanense* and *P. lilacinum*. A total of 1,154 unaligned fragments were identified by genome-wide alignment, ten DGRs were randomly selected for specific primer design, and the specificity of the primer sets was tested. The results showed that all primer sets could serve as specific molecular markers for PI-NCHU-NPUST-175. Among the ten primer sets, three primer sets were selected for the soil-sample detection test. The qualitative PCR results indicate that all three isolate-specific primer sets (pl_1, pl_8, and pl_10) detected 10^5 conidia g^{-1} soil in the natural soil samples for up to 14 days. However, qPCR improve the accuracy and sensitivity of detecting Low fungal concentrations 10^3

†Zhi-Yu Yeh and Kusum Mushyakhwo contributed equally to this work.

*Correspondence:
Chuen-Fu Lin
cflin2283@mail.npust.edu.tw
Yu-Shin Nai
ysnai@nchu.edu.tw

Full list of author information is available at the end of the article

conidia g^{-1} soil in complex environmental matrices like soil. These findings suggest that these primers set have potential for future applications in monitoring specific fungal isolates in the field over period of time.

Conclusions The genome sequence of Pl-NCHU-NPUST-175 was decoded and compared with those of closely related isolates and species. Based on a developed comparative genomics approach, we highlight the feasibility of developing specific molecular markers of EPF, and this method could be further applied to commercialize EPF products to address sustainability in the natural environment.

Keywords Comparative genomics, Molecular markers, *Purpureocillium lilacinum*, Oxford Nanopore Technologies (ONT), Distinctive genomic regions (DGRs)

Introduction

Entomopathogenic fungi (EPF) are natural pathogens of insects that can persist in the soil and lead to a saprophytic lifestyle. The isolation of EPF from soil is a common method worldwide for the development of microbial control agents [1–3]. Currently, numerous commercial EPF biopesticides are readily available on the market [4]. Most of these biopesticides are made of fungi in the genera *Beauveria*, *Metarhizium*, and *Isaria* [5], many of which function as biological control agents against a wide range of agricultural and economically important pests [6]. The genus *Purpureocillium,* belonging to the order Hypocreales order, is also reported as EPF with a broad host range and significant potential for the biological control of insect pests [7]. *Purpureocillium takamizusanense* has demonstrated pathogenicity toward adult *Tessaratoma papillosa* (litchi stink bug), with effective infection observed by 14 days post-inoculation [8]. Similarly, under laboratory conditions, *Purpureocillium lilacinum* exhibited pathogenic activity that negatively affected the reproduction of *Aphis gossypii* (cotton aphid). Furthermore, *P. lilacinum* (Pl-NCHU-NPUST-175) significantly reduced the egg-hatching and adult emergence rates of *Forcipomyia taiwana* (little black mosquito) in laboratory bioassays, indicating its potential as a biocontrol agent [9]. *Purpureocillium lavendulum* and *P. lilacinum* have also been reported as endophytes in maize, bean, and soybean plants [10]. Among *Purpureocillium* spp., *Purpureocillium lilacinum* (formerly known as *Paecilomyces lilacinus*) has been commercialized as a biological control agent for plant-parasitic nematodes (*Meloidogyne* spp.) [11]. These all show its importance as a biological control agent.

The isolate, Pl-NCHU-NPUST-175, known for its potential in controlling *F. taiwana*, is anticipated for application in field trials. During field applications, monitoring long-term presence of this specific biocontrol fungal is essential [12]. Accurate identification at both the species and strain levels can provide evidence regarding their safety for nontarget organisms in the environment and their insecticidal efficacy against target pests. Additionally, it is of paramount importance to assess the persistence of applied fungi in the environment. Prolonged conidial survival enhances the likelihood of host infection [13], thereby contributing to sustainable pest management [14].

To monitor the prevalence of EPF in nature, several methods can be applied, including isolating fungi from soil or diseased insects, baiting them with insects, and identifying soil DNA with PCR [15]. However, distinguishing different isolates within the same species of EPF by morphological identification is challenging, as many fungi are microscopic and have complex life cycles, making molecular identification essential. Previous studies have shown that different culture media affect the morphological characteristics of *P. lilacinum* [16]. However, to our knowledge, there is no report has been published on the development of specific biomarkers for *P. lilacinum*. For fungal molecular identification, internal transcribed spacer (ITS) is a commonly used and standard DNA barcodes for fungi [17]. However, ITS regions may not always distinguish closely related fungal isolates, necessitating the use of additional molecular markers [18]. Molecular markers such as Restriction Fragment Length Polymorphism (RFLP), Random Amplified Polymorphic DNA (RAPD), Amplified Fragment Length Polymorphism (AFLP), Simple Sequence Repeats (SSR), and Inter Simple Sequence Repeats (ISSR) have been used as effective tools to detect genetic variability [19]. Nevertheless, existing markers often lack sufficient resolution and may be influenced by homoplasy, homologous recombination, and lateral gene transfer [20, 21]. Recent advancements in sequencing technologies have significantly increased genome availability and enhanced investigation of genome diversity, which enables a more comprehensive approach to selecting marker genes [22]. Whole-genome sequencing (WGS) and comparative genome analysis offer a more precise and reliable approach for developing isolate-specific molecular markers [23]. By analyzing the complete genomic information of different isolates, distinct genetic regions can be identified and targeted for marker development, ensuring specificity and reproducibility.

In this study, Pl-NCHU-NPUST-175 was subjected to genome sequencing by using long-read sequencing (Oxford Nanopore Technologies, ONT), which enables

long-read sequencing to improve contiguity of genome assembly. Through the annotation of the Pl-NCHU-NPUST-175 genome, the gene locations were predicted, and the virulence-associated genes, secondary metabolites, and carbohydrate-active enzymes were further annotated to understand the biochemical properties of the organisms. Additionally, phylogenetic analysis of the core genes of the nuclear genome and mitogenome was performed. A comparative genomics approach was subsequently used to compare the genome sequences of Pl-NCHU-NPUST-175 with those of two fungal species belonging to the *Purpureocillium* genus (*P. takamizusanense* and *P. lilacinum*). This comparison aimed to identify highly specific regions in Pl-NCHU-NPUST-175 which can serve as molecular markers for further application in EPF isolate- and species-level identification.

Methods

Entomopathogenic fungi

The entomopathogenic fungi Pl-NCHU-152 and Pl-NCHU-NPUST-175 were provided by the Insect Pathology and Pathogen Genomics Lab (IPG) of National Chung Hsing University (NCHU), Taiwan. Pl-NCHU-152 was isolated from *T. papillosa* (litchi stink bug), and Pl-NCHU-NPUST-175 was isolated from diseased pupae of *F. taiwana* [9]. These two isolates were recovered at −80 °C and cultured on 55-mm quarter strength Sabouraud dextrose agar (¼ SDA) media at 25 °C in an incubator with a 24-h dark cycle for 14 days for genomic DNA (gDNA) extraction.

Fungal genomic DNA extraction

The gDNA of Pl-NCHU-152 and Pl-NCHU-NPUST-175 were extracted via the Presto™ Mini gDNA Yeast Kit (Geneaid, New Taipei, Taiwan). In addition, the gDNA of two isolates of *P. takamizusanense* (TCTeb01 and TCP05), which were provided by Taichung District Agricultural Research and Extension Station, Taiwan, was extracted by a Genomic DNA Isolation Kit (FairBiotech, Taiwan). The concentration of the DNA samples was measured via an EzDrop 1000 microvolume spectrophotometer (Blue-ray Biotech, New Taipei, Taiwan) and a Qubit instrument (Thermo Fisher Scientific, MA, USA).

Long-read sequencing by Oxford Nanopore Technologies (ONT)

A total of 2 μg of Pl-NCHU-NPUST-175 DNA was used to prepare the ONT sequencing library. The ONT sequencing library was constructed from genomic DNA using SQK-LSK109 (Oxford Nanopore Technologies, Cambridge, UK) following standard protocols. Genomic sequences were generated using an R9.4.1 SpotON flow cell and processed on a MinION device with ONT MinKNOW software version 22.05.5.

Genome assembly, annotation, and repeat region prediction

The genome assembly pipeline is summarized in Supplementary Fig. 1. The FAST5 files were basecalled by Guppy (v6.1.7) and obtained from the ONT community site (https://community.nanoporetech.com) for further analysis, and sequencing QC was performed by NanoPlot [24]. Porechop (v0.2.4, https://github.com/rrwick/Porechop) was subsequently used for barcode and adapter trimming from the basecalled reads [25, 26]. The QC and trimmed reads were subjected to genome assembly by using Canu, and then BWA and SAMtools were used to index the original FASTA files. The BWA files were subsequently input into Nanopolish to polish the assembled draft genome and improve its quality [27]. The contigs containing the mitochondrial and bacterial genomes were identified with BLASTn. To evaluate the quality of the genome assembly, the assembled genome was checked by using QUAST (v5.1.0) [28]. In addition, the completeness of the genome assembly was assessed with BUSCO on the basis of the hypocreales_odb10 database [29]. To identify the repetitive regions and transposable element (TE) regions, RepeatMasker and RepeatModeler2, respectively, were used [30, 31]. Genomic regions affected by the repeat-induced point (RIP) method were identified with RIPper [32]. For Pl-NCHU-NPUST-175 and the two closely related *Purpureocillium* spp. from the NCBI annotation, *P. takamizusanense* (GCA_022605165.1) and *P. lilacinum* (GCA_001653265.1). FunGap was used on the basis of the proteome database (GCA_023168115.2 for *P. lilacinum*, GCA_029168975.1 for *P. takamizusanense*) and RNA-seq data (SRR8940714 for *P. lilacinum*, SRR16888993 for *P. takamizusanense*) of related strains in the NCBI database [33]. The completeness of genome annotation was assessed with BUSCO based on the hypocreales_odb10 database [29]. The predicted protein from FunGap conjunct with AUGUSTUS v3.4.0 was further subjected to functional annotated via InterProScan to identify the families and the domains and sites it contains. Pathogenicity, virulence and effector genes from fungal, bacterial and protist pathogens that infect animal, plant, fish, insect and fungal hosts were identified with BLASTp (e-value < 1E-10) against the pathogen–host interaction database (PHI-base) 4.17 [34]. The secondary metabolites were annotated by antiSMASH 7.1.0 [35]. The carbohydrate-active enzymes (CAZymes) were identified and classified using dbCAN3, which integrates multiple tools including HMMER, DIAMOND, and dbCAN-sub [36]. To improve annotation accuracy and reduce false positives, we adopted a conservative strategy: only genes supported by at least HMMER combined with either dbCAN-sub or DIAMOND were considered as confident CAZymes. In cases of conflicting annotations among the tools, we prioritized HMMER-based

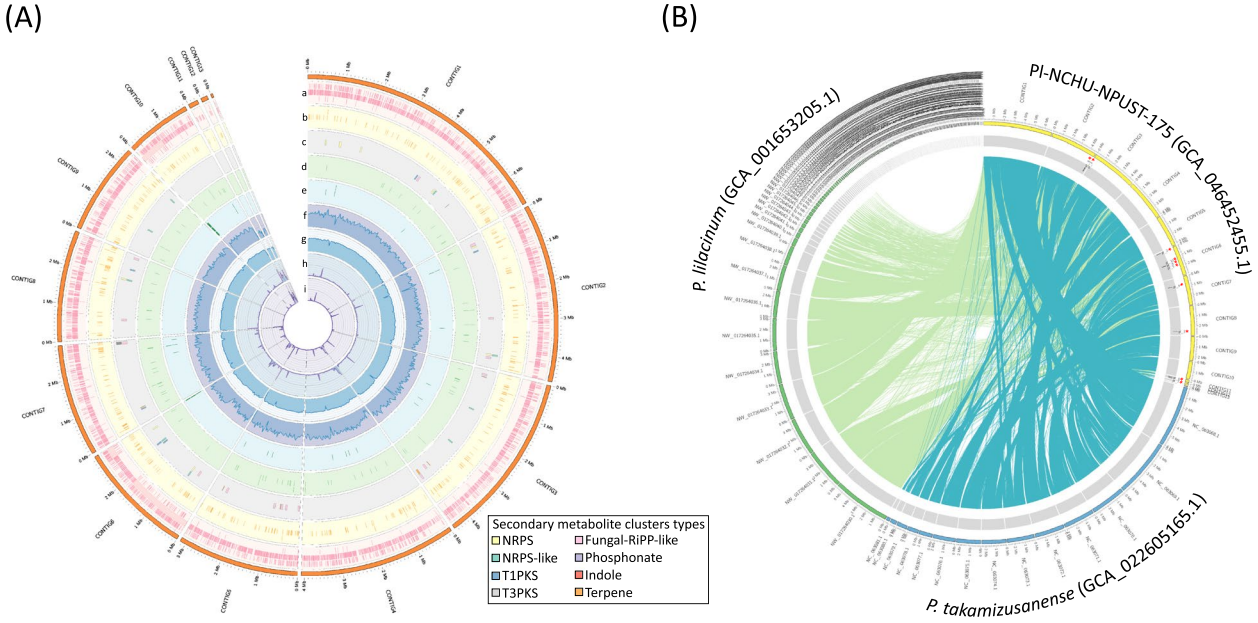

Fig. 1 Genomic structure and comparison of PI-NCHU-NPUST-175 with other *P. lilacinum* isolates and *P. takamizusanense*.**A** The genomic features of PI-NCHU-NPUST-175. The genomic map was visualized using Circos plots, employing a 50-kbp window, to illustrate the regions of (a) PHI, (b) CAZyme, (c) secondary metabolite clusters, (d) tRNAs, and (e) rRNAs, and percentages of (f) gene density, (g) GC%, (h) repeat density, and (i) TE density within each contig. **B** Syntenic CIRCOS plots of PI-NCHU-NPUST-175, *P. takamizusanense* (GCA_022605165.1), and *P. lilacinum* (GCA_001653265.1). The line on the plots represents the alignment region. The locations of the isolate-specific primer set were marker and noted with red-star markers

predictions because HMMER identifies CAZyme-related protein domains directly, providing a more reliable domain-level functional inference. The locations of ribosomal RNA genes and transfer RNA genes (tRNAs) in genomes were predicted by Barrnap (https://github.com/nigyta/dfast_core/blob/master/bin/Darwin/barrnap-0.8) and tRNAscan-SE (v.2.0) [37].

Mitochondrial genome assembly and annotation

The genome assembly and annotation pipeline is summarized in Supplementary Fig. S2. The mitochondrial contig assembly was circularized by Circlator [38]. Strand asymmetries were calculated via the following formulas: AT skew = [A—T]/[A + T], and GC skew = [G—C]/[G + C]. The complete mitochondrial genome was initially annotated by MITOS2 and MFannot (https://megasun.bch.umontreal.ca/apps/mfannot/), both of which are based on genetic code 4 (the Mold, Protozoan, and Coelenterate Mitochondrial Code). The tRNAs were also annotated via the RNAweasel (https://megasun.bch.umontreal.ca/apps/rnaweasel/) and tRNAScan-SE (v.2.0) web servers under genetic code 4. These initial annotations of protein-coding genes (PCGs), rRNA genes and tRNA genes were then modified by the NCBI Open Reading Frame (ORF) Finder (https://www.ncbi.nlm.nih.gov/orffinder). Finally, the annotated gene boundaries were curated manually. All annotated genes of the mitochondrial genome were illustrated using Circos [39].

Phylogenetic analysis

For phylogenetic analysis, the core genes of the fungal species of interest were searched with the UFCG v1.0 profile module, employing AUGUSTUS v3.4.0 [40] pretrained block profile HMMs and MMseqs2 v13.45111 to validate the sequences. The genome sequences of 34 fungal species were sourced from the NCBI database (http://www.ncbi.nlm.nih.gov/) (Supplementary Table S1). Among these genomes, 32 belonged to the fungal order Hypocreales, including *Metarhizium anisopliae*, *Metarhizium rileyi*, *Beauveria bassiana*, *Cordyceps cicadae*. Additionally, *Aspergillus nidulans* belonging to Eurotiales served as the outgroup in the analysis. A total of 59 conserved genes (Supplementary Table S2) were found among 34 fungal isolates, and were subjected to phylogenetic analysis by using UFCG [41]. The 59 conserved genes were locally aligned via the Smith–Waterman algorithm with water. The UFCG utilized multiple alignment and concatenation of core genes for phylogenetic tree analysis. The tree module collected JSON files generated from the profile module, formed MSAs for shared marker genes using MAFFT v7.310, eliminated alignment columns with ≥ 50% gaps, and constructed the ML tree from the concatenation using the JTT model. IQ-TREE (v 2.2.5) [42] was employed for tree building. The resulting tree was tested by bootstrapping with 100 replicates. Additionally, the module computed a concatenation tree annotated with the gene support index (GSI)

for branch support, displaying the number of individual gene trees backing each branch as support values.

To construct a mitochondrial genome phylogenetic tree, complete mitochondrial genomes of 29 Hypocreales species were downloaded from the NCBI. In addition, *Aspergillus nidulans* was used as an outgroup. The 14 core genes (*atp6*, *atp8*, *atp9*, *cob*, *cox1*, *cox2*, *cox3*, *nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5*, and *nad6*) were used for phylogenetic analysis. The nucleotide sequences of the 14 core genes were individually aligned using MAFFT (v.7.453), and multiple sequence alignments were concatenated via SequenceMatrix (v.1.7.8) [43]. The concatenated protein alignment was used to infer a phylogenetic tree with IQ-TREE (v2.4.0) [42]. JTT substitution model was specified for protein evolution. The analysis was performed with 1,000 bootstrap replicates, and the bootstrap values are displayed on the nodes of the tree.

Identification of fungal distinctive genomic regions (DGRs)

The pipeline of comparative genomics is summarized in Supplementary Fig. S3. The genome sequence of Pl-NCHU-NPUST-175 was compared with two closely related reference genomes of *Purpureocillium* spp. from the NCBI database, namely, *Purpureocillium takamizusanense* (GCA_022605165.1) and *P. lilacinum* (GCA_001653265.1), which contain comprehensive information for further analysis. For comparison, MUMmer was used for sequence alignment analysis [44]. After alignment analysis, the regions that did not match the reference genome were selected, and the resulting data are presented as circle plots created with Circos [39].

The pipeline of fungal DGR identification is summarized in Supplementary Fig. S3. The MUMmer alignment was performed as above, and the'subtract'function in bedtools software was used to remove regions with high similarity from the reference genome and output the unmatched fragments. Fragments that did not match the reference genome with lengths greater than 5000 bp were selected, followed by BLAST analysis. Fragments with a query coverage of less than 20% were chosen. The corresponding regions were then removed, and another BLAST analysis was performed using the NCBI database. The unaligned fragments were subsequently selected using the graphic summary. The results were subsequently confirmed to show no matches to any data in the NCBI database. The filtered fragments were subjected to primer design using Primer3 (v 4.1.0) [45] and NetPrimer (https://www.premierbiosoft.com/netprimer/). A total of 10 sets of specific primers (pl_1-pl_10) (Supplementary Table S3) that generate 300–400 bp products were designed.

Validation of fungal genomic DGRs of Purpureocillium species

PCR amplification was employed to confirm the specificity of the 10 primer sets inter- and intraspecies, including Pl-NCHU-152 and two strains of *P. takamizusanense* (TCTeb01 and TCP05) (Supplementary Table S3). The PCR mixture comprised 10 µL of 2X SuperRed Master Mix (Tools, New Taipei, Taiwan), 1 µL of each primer (10 µM), 1 µL of fungal genomic DNA, and 7 µL of ddH_2O, resulting in a total reaction volume of 20 µL. PCR was conducted using the following program: initial denaturation at 95 °C for 10 min; 35 cycles of denaturation at 95 °C for 15 s, annealing at 62 °C for 30 s, and extension at 72 °C for 90 s; and a final extension at 72 °C for 10 min. Subsequently, the PCR products were analyzed via electrophoresis using a 1% agarose gel in TAE buffer, followed by Sanger sequencing of the amplicons generated by 10 primer sets to confirm amplification of the intended target regions. The experiment was repeated 8 times, and the primer set that was most sensitive for detecting Pl-NCHU-NPUST-175 was selected for soil-based detection.

Soil-based fungal DNA detection assay

The primer sets that were highly sensitive for detecting intra- and interspecies DNA were selected for the soil-based fungal DNA detection assay. The sensitivity of the primers was evaluated by conventional PCR. Fungal inoculation was done in soil samples from agriculture origin (GPS: 24°07′07.9"N 120°40′39.9"E) and the assay method was modified from Zhang et al., 2019 [46]. Spore suspension of Pl-NCHU-NPUST-175 of known concentration (10^6 conidia mL^{-1}) was serially diluted 1:10 (from 10^6 up to 10^3) and inoculated 3 ml in glass container (35 mm in height, 45 mm in inner diameter) prefilled with 3 g of agriculture soil. The control was inoculated with equal volumes of water. All the samples were incubated under darkness at 25 °C. After 1-, 7- and 14-days incubation, total genomic DNA was extracted using the Presto™ Soil DNA Extraction Kit (Geneaid, New Taipei, Taiwan), followed by PCR amplification using specific primers described above.

The detection of the soil-based fungal DNA by absolute quantitative PCR (qPCR) was performed with qPCR primer sets (Supplementary Table S3), which were designed using Primer Express v3.0 (Applied Biosystems). The qPCR amplification was performed in a 20 µl reaction volume using the SensiFASTTM SYBR® Hi-ROX Kit (Meridian Bioscience, USA) in a 96-well Applied Biosystems™ Real-time PCR System (Thermo-Fisher Scientific, USA). The qPCR mixture comprised 10 µL of 2X SuperRed Master Mix (Tools, New Taipei, Taiwan), 1 µL of each primer (10 µM), 1 µL of DNA template, and 7 µL of ddH_2O, resulting in a total reaction volume of 20 µL,

following program: 95 °C for 2 min, followed by 40 cycles of 95 °C for 5 s and 60 °C for 30 s. For absolute qPCR, a standard curve of fungal spore DNA (approximately 10^4–10^8 conidia) was established by using tenfold serial dilution. Total fungal genome copies in soil samples were calculated from the standard curves by plotting the CT value versus the $\log_{10}$ of each fungal spore number. The standard curve equations were: $y = -3.8674x + 50.107$, $R^2 = 0.9936$ for pl_1, $y = -4.0358x + 50.474$, $R^2 = 0.9976$ for pl_8 and $y = -3.778x + 48.481$, $R^2 = 0.9957$ for pl_10. Each sample was assessed in three replicates. The results were expressed as lg (copies g^{-1}) dry soil. Statistical analysis was performed in R 4.4.1 [47] using one-way ANOVA, and mean comparisons were conducted with Tukey's test. Differences were considered statistically significant at $p < 0.05$. The PCR inhibition in soil-extracted DNA was performed by adding the Black Queen Cell Virus (BQCV) plasmid (9.97×10^7 copies) in the 10 ng/mL of the soil-based fungal DNA to reach 50%, 10%, 1%, and 0.1% of the whole DNA sample, respectively. Plasmid DNA of the same concentration was also prepared for generating the standard curve. A 1 µL aliquot from each mixture was then subjected to qPCR amplification using the specific primer set (BQCV/Fq and BQCV/Rq) (Supplementary Table S3). The qPCR mixture and program were as described above. The ΔCt shifts for spike-in DNA in soil DNA were calculated by subtracting Ct value of plasmid and Ct value of spike-in DNA in soil DNA.

Results

Sequencing summary and genome assembly

The long-read sequencing (ONT) data of Pl-NCHU-NPUST-175 are listed in Supplementary Table S4. Long-read sequencing (ONT) generated 189,591 reads with an N50 of 15,694 bp and a total of 1.0 Gb for the Pl-NCHU-NPUST-175 genome. The information concerning the Pl-NCHU-NPUST-175 genome assembly and annotation is summarized in Table 1. The genome of Pl-NCHU-NPUST-175 was assembled into 13 contigs with 32 × coverage. The nuclear genome size is 36.55 Mb. Among these contigs, the largest contig was 6.79 Mb, and the N50 contig size was 4.08 Mb. Compared with the sequenced *P. lilacinum* PLFJ-1 (GCA_001653265.1) and *Purpureocillium takamizusanense* PT3 (GCA_022605165.1), the genome size is smaller than that of *P. lilacinum* PLFJ-1 (38.53 Mb) but larger than that of *P. takamizusanense* PT3 (35.57 Mb) (Table 1). The GC content of Pl-NCHU-NPUST-175, which was approximately 58.64% (Table 1), was similar to that of *P. lilacinum* PLFJ-1 and *P. takamizusanense* PT3. The percentage of complete BUSCO marker genes was 94.70%. The percentage of fragmented BUSCO marker genes was 0.60%. The percentage of missing BUSCO marker genes was 4.70%. Notably, the sequencing technology of Pl-NCHU-NPUST-175 and *P. takamizusanense* PT3 involved Long-read sequencing, which revealed only 13 and 14 contigs, respectively, whereas the *P. lilacinum* PLFJ-1 genome consisted of more than 800 contigs, indicating the advantage of long-read data for genome assembly.

Genome features

The genomic features of Pl-NCHU-NPUST-175 are shown in Fig. 1A. A total of 14,069 protein-coding genes were annotated in the genome of Pl-NCHU-NPUST-175, resulting in a gene density of 384.90 genes per Mb. The analysis of the completeness of gene annotation using BUSCO marker genes revealed 94.70% completeness, whereas the numbers of protein-coding genes in the genomes of *P. lilacinum* PLFJ-1 and *P. takamizusanense* PT3 were 11,763 (99.50% completeness BUSCO value) and 11,855 (98.90% completeness BUSCO value), which are lower than those of Pl-NCHU-NPUST-175 (Table 1). Therefore, a lower gene density of *P. lilacinum* PLFJ-1 (342.68 genes per Mb) and *P. takamizusanense* PT3 (340.84 genes per Mb) was also detected (Table 1). Furthermore, the number of exons was also greater than that of the two related *Purpureocillium* spp. genes, comprising 42,768 exons, with the median of 3 exons per gene in the Pl-NCHU-NPUST-175 genome (Table 1). Among these annotated genes, 8,371 genes encode proteins with a conserved Pfam domain (total 12,948 regions) in their primary sequence based on functional annotations with InterProScan (Supplementary Table S5).

According to the results of PHI database mapping, the genes whose relevance to fungal pathogens were annotated. A total of 4,315 genes (30.60%) were predicted to affect the outcome of pathogen−host interactions. Notably, the number of assumed pathogen-host interaction genes associated with Pl-NCHU-NPUST-175 was greater than that associated with *P. lilacinum* PLFJ-1 (2,892 genes, 24.60%) (Table 1).

Among genome, 42 regions are predicted to encode key enzymes for the biosynthesis of secondary metabolites, including 5 polyketide synthases (PKSs), 6 nonribosomal peptide synthetases (NRPSs), 5 NRPS-like fragments, 14 fungal-RiPP-like, 4 terpene synthases, 1 phosphonate synthase, 1 indole synthase, 3 NRPS + NRPS-like, 1 fungal-RiPP-like + terpene, 1 NRPS-like + T1PKS (Type I PKS), and 1 T3PKS (Type III PKS) + fungal-RiPP-like (Fig. 1A; Supplementary Table S6).

In addition to the key enzymes for the biosynthesis of secondary metabolites, the dbCAN3 annotation results revealed a total of 423 genes encoding CAZymes, including 231 glycoside hydrolases (GHs), 92 glycosyltransferases (GTs), 16 carbohydrate esterases (CEs), 64 auxiliary activity enzymes (AAs), 4 polysaccharide lyases (PLs), 4 CAZymes with single carbohydrate binding domains

Table 1 Summary of the genome assembly and comparison of Purpureocillium spp

Genome feature	Purpureocillium lilacinum-NCHU-NPUST-175	*Purpureocillium lilacinum* PLFJ-1 (GCA_001653265.1)*	*Purpureocillium takamizusanense* PT3 (GCA_022605165.1)*	Note
Contigs	13	818	14	Genome assembly
Total length (bp)	36,552,814	38,534,601	35,574,015	
Largest contig (Mb)	6,785,927	5.70	6.10	
Coverage	32	152	282	
GC (%)	58.64	58.4	57.76	
N50 (bp)	4,081,770	150,141	3,094,299	
L50	4	76	4	
Complete BUSCO marker genes	94.70%	99.50%	98.90%	
Fragmented BUSCO marker genes	0.60%	0.00%	0.10%	
Missing BUSCO marker genes	4.70%	0.50%	1.00%	
Sequencing technology	Oxford Nanopore MinION	Illumina HiSeq	Illumina HiSeq; Oxford Nanopore PromethION	
Total protein-coding genes	14,069	11,763	11,855	Genome annotation
Transcript length (avg/med)(bp)	1,429.1/1,192.0	1,529.8/1,248.0	1,543.3/1,309.0	
CDS length (avg/med)(bp)	1,235.4/1,020.0	1,359.7/1,095.0	1,372.5/1,146.0	
Protein length (avg/med)(aa)	411.8/340.0	453.2/365.0	457.5/382.0	
Exon length (avg/med)(bp)	406.4/265.0	523.7/307.0	504.6/273.0	
Intron length (avg/med)(bp)	95.0/62.0	106.6/63.0	99.3/67.0	
Spliced genes	11,192 (79.55%)	9,475 (71.75%)	8,903 (73.43%)	
Number of introns	28,699	21,082	20,854	
Number of introns per gene (med)	2	2	2	
Number of exons	42,768	34,287	32,979	
Number of exons per gene (med)	3	2	2	
Gene density (genes/Mb)	384.90	342.68	340.84	
Repeat sequence density (%)	3.7	6.0	4.6	
Secreted protein	NA	1448	1,089	Gene group of interest
Glycoside hydrolases	459	253	NA	
Carbohydrate esterases	53	32	NA	
Protease	NA	443	NA	
PHI genes	4315	2892	NA	
Secondary metabolies genes	48	41	36	
Length (bp)	23,488	NA	33,113	Mitogenome
Protein-coding genes	16	NA	24	
rRNA genes	2	NA	2	
tRNA genes	25	NA	35	

*The information refers to the references Listed below: 1. Wang G, Liu Z, Lin R, Li E, Mao Z, Ling J, et al. (2016) Biosynthesis of Antibiotic Leucinostatins in Bio-control Fungus Purpureocillium lilacinum and Their Inhibition on Phytophthora Revealed by Genome Mining. PLoS Pathog 12(7): e1005685. https://doi.org/10.1371/journal.ppat.1005685. 2. Nguyen N.-H., Tamura T., Shiminori K. (2022) Draft genome sequence of Purpurreocilium takamizusanense, a potential bioinsecticide. Microbiology Resource Announcements 11(7): e0026822. https://doi.org/10.1128/mra.00268-22

avg average

med median

(CBMs), 12 CBM + GH (Fig. 1A; Supplementary Table S7).

In the Pl-NCHU-NPUST-175 genome, the repeat region accounted for 3.70% of the genome, which was lower than those of *P. lilacinum* PLFJ-1 (6.00%) and *P. takamizusanense* PT3 (4.60%) (Table 1). De novo identification of the repetitive regions and transposable element (TE) regions in the genome of Pl-NCHU-NPUST-175 revealed that 1.95% of the genome (0.71 Mb) is composed of interspersed repetitive DNA. Most repeats were classified as retrotransposons (0.33 Mb; 0.92% of the genome), which was more than unclassified repeats (0.32 Mb; 0.89% of the genome) and DNA transposons (0.06 Mb; 0.17% of the genome). The retroelements included SINEs (0.003 Mb; 0.01%), LINEs (0.22 Mb; 0.61%), PLE/Chlamys (0.006 Mb; 0.02%) and LTR elements (0.10 Mb; 0.28%) (Supplementary Table S8). Comparative analysis revealed that most repetitive sequences presented a nucleotide sequence divergence of less than 30% compared to corresponding consensus

sequences (Fig. 2). The results of repeat-induced point (RIP) analysis revealed that 1.73% of the genome of Pl-NCHU-NPUST-175 was affected by RIP mutations, and there were 49 large RIP-affected regions (LRARs), which were affected by a RIP at least 4,000 bp (average size of 11.50 kb) (Supplementary Table S9).

The comparative genomics of Pl-NCHU-NPUST-175, *P. lilacinum* (GCA_001653265.1) and *P. takamizusanense* (GCA_022605165.1) are presented as syntenic CIRCOS plots utilizing a 50-kbp window (Fig. 1B). The line on the plots represents the alignment region, indicating the similarity of the three genomes. The comparison revealed the conserved regions among Pl-NCHU-NPUST-175, *P. lilacinum* and *P. takamizusanense*, while the sequencing method used might influence the genome assembly and result in differences in comparison.

Phylogenetic analysis

To better understand the classification status of Pl-NCHU-NPUST-175 and other *Purpureocillium* spp., a total of 59 conserved genes of Pl-NCHU-NPUST-175 were identified by UFCG (Supplementary Table S2). The pairwise percent identities between Pl-NCHU-NPUST-175 and *P. lilacinum* (GCA_001653205.1) and *P. takamizusanense* (GCA_022605165.1) were as high as 80%, indicating the conservation of these conserved genes in the three *Purpureocillium* spp. (Supplementary Fig. S4). In addition, on the basis of their nucleotide identities, 7 conserved genes were similar to those of *P. takamizusanense* (GCA_022605165.1), and 24 conserved genes presented lower identities to those of *P. lilacinum* (GCA_001653205.1) and *P. takamizusanense* (GCA_022605165.1), suggesting the distinct genome features of Pl-NCHU-NPUST-175 (Supplementary Fig. S4; Supplementary Table S2).

The phylogenetic analysis was performed on the basis of 59 conserved genes of Pl-NCHU-NPUST-175 and included 33 fungal strains or species belonging to 7 families within Hypocreales (Ophiocordycipitaceae, Clavicipitaceae, Bionectriaceae, Nectriaceae, Stachybotryaceae, Hypocreaceae and Cordycipitaceae), and *Aspergillus nidulans* served as an outgroup, which formed a separate node in the phylogenetic tree. (Fig. 3). The results revealed that Pl-NCHU-NPUST-175 belongs to Ophiocordycipitaceae and is closely related to *P. lilacinum* TERIBC 1 and *P. lilacinum* 36–1 (Fig. 3). Moreover, *Purpureocillium* spp. formed a clade with the other fungal species *Drechmeria coniospora* and *Hirsutella* spp., including *H. rhossiliensis* and *H. minnesotensis*; *Ophiocordyceps sinensis*; and *Tolypocladium* spp., including *T. ophioglossoides*, *T. cylindrosporum* and *T. inflatum* (Fig. 3). Within this cluster, most species are pathogenic to insects, while *P. lilacinum, P. lavendulum*, *D. coniospora*, *H. rhossiliensis* and *H. minnesotensis* are also considered nematophagous fungi, indicating that nematocidal and insecticidal fungi might have a common ancestor [48].

Pl-NCHU-NPUST-175 Mitogenome

The mitogenome of Pl-NCHU-NPUST-175 is composed of circular DNA molecules with a size of 23,488 bp and a GC content of 28.50%. Moreover, the GC skew was positive, and the AT skew was negative (Table 1; Supplementary Table S10).

The Pl-NCHU-NPUST-175 mitogenome consists of 16 protein-coding genes (*atp6*, *atp8*, *atp9*, *cob*, *cox1*, *cox2*, *cox3*, *nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5*, *nad6* and 2 hypothetical proteins) (Fig. 4; Supplementary Table S11). The hypothetical proteins were subjected to BLAST searches against the ribosomal protein S3 (*rps3*). The

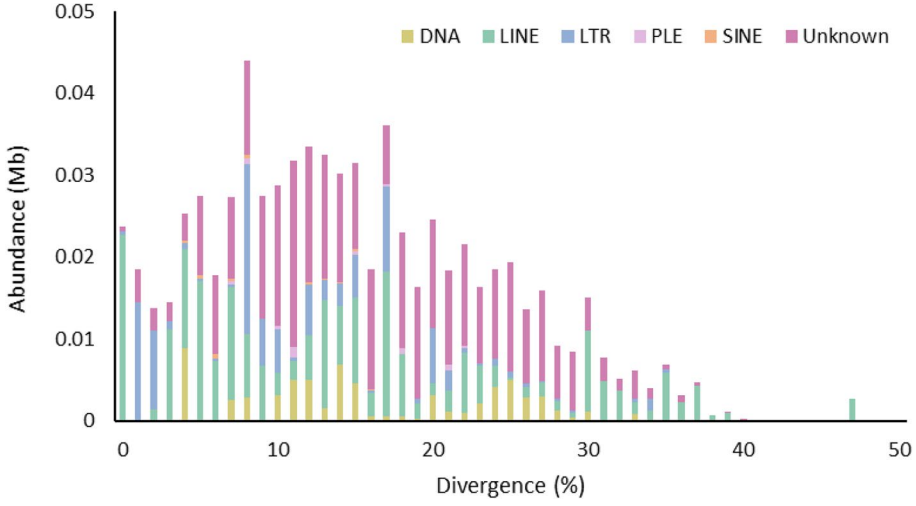

Fig. 2 The number of bases covered by the different types of predicted transposable elements (TEs). DNA = DNA transposons, LINE = long interspersed nuclear elements, LTR = long terminal repeats, PLE = Penelope-like elements, SINE = short interspersed nuclear elements, and Unknown = unclassified TEs

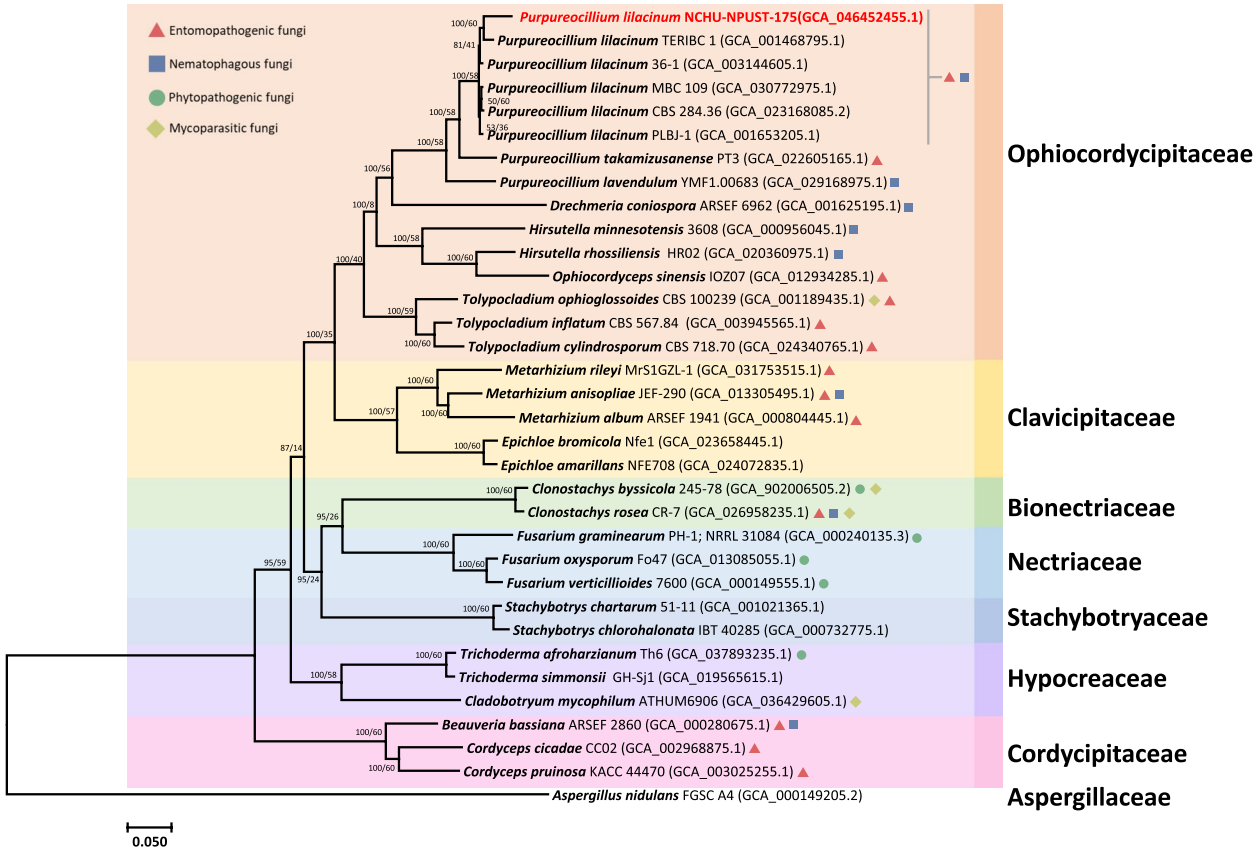

Fig. 3 Maximum likelihood (ML) tree of the concatenated alignment of UFCG conserved genes. Thirty-four fungal isolates were used to construct a phylogenetic tree. Branches of the resulting tree were annotated by their bootstrap support and gene support index (GSI) values. Bootstrap support and the GSI for branch support, which display the number of individual gene trees backing each branch as support values, are shown next to the branches (bootstrap support/GSI). The marks behind the species indicate the pathogenicity to different species

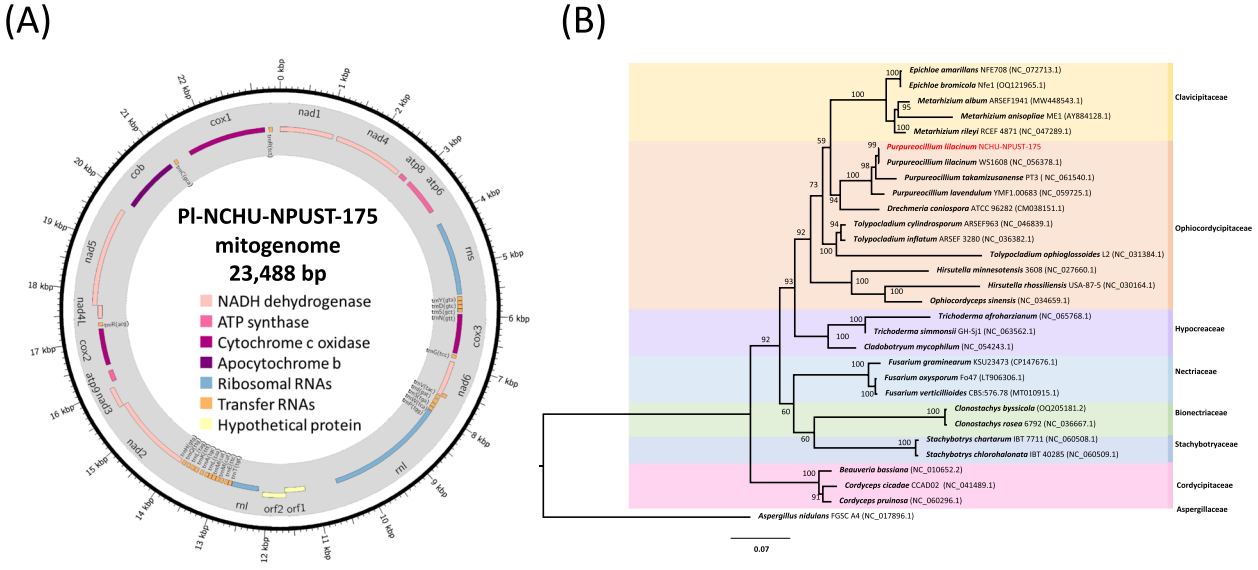

Fig. 4 Mitogenome analysis of PI-NCHU-NPUST-175. **A** The features and annotations of the PI-NCHU-NPUST-175 mitogenome were visualized by a CIRCOS plot. **B** Maximum likelihood (ML) tree of the mitogenome was constructed based on the concatenated alignment of 14 core genes using IQ-TREE (v2.4.0). Thirty fungal species were included in the phylogenetic tree. Branches represent the bootstrap support values

mitogenomes of Pl-NCHU-NPUST-175 contained two rRNA genes, the large subunit ribosomal RNA (*rnl*) gene and the small subunit ribosomal RNA gene. Twenty-five tRNA genes were detected in the mitogenome, including 3 tRNAs with different anticodons coding for methionine; 2 tRNAs with different anticodons coding for arginine, leucine, and serine; and tRNAs coding for alanine, asparagine, aspartate, cysteine, glutamine, glutamate, glycine, histidine, isoleucine, lysine, phenylalanine, proline, threonine, tryptophan, tyrosine, and valine. All 40 genes detected were Located on the same direction strand. Protein-coding regions accounted for 59.98% of the mitogenome, and RNA-coding regions accounted for 27.35%. The phylogenetic tree constructed with 14 core genes of the mitogenome is shown in Fig. 4B. We compared the phylogenetic relationships inferred from mitochondrial genomes with those obtained from the nuclear genome. The two trees show generally similar topologies, indicating the close relationship between *P. lilacinum* and *P. takamizusanense*. However, Clavicipitaceae branches off earlier in the mitochondrial phylogeny. This suggests that, based on the 14 conserved mitochondrial genes analyzed, Clavicipitaceae shares lower sequence similarity with Ophiocordycipitaceae, the family to which *Purpureocillium* belongs. These differences may reflect the choice of gene markers (mitochondrial vs. nuclear) can influence phylogenetic resolution, especially for closely related taxa.

Development of a Pl-NCHU-NPUST-175-specific Molecular Marker

The results of the comparative genomics analysis were used to identify distinctive genomic regions (DGRs) in Pl-NCHU-NPUST-175 by comparing them with those of two *Purpureocillium spp*. (*P. takamizusanense* and *P. lilacinum*). The analysis revealed a total of 1154 unaligned DGRs (> 5000 bp) (Supplementary Table S12). Subsequently, ten DGRs were randomly selected for specific primer design and subjected to further testing (Supplementary Table S3). Among these primer sets, all primer sets were found to be specific amplified for Pl-NCHU-NPUST-175 as confirmed by Sanger sequencing (Supplementary Fig. S5); however, four primer sets, including pl_4, pl_6, pl_7 and pl_9, presented a relatively high frequency of nonspecific binding to *P. lilacinum*-NCHU-152 (Pl-NCHU-152), which was observed in 6, 5, 6 and 7 of the 8 times tested, respectively (Fig. 5). For the other primer sets, pl_1 resulted in nonspecific amplification only 2 times, nonspecific binding of pl_10 was observed 3 times, and pl_2, pl_3, pl_5, and pl_8 resulted in 4 nonspecific amplifications of Pl-NCHU-152; however, there was no nonspecific amplification of these primer sets to the two isolates (TCTeb01 and TCP05) of *P. takamizusanense*, indicating that all ten primer sets were suitable for species-level identification (Fig. 5). In summary, three primer sets, including pl_1, pl_8,

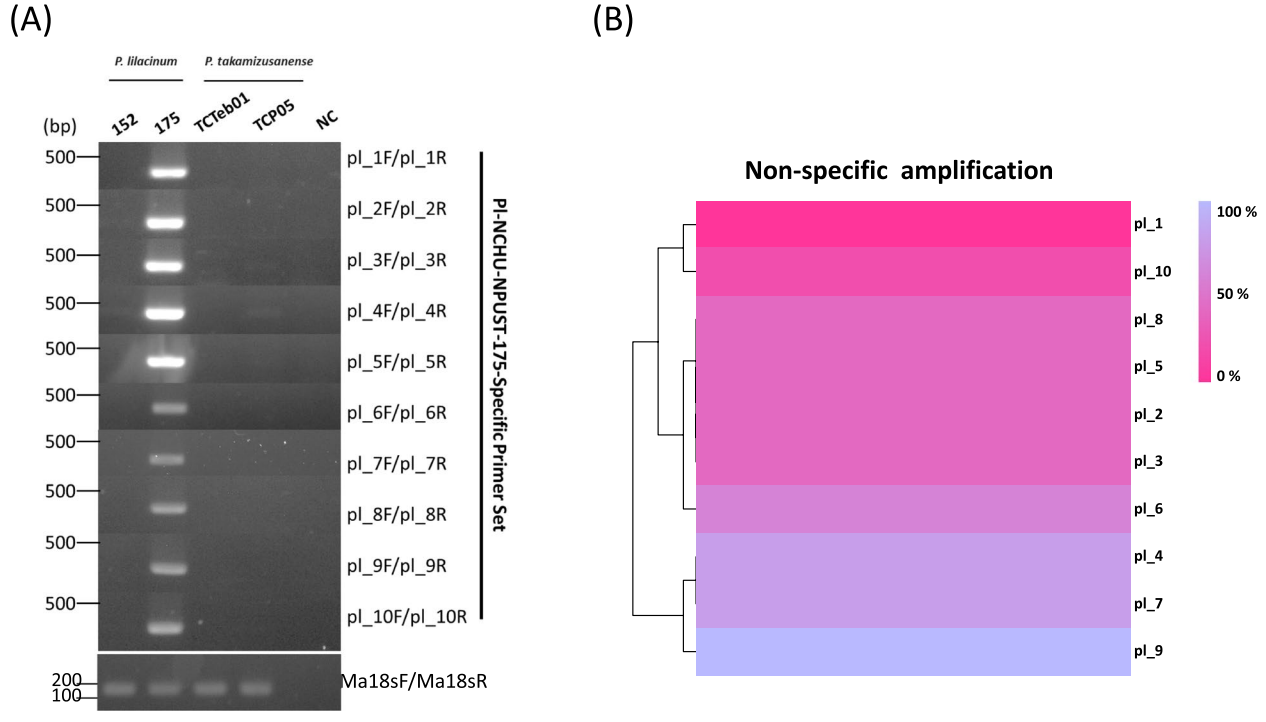

Fig. 5 Validation and selection of Pl-NCHU-NPUST-175 isolate-specific primer sets with one *P. lilacinum*-NCHU-152 (Pl-NCHU-152) and two isolates of *P. takamizusanense* (TCTeb01 and TCP05). **A** PCR was performed with ten primer sets. **B** The frequency of nonspecific amplification. Three primer sets (pl_1, pl_8, and pl_10) showed greater stability over eight repeated tests and a lower probability of nonspecific amplification

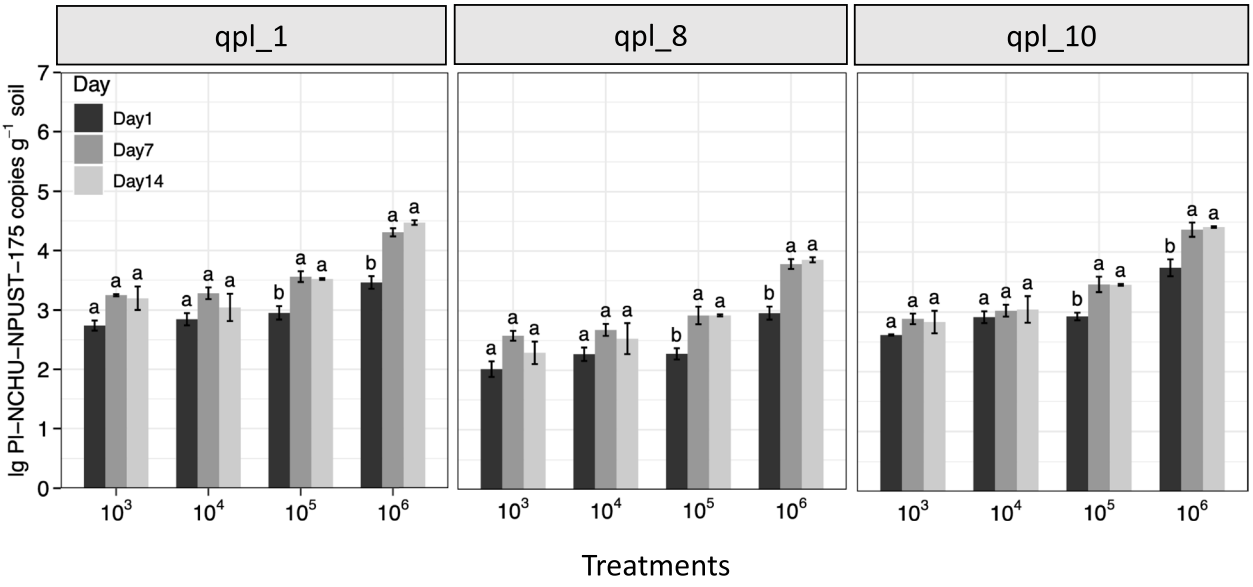

Fig. 6 Amounts of *P. lilacinum*-NCHU-NPUST-175 detected by qPCR using the primer pairs qpl_1, qpl_8 and qpl_10 in soils amended with different inoculation concentrations. Data represent the mean values from three replicates per soil treatment. Different letters indicate statistically significant differences among different time point based on Tukey's test

and pl_10, which presented greater stability over eight repeated tests, were selected for subsequent soil testing.

Detection of Pl-NCHU-NPUST-175 in soil samples via conventional PCR

Based on the result, the three selected primer sets successfully to detect the sample of 10^6 conidia g^{-1} soil. The faintest recoverable signal was detected at 10^5 conidia g^{-1} soil with primer sets pl_1 and pl_10 at one day post-inoculation (Supplementary Fig. S6). However, the intensity of the PCR signal on agarose gel increased at 7 and 14 days post-inoculation for soil with 10^5 conidia g^{-1} soil when tested with all primer sets (Supplementary Fig. S6). Nevertheless, the absence of signals in water control (NC) confirmed the isolate-specificity of these primers towards Pl-NCHU-NPUST-175. Thus, primer sets pl_1, pl_8, and pl_10 demonstrated the ability to effectively monitor the target species in natural soil, with a limitation of10^5 conidia g^{-1} soil.

Detection of the soil-based fungal DNA by qPCR

It was noted that qPCR revealed high sensitivity to detect from 10^3 conidia g-1 soil to 10^6 conidia g^{-1} soil, besides, the results of qPCR indicated that all primers exhibited an increasing trend in response to the added spore concentrations (Fig. 6). Despite there was no significant differences were observed over time at 10^3 conidia g-1 soil and 10^4 conidia g-1 soil, a significant increase in the copy number of the inoculated fungus was detected at 7 and 14 days by all three primer sets were found at 10^5 and 10^6 conidia g^{-1} soil. The primer qpl_8 consistently detected lower copy numbers compared to those of primer qpl_1 and qpl_10, suggesting that the two primer sets are the most effective primers for detecting Pl-NCHU-NPUST-175 in the soil matrix. To evaluate the inhibition effect in a complex background, BQCV plasmid DNA was spiked into 10 ng/μL of soil DNA at final proportions of 50%, 10%, 1%, and 0.1%. There was a slight shift in the Ct value of the spike in DNA when added to soil DNA (Supplementary Table S13). It ranged from −1.01 to + 1.31, indicating minimal to moderate PCR inhibition.

Discussion

The length of the nuclear genome assembly and comparative analysis of Pl-NCHU-NPUST-175 generated from nanopore sequencing was 36.55 Mb, which conformed to the expected genome size for the Ascomycota phylum (average 36.90 Mb) [49]. Compared with those of the *Purpureocillium* spp. in the NCBI database, the Pl-NCHU-NPUST-175 genome reveals a larger genome size than that of *P. takamizusanense* (GCA_022605165.1) (35.57 Mb) and a smaller genome size than that of the closely related species *P. lilacinum* (GCA_001653265.1) (38.53 Mb). BUSCO assesses genome assembly completeness by detecting conserved single-copy orthologs using gene prediction tools. The genome of NCHU-NPUST-175 showed 94.70% complete BUSCOs, indicating acceptable genome completeness. However, the actual roles of these genes still need to be further verified to identify the specific duplicated genes, the resulting gene structure, and the impact of duplication events.

The regions of secondary metabolites were analyzed via antiSMASH. Secondary metabolites exhibit a wide range of biological activities, and studies have shown that they

can serve as adjuncts to chemical pesticides [50]. Therefore, further investigations into secondary metabolites could contribute to the discovery of novel compounds and a deeper understanding of their biological activities. There were 42 regions predicted to be related to the biosynthesis of secondary metabolites; this result is similar to that previously reported [51]. The carbohydrate-active enzymes (CAZymes) were identified and classified with dbCAN3. The cuticular egg envelope components of insects include proteins, lipids, waxes, and chitin [52]. To penetrate structural polysaccharides such as chitin on insect eggs, CAZymes are secreted by fungi to degrade polysaccharides. In this study, many types of CAZymes were predicted in the Pl-NCHU-NPUST-175 genome, and chitinases (GH18 class) were the most abundant glycoside hydrolases. The number of GH18 genes is comparable to that found in other *P. lilacinum* strains, suggesting that GH18 enrichment is a conserved feature within this genus [51]. GH18 chitinases are well known for their ability to hydrolyze β–1,4-linkages in chitin, a major structural component of insect cuticles. A similar number of GH18 genes has also been reported in other entomopathogenic fungi, such as *Metarhizium anisopliae* [53]. While gene expression data are not available in this study, the conserved presence and abundance of GH18 genes across entomopathogenic fungi highlight their potential functional importance in host infection.

In addition to the nuclear genome, the mitogenome was analyzed in this study, and the size and the GC content were similar to those of other species and *P. lavendulum* [54, 55]. The size variations of fungal mitogenomes might vary greatly and are due mainly to repetitive regions, changes in introns, differences in intergenic sequences and horizontal gene transfer (HGT) [56]. There were some conserved characteristics in the Hypocreales mitogenomes, including 14 core PCGs and 2 rRNA genes and the joining pattern of *nad4L* and *nad5* (i.e., one base overlapping). These characteristics were also observed for the Pl-NCHU-NPUST-175 mitogenome.

Previous study has shown that *P. lilacinum* has the potential to establish itself in both field and greenhouse environments and act as an endophytic fungus for pest control [57]. Similarly, previous study demonstrated the potential applications of Pl-NCHU-NPUST-175 in the biological control of *F. taiwana* [9]. Therefore, a rapid and sensitive diagnostic test to confirm the persistence of the Pl-NCHU-NPUST-175 inoculant in the environment would be beneficial for field monitoring. In the case of EPF belonging to the *Purpureocillium* genus, their morphology closely resembles that of phylogenetically related species, making classification on the basis solely of morphology impractical [58]. In terms of precise classification, molecular identification techniques are valuable for distinguishing not only inter- but also intraspecies characteristics that are difficult to identify morphologically and for understanding the genetic relatedness among different isolates [59]. Therefore, reliable molecular identification plays a pivotal role in the aforementioned research.

EPF can be identified using molecular markers, such as ITSs. The Consortium for the Barcode of Life (CBOL) recommends ITSs as the preferred DNA barcode for fungi [60]. However, ITS variation has been confirmed in populations and individuals [18]. Therefore, relying solely on ITS may not be sufficient to identify each isolate at the species level; thus, other molecular markers must be designed. In this study, Pl-NCHU-NPUST-175 was tracked in the environment via the use of specific primer sets, which were explored on the basis of comparative genome analysis to investigate Pl-NCHU-NPUST-175. Unlike other molecular tools such as restriction fragment length polymorphism (RFLP) or random amplification of polymorphic DNA (RAPD) which lack the tendency to differentiate in strain level [19], comparative genomic sequencing allows for genome-wide identification of unique isolate-specific variations, ensuring precise differentiation [23]. While comparative genomics analysis offers high-resolution marker development, it has certain limitations. Differences in sequencing platforms (e.g., Illumina vs. ONT) can introduce errors and bias, affecting marker selection. Despite this limitation, genome-based molecular marker development remain a powerful approach, particularly for differentiation and functional genomics applications [46]. A total of 1154 unaligned DGRs were identified in this study, and ten primers were designed from 10 DGRs.

Experimental validation using 10 PCR primer sets designed from randomly selected target regions successfully distinguished Pl-NCHU-NPUST-175 from its closely related species, *P. takamizusanense*, or other *P. lilacinum* isolates, indicating higher primer specificity. Qualitative and quantitative PCR also demonstrated the potential of the primer sets (pl_1, pl_8, and pl_10) for detecting the fungus in agricultural soil samples. However, PCR amplification using DNA directly extracted from soil samples has been reported to be challenging [61]. While the PCR result showed that the sensitivity of primers was Limited to 10^5 conidia g^{-1} soil in natural soil, qPCR successfully detected the number of Pl-NCHU-NPUST-175 even in 10^3 conidia g^{-1} soil. There was increasing in signal intensity at 7 and 14 days, particularly at 10^5 conidia g^{-1}, assuming the fungi might proliferate in the soil over time, however, the process of fungal proliferate needs to be further evaluated.

In terms of potential PCR inhibition from soil sample, the ΔCt values ranged from −1.01 to + 1.31 was detected, indicating minimal to moderate inhibition. These shifts reflect the presence of inhibitory substances in the soil

DNA but confirm that amplification was still robust [62]. This finding aligns with previous observations that DNA extracted from soil often yields less reproducible amplification compared to pure DNA. Such variability is commonly attributed to PCR inhibitors, such as humic acids, and other organic compounds, as well as the presence of high-background microbial DNA in environmental samples [63]. Although this level of inhibition did not prevent amplification, it suggests a partial suppression of qPCR efficiency. Similarly, when comparing Ct values from DNA extracted from fungal-inoculated soil with those from pure fungal DNA, slight shifts in Ct values were observed. This further supports the presence of PCR inhibitors in environmental samples. Despite this, the qPCR assay remained effective, and the primer sets qpl-1 and qpl_10 demonstrated high specificity and consistent amplification, indicating the robustness of the assay under complex environmental conditions. Three DGR markers detected target fungi at different time intervals, indicating that Pl-NCHU-NPUST-175 is persistent in the environment and can be tracked in soil samples by using the specific markers designed by genome-wide comparison strategies. A similar strategy has been applied to identify strain-specific qPCR markers to detect *Trichoderma guizhouense* strain NJAU 4742 (transformed with the gfp gene) in the soil [46], indicating its applicability to monitor the population dynamics of different strains based on the designed primers. However, further validation under more variable environmental settings, such as greenhouse or field conditions, is necessary to fully establish their robustness and applicability for field use.

Conclusion

In this study, the genome sequence of Pl-NCHU-NPUST-175 was decoded and compared with those of closely related isolates and species. The DGRs specific to this fungal isolate were identified through a comparative genomics approach and subsequently utilized for the design of 10 specific primer sets (pl_1 to pl_10). They were successful to distinguish Pl-NCHU-NPUST-175 from other isolates or similar species. The qPCR primer set, qpl_1, qpl_8, and qpl_10, could detect the number of Pl-NCHU-NPUST-175 inoculated into the soil over period of time. Hence, the findings of this study represent a foundational step toward establishing reliable molecular tools for *P. lilacinum* detection.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-025-12018-6.

Supplementary Material 1. Supplementary Table S1. Fungal species used for phylogenetic analysis."-": no strain names."x": no presented in the mitogenome tree. Supplementary Table S2. Conserved genes of ***Purpureocillium lilacinum***-NCHU-NPUST-175 (Pl-NCHU-NPUST-175) and other fungal species. *SGD: Saccharomyces Genome Database. **CDD: Conserved Domain Database. ***COG: Clusters of Orthologous Group. Supplementary Table S3. Information on ten specific primer sets in the distinctive genomic regions (DGRs) of Pl-NCHU-NPUST-175. *Primer sets selected for further test. £Primer set for internal control referenced from Fang, W., & Bidochka, M. J. (2006). Expression of genes involved in germination, conidiogenesis and pathogenesis in ***Metarhizium anisopliae*** using quantitative real-time RT-PCR. Mycological Research, 110(10), 1165-1171. Supplementary Table S4. Nanopore sequencing summary of ***Purpureocillium lilacinum***-NCHU-NPUST-175 (Pl-NCHU-NPUST-175). Supplementary Table S5. The output of InterProScan. Supplementary Table S6. Key enzymes involved in secondary metabolism in the ***Purpureocillium lilacinum***-NCHU-NPUST-175 (Pl-NCHU-NPUST-175) genome. Supplementary Table S7. CAZyme analysis of the ***Purpureocillium lilacinum***-NCHU-NPUST-175 (Pl-NCHU-NPUST-175) genome. Supplementary Table S8. Repeat region analysis of the ***Purpureocillium lilacinum***-NCHU-NPUST-175 (Pl-NCHU-NPUST-175) genome. SINE= short interspersed nuclear elements. LINE= long interspersed nuclear elements. LTR= long terminal repeats. Low complexity DNA= It refers to sequences with nucleotide compositions that are biased towards simplicity or uniformity but do not strictly form tandem repeats. (eg: primarily poly-purine/poly-pyrimidine stretches, or regions of extremely high AT or GC content). Supplementary Table S9. Repeat-Induced Point analysis of ***Purpureocillium lilacinum***-NCHU-NPUST-175 (Pl-NCHU-NPUST-175) genome. Supplementary Table S10. Mitogenome assembly and annotation of ***Purpureocillium lilacinum***-NCHU-NPUST-175 (Pl-NCHU-NPUST-175). Supplementary Table S11. The annotation of the ***Purpureocillium lilacinum***-NCHU-NPUST-175 (Pl-NCHU-NPUST-175) mitogenome. *The core genes used for phylogenetic analysis. Supplementary Table S12. Information on 1154 unaligned genomic fragments. Supplementary Table S13. ΔCt Comparison Between Plasmid-Only and Plasmid + Soil DNA at different concentration to assess PCR inhibition.

Supplementary Material 2. Supplementary Figure S1. Genome assembly and annotation pipeline.

Supplementary Material 3. Supplementary Figure S2. Mitogenome assembly and annotation pipeline.

Supplementary Material 4. Supplementary Figure S3. Comparative genome analysis and primer development process.

Supplementary Material 5. Supplementary Figure S4. Pairwise alignment between marker genes predicted by UFCG of Pl-NCHU-NPUST-175, ***P. takamizusanense*** (pt) (GCA_022605165.1), and ***P. lilacinum*** (pl) (GCA_001653205.1).

Supplementary Material 6. Supplementary Figure S5. Representative Sanger sequencing alignment results of PCR amplicons generated by 10 primer sets. The dot indicates that the sequence is identical to the corresponding position in Sanger sequences.

Supplementary Material 7. Supplementary Figure S6. Specificity of the isolate-specific markers in natural soil via conventional PCR. A total of 50 µL of the conidial suspension was added to 1.5 mL centrifuge tubes containing 0.25 g of sterilized dry soil (approximately 20% soil moisture). Concentrations of 10^5, 10^4, 10^3, 10^2 and 10^1 conidia per gram of sterile soil were tested with three biological replicates. Concentrations of 10^6, 10^5, 10^4, and 10^3 conidia per gram of natural soil were tested with three biological replicates at 1, 7 and 14 days post-inoculation. Rep.= Experimental replications.

Acknowledgements

We thank Ms. Ju-Chun Chang for gDNA sample preparation and ONT sequencing and Ms. Yi-Hsuan Li for the and data consulting.

Authors' contributions

ZYY conducted all the experiments, bioinformatic analysis and drafted the manuscript. KM performed the fungal identify experiment and review the manuscript. NTN and PHL performed the fungal cultivation, fungal gDNA preparation and review the manuscript. CFL and YSN as corresponding authors, provided experimental conception, guidance and drafted the manuscript.

Funding
This research was supported by 113–2313-B-005–024-MY3 and 113–2327-B-020–001- from the National Science and Technology Council (NSTC).

Data availability
The long-read sequencing (ONT) raw data and the annotated genome information of *P. lilacinum* -NCHU-NPUST-175 genome project were deposited in NCBI under BioProject accession number: PRJNA1149534 and GenBank accession number: GCA_046452455.1. The Sanger sequencing data generated for the validation of 10 primer sets have been deposited in the NCBI database under accession numbers PX098487–PX098496.

Declarations

Ethics approval and consent to participate
This is not applicable in this study.

Consent for publication
This is not applicable in this study.

Competing interests
The authors declare no competing interests.

Author details
[1]Department of Entomology, National Chung Hsing University, Taichung 40227, Taiwan
[2]International Doctoral Program in Agriculture, National Chung Hsing University, Taichung 40227, Taiwan
[3]Taichung District Agricultural Research and Extension Station, Ministry of Agriculture, Changhua 515008, Taiwan
[4]Department of Veterinary Medicine, College of Veterinary Medicine, National Pingtung University of Science & Technology, Pingtung 912, Taiwan

Received: 28 December 2024 / Accepted: 19 August 2025
Published online: 26 September 2025

References

1. Thakur N, Tomar P, Kaur S, Jhamta S, Thakur R, Yadav AN: Entomopathogenic soil microbes for sustainable crop protection. Soil Microbiomes for Sustainable Agriculture: Functional Annotation. 2021:529–571.
2. Alfiky A. Screening and identification of indigenous entomopathogenic fungal isolates from agricultural farmland soils in Nile Delta. Egypt J Fungi (Basel). 2022;8(1):54.
3. Keçili S, Bakır A, Kutalmış A, Çelik T, Sevim A. Soil isolation, identification, and virulence testing of Turkish entomopathogenic fungal strains: a potential native isolate of *Beauveriabassiana* for the control of *Leptinotarsadecemlineata*. Biocontrol. 2022;67(6):593–603.
4. Sharma A, Sharma S, Yadav PK. Entomopathogenic fungi and their relevance in sustainable agriculture: a review. Cogent Food Agric. 2023;9(1):2180857.
5. Chandler D: Basic and applied research on entomopathogenic fungi. In: Microbial control of insect and mite pests. Elsevier; 2017: 69–89.
6. Mora MAE, Castilho AMC, Fraga ME: Classification and infection mechanism of entomopathogenic fungi. Arq Inst Biol. 2018; 84.
7. Desoky SMA, Abdelall MF, Ahmed Y. Isolation, identification, evaluation of *Purpureocilliumlilacinum* Egyptian isolate toxicity test in vitro and analysis its bioactive products. Journal of the Advances in Agricultural Researches. 2022;27(4):602–17.
8. Lo P, Yu Y, Pai K. First report of *Purpureocillium takamizusanense* as an omopathogenic fungus infecting *Tessaratoma papillosa* (Drury) in Taiwan. J Plant Med. 2019;61(2/3):27–30.
9. Ni NT, Wu SS, Liao KM, Tu WC, Lin CF, Nai YS. Evaluation of the potential entomopathogenic fungi *Purpureocillium lilacinum* and *Fusarium verticillioides* for biological control of *Forcipomyia taiwana* (Shiraki). J Fungi. 2022;8(8):861.
10. Baron NC, de Souza Pollo A, Rigobelo EC. *Purpureocillium lilacinum* and *Metarhizium marquandii* as plant growth-promoting fungi. PeerJ. 2020;8:e9005.
11. Atkins SD, Clark IM, Pande S, Hirsch PR, Kerry BR. The use of real-time PCR and species-specific primers for the identification and monitoring of *Paecilomyces lilacinus*. FEMS Microbiol Ecol. 2005;51(2):257–64.
12. Atkins S, Clark I, Sosnowska D, Hirsch P, Kerry B. Detection and quantification of *Plectosphaerella cucumerina*, a potential biological control agent of potato cyst nematodes, by using conventional PCR, real-time PCR, selective media, and baiting. Appl Environ Microbiol. 2003;69(8):4788–93.
13. Ding JL, Lin HY, Hou J, Feng MG, Ying SH. The entomopathogenic fungus *Beauveria bassiana* employs autophagy as a persistence and recovery mechanism during conidial dormancy. mBio. 2023;14(2):e03049-03022.
14. Garrido-Jurado I, Ruano F, Campos M, Quesada-Moraga E. Effects of soil treatments with entomopathogenic fungi on soil dwelling non-target arthropods at a commercial olive orchard. Biol Control. 2011;59(2):239–44.
15. Chang JC, Wu SS, Liu YC, Yang YH, Tsai YF, Li YH, Tseng CT, Tang LC, Nai YS. Construction and selection of an entomopathogenic fungal library from soil samples for controlling spodoptera litura. Front Sustain Food Syst. 2021;5.
16. Riaz M, Chen W-H, Kafle L, Tseng M-N. Morphological and molecular characterization of *Purpureocillium lilacinum* along with its biopesticidal effect against fall armyworm (*Spodoptera frugiperda*) in Southern Taiwan. Egypt J Biol Pest Control. 2024;34(1):60.
17. Schoch CL, Seifert KA, Huhndorf S, Robert V, Spouge JL, Levesque CA, et al. Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for fungi. Proc Natl Acad Sci U S A. 2012;109(16):6241–6.
18. Smith ME, Douhan GW, Rizzo DM. Intra-specific and intra-sporocarp ITS variation of ectomycorrhizal fungi as assessed by rDNA sequencing of sporocarps and pooled ectomycorrhizal roots from a *Quercus* woodland. Mycorrhiza. 2007;18:15–22.
19. Agrawal PK, Shrivastava R: Molecular markers. In: Advances in Biotechnology. Springer; 2013: 25–39.
20. Achtman M, Wagner M. Microbial diversity and the genetic nature of microbial species. Nat Rev Microbiol. 2008;6(6):431–40.
21. Bapteste E, Boucher Y, Leigh J, Doolittle WF. Phylogenetic reconstruction and lateral gene transfer. Trends Microbiol. 2004;12(9):406–11.
22. Lofgren LA, Stajich JE. Fungal biodiversity and conservation mycology in light of new technology, big data, and changing attitudes. Curr Biol. 2021;31(19):R1312–25.
23. Qi G, Hao L, Xin T, Gan Y, Lou Q, Xu W, et al. Analysis of whole-genome facilitates rapid and precise identification of fungal species. Front Microbiol. 2024;15:1336143.
24. De Coster W, Rademakers R. Nanopack2: population-scale evaluation of long-read sequencing data. Bioinformatics. 2023;39(5):btad311.
25. Lanfear R, Schalamun M, Kainer D, Wang W, Schwessinger B. MinIONqc: fast and simple quality control for MinION sequencing data. Bioinformatics. 2019;35(3):523–5.
26. Wick RR, Judd LM, Gorrie CL, Holt KE. Completing bacterial genome assemblies with multiplex MinION sequencing. Microb Genom. 2017. https://doi.org/10.1099/mgen.0.000132.
27. Loman NJ, Quick J, Simpson JT. A complete bacterial genome assembled *de novo* using only nanopore sequencing data. Nat Methods. 2015;12(8):733–5.
28. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. Bioinformatics. 2013;29(8):1072–5.
29. Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. Mol Biol Evol. 2021;38(10):4647–54.
30. Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, et al. Repeatmodeler2 for automated genomic discovery of transposable element families. Proc Natl Acad Sci U S A. 2020;117(17):9451–7.
31. Chen N. Using repeat masker to identify repetitive elements in genomic sequences. Curr Protoc Bioinformatics. 2004;5(1):4.10. 11–14.10. 14.
32. Van Wyk S, Harrison CH, Wingfield BD, De Vos L, van Der Merwe NA, Steenkamp ET. The ripper, a web-based tool for genome-wide quantification of repeat-induced point (RIP) mutations. PeerJ. 2019;7:e7447.
33. Min B, Grigoriev IV, Choi I-G. FunGAP: fungal genome annotation pipeline using evidence-based gene model evaluation. Bioinformatics. 2017;33(18):2936–7.
34. Urban M, Cuzick A, Seager J, Wood V, Rutherford K, Venkatesh SY, et al. PHI-base in 2022: a multi-species phenotype database for pathogen–host interactions. Nucleic Acids Res. 2022;50(D1):D837–47.
35. Blin K, Shaw S, Augustijn HE, Reitz ZL, Biermann F, Alanjary M, et al. AntiSMASH 7.0: new and improved predictions for detection, regulation, chemical structures and visualisation. Nucleic Acids Res. 2023;51(W1):W46–50.

36. Zheng J, Ge Q, Yan Y, Zhang X, Huang L, Yin Y. dbCAN3: automated carbohydrate-active enzyme and substrate annotation. Nucleic Acids Res. 2023;51(W1):W115-21.
37. Chan PP, Lin BY, Mak AJ, Lowe TM. tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes. Nucleic Acids Res. 2021;49(16):9077–96.
38. Hunt M, Silva ND, Otto TD, Parkhill J, Keane JA, Harris SR. Circlator: automated circularization of genome assemblies using long sequencing reads. Genome Biol. 2015;16:1–10.
39. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, et al. Circos: an information aesthetic for comparative genomics. Genome Res. 2009;19(9):1639–45.
40. Stanke M, Keller O, Gunduz I, Hayes A, Waack S, Morgenstern B. AUGUSTUS: ab initio prediction of alternative transcripts. Nucleic Acids Res. 2006;34(suppl_2):W435–9.
41. Kim D, Gilchrist CL, Chun J, Steinegger M. UFCG: database of universal fungal core genes and pipeline for genome-wide phylogenetic analysis of fungi. Nucleic Acids Res. 2023;51(D1):D777–84.
42. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, Von Haeseler A, et al. Iq-tree 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol. 2020;37(5):1530–4.
43. Vaidya G, Lohman DJ, Meier R. Sequencematrix: concatenation software for the fast assembly of multi-gene datasets with character set and codon information. Cladistics. 2011;27(2):171–80.
44. Marçais G, Delcher AL, Phillippy AM, Coston R, Salzberg SL, Zimin A. MUMmer4: a fast and versatile genome alignment system. PLoS Comput Biol. 2018;14(1):e1005944.
45. Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, et al. Primer3—new capabilities and interfaces. Nucleic Acids Res. 2012;40(15):e115–e115.
46. Zhang Y, Wang X, Pang G, Cai F, Zhang J, Shen Z, et al. Two-step genomic sequence comparison strategy to design *Trichoderma* strain-specific primers for quantitative PCR. AMB Express. 2019;9:1–10.
47. Team RC. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2024.
48. Flores Francisco BG, Ponce IM, Plascencia Espinosa MÁ, Mendieta Moctezuma A, López y López VE. Advances in the biological control of phytoparasitic nematodes via the use of nematophagous fungi. World J Microbiol Biotechnol. 2021;37(10):180.
49. Mohanta TK, Bae H. The diversity of fungal genome. Biol Proced Online. 2015;17:1–9.
50. Wang G, Liu Z, Lin R, Li E, Mao Z, Ling J, et al. Biosynthesis of antibiotic leucinostatins in bio-control fungus *Purpureocillium lilacinum* and their inhibition on *Phytophthora* revealed by genome mining. PLoS Pathog. 2016;12(7):e1005685.
51. Prasad P, Varshney D, Adholeya A. Whole genome annotation and comparative genomic analyses of bio-control fungus *Purpureocillium lilacinum*. BMC Genomics. 2015;16:1–14.
52. Rezende GL, Vargas HCM, Moussian B, Cohen E: Composite eggshell matrices: chorionic layers and sub-chorionic cuticular envelopes. Extracellular composite matrices in arthropods. 2016:325–366.
53. Junges Â, Boldo JT, Souza BK, Guedes RLM, Sbaraini N, Kmetzsch L, et al. Genomic analyses and transcriptional profiles of the glycoside hydrolase family 18 genes of the entomopathogenic fungus *Metarhizium anisopliae*. PLoS One. 2014;9(9):e107864.
54. Li J, Zhang G, Yu H, Huang L, Zeng W, Wang Y. Complete mitochondrial genome of the important bio-control fungus *Purpureocillium lilacinum* (Ophiocordycipitaceae, Hypocreales) and its phylogenetic analysis. Mitochondr DNA Part B. 2020;5(1):240–2.
55. Liang LM, Zhang Y, Xu J, Zhang KQ, Cao YR. Characterization of the complete mitochondrial genome of the nematophagous fungus *Purpureocillium lavendulum*. Mitochondr DNA B Resources. 2021;6(1):33–5.
56. Li Q, Yang L, Xiang D, Wan Y, Wu Q, Huang W, et al. The complete mitochondrial genomes of two model ectomycorrhizal fungi (Laccaria): features, intron dynamics and phylogenetic implications. Int J Biol Macromol. 2020;145:974–84.
57. Castillo Lopez D, Zhu-Salzman K, Ek-Ramos MJ, Sword GA. The entomopathogenic fungal endophytes *Purpureocillium lilacinum* (formerly *Paecilomyces lilacinus*) and *Beauveria bassiana* negatively affect cotton aphid reproduction under both greenhouse and field conditions. PLoS One. 2014;9(8):e103891.
58. Tigano-Milani MS, Samson RA, Martins I, Bruno W. DNA markers for differentiating isolates of *Paecilomyces lilacinus*. Microbiology. 1995;141(1):239–45.
59. Peay KG, Kennedy PG, Bruns TD. Fungal community ecology: a hybrid beast with a molecular master. Bioscience. 2008;58(9):799–810.
60. Pawlowski J, Audic S, Adl S, Bass D, Belbahri L, Berney C, et al. Cbol protist working group: barcoding eukaryotic richness beyond the animal, plant, and fungal kingdoms. PLoS Biol. 2012;10(11):e1001419.
61. Cullen D, Hirsch PR. Simple and rapid method fordirect extraction of microbial DNA fromsoil for PCR. Soil Biol Biochem. 1998;30(8–9):983–93.
62. Bott T, Shaw G, Gregory S. A simple method for testing and controlling inhibition in soil and sediment samples for qPCR. J Microbiol Methods. 2023;212:106795.
63. Wnuk E, Waśko A, Walkiewicz A, Bartmiński P, Bejger R, Mielnik L, et al. The effects of humic substances on DNA isolation from soils. PeerJ. 2020;8:e9378.

Publisher's Note