scientific reports

Check for updates

OPEN

Genomic and metabolomic insights into *Trichoderma harzianum* T9, a resilient biocontrol fungus from arid environments

Francisco Vargas-Gasca[1], Enrique Pola-Sánchez[2], Ana Valeria García-Lartigue[2], Alan D. Gomez-Vargas[3], Pablo Cruz-Morales[4], Ana Calheiros de Carvalho[4], Daniela Rago[4], Linda Ahonen[4], Elva Teresa Aréchiga-Carvajal[5], José Manuel Villalobos-Escobedo[2]✉ & Vianey Olmedo-Monfil[1]✉

The search for sustainable agricultural solutions to reduce pesticide dependence is urgent, particularly under the growing pressures of climate change. Although *Trichoderma* species are well-known biocontrol agents, many strains perform poorly in extreme soils characterized by high salinity or alkaline pH. Here, we characterize *Trichoderma harzianum* T9, an isolate from the alkaline desert soils of Nuevo León, Mexico, that exhibits exceptional ecological resilience. A competition assay revealed that T9 exhibits significantly higher biocontrol activity against the phytopathogenic fungi *Macrophomina* sp., *Neopestaloptiosis rosae*, and two strains of *Fusarium* sp., all major threats to strawberry production in Mexico, outperforming other *Trichoderma* strains. Genome sequencing and phylogenomics confirmed that T9 belongs to the *T. harzianum* species. SNP-based variant analysis revealed numerous genes involved in secondary metabolism with elevated nucleotide substitution rates, including a highly diverged polyketide synthase (PKS). Metabologenomics analysis predicted chemical variation primarily in peptaibols, while biosynthetic gene cluster (BGC) prediction identified six additional putative metabolites likely contributing to its strong antifungal activity. In this work, we also delineate the biosynthetic pathways underlying these seven compounds. Together, these findings position *T. harzianum* T9 as a highly promising biocontrol agent for managing phytopathogens in degraded soils, providing an eco-friendly strategy that supports sustainable agriculture. The distinctive genomic and metabolic traits of T9 highlight the untapped potential of microorganisms from extreme environments to drive innovative approaches for crop protection and soil restoration.

Keywords Biocontrol, Extreme environments, Secondary metabolism, Peptaibols, Biosynthetic gene clusters

Filamentous fungi of the genus *Trichoderma* are widely distributed across diverse ecosystems and are renowned for their rapid growth, substrate versatility, and antagonistic activity against phytopathogenic fungi[1]. Some species also establish beneficial associations with plants, promoting growth and nutrient uptake[2–4], which underscores their value as biocontrol agents in agriculture. Despite these advantages, many *Trichoderma* strains are sensitive to harsh soil conditions such as high salinity and alkaline pH, which restricts their application in challenging environments[5].

Microorganisms adapted to extreme habitats, including deserts, often display unique genomic and phenotypic traits shaped by selective pressures such as heat, water scarcity, and nutrient limitation. These adaptations, including altered cell membranes, stress-related proteins, and broad metabolic flexibility, make extremophilic fungi promising resources for biotechnology[6]. However, widely used commercial strains like *T. harzianum* T22

[1]División de Ciencias Naturales y Exactas, Departamento de Biología, Universidad de Guanajuato, Guanajuato, México. [2]Tecnológico de Monterrey, Institute for Obesity Research, Ave. Eugenio Garza Sada 2501, Monterrey 64849, N.L, Mexico. [3]Departamento de Ingeniería Genética, CINVESTAV Irapuato, Km. 9.6 Libramiento Norte Carretera Irapuato-León, Guanajuato, Irapuato 36821, México. [4]The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kgs. Lyngby DK-2800, Denmark. [5]Unidad de Manipulación Genética, Laboratorio de Micología y Fitopatología, Facultad de Ciencias Biológicas, Universidad Autónoma de Nuevo León, San Nicolás de los Garza 66451, México. ✉email: jose.villalobos@tec.mx; vg.olmedo@ugto.mx

and *T. atroviride* IMI206040 show reduced performance in saline or alkaline soils, highlighting the need for novel strains capable of thriving under such conditions.

In a previous study, we isolated *Trichoderma* strains from the alkaline desert soils of Mina, Nuevo León, México. Among them, *T. harzianum* T9[7] demonstrated exceptional tolerance to pH levels up to 9 and improved plant resilience under water stress. These features position T9 as a promising candidate for biocontrol and soil remediation in extreme agricultural settings.

Notably, *Trichoderma atroviride* is known to combat phytopathogens such as *Fusarium*, *Botrytis*, *Rhizoctonia*, and *Alternaria*[8,9]. However, the emergence of highly aggressive fungi resistant to both chemical fungicides and conventional biocontrol strains has created serious challenges. Strawberry producers in Mexico, for example, face devastating outbreaks caused by *Macrophomina* spp. and *Neopestalotiopsis rosae*, which can reduce yields by up to 50%[10]. Current management practices have proven insufficient, underscoring the urgent need for novel, robust biocontrol solutions.

In this study, we collected strawberry plants showing apparent symptoms of fungal infection from commercial fields in Guanajuato, Mexico. Fungal isolates were obtained from symptomatic tissues, and ITS sequencing revealed that the pathogens corresponded to *Macrophomina* sp. and *Fusarium* sp. In addition, we included two phytopathogenic strains—*Neopestalotiopsis rosae*[11] and *Fusarium* sp.-UANL—to further evaluate the biocontrol potential of *Trichoderma harzianum* T9 against a broader set of agriculturally relevant pathogens.

Our results showed that T9 exhibits markedly more potent inhibition of emerging strawberry pathogens, as well as other fungal threats, compared to reference strains such as *T. atroviride* IMI206040 and *T. harzianum* M10[12,13]. Using an evolutionary framework and SNP-based analyses, we identified several genetic variants in T9 associated with secondary metabolism, including notable divergence in the PKS1 gene. These findings prompted a deeper genomic and metabolomic characterization.

To assess whether these genomic differences translated into changes in metabolite production, we performed an untargeted metabolomic analysis. This approach revealed apparent variation in peptaibol profiles—one of the most abundant metabolite families in T9 with known antimicrobial properties—as well as six additional secondary metabolites detectable under standard laboratory conditions.

Our combined functional genomics and metabologenomics approach allowed us to characterize *T. harzianum* T9 in depth and propose it as a promising biocontrol agent capable of thriving under harsh agricultural conditions, including low humidity, high pH, and elevated temperatures—conditions that are increasingly common in modern crop production systems. We further argue that intentional efforts to discover new microbial strains from deserts and other extreme environments, coupled with integrative analyses such as those presented here, can provide valuable alternatives for disease management. In this context, our study serves as an early example of how exploring extreme environments can reveal microorganisms with exceptional potential for sustainable agriculture.

Methods

Trichoderma strains

The *T. harzianum* T9 strain was previously isolated in a prior study[8]. We thawed this strain from glycerol stock stored at −80 °C. The same for *T. harzianum* M10 and *T. atroviride* IMI206040. These two strains were donated by the laboratory of Dr. Louise Glass at UC Berkeley.

Isolation and identification of pathogens

The sampling for the isolation of phytopathgens *Fusarium* sp UG strain. and *Macrophomina* sp. (Supplementary Fig. 1-A) was conducted in the Irapuato region, of the state of Guanajuato, México, to isolate pathogens at coordinates 20.792689 and −101.362693. Strawberry plants of the *Fragaria x ananassa* variety, exhibiting severe wilting, were collected. The plants were sorted to the laboratory, where leaf, crown, and root sections were excised. The tissues were rinsed three times in 50 mL of a 5% sodium hypochlorite solution, followed by three washes with sterile distilled water. The disinfected tissues were then placed on plates containing PDA medium, supplemented with 34 mg/mL chloramphenicol to eliminate bacterial contamination, and incubated for 48 h at 28 °C.

The fungal colonies that grew were isolated by transferring a block from the edge of the colony onto fresh PDA plates. Axenic fungal cultures were then cultivated in PDB medium for 24 h with agitation at 140 rpm and 28 °C. The mycelium was collected by filtration, and the genomic DNA was extracted following the protocol described by Sambrook (2006)[14].

To identify the fungal isolates, we amplified the ITS region by PCR using the primers ITS1 (TCCGTAGGTG AACCTGCGG) and ITS4 (TCCTCCGCTTATTGATATGC)[15]. The resulting PCR fragments were subsequently sent for sequencing at the "Laboratorio de Servicios Genómicos" (LABSERGEN), Langebio–Cinvestav, Irapuato, Mexico. The amplicons were sequenced bidirectionally using the Sanger method, and sequence identities were determined using BLAST searches against the GenBank database, allowing taxonomic assignment based on sequence similarity (Supplementary Fig. 1C).

The *Neopestaloptosis rosae* strain was kindly provided by Dr. Angel Rebollar-Alviter[11]. A second *Fusarium* strain was isolated from stored sorghum seeds and morphologically characterized by Dra. Elva Aréchiga (unpublished data), it will be referred as *Fusarium* sp-UANL strain.

Confrontation assays

Confrontation assays were performed on PDA medium to assess the antagonistic ability of *T. harzianum* T9. All fungal strains were re-streaked on PDA medium and incubated for 48 h at 28 °C before evaluation. After incubation, 5 mm diameter plugs were excised from the colony edges and placed at the center of fresh PDA plates, followed by another 48-hour incubation.

We observed the colonial growth of each phytopathogenic fungal strain every 24 h, to have a comparison of its growth rate (Figure supplementary 1-D). For the antagonism assay, 5-mm plugs of actively growing mycelium were 1 cm from the plate edges on the PDA medium. In the case of *Fusarium* strains, the plugs were inoculated 48 h before *Trichoderma* inoculation, as *Fusarium* strains grow more slowly than *Trichoderma* strains. In contrast, *Macrophomina sp* and *N. rosae* strains were inoculated simultaneously with *Trichoderma* strains.

To analyze the antagonistic ability under alkaline conditions, the PDA medium was adjusted at pH 8.5 using 15 mM *4-(2-hydroxyethyl)−1-piperazineethanesulfonic acid* (HEPES). Same media was used for growth assays under alkaline conditions. In all cases two biological replicates were performed, each with three technical replicates. Each biological replica consisted of independent experiments, performed on different days form independent starter cultures. Plates were photographed every 24 h, and colony areas were measured using ImageJ software. Data were analyzed using a two-way ANOVA (ordinary model, $\alpha = 0.05$). Post hoc Tukey's multiple comparison tests were performed to assess pairwise differences between treatments at each time point. Differences were considered significant when $p < 0.05$. Statistical analyses were conducted using GraphPad software, version 8[16].

Genome sequencing and assembly

The genome of *T. harzianum* T9 was sequenced using the Illumina Novaseq platform (Illumina Inc., San Diego, CA, USA). Fragment libraries were prepared using the TruSeq Nano DNA preparation kit (Illumina Inc., San Diego, CA, USA) with paired end reads of 150 bp. Quality control assessment was performed using FastQC v0.11.5[17], followed by adapter and short sequence removal using Trimmomatic v0.39[18]. The resulting clean reads were used for *de novo* assembly using SPAdes v.3.15.2 with *k*-mer values ranging from 21 to 127[19]. To eliminate potential contaminant sequences, FCS-GX v0.5.5, a tool from the NCBI Foreign Contamination Screen (FCS) suite[20], was used. The assembly statistics were determined using QUAST[21], yielding a genome size of 39.85 Mb. Analysis with Benchmarking Universal Single-Copy Orthologs (BUSCOs) revealed genome completeness of 99.3%, according to the fungi lineage[22]. All these assembly statistics, along with several additional metrics, are presented in Table S2.

Structural and functional annotation

Prior to annotation, contigs with less than 500 bp and duplicated were filtered out, subsequently, the remaining contigs were sorted and renamed using a toolkit via funannotate (https://funannotate.readthedocs.io/en/latest/install.html). Transposable elements (TEs) were identified using Earl Grey TE annotation pipeline (v4.3.0)[23], with the Ascomycota library from Dfam release 3.7[24]. The generated soft-mask genome was used for annotation with BRAKER2[25], using as evidence a fungal-specific protein partition, obtained from OrthoDB v11[26]. The annotation stats were generated with JCVI toolkit[27] via Galaxy Europe Server (The Galaxy Community, 2024). Functional annotation was carried out with eggNOG Mapper v4.5[28] and InterProScan 5[29].

CAZymes identification

Carbohydrate-active enzymes (CAZymes) were identified by using the dbCAN3 server (https://bcb.unl.edu/dbCAN2/blast.php)[30]. First, only the longest protein isoform sequences obtained by BRAKER2 were selected for the analysis. Afterward, the protein file was submitted to dbCAN3 under the following configuration: protein sequence mode, HMMER (E-value < 1e-15, coverage > 0.35), DIAMOND (E-value < 1e-102), and dbCAN_sub (E-value < 1e-15, coverage > 0.35). To assure the accuracy of enzyme identification, the sequences that fulfilled the selected analysis criteria by HMMER, DIAMOND, and dbCAN_sub were chosen to determine the CAZymes classification.

Phylogenetic analysis

Phylogenetic reconstruction was determined by using 29 fungal genomes from public databases[31–46] and the *T. harzianum* T9 genome. The webpage links to the genomes that were obtained from open databases can be consulted in the Supplementary Table 1. All genomes were annotated *de novo* by using BRAKER2[25]. Proteomes of the following strains: *Fusarium graminearum*, *Fusarium oxysporum*, *T. harzianum* T9, *T. harzianum* CBS, *T. harzianum* M10, *T. harzianum* T22, *T. atroviride* IMI206040 and *Trichoderma hamatum* FBL_587 were then filtered to retain only the longest isoform from each gene. These single-copy sequences were used to construct the phylogenetic tree through OrthoVenn3[47]. Single-copy sequences were aligned with Muscle[48], and conserved regions were obtained with TrimAL[49]. A phylogenetic tree was inferred by the Maximum Likelihood using FastTree[50] and the JTT + CAT model. *Fusarium* species were added to the analysis as an external group. A Newick format was obtained, and it was further used as a reference for the ROADIES (Reference-free Orthology-free Annotation-free DIscordance-aware Estimation of Species tree) software[51], which was executed with the 30 genomes in an accurate and convergence mechanism mode to obtain a solid, well-defined roots phylogenetic tree. In RODIES software a GENE_COUNT of 2000, IDENTITY of 65%, and COVERAGE of 85% was established to optimize the analysis for fungal genomes.

Sequence alignment and genetic variant identification

Clean reads were aligned against the reference genomes of *T. harzianum* strains M10, T22, Th0179, Th3844, TR274, *T. afroharzianum* Th6 and *T. atroviride* IMI206040 using the BWA-MEM algorithm[52] via Bash command-line interface. SAMtools and BCFtools[53] were employed to sort, index, and identify SNPs in the alignment files. Variant filtering was subsequently performed using VCFtools[54] to generate variant position files for each comparison between *T. harzianum* T9 and the reference genomes. To assess gene-level variability, the number of SNPs per gene was normalized by gene length or relative abundance using R. Statistical normalization was performed with boxplot analysis, using *T. atroviride* IMI206040 as the control to establish standard variation thresholds and remove bias. Genes with a percentage of variation exceeding the interquartile range were classified

as highly divergent. Orthologous relationships between T9 and other *harzianum* clade strains were determined with *OrthoFinder*[55], based on protein sequence comparisons from each strain's genome annotation, enabling the identification of shared genes and the assessment of whether T9 variants occurred in highly conserved genes. Functional annotation of orthogroups was performed with *eggNOG-mapper*[28], to interpret the potential biological impact of detected variants. Finally, divergent genes were analyzed within their evolutionary and functional context, evaluating whether they belonged to conserved orthogroups under selective pressure and identifying functional patterns linked to genetic variability.

Metabolite extraction

After 4 days of cultivation in Petri dishes with PDA medium, the strains were sectioned into 6 mm diameter plugs. Five of these plugs were added to vials. In each vial, 0.6 mL of the solvent mixture MeOH: CH_2Cl_2:EtOAc (1:2:3, v/v/v) + 0.1% formic acid was added. The vials were placed in an ultrasonic bath for 60 min. The supernatant was transferred to a 2 mL microtube and evaporated to dryness using an Eppendorf Concentrator®. The samples were then resuspended in 1 mL of MeOH (Optima™ LC/MS Grade, Fisher Chemical™) centrifuged at 15,000 rpm for 10 min, and transferred to 1.5 mL glass vials for LC-MS/MS analysis.

LC-MS/MS analysis

Untargeted LC-MS/MS metabolomics analyses of the extracts were performed on a Vanquish Duo UHPLC binary system (Thermo Fisher Scientific, Waltham, MA, USA) coupled to an IDX-Orbitrap Mass Spectrometer (Thermo Fisher Scientific). The chromatographic separation of the analytes was achieved using two different methods:

1. Waters ACQUITY BEH C18 (10 cm × 2.1 mm, 1.7 µm) (Waters™, Milford, MA, USA) column equipped with an ACQUITY BEH C18 guard column kept at 40 °C. The mobile phase consisted of MilliQ water + 0.1% formic acid (v/v) (A) and acetonitrile + 0.1% formic acid (v/v) (B) (both sourced from HiPerSolv CHROMANORM®, HPLC and LC-MS grade, VWR Chemicals BDH®). The mobile phase gradient composition was as follows: 0–0.8 min 2% B, 0.8–3.3 min 2% to 5% B, 3.3–10 min 5% to 100% B, 10–11 min 100% B. The column was then re-equilibrated at 2% B for 2.7 min. Flow rate was set at 0.35 mL/min. The injection volume was 1µL.
2. Waters ACQUITY BEH C18 (5 cm × 2.1 mm, 1.7 µm) column (Waters™, Milford, MA, USA), equipped with an ACQUITY BEH C18 guard column kept at 70 °C. Mobile phase consisted of H_2O + 10 mM (v/v) ammonium acetate (adjusted to pH 9.2 with ammonium hydroxide) as eluent A and acetonitrile + 0.1% formic acid (v/v) as eluent B. Gradient elution was applied at a flow rate of 0.5 mL min^{-1} according to the following: 0–0.8 min 40% B, 0.8–6.5 min 40% to 100% B, 6.5–8.5 min 100% B. The column was then re-equilibrated at 40% B for 1.5 min. The injection volume was 1 uL.

The MS measurements were performed using heated electrospray ionization (HESI) mode in positive and negative ion mode in the first method. The second method utilized heated electrospray ionization (HESI) in positive ion mode only. Both methods used a voltage of 3500 V in positive mode and 2500 V in negative mode, acquiring in full MS/MS spectra (Data dependent Acquisition-driven MS/MS, DDA) in the mass range of 70–1000 Da in the first method and 100–1500 Da in the second method. The mass resolution was set to 120,000 for full scan MS and 30,000 for MS/MS events. Precursor ions were fragmented by stepped High-energy Collision Dissociation (HCD) using collision energies of 20, 40, and 55. The automatic gain control (AGC) target value set at 4×10^5 for the full MS and 5×10^4 for the MS/MS spectral acquisition.

LC-MS/MS data analysis

For MS/MS dereplication MS/MS data were converted to mzXML format using the MS-Convert software, which is part of ProteoWizard (Palo Alto, CA, USA). Feature detection was performed with MZmine3 version 3.3.0 and 3.6.0[56]. Default values were used on the Processing wizard and MoNA spectral library was added for compound annotation. The resulting feature table (.csv) and MS/MS spectra files (.mgf) were exported and used for the FBMN analysis on GNPS. For FBMN analysis the GNPS Super-Quick Start Interface was used and the resulting feature table (.csv), MS/MS spectra files (.mgf) exported from MZmine3.

Comparative genomic insights into secondary metabolite biosynthetic gene clusters

To assess the presence and evolutionary relationships of the BGC encoding peptaibols, across different *Trichoderma* species, we performed a phylogenomic analysis using a curated panel of *Trichoderma* genomes. For this, we employed an optimized version of the CORASON pipeline[57], specifically adapted for fungal genomes and BGC identification (fungison, available at https://github.com/WeMakeMolecules/fungison).

The CORASON workflow integrates synteny-based gene cluster comparisons with phylogenetic reconstruction of core biosynthetic genes, allowing the identification of orthologous clusters across multiple genomes. Using this approach, we linked BGC sequences from each *Trichoderma* genome to phylogenetic trees of key biosynthetic enzymes, enabling us to infer the evolutionary relationships and potential conservation of these clusters across species.

This methodology provided insights into the distribution and evolutionary trajectories of peptaibol biosynthetic pathways in *Trichoderma*, shedding light on their diversification within the genus.

Results

Biocontrol efficiency of the *T. harzianum* T9 strain

To analyze the antagonistic capacity of strain *T. harzianum* T9, as a characteristic related to its production of metabolites with antimicrobial activity, we selected phytopathogenic fungi belonging to genera reported to have negative effects on agricultural crops, to carry out direct confrontations in plate. We use *N. rosae*, as one of the most important emerging pathogens in agriculture, associated with strawberry plants[11]. Two of the strains were isolated from diseased strawberry plants and were identified as *Macrophomina sp*., with a 99.8% similarity and 100% overlap, and *Fusarium sp*. (designated as *Fusarium* UG), with a 99.58% similarity and 100% overlap. We used the mix of both fungal strains to co-cultivate with strawberry seedlings, observing disease symptoms in roots and crown mainly (see Supplementary Fig. 1-A to -C).

The sequences corresponding to *Fusarium sp*. and *Macrophomina sp*. were deposited in the NCBI database, with the identifiers GenBank: PX512993 and GenBank: PX513168, respectively. The genus *Fusarium* is widely known as a plant pathogen, with various specialized forms. Meanwhile, the genus *Macrophomina* is reported as an emerging pathogen of many plant species, exhibiting high persistence in soil[58]. We included a second strain of *Fusarium* (UANL strain), as a potential pathogen of stored grains, since the strain was isolated from packaged sorghum seeds.

Co-culture experiments revealed that *T. harzianum* T9 has better antagonistic activity than the well-known biocontrol *Trichoderma* strains, *T. atroviride* IMI 206,040 and *T. harzianum* M10, which showed limited efficacy against the pathogens tested. The *T. harzianum* T9 strain significantly inhibited the growth of the four pathogens tested after 120 h (Fig. 1A). Notably, exhibited significant biocontrol activity as early as 72 h (Fig. 1D,E). Furthermore, observations on the backside of petri dishes showed that *T. harzianum* T9 produced an orange-brown pigment, which appeared to invade and dominate the area occupied by the pathogenic fungi (Fig. 1B). This strain was particularly effective against the aggressive pathogens *Macrophomina sp* and *N. rosae* (Fig. 1C).

T. harzianum T9 demonstrates high biocontrol efficiency under extreme alkaline conditions

In a previous study, we demonstrated that *T. harzianum* T9 is highly resistant to salinity, either low or high pH levels, and capable of promoting growth in sorghum plants[8]. This adaptability is likely due to its origin from an arid region, which may have driven its resilience. However, high stress tolerance might also provide an advantage when competing with pathogens, which often secrete compounds that either acidify or alkalize their environment. It is likely that *T. harzianum* T9 is a strong competitor against such pathogens even under extreme conditions.

T. harzianum T9 grew effectively in alkaline media (pH 8.5), while the growth of the other two *Trichoderma* and pathogenic strains was significantly impaired at the same time point (Supplementary Fig. 2). Interestingly, in the presence of a pH indicator (litmus, blue color), *T. atroviride* acidified the medium (turning it yellow), whereas *T. harzianum* T9 did not, despite its ability to grow (Supplementary Fig. 2).

T. harzianum T9 showed, at a pH of 8.5, high effectiveness in controlling *Macrophomina*, unlike the other two *Trichoderma* strains (Fig. 2A). This observation was consistent across two independent biological replicates, and the reduction in the diameter of *Macrophomina* was statistically significant by Tukey-HSD test ($p < 0.05$) when competing against *T. harzianum* T9 (Fig. 2B). Notably, at this pH, *T. harzianum* T9 was able to overgrow the pathogenic strain, with a distinct necrotic area visible after 96 and 120 h on the backside of the petri dishes. We observed the same in *N. rosae* at this pH (Fig. 2C). The graphs show significant degree of growth inhibition exerted by the *T. harzianum* T9 strain compared to that of the strains *T. atroviride* IMI 206,040 and *T. harzianum* M10 (Fig. 2D).

Assembly and prediction of genome elements

To delve into the particularities of the *T. harzianum* T9 strain and try to correlate part of its physiology with its genetic characteristics, we decided to get its genome to look for traits linked to its activity. Therefore, we sequenced, assembled and annotated the genome. First, we used the illumina reads to obtain a genome profile using tools Jellyfish version 2.2.10[59] and GenomeScope version 1[60] to analyze k-mer frequencies from the genome reads. This analysis (Fig. 3A), allowed us to estimate a genome size of 39.85 Mb, which was consistent with the genome assembly obtained in 134 contigs using SPADES. This genome has a GC content of 48.24%, an N50 of 903,272 (Supplementary Table 2), and an estimated completeness of 99.3%. These results confirm the high quality of the assembly (Fig. 3B). The genome size and GC content of the *T. harzianum* T9 assembly were comparable to those reported for other *T. harzianum* strains sequenced using both short- and long-read technologies[61–63]. This suggests that the primary genomic characteristics of the species remain consistent.

Additionally, the content TEs was analyzed using the Earl Grey tool[23]. This analysis revealed the presence of various types of TEs, including DNA transposons, LINEs (Long Interspersed Nuclear Elements), LTRs (Long Terminal Repeat retrotransposons), and other repetitive elements such as simple repeats, microsatellites, and RNA (Fig. 3C). LTR retrotransposons were the most prevalent type of TEs, representing 2.30% of the genome (377 elements). In contrast, the largest number of individual elements belonged to the "Other" category, which included simple repeats, RNA, and microsatellites, accounting for 9851 elements and 1.07% of the genome. DNA transposons and LINE elements were less abundant, representing 0.18% and 0.45% of the genome, respectively. Overall, TEs constituted approximately 5.56% of the genome, distributed across 2661 distinct subcategories (Supplementary Table 3).

Genome annotation

Gene prediction in the assembled genome was performed using BRAKER2, resulting in 12,058 predicted genes, with an average gene size of 1,668 bp and a median gene size of 1,393 bp. The average exon size was 527 bp, and the median exon size was 284 bp (Fig. 4A,C,D). These results are comparable to other *Trichoderma*

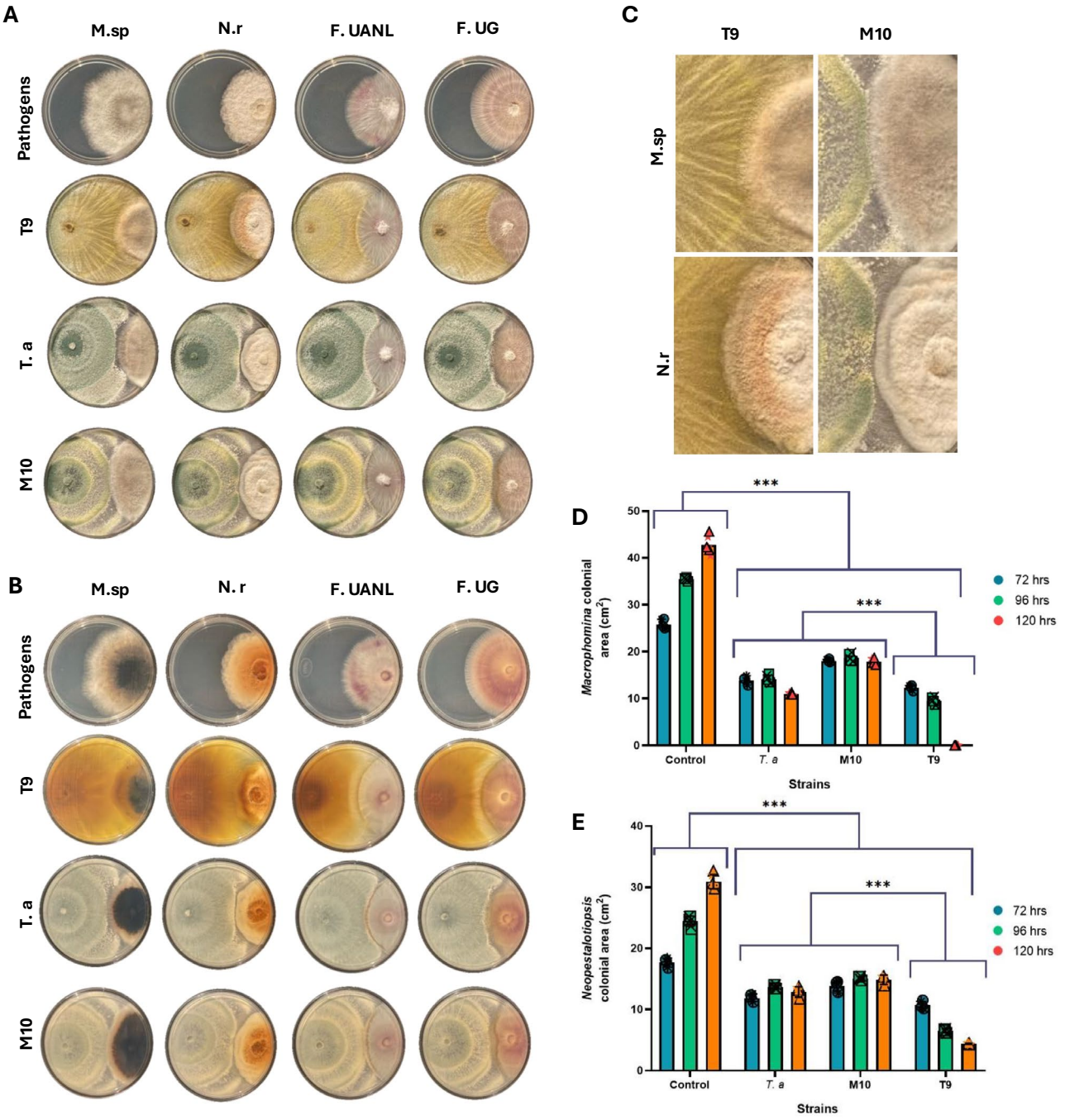

Fig. 1. *T. harzianum* T9 demonstrates enhanced biocontrol efficacy against phytopathogenic fungi. (**A**) Top and (**B**) bottom view of plates showing interactions between the phytopathogens *Macrophomina* sp. (*M.sp*), *N. rosae* (*N.r*), *Fusarium* sp. (*F.sp*), and *Fusarium UG* (*F.UG*) with the *Trichoderma* strains *T. harzianum* T9, *T. harzianum* M10, and *T. atroviride* IMI206040 (*T.a*). Phytopathogens are positioned on the right side, while *Trichoderma* strains are on the left side of the plate. The photographs were taken after 96 h of interaction. (**C**) Close-up of the interaction zone between *M.sp* or *N. r* with the *Trichoderma* strains T9 and M10. (**D**) Average colony area of *M.sp* and (**E**) *N.r* in interaction with different *Trichoderma* strains. The control represents the average colony size of each pathogen growing alone. Asterisks indicate statistical differences according to the Tukey-HSD test at a significance level of $\alpha < 0.05$, with $^{*}p < 0.05$. $n = 3$.

genome annotations, where gene counts and BUSCO completeness statistics generally fall within a similar range, further supporting the robustness of the annotation. The quality of the annotation was assessed using BUSCO in transcriptome mode. To accomplish this, the transcripts of the genes predicted by BRAKER2 were extracted using GFFread v0.12.7[64] and subjected to analysis. BUSCO results showed that 100% of the predicted gene transcripts were complete, with 96% identified as single-copy genes. No fragmented or missing transcripts were

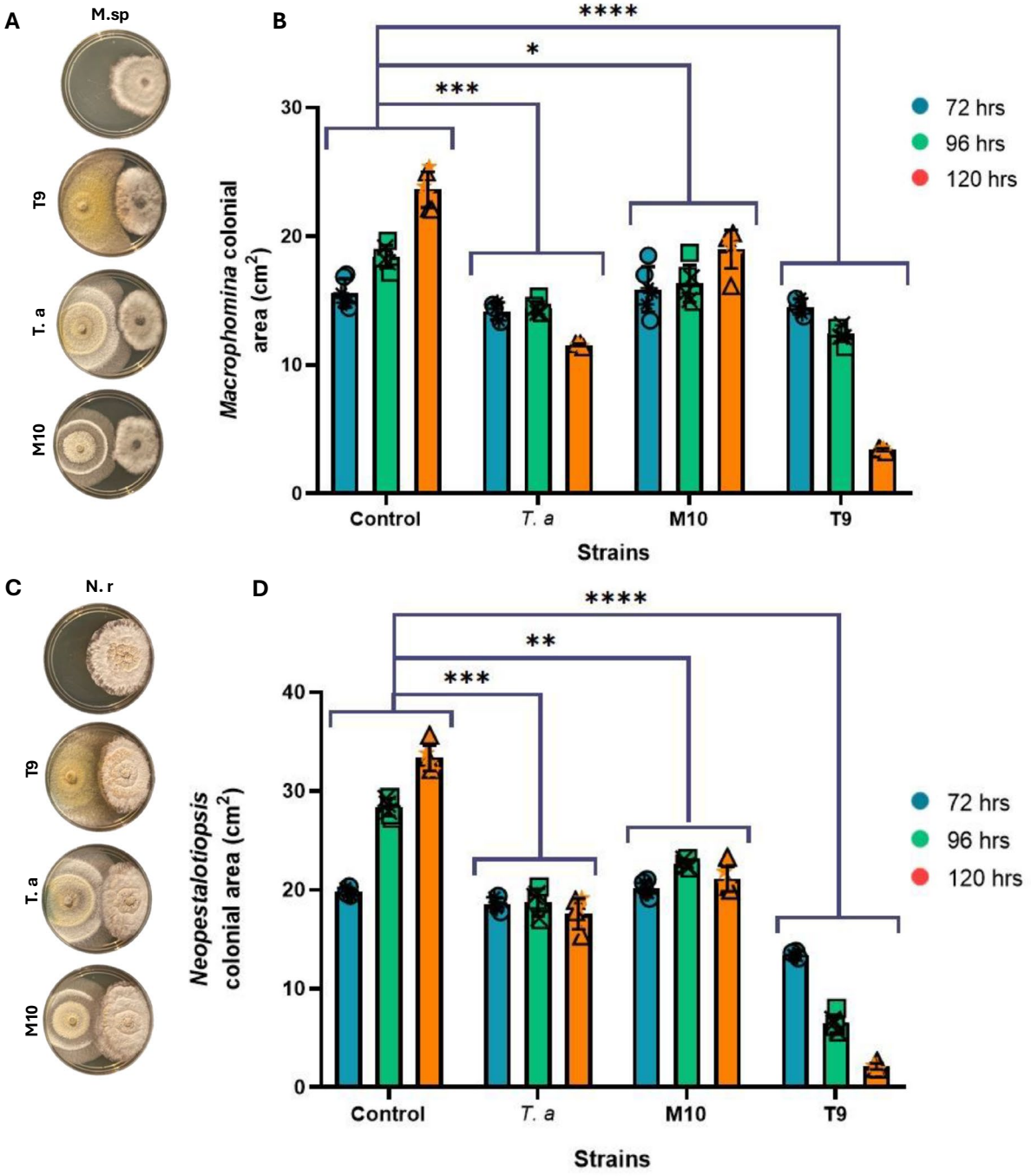

Fig. 2. *T. harzianum* T9 strain maintains its high antagonistic capacity under extreme alkaline conditions. (**A**) Top view of plates showing interactions between *T. harzianum* T9, *T. harzianum* M10, and *T. atroviride* IMI206040 (*T.a*) with *Macrophomina* sp. (*M.sp*) or (**C**) *N. rosae* (*N.r*), cultivated on PDA medium at pH 8.5. Phytopathogens are showed on the right side, while *Trichoderma* strains are on the left. Photographs were taken after 96 h of interaction. (**B**) Average colony area of *M.sp* and (**D**) *N.r* in interaction with different *Trichoderma* strains. The control represents the average colony size of each pathogen growing alone. Asterisks indicate statistical differences according to the Tukey-HSD test at a significance level of $\alpha < 0.05$, with $^{*}p < 0.05$. $n = 3$.

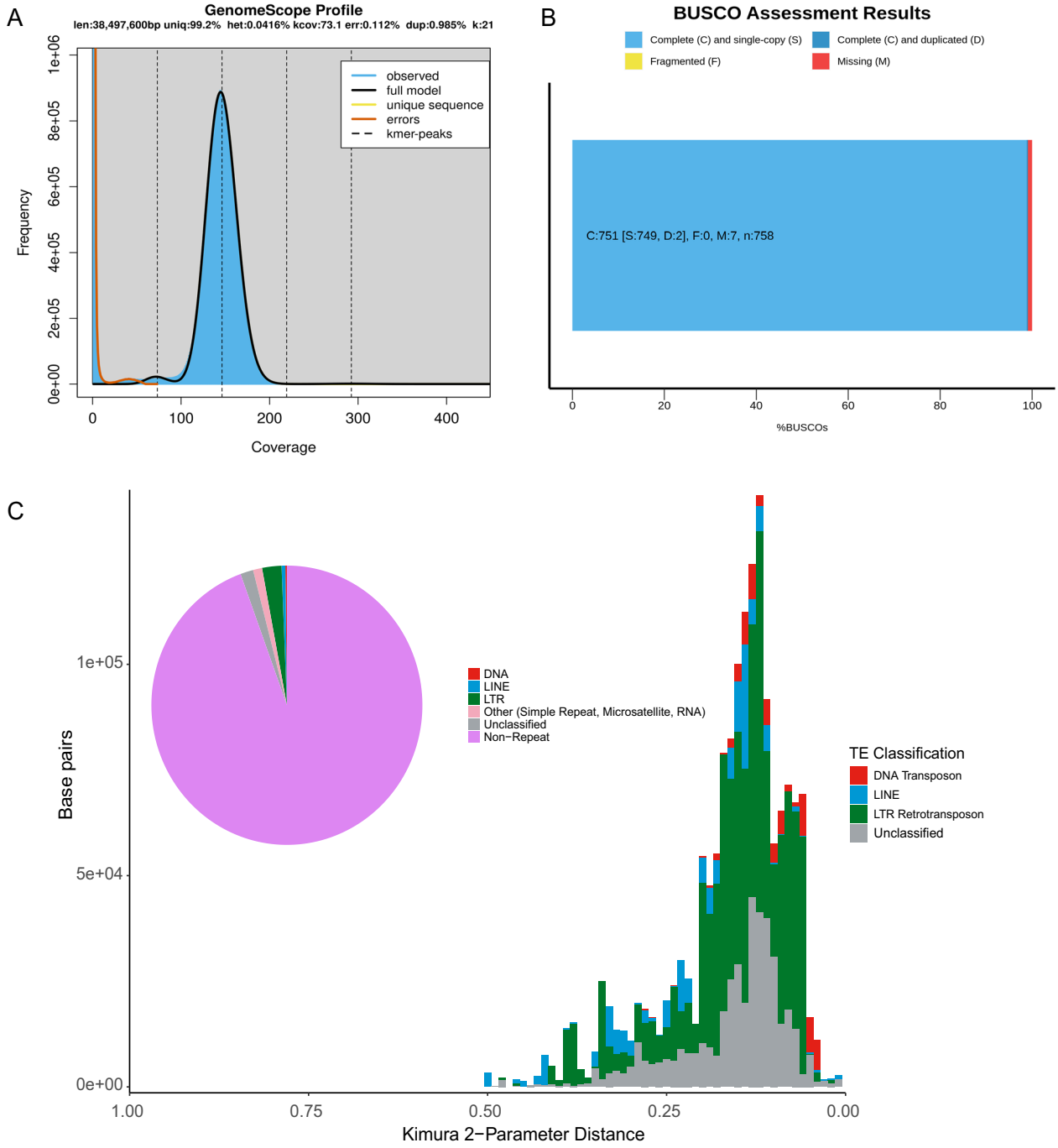

Fig. 3. Genomic assembly of *T. harzianum* T9. (**A**) K-mer spectra analysis using Jellyfish with k = 21 for k-mer counting, and GenomeScope 1.0 for fitting models to estimate genome size, heterozygosity, and repetitiveness. (**B**) Assessment of assembly completeness using the BUSCO (Benchmarking Universal Single-Copy Orthologs) program with the fungi dataset. (**C**) Prediction and quantification of low complexity regions and repetitive elements in the genome.

detected, indicating that the structural annotation of the genome is of high quality (Fig. 4B). This indicates that we have an excellent assembly, covering the largest possible number of coding regions, which provides strong support for our future functional conclusions using this genome.

CAZymes

Carbohydrate-active enzymes (CAZymes) mediate the breakdown of plant cell walls, facilitating fungal interaction with the environment. In *Trichoderma*, they play a key role in mycoparasitism and organic matter degradation. Therefore, when analyzing a new *Trichoderma* genome, it is essential to predict how many and what types of CAZymes are encoded, as this reveals its ecological and biotechnological potential.

A

Characteristics	Number
Max number of transcripts per gene	5
Mean exon size	527.9
Mean gene locus size (first to last exon)	1668.6
Mean number of distinct exons per gene	3.0
Mean number of transcripts per gene	1.1
Mean transcript size (UTR, CDS)	1532.6
Number of distinct exons	35,896
Number of genes	12,058
Number of genes with alternative transcript variants	957 (7.9%)
Number of multi-exon genes	9,324 (77.3%)
Number of predicted transcripts	13,147
Number of single-exon genes	2,734 (22.7%)
Average size of exon (bp)	527
Median size of exon (bp)	284
Average size of gene (bp)	1,668
Median size of gene (bp)	1,393

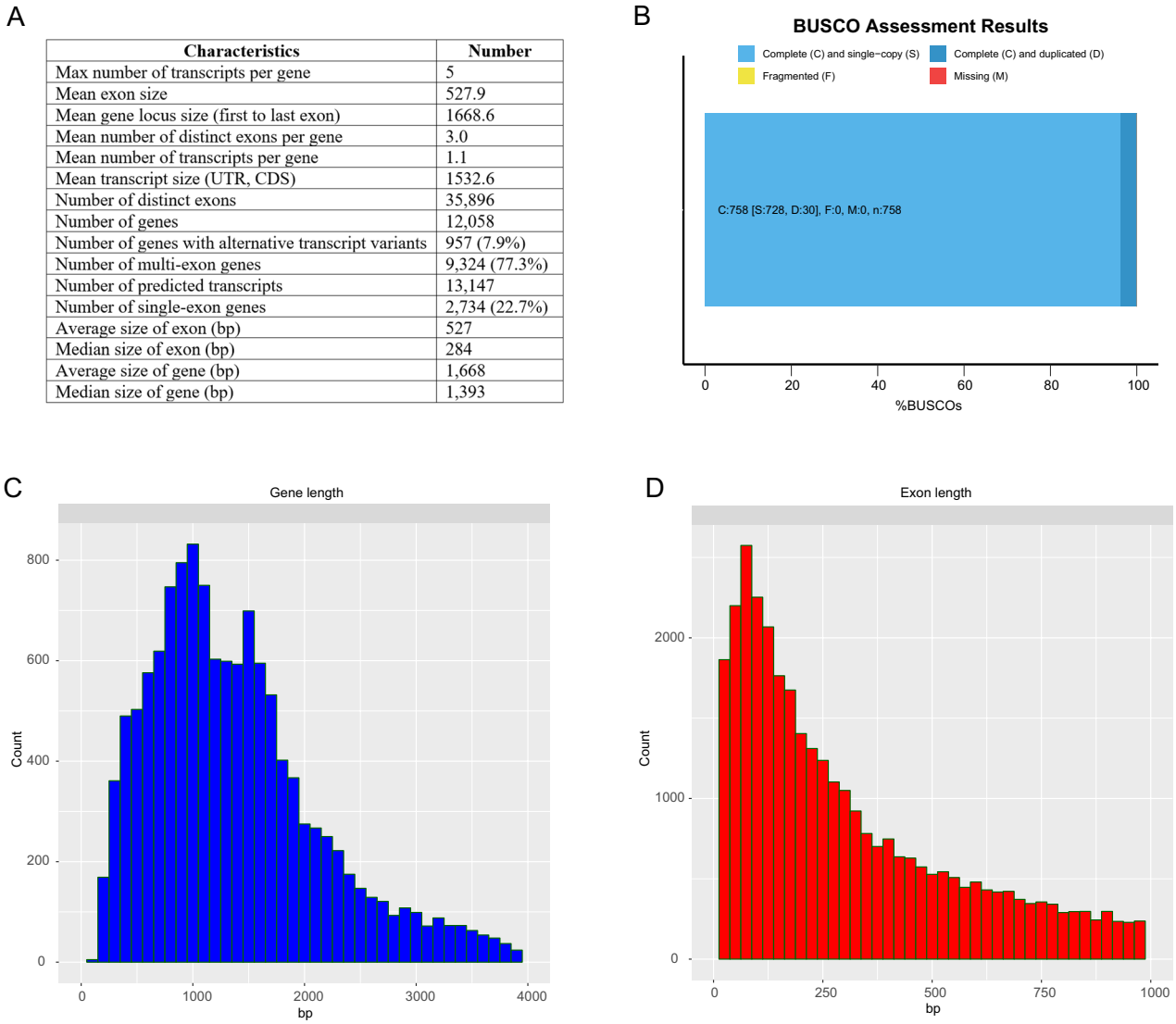

Fig. 4. Structural annotation of *Trichoderma harzianum* T9 genome. (**A**) Table of genome annotation statistics. (**B**) Completeness of the annotation using BUSCO (Benchmarking Universal Single-Copy Orthologs) with the fungi dataset, based on transcripts extracted with GFFread. (**C**) Gene length distribution. (**D**) Exon length distribution.

The analysis of CAZymes, classified a total of 424 proteins as CAZymes, representing 3.5% of the 12,058 proteins used as input (Fig. S4A). Among these, 230 GH (glycoside hydrolases), 86 GT (glycosyltransferases), 51 AA (auxiliary activities), 20 CE (carbohydrate esterases), 8 PL (polysaccharide lyases), 1 CBM (carbohydrate-binding module), and 28 GH + CBM were identified. Out of the 424 CAZymes, 197 (46.5%) contained a signal peptide, suggesting that these proteins might be secreted (Fig. S4B; Supplementary Dataset 1).

Phylogenetic analysis

To move beyond the traditional descriptions based on phylogenies inferred from single molecular markers such as ITS, we generated a phylogenomic tree using all orthologous genes shared among 30 full-length genomes from other *Hypocreales* species. Out of the 30 genomes analyzed, 28 correspond to *Trichoderma* species, while the remaining two belong to *Fusarium* species, which were included as outgroups. All genomes underwent *de novo* annotation using BRAKER2. Prior to the phylogenetic analysis, genome annotations and quality were assessed using BUSCO in both genome and transcriptome modes. The BUSCO results demonstrated high-quality structural and functional annotations for most of the genomes (Supplementary Fig. 3).

The phylogenomic tree inference was performed with ROADIES[50] revealing robust clustering of the *Trichoderma* genus, supported by high bootstrap values, indicating strong evolutionary relationships among fungal species (Fig. 5A). The phylogenetic tree showed well-defined clades, such as the group containing *Trichoderma hamatum*, *Trichoderma asperellum*, and *T. atroviride*, which displayed notable divergence from other genus members, suggesting specific evolutionary adaptations. Additionally, *Trichoderma virens* clustered

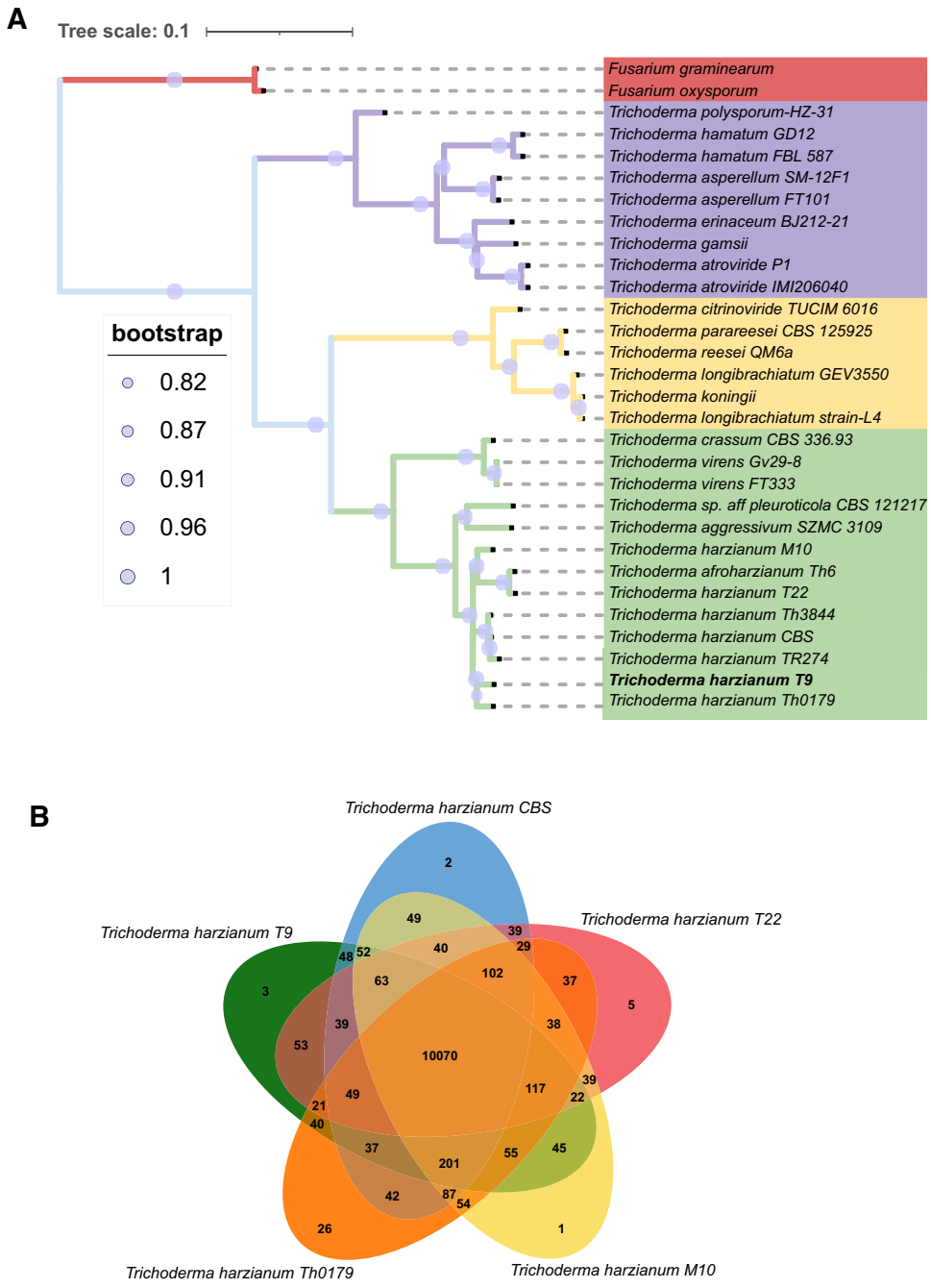

Fig. 5. Phylogenetic analysis and orthologous clusters analysis in *Trichoderma* species and annotation of biosynthetic gene clusters (BGCs). (**A**) Rooted phylogenetic tree of 30 fungal species generated using OrthoVenn3 and ROADIES. *Fusarium oxysporum* and *Fusarium graminearum* were used as outgroup species in the plot. (**B**) Venn diagram showing unique and shared orthologous clusters families among five *Trichoderma* species, including the sequenced *T. harzianum* T9, generated using OrthoVenn3.

with *Trichoderma aggressivum*, suggesting shared phenotypic and ecological traits. The clade comprising multiple *T. harzianum* strains included the T9 strain, which was most closely related to *T. harzianum* Th0179, while remaining within the *harzianum* group. This indicates a close evolutionary relationship among these strains. Our analysis clearly indicates that T9 belongs to *T. harzianum*.

Additionally, we identified clusters of orthologous genes shared among *T. harzianum* strains (T9, *M10*, T22, Th0179 and CBS). This analysis revealed a conserved core of 10,208 orthologous genes shared among all five strains, suggesting a highly conserved genetic basis (Fig. 5B).

Secondary metabolism genes show a high number of nucleotide changes

Since the *T. harzianum* T9 strain shares many orthologs with other *T. harzianum* strains and containsrelatively few strain-specific genes (Fig. 6A), we hypothesized that most differences occur at the nucleotide level in gene variants. For this reason, we performed a variant analysis, comparing all orthologous genes between strain T9 and *T. harzianum* strains Th6, M10, T22, Th0179, Th3844, and TR274, as well as *T. atroviride* IMI206040 as a control due to its evolutionary distance, as shown in the phylogenetic tree in Fig. 6A, where *T. atroviride* forms a separate clade from *T. harzianum*.

Our initial SNP analysis revealed that the comparison between *T. harzianum* T9 and *T. atroviride* displayed a very broad variation, indicating that almost all genes exhibit changes, an expected result since they are different species. In contrast, comparisons between *T. harzianum* T9 and the other *T. harzianum* strains showed a

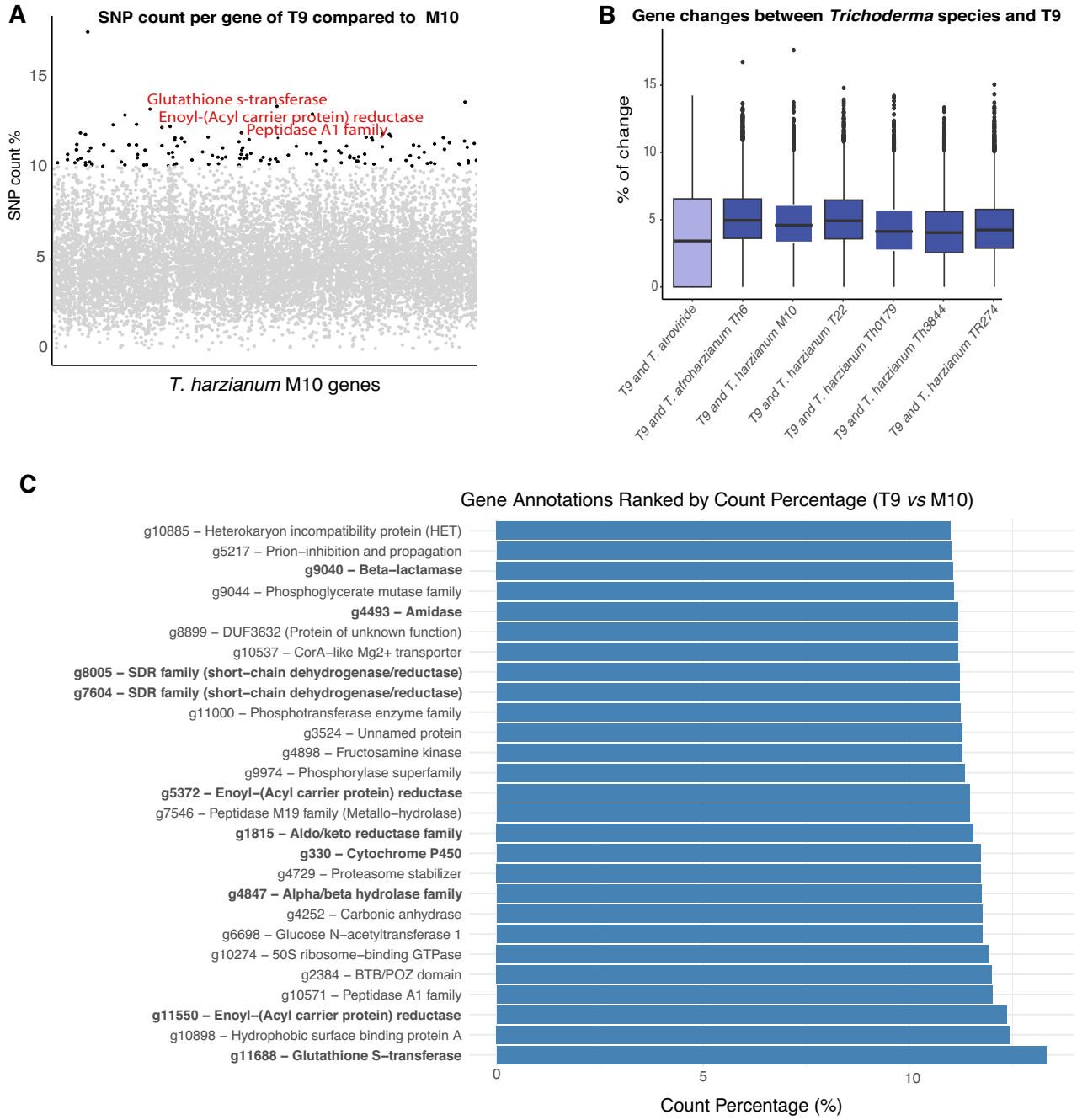

Fig. 6. Analysis of nucleotide variation in T9 genes compared with other *T. harzianum* strains. (**A**) Boxplot showing the percentage of SNPs in T9 orthologous genes compared with *T. harzianum* strains Th6, M10, T22, Th0179, Th3844, and TR274, using *T. atroviride* IMI206040 as a control. (**B**) Scatterplot displaying the number of SNPs across orthologous genes in the comparison between T9 and the M10 strain. (**C**) Annotation of genes with the highest SNP percentages (> 11%) in the comparison of T9 versus M10.

narrower distribution, with an average of ~5% variation per gene (Fig. 6A). However, some genes appeared as outliers, showing more than 10% nucleotide changes across their coding sequences.

We therefore focused on differences between *T. harzianum* T9 and the *T. harzianum* M10 genome, a *T. harzianum* strain that we previously used in competition assays against phytopathogenic fungi. In our initial inspection, certain genes clearly displayed a substantial number of SNPs throughout their coding regions (Fig. 6B). While some of these genes lacked functional annotation, it was noteworthy that several of the most divergent genes were involved in secondary metabolism, such as Glutathione S-transferase[65], which showed 13% variation, and a T1PKS[66], with 12% variation.

Further annotation of the genes with the highest number of changes revealed the presence of enzymes relevant to secondary metabolite biosynthesis and modification (Fig. 6C), including a cytochrome P450, an enoyl reductase, which participates in both polyunsaturated fatty acid (PUFA) and polyketide biosynthesis in bacteria and fungi[67,68], two short-chain dehydrogenase/reductases, typically components of large multi-domain enzymes such as mammalian fatty acid synthases or bacterial polyketide synthases[69], as well as an amidase, a beta-lactamase, an aldo/keto reductase, and an alpha/beta hydrolase.

The high density of variants observed in the coding regions of these genes suggests a complex evolutionary scenario in which most changes may be selectively neutral, but a subset could be under positive selection, driving adaptive modifications. Given their association with secondary metabolism and the enhanced biocontrol ability of strain *T. harzianum* T9, these variants may contribute to increased production or diversification of bioactive metabolites, potentially leading to chemical innovations that improve competitive interactions with other organisms. This finding underscores the need for an in-depth functional characterization to identify which of these changes have adaptive value.

BGC diversity and distribution analysis

To investigate whether *T. harzianum* T9 harbors novel biosynthetic pathways for metabolites that could contribute to the control of phytopathogenic fungi, we annotated the presence of 76 BGCs in its genome, using AntiSMASH (Fig. 7). We then examined whether any of the genes with high nucleotide variation rates were part of the identified BGCs. As shown in Table 1, we found 16 genes with more than 10% nucleotide variation that are located within BGCs, including gene g11550, which encodes the core enzyme of cluster 8.2, a polyketide synthase (PKS) for which no associated metabolite has yet been identified, and gene g2166, the core of cluster 14.2, a nonribosomal peptide synthetase (NRPS) potentially involved in the biosynthesis of dichlorodiaporthin. In the table, genes highlighted in bold indicate the central core genes of the BGCs, as well as the compounds identified by antiSMASH.

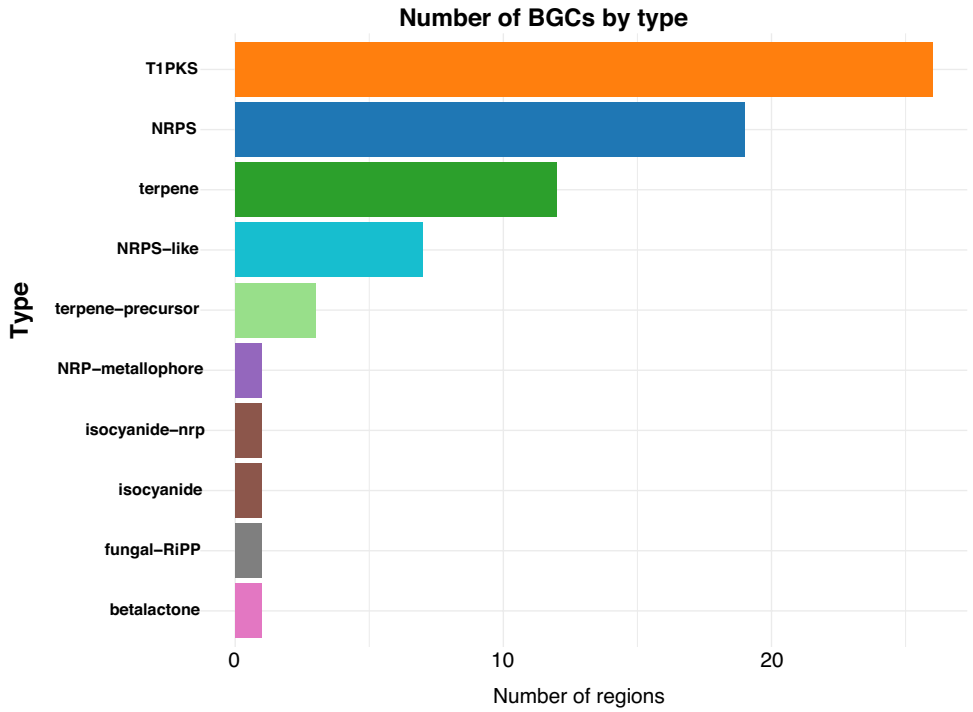

Fig. 7. Number of predicted BGCs by family in the genome of *T. harzianum* T9. This prediction was performed using fungiSMASH (the fungal analysis version of AntiSMASH). Ten cluster families were identified, with T1PKS, NRPS, and terpene being the three most abundant.

gene_id	Percentage	BGC	Description	FungiSMASH	Compounds
g7952	13.713	4.2	Hypotetical protein	Terpene	**HEx-tc1 terpene**
g11550	**12.36**	**8.2**	**T1PKS-core**	**T1PKS**	Unidentified
g7546	11.466	37.2	Belongs to the metallo-dependent hydrolases superfamily. Peptidase M19 family	NRPS	**Verticillin**
g3049	10.304	18.1	Hypotetical protein	NRPS-like	Unidentified
g3053	11.4	18.1	Hypotetical protein	NRPS-like	Unidentified
g8005	11.212	4.3	Belongs to the short-chain dehydrogenases reductases (SDR) family	T1PKS	**Trichoxide**
g9221	11.096	5.1	Hypotetical protein	NRPS, T1PKS	Unidentified
g7638	10.928	38.1	Calcineurin-like phosphoesterase	Terpene	Unidentified
g9001	10.868	47.2	Protein kinase domain	T1PKS	Unidentified
g9011	10.187	47.2	Protein of unknown function (DUF3645)	T1PKS	Unidentified
g4667	10.822	22.1	Serine hydrolase (FSH1)	T1PKS	Unidentified
g4671	10.522	22.1	Hypotetical protein	T1PKS	Unidentified
g9073	10.522	48.1	Acetyltransferase (GNAT) domain	NRPS, T1PKS	**Fujikurin A**, **B**,**C**, **D**
g7205	10.485	34.2	Carbon-nitrogen hydrolase	NRPS, T1PKS	unidentified
g2166	**10.262**	**14.2**	**NRPS-core**	**NRPS, T1PKS**	**Dichlorodiaporthin**
g4025	10.234	2.4	GAL4-like Zn(II)2Cys6 (or C6 zinc) binuclear cluster DNA-binding domain	Terpene-precursor	Unidentified

Table 1. Genes with the highest number of nucleotide changes identified within biosynthetic gene clusters (BGCs). *Percentage* indicates the proportion of nucleotide changes normalized by gene length. *BGC* is the cluster identifier in the genome. *Description* is the functional annotation of the gene. *FungiSMASH* denotes the cluster family, and *Compounds* shows the predicted metabolite synthesized by each BGC. Significant values are in bold.

Untargeted metabolomic analysis reveals divergent peptaibols in *T. harzianum* T9

Given that our genomic analysis suggests that many of the differences between *T. harzianum* T9 and other *T. harzianum* species may be related to secondary metabolism, we sought to characterize the chemical diversity produced under the growth conditions tested in this study. To profile the metabolites synthesized by *T. harzianum* T9, the strain was cultured on PDA plates, and the resulting metabolites were extracted for UHPLC-MS/MS analysis. The spectral data were processed using MZmine3 and submitted to Feature-Based Molecular Networking (FBMN), a workflow available through the Global Natural Products Social Molecular Networking (GNPS) platform. Compound annotations obtained from the FBMN workflow were then manually curated.

We used BGC annotation data to corroborate the mass spectrometry-based structural annotations while assigning known or putative BGCs to most of the detected molecules. Among the annotated compounds, we found an 11-residue amino acid peptaibol (Harzianin HB I) with *m/z* of 1189.79 (Fig. 8A), for which we annotated and proposed a chemical structure based on fragmentation patterns (Fig. 8B). MS-MS network analysis linked this ion to other less abundant molecules, including 14-residue amino acid peptaibols (Supplementary Fig. 5). We proposed structures for the most abundant ions correspondent to m/z = 1175.7762 $[M + H]^+$ (Fig. 8B), m/z = 1189.7918 $[M + H]^+$ (Fig. 8C), and m/z = 1442.9344 $[M + H]^+$ (Fig. 8D; Supplementary Table 4). The MS2 fragment analysis showed several shared ions, indicating that the peptaibol products share most of their amino acid sequences which suggested that they are all derived from a single Non-Ribosomal Peptide Synthetase (NRPS) involving a module skipping mechanism. Such NRPS system (Tex2) has already been reported for other peptaibols in *T. virens* Gv29-8[70]. Thus, *T. harzianum* T9 must have a NRPS featuring 14 modules, an N-terminal acyl loading module and a C-terminal reductase domain. Only one NRPS with this characteristic was found in the genome of *T. harzianum* T9, while no NRPS capable of generating 11-amino-acid peptaibols was found.

To investigate whether this BGC is present in other *Trichoderma* species and shed light on the evolution of these peptaibols, we performed a phylogenomic analysis using our panel of *Trichoderma* genomes. For this purpose, we used an improved version of the CORASON pipeline, optimized for fungal genomes and BGC (https://github.com/WeMakeMolecules/fungison). The analysis revealed that Tex2 from *T. virens* and the T9 NRPS (t9_Tex2) are orthologs, and that their BGC is highly syntenic among *T. harzianum*, *T. virens*, *Trichoderma crassum*, and *T. aggressivum* (Fig. 9A) while in other lineages Tex2 orthologs are lost but the rest gene neighborhood remains conserved.

We also identified additional secondary metabolites produced by *T. harzianum* T9 when grown under laboratory conditions on PDA medium, including harzianic acid (m/z = 364.1762; Fig. S6), harzianopyridone (m/z = 282.1340; Fig. S7), desferrichrocin (m/z = 718.3364; Fig. S8), pachybasin (m/z = 239.0705; Fig. S9), dimerumic acid (m/z = 485.2608; Fig. S10) and tricholignan (m/z = 221.1173; Fig. S11). Their structures were elucidated using biosynthetic gene clusters (BGCs) 19.2, 4.2, 11.1, 27.1, and 7.2 as references (Table S5).

Discussion

Most biocontrol agents currently used in agriculture have been isolated from agricultural soils[71,72], which are typically non-extreme and less challenging environments. It can be hypothesized that microorganisms from extreme environments not only develop mechanisms to resist environmental stresses, such as high temperatures, drought, UV radiation and osmotic stress, but also need to compete for the scarce nutrients available in these

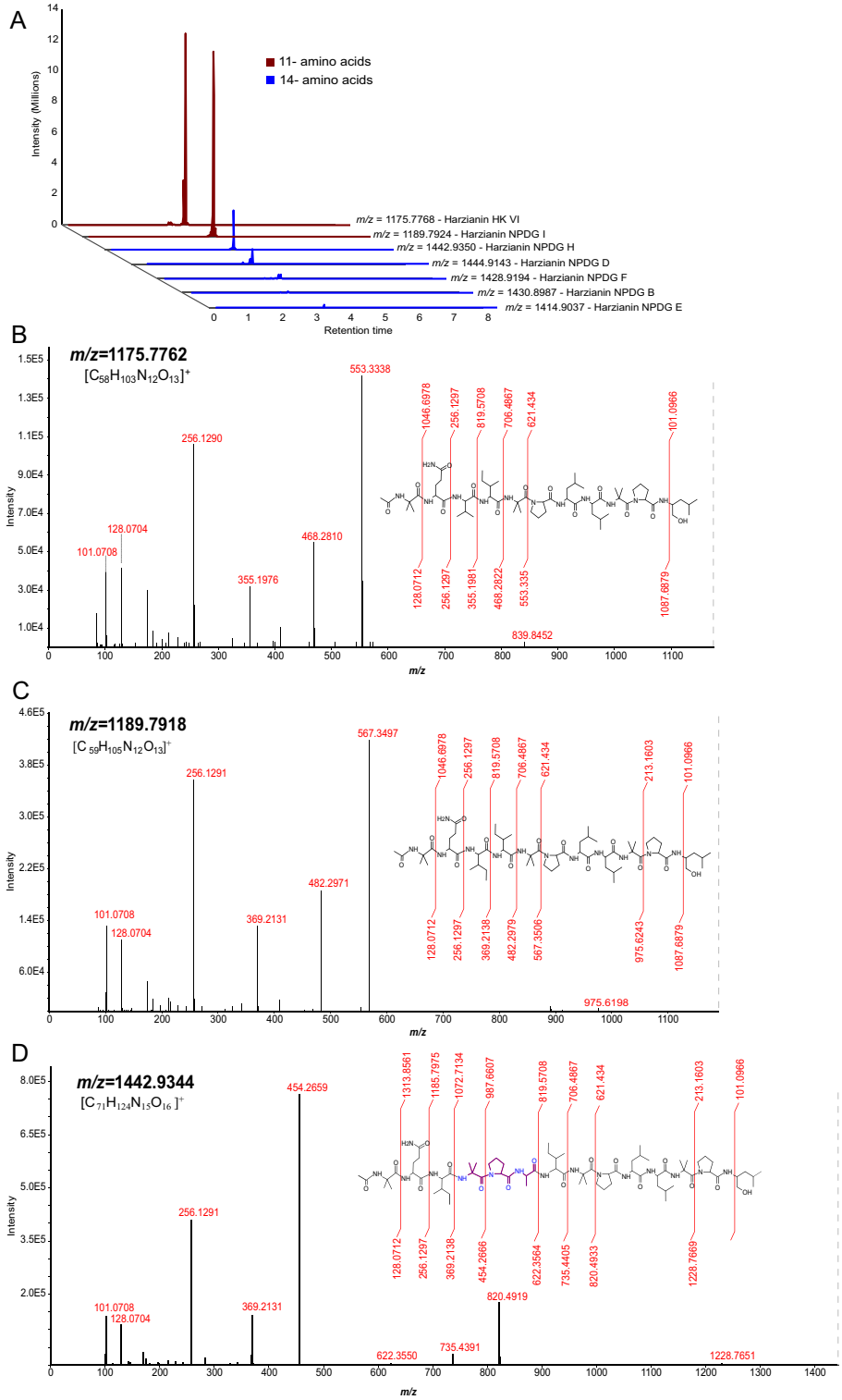

Fig. 8. Extracted ion chromatograms of m/z values referring to the peptaibols detected in *T. harzianum* T9 extracts cultivated in PDA medium. (**A**) Extracted ion chromatogram of (we have add here the m/z of each extracted ion chromatogram). and (**B**) Fragmentation spectra of the precursor ion with m/z = 1175.7762 $[M+H]^+$ corresponding to 11-residue amino acid peptaibol. (**C**) Fragmentation spectra of the precursor ion with m/z = 1189.7918 $[M+H]^+$ corresponding to 11-residue amino acid peptaibol. (**D**) Fragmentation spectra of the precursor ion with m/z = 1442.9344 $[M+H]^+$ corresponding to 14-residue amino acid peptaibol.

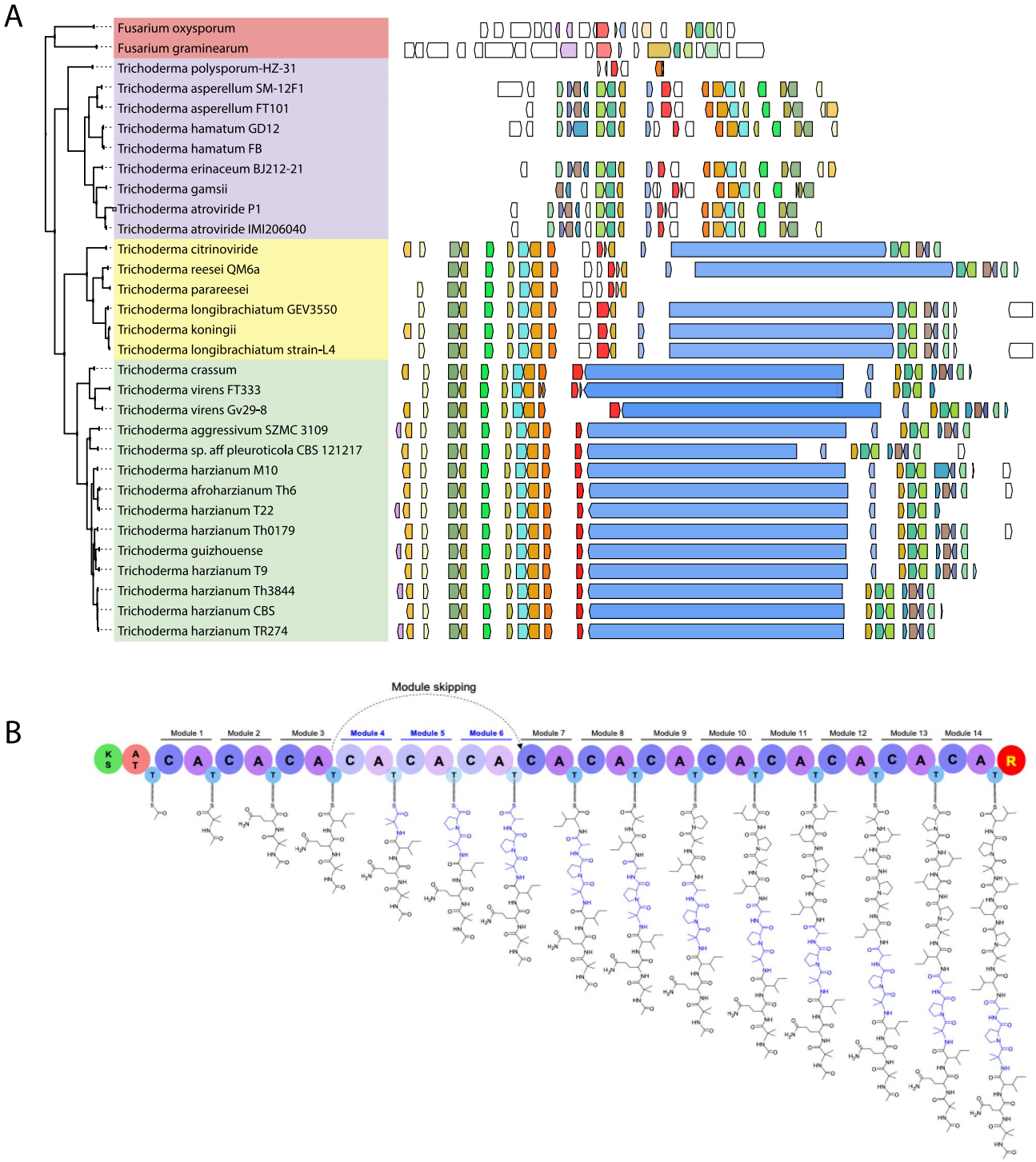

Fig. 9. Metabologenomic analysis of the most abundant BGCs producing peptaibols in *T. harzianum* T9. (**A**) CORASON phylogenetic reconstruction using "gene-name" as the query gene and the *T.harzianum* NRPS cluster responsible for 11- and 14-residue peptaibols. Genes absent in the reference cluster are highlighted and color-coded based on BLAST analysis. (**B**) Biosynthetic pathway of the 12- and 14-residue peptaibols. The skipped residues during the biosynthesis process, caused by NRPS functionality, are highlighted in blue.

ecosystems. One case is that of microorganisms capable of thriving in deserts. Deserts impose multiple challenges, including physical barriers and competition with other microorganisms for limited nutrients. Therefore, these microorganisms must establish associations with plants to acquire nutrients such as photosynthates, while also effectively competing with other organisms for these resources. At the same time, they may promote plant growth and protect the host against pathogens, resulting in a commensal relationship.

In the case of the genus *Trichoderma*, it is known that these fungi can parasitize other fungi[9], and the application of this mechanism is useful to control the growth of phytopathogenic fungi, which benefits agricultural crops[73]. However, with increasing temperatures due to climate change, several phytopathogenic fungi have developed more efficient competitive mechanisms[74] enabling them to resist attacks from commercially available *Trichoderma* species.

The desert-derived strain *T. harzianum* T9, isolated from the rhizosphere of *Agave lechuguilla* in Nuevo León, Mexico[8], exhibited highly effective biocontrol activity against aggressive phytopathogenic fungi affecting strawberry crops in Mexico. For comparison, the performance of *T. atroviride* IMI206040 (widely recognized for its efficacy against pathogens of the genera *Fusarium*, *Botrytis*, and *Alternaria*[9] and *T. harzianum* M10 was also evaluated. Both strains displayed considerably lower efficiency against strawberry-associated pathogens. These results indicate that the genetic repertoire of *T. harzianum* T9 provides a distinct competitive advantage, enabling it to outperform other strains of the same species and suggesting specific genetic adaptations underlying its superior biocontrol potential. Thus, T9 emerges as a promising option for the management of these highly damaging phytopathogens, which are often not effectively controlled by other *Trichoderma* strains in strawberry fields.

To determine the genetic characteristics responsible of these properties in strain *T. harzianum* T9, we performed complete genome sequencing. Our phylogenomic analysis confirmed that T9 strain belongs to the *harzianum* species, validating previous ITS-based results[8]. Interestingly, *T. harzianum* T9 shares approximately 85% of orthologous genes with the *harzianum* clade, indicating that differences between strains may stem from a small number of specific genes and amino acid point mutations.

Based on this, we conducted a species-wide polymorphism analysis using all high-quality *T. harzianum* genomes available. This approach allowed us to identify genes with the highest SNP density across their ORFs. Interestingly, many of the genes with elevated polymorphism rates were involved in secondary metabolism. Such a pattern has been previously reported in fungi; for instance, in the wheat pathogen *Zymoseptoria tritici*, a study analyzing 102 isolates found that significantly associated SNPs were mostly located in coding and gene regulatory regions, with the associated genes primarily encoding transport and catalytic activities[75]. This trend closely mirrors our observations in *T. harzianum* T9 when compared with other *T. harzianum* strains, where genes related to secondary metabolite biosynthesis exhibited a high frequency of nucleotide changes. These observations led us to hypothesize that selective pressures in desert environments have driven a specialization in the secondary metabolism profile of *T. harzianum* T9. For this reason, we set out to explore the metabolomic profile of this strain, aiming to identify the compounds it produces and to determine whether any metabolic variants could be detected.

Through our metabologenomic analysis we were able to correlate several metabolites to their correspondent BGC. Interestingly, we were able to correlate seven molecules with their respective BGCs. Moreover, we observed that the most prominent peak corresponded to an 11-residue peptaibol, while a 14-residue peptaibol was also detected. Our evidence, along with a previous study by Mukherjee et al., (2011), suggests that this is synthesized by the same NRPS but with certain skipped steps, resulting in a diversity of peptaibols. These molecules were previously identified in *T. virens*, a species highly virulent during mycoparasitism[70,76], suggesting that *T. harzianum* T9 may leverage this diversity of peptaibols as an advantage during mycoparasitism. Together, these findings indicate that, in addition to enhanced stress resistance capabilities, *T. harzianum* T9 can produce a wide variety of secondary metabolites that may contribute to its ability to inhibit the growth of phytopathogenic fungi. Moreover, we were surprised to find that, in this strain, the skipping system bypassing the synthesis of the 14-amino acid peptaibol was highly effective, leading to a significantly higher production of the 11-amino acid peptaibol. This could clearly contribute to differences in the biological activity of this compound. Further research is needed on this 11-residue peptaibol and the evolution of the skipping process in fungal NRPS.

These metabolomic findings are highly relevant from an evolutionary perspective, as they suggest that certain fungi with strong biocontrol capabilities share chemical repertoires that enable them to perform this activity efficiently. However, it is particularly interesting that *T. harzianum* T9 exhibits an increased production of an 11-amino acid compound compared to Tex2 and presumably other *T. harzianum* strains. This may be due to modifications in transporters or even in the skipping system (Fig. 9B), which could contribute to its remarkable biocontrol ability by synthesizing this peptaibol in great abundance.

Such modifications, or even alterations in other yet unidentified metabolites, are also likely to contribute to the observed phenotype. The shift in synthesis preference from an 11 amino acid to a 14 amino acid molecule exemplifies the type of chemical variation that may arise. Given the extensive SNP profile observed, it is reasonable to expect that multiple additional variations of this kind occur, potentially influencing the metabolic capabilities and biocontrol properties of *T. harzianum* T9.

Our in-depth phylogenomic analysis of *Trichoderma* genomes suggests that it is possible to identify the key mechanisms of competition and predict some secondary metabolic features from assembled genomes. This could represent a novel strategy to identify highly efficient fungal biocontrol agents based on their genetic characteristics. This approach is particularly relevant in the post-genomic era, where the cost of genome sequencing is relatively low, making routine genomic screening for new strains with desirable biotechnological applications feasible even before experimental validation.

Conclusions

Our study demonstrates that *T. harzianum* T9 is a highly versatile strain, capable of thriving in extreme environmental conditions while effectively controlling common strawberry pathogens in México. We propose this strain as a promising biocontrol agent for crops cultivated in high-pH soils and under drought stress.

In addition, our genomic analysis enabled a detailed characterization of SNP distribution across the *T. harzianum* T9 genome, revealing specific changes within biosynthetic clusters. Notably, we observed alterations in

the synthesis preferences of a peptaibol, underscoring the strain's versatility in producing antifungal compounds. This biosynthetic pathway is also present in *T. virens*, a species recognized for its strong antifungal capacity. By identifying such patterns, our metabologenomics approach highlights its potential for characterizing novel strains with biocontrol potential.

Finally, all our biocontrol assays were performed in vitro under controlled laboratory conditions. We therefore propose that future experiments should be conducted under greenhouse conditions and, subsequently, in open-field trials to determine whether the enhanced biocontrol activity observed against phytopathogenic strains is maintained in more realistic agricultural environments, including assessing *in planta* colonization. This step is essential for supporting the designation of T9 as a high-performance biocontrol strain. Moreover, given the challenges currently faced by farmers in Mexico, our long-term goal is to develop an agricultural product based on *T. harzianum* T9 that can be applied in fields already affected by fungal diseases. We believe that this final translational step is as crucial as the research itself to ensure that these biotechnological solutions effectively reach and benefit the agricultural sector.

Data availability

The genome assembly (FASTA) and annotation files for *Trichoderma harzianum* T9 have been deposited under NCBI BioProject PRJNA1231633 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1231633). The raw sequenci ng reads (FASTQ) have been deposited in Zenodo (https://zenodo.org/records/16968646).

Code availability

All codes used for the comparative genomics analyses are available in our GitHub repository at the following link: https://github.com/AnaValeriaG01/trichodermaT9.

Received: 25 August 2025; Accepted: 18 December 2025
Published online: 08 January 2026

References

1. Sariah, M., Choo, C. W., Zakaria, H. & Norihan, M. S. Quantification and characterisation of *Trichoderma* spp. From different ecosystems. *Mycopathologia* **159**, 113–117 (2005).
2. Woo, S. L., Hermosa, R., Lorito, M. & Monte, E. *Trichoderma*: a multipurpose, plant-beneficial microorganism for eco-sustainable agriculture. *Nat. Rev. Microbiol.* **21**, 312–326 (2023).
3. Guzmán-Guzmán, P., Porras-Troncoso, M. D. & Olmedo-Monfil, V. Herrera-Estrella, A. *Trichoderma* species: versatile plant symbionts. *Phytopathology* **109**, 6–16 (2019).
4. Tyśkiewicz, R., Nowak, A., Ozimek, E. & Jaroszuk-Ściseł, J. *Trichoderma*: the current status of its application in agriculture for the biocontrol of fungal phytopathogens and stimulation of plant growth. *Int. J. Mol. Sci.* **23**, 2329 (2022).
5. Yao, X. et al. *Trichoderma* and its role in biological control of plant fungal and nematode disease. *Front. Microbiol.* **14**, 1160551 (2023).
6. Guo, R., Wang, Z., Huang, Y., Fan, H. & Liu, Z. Biocontrol potential of saline- or alkaline-tolerant *Trichoderma asperellum* mutants against three pathogenic fungi under saline or alkaline stress conditions. *Braz J. Microbiol.* **49**, 236–245 (2018).
7. de Bueno, C. P. et al. Microbial ecology and site characteristics underlie differences in salinity–methane relationships in coastal wetlands. *J. Geophys. Res. Biogeosci.* **129**, e2024JG008133 (2024).
8. Cabral-Miramontes, J. P., Olmedo-Monfil, V., Lara-Banda, M. & Zúñiga-Romo, E. R. Aréchiga-Carvajal, E. T. Promotion of plant growth in arid zones by selected *Trichoderma* spp. Strains with adaptation plasticity to alkaline pH. *Biology (Basel)*. **11**, 1206. https://doi.org/10.3390/biology11081206 (2022).
9. Enriquez-Felix, E. E. et al. Argonaute and Dicer are essential for communication between *Trichoderma atroviride* and fungal hosts during mycoparasitism. *Microbiol. Spectr.* **12**, e0316523. https://doi.org/10.1128/spectrum.03165-23 (2024).
10. Özkale, E., Yörük, E., Budak, M. & Korkmaz, E. M. *Trichoderma atroviride* suppresses *Fusarium graminearum* by altering primary and secondary metabolite biosynthesis profiling. *Plant. Pathol.* **72**, 1428–1441 (2023).
11. Rebollar-Alviter, A. et al. An emerging strawberry fungal disease associated with root rot, crown rot and leaf spot caused by *Neopestalotiopsis Rosae* in Mexico. *Plant. Dis.* **104**, 2054–2059 (2020).
12. Chaverri, P. et al. Systematics of the *Trichoderma Harzianum* species complex and the re-identification of commercial biocontrol strains. *Mycologia* **107**, 558–590 (2015).
13. Al-Salihi, S. A. & Alberti, F. Genomic based analysis of the biocontrol species *Trichoderma harzianum*: a model resource of structurally diverse pharmaceuticals and biopesticides. *J. Fungi.* **9**, 895 (2023).
14. Sambrook, J. & Russell, D. W. Purification of nucleic acids by extraction with phenol: chloroform. *Cold Spring Harb. Protoc.* **2006**, pdb-prot4455 (2006).
15. White, T. J., Bruns, T., Lee, S. J. W. T. & Taylor, J. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In *PCR Protocols: A Guide To Methods and Applications*. 315–322 (Academic, 1990).
16. Swift, M. L. & GraphPad Prism, data analysis, and scientific graphing. *J. Chem. Inf. Comput. Sci.* **37**, 411–412 (1997).
17. Andrews, S. FastQC: a quality control tool for high throughput sequence data. *Babraham Bioinformatics*. http://www.bioinformatics.babraham.ac.uk/projects/fastqc (2010).
18. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
19. Bankevich, A. et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455–477 (2012).
20. Astashyn, A. et al. Rapid and sensitive detection of genome contamination at scale with FCS-GX. *Genome Biol.* **25**, 60. https://doi.org/10.1186/s13059-024-03198-7 (2024).
21. Gurevich, A., Saveliev, V., Vyahhi, N. & Tesler, G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* **29**, 1072–1075. https://doi.org/10.1093/bioinformatics/btt086 (2013).
22. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).
23. Baril, T., Galbraith, J. G. & Hayward, A. Earl grey: a fully automated user-friendly transposable element annotation and analysis pipeline. *Mol. Biol. Evol.* **41**, msae068. https://doi.org/10.1093/molbev/msae068 (2024).
24. Hubley, R. et al. The Dfam database of repetitive DNA families. *Nucleic Acids Res.* **44**, D81–D89 (2016).

25. Bruna, T., Hoff, K. J., Lomsadze, A., Stanke, M. & Borodovsky, M. BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP + and AUGUSTUS supported by a protein database. *NAR Genom Bioinform*. **3**, lqaa108. https://doi.org/10.1093/nargab/lqaa108 (2021).
26. Kuznetsov, D. et al. OrthoDB v11: annotation of orthologs in the widest sampling of organismal diversity. *Nucleic Acids Res.* **51**, D445–D451. https://doi.org/10.1093/nar/gkac998 (2023).
27. Tang, H. et al. JCVI: a versatile toolkit for comparative genomics analysis. *Imeta* **3**, e211 (2024).
28. Huerta-Cepas, J. et al. Fast genome-wide functional annotation through orthology assignment by eggNOG-Mapper. *Mol. Biol. Evol.* **34**, 2115–2122. https://doi.org/10.1093/molbev/msx148 (2017).
29. Jones, P. et al. InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240. https://doi.org/10.1093/bioinformatics/btu031 (2014).
30. Zheng, J. et al. dbCAN3: automated carbohydrate-active enzyme and substrate annotation. *Nucleic Acids Res.* **51**, W115–W121 (2023).
31. Cuomo, C. A. et al. The *Fusarium graminearum* genome reveals a link between localized polymorphism and pathogen specialization. *Science* **317**, 1400–1402. https://doi.org/10.1126/science.1143708 (2007).
32. Ma, L. J. et al. Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. *Nature* **464**, 367–373. https://doi.org/10.1038/nature08850 (2010).
33. He, Y. & Zhu, H. Whole genome sequencing and analysis of the weed pathogen *Trichoderma polysporum* HZ-31. *Sci. Rep.* **14**, 15228. https://doi.org/10.1038/s41598-024-66041-w (2024).
34. Studholme, D. J. et al. Investigating the beneficial traits of *Trichoderma Hamatum* GD12 for sustainable agriculture—insights from genomics. *Front. Plant. Sci.* **4**, 258. https://doi.org/10.3389/fpls.2013.00258 (2013).
35. Davolos, D. et al. A genomic and transcriptomic study on the DDT-resistant *Trichoderma Hamatum* FBL 587: first genetic data into mycoremediation strategies for DDT-polluted sites. *Microorganisms* **9**, 1680. https://doi.org/10.3390/microorganisms9081680 (2021).
36. Li, L. et al. Genome sequence of the fungus trichoderma asperellum SM-12F1 revealing candidate functions of growth promotion, biocontrol, and bioremediation. *PhytoFrontiers* **1**, 171–173. https://doi.org/10.1094/PHYTOFR-12-20-0052-A (2021).
37. Li, W. C. et al. Complete genome sequences and genome-wide characterization of *Trichoderma* biocontrol agents provide new insights into their evolution and variation in genome organization, sexual development, and fungal–plant interactions. *Microbiol. Spectr.* **9**, e0066321. https://doi.org/10.1128/Spectrum.00663-21 (2021).
38. Baroncelli, R., Zapparata, A., Piaggeschi, G., Sarrocco, S. & Vannacci, G. Draft whole-genome sequence of *Trichoderma gamsii* T6085, a promising biocontrol agent of fusarium head blight on wheat. *Genome Announc*. **4**, e01747–e01715. https://doi.org/10.1128/genomeA.01747-15 (2016).
39. Kubicek, C. P. et al. Comparative genome sequence analysis underscores mycoparasitism as the ancestral lifestyle of Trichoderma. *Genome Biol.* **12**, R40. https://doi.org/10.1186/gb-2011-12-4-r40 (2011).
40. Druzhinina, I. S. et al. Massive lateral transfer of genes encoding plant cell wall-degrading enzymes to the mycoparasitic fungus *Trichoderma* from its plant-associated hosts. *PLoS Genet.* **14**, e1007322. https://doi.org/10.1371/journal.pgen.1007322 (2018).
41. Kubicek, C. P. et al. Evolution and comparative genomics of the most common *Trichoderma* species. *BMC Genom.* **20**, 485. https://doi.org/10.1186/s12864-019-5680-7 (2019).
42. Steindorff, A. S. et al. Identification of mycoparasitism-related genes against the phytopathogen *Sclerotinia sclerotiorum* through transcriptome and expression profile analysis in *Trichoderma Harzianum*. *BMC Genom.* **15**, 204. https://doi.org/10.1186/1471-2164-15-204 (2014).
43. Yang, D. et al. Genome sequence and annotation of *Trichoderma parareesei*, the ancestor of the cellulase producer *Trichoderma reesei*. *Genome. Announc.* **3**, e00885-15. https://doi.org/10.1128/genomeA.00885-15 (2015).
44. Li, W. C. et al. *Trichoderma Reesei* complete genome sequence, repeat-induced point mutation, and partitioning of cazyme gene clusters. *Biotechnol. Biofuels*. **10**, 170. https://doi.org/10.1186/s13068-017-0825-x (2017).
45. Suazo Tejada, A. K. et al. Genome sequencing and de Novo assembly of *Trichoderma longibrachiatum* isolate collected from Florida agricultural soils. *Microbiol. Resour. Announc*. **13**, e00906–e00923. https://doi.org/10.1128/mra.00906-23 (2024).
46. Rosolen, R. R. et al. Whole-genome sequencing and comparative genomic analysis of potential biotechnological strains of *Trichoderma harzianum*, *Trichoderma atroviride*, and *Trichoderma reesei*. *Mol. Genet. Genomics* **298**, 735–754. https://doi.org/10.1007/s00438-023-02013-5 (2023).
47. Sun, J. et al. OrthoVenn3: an integrated platform for exploring and visualizing orthologous data across genomes. *Nucleic Acids Res.* **51**, W397–W403 (2023).
48. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797. https://doi.org/10.1093/nar/gkh340 (2004).
49. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. TrimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973. https://doi.org/10.1093/bioinformatics/btp348 (2009).
50. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS One*. **5**, e9490 (2010).
51. Gupta, A., Mirarab, S. & Turakhia, Y. Accurate, scalable, and fully automated inference of species trees from raw genome assemblies using ROADIES. *Proc. Natl Acad. Sci. USA* **122**, e2500553122. https://doi.org/10.1073/pnas.2500553122 (2025).
52. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *ArXiv* https://doi.org/10.48550/ArXiv.1303.3997 (2013).
53. Li, H. et al. The sequence Alignment/Map format and samtools. *Bioinformatics* **25**, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352 (2009).
54. Danecek, P. et al. The variant call format and vcftools. *Bioinformatics* **27**, 2156–2158. https://doi.org/10.1093/bioinformatics/btr330 (2011).
55. Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 238. https://doi.org/10.1186/s13059-019-1832-y (2019).
56. Heuckeroth, S. et al. Reproducible mass spectrometry data processing and compound annotation in MZmine 3. *Nat. Protoc.* **19**, 2597–2641 (2024).
57. Navarro-Muñoz, J. C. et al. A computational framework to explore large-scale biosynthetic diversity. *Nat. Chem. Biol.* **16**, 60–68 (2020).
58. Kaur, S. et al. Emerging phytopathogen *Macrophomina phaseolina*: biology, economic importance and current diagnostic trends. *Crit. Rev. Microbiol.* **38**, 136–151 (2012).
59. Marçais, G. & Kingsford, C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27**, 764–770 (2011).
60. Vurture, G. W. et al. GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics* **33**, 2202–2204 (2017).
61. Baroncelli, R. et al. Draft whole-genome sequence of the biocontrol agent trichoderma Harzianum T6776. *Genome Announc*. **3**, e00647–e00615. https://doi.org/10.1128/genomeA.00647-15 (2015).
62. Fanelli, F., Liuzzi, V. C. & Logrieco, A. F. Genomic characterization of *Trichoderma atrobrunneum* (*T. harzianum* species complex) ITEM 908: insight into the genetic endowment of a multi-target biocontrol strain. *BMC Genom.* **19**, 662. https://doi.org/10.1186/s12864-018-5049-3 (2018).

63. Beijen, E. P. W. & Ohm, R. A. Genome annotations for the ascomycete fungi *Trichoderma harzianum*, *Trichoderma aggressivum* and *Purpureocillium lilacinum*. *Microbiol. Resour. Announc.* **13**, e01153–e01123. https://doi.org/10.1128/mra.01153-23 (2024).
64. Pertea, G. & Pertea, M. GFF utilities: GffRead and GffCompare. *F1000Research* 9, ISCB-Comm (2020).
65. Dixon, D. P., Skipsey, M. & Edwards, R. Roles for glutathione transferases in plant secondary metabolism. *Phytochemistry* **71**, 338–350 (2010).
66. Baker, S. E., Perrone, G., Richardson, N. M., Gallo, A. & Kubicek, C. P. Phylogenomic analysis of polyketide synthase-encoding genes in *Trichoderma*. *Microbiology* **158**, 147–154 (2012).
67. Bumpus, S. B., Magarvey, N. A., Kelleher, N. L., Walsh, C. T. & Calderone, C. T. Polyunsaturated fatty-acid-like trans-enoyl reductases utilized in polyketide biosynthesis. *J. Am. Chem. Soc.* **130**, 11614–11616 (2008).
68. Yao, T., Wang, X. & Chen, F. The role of Enoyl reductase in the Monacolin K biosynthesis pathway in *Monascus* spp. *J. Fungi.* **11**, 199 (2025).
69. Moummou, H., Kallberg, Y., Tonfack, L. B., Persson, B. & van Der Rest, B. The plant short-chain dehydrogenase (SDR) superfamily: genome-wide inventory and diversification patterns. *BMC Plant. Biol.* **12**, 219 (2012).
70. Mukherjee, P. K. et al. Two classes of new peptaibols are synthesized by a single non-ribosomal peptide synthetase of *Trichoderma virens*. *J. Biol. Chem.* **286**, 4544–4554 (2011).
71. Nandini, B., Puttaswamy, H., Saini, R. K., Prakash, H. S. & Geetha, N. Trichovariability in rhizosphere soil samples and their biocontrol potential against downy mildew pathogen in Pearl millet. *Sci. Rep.* **11**, 9517. https://doi.org/10.1038/s41598-021-89061-2 (2021).
72. Xue, M. et al. Screening and identification of Trichoderma strains isolated from natural habitats in China with potential agricultural applications. *Biomed. Res. Int.* **2021**, 7913950. https://doi.org/10.1155/2021/7913950 (2021).
73. Mukherjee, P. K., Mendoza-Mendoza, A., Zeilinger, S. & Horwitz, B. A. Mycoparasitism as a mechanism of *Trichoderma*-mediated suppression of plant diseases. *Fungal Biol. Rev.* **38**, 15–33. https://doi.org/10.1016/j.fbr.2021.11.004 (2022).
74. Timmusk, S., Nevo, E., Ayele, F., Noe, S. & Niinemets, Ü. Fighting *Fusarium* pathogens in the era of climate change: a conceptual approach. *Pathogens* **9**, 419 (2020).
75. Singh, N. K., Tralamazza, S. M., Abraham, L. N., Glauser, G. & Croll, D. Genome-wide association mapping reveals genes underlying population-level metabolome diversity in a fungal crop pathogen. *BMC Biol.* **20**, 224 (2022).
76. Mukherjee, P. K., Horwitz, B. A. & Kenerley, C. M. Secondary metabolism in *Trichoderma* – a genomic perspective. *Microbiology* **158**, 35–45 (2012).

Acknowledgements

We thank Dr. Rebollar-Alviter for the kind donation of the N. rosae strain, Axel Sánchez-Fonseca, and Juan Ignacio Macías for their technical support in this work. We also thank Dr. Louise Glass at UC Berkeley for donating two model Trichoderma strains, which we used for comparison against our T. harzianum T9 strain.

Author contributions

J.M.V.E. and V.O.M. conceived and designed the intellectual content; F.V.G. conducted strain isolation experiments, genetic identification, and biocontrol assays. Additionally, they contributed to the manuscript writing; E.P.S. performed the sequencing experiments and genomic analyses presented in this study; A.V.G.L. performed the bioinformatic analyses for gene variant identification; A.V.G.L., A.D.G.V., P.C.M., A.C.C., D.R., L.A., and E.T.A.C. contributed to data collection, analysis, and manuscript revision; J.M.V.E. and V.O.M. secured funding for the project and drafted the manuscript; All authors have read and approved the final manuscript and agree to be responsible for all aspects of the work.

Funding

The work of ACC, PCM, DR was funded by the Novo Nordisk Foundation with grant NNF20CC0035580. The antagonism tests were carried out thanks to the resources derived from the CIIC 138/2025 project of the University of Guanajuato, under the responsibility of Vianey Olmedo Monfil.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-33347-2.

Correspondence and requests for materials should be addressed to J.M.V.-E. or V.O.-M.