Article

Not peer-reviewed version

Comparative Genomics Analysis Reveals the Genomic Basis of S8 Proteases, CAZymes, and Secondary Metabolism Associated with Nematode Biocontrol in *Purpureocillium lilacinum*

Xiaoxi Cheng , Li Liu , Zhimin Zhu , Minghao Chen , Wenbo Wang , Jialin Li , Ramon Santos Bermudez , Xiujun Zhang * , Wenxing He *

Posted Date: 15 April 2026

doi: 10.20944/preprints202604.1074.v1

Keywords: *Purpureocillium lilacinum*; comparative genomics; plant-parasitic nematodes; CAZymes; S8 serine protease; secondary metabolite gene clusters

 doi:10.20944/preprints202604.1074.v1

Article

Comparative Genomics Analysis Reveals the Genomic Basis of S8 Proteases, CAZymes, and Secondary Metabolism Associated with Nematode Biocontrol in *Purpureocillium lilacinum*

Xiaoxi Cheng, Li Liu, Zimin Zhu, Minghao Chen, Wenbo Wang, Jialin Li, Ramon Santos Bermudez, Xiujun Zhang * and Wenxing He *

School of Biological Science and Technology, University of Jinan, Jinan, China

* Correspondence: bio_zhangxj@ujn.edu.cn (X.Z.); chm_hewx@ujn.edu.cn (W.H.)

Abstract

Biological control fungi play an important role in the management of plant-parasitic nematodes; however, the molecular basis underlying their diverse biocontrol strategies remains incompletely understood. In this study, a comparative genomic analysis was performed on four representative biocontrol fungi, *Purpureocillium lilacinum* PLFJ-1, *Trichoderma harzianum* CBS 226.95, *Pochonia chlamydosporia* 170, and *Aspergillus niger* CBS 513.88. Genome comparison revealed substantial variation in genome size, GC content, and gene composition among the four fungi. Phylogenetic analysis divided these fungi into two major clades corresponding to distinct evolutionary lineages. Orthogroup analysis identified both a conserved core gene set and species-specific gene repertoires. Functional annotation based on KEGG, KOG, and GO indicated a high degree of conservation in core metabolic processes, catalytic activities, and cellular components, accompanied by distinct differences in specific functional categories. Further comparative analyses demonstrated pronounced variation in the composition and abundance of carbohydrate-active enzymes (CAZymes) and peptidases, as well as a notable expansion and enrichment of S8 subtilisin-like serine peptidases in the nematode-parasitic fungi *P. lilacinum* and *P. chlamydosporia*. Secondary metabolite analysis identified diverse biosynthetic gene clusters, with differences in PKS and NRPS distribution among species. Together, these findings suggest that distinct enzymatic and metabolic gene repertoires, particularly expansions of S8 serine peptidases and specific CAZyme families, may contribute to the biocontrol potential of these fungi.

Keywords: *Purpureocillium lilacinum*; comparative genomics; plant-parasitic nematodes; CAZymes; S8 serine protease; secondary metabolite gene clusters

1. Introduction

Over recent decades, rapid global population growth has placed increasing pressure on agricultural production, creating an urgent need for efficient and sustainable crop protection strategies. Root-knot nematodes (RKNs) are major plant-parasitic nematodes worldwide, infecting a wide range of crops, including staple foods, vegetables, and medicinal plants, and causing substantial economic losses. It is estimated that four major species-*Meloidogyne incognita*, *Meloidogyne arenaria*, *Meloidogyne javanica*, and *Meloidogyne hapla*-collectively account for approximately US$100 billion in annual global agricultural yield losses, with economic damage exceeding US$1 billion per year in the United States [1]. In many agricultural systems, RKNs infestation causes substantial yield losses and quality deterioration across diverse crops, with infection rates exceeding 80% and yield reductions reaching up to 50–80% in severe cases [2–4].

 doi:10.20944/preprints202604.1074.v1

Current control strategies for RKNs remain limited. Agronomic practices provide only partial suppression due to the broad host range and environmental adaptability of these nematodes, while chemical nematicides raise concerns regarding environmental contamination, human health risks, and resistance development. Physical control methods, such as soil solarization [5], are environmentally friendly but require prolonged treatment and are difficult to implement at large scales [5–10].Given these limitations, biological control has emerged as a viable alternative [11]. Nematophagous fungi, as natural enemies of nematodes, suppress nematode populations through secreting extracellular enzymes and trapping or parasitic mechanisms, and can be classified into four groups: (i) predatory fungi, (ii) endoparasitic fungi, (iii) egg-parasitic fungi, and (iv) toxin-producing fungi [11–14].

Among them, *Purpureocillium lilacinum* (formerly known as *Paecilomyces lilacinus*) is one of the most widely studied and commercially applied biocontrol fungi. Previous research has demonstrated that *P. lilacinum* infects the eggs and adult females of root-knot nematodes, secretes extracellular hydrolytic enzymes, such as proteases and chitinases, and produces secondary metabolites that enhance nematode mortality [15–19].

However, to date, the molecular mechanisms underlying the pathogenicity of *P. lilacinum* toward nematodes remain poorly elucidated. Previous studies have demonstrated the involvement of extracellular enzymes, including proteases, collagenases, and chitinases, and secondary metabolites/bioactive compounds that are essential for their biocontrol attributes [20,21]. However, systematic molecular characterization remains limited, with only a small number of individual genes, including one serine protease [20] and one keratinase [22], having been functionally characterized to date.

In recent years, rapid advances in genome sequencing and comparative genomics have provided powerful approaches for elucidating the mechanisms underlying biological control. Functional genome annotation and analyses of metabolic potential across diverse biocontrol fungi can reveal differences in secondary metabolism, enzyme secretion, and signal transduction, thereby offering a theoretical framework for the identification and development of highly effective biocontrol strains.

In addition to *P. lilacinum*, several fungi, including *T. harzianum*, *P. chlamydosporia*, and *A. niger*, have also been reported to exhibit nematode antagonistic activity. However, systematic comparative genomic analyses among these species remain limited, particularly with respect to key functional modules such as carbohydrate-active enzymes, serine proteases, and secondary metabolite biosynthetic gene clusters.

In this study, we performed a comparative genomic analysis of *P. lilacinum* PLFJ-1, *T. harzianum* CBS 226.95, *P. chlamydosporia* 170, and *A. niger* CBS 513.88. By integrating genome assembly, phylogenetic analysis, orthogroup comparison, functional annotation, and detailed investigation of CAZymes, peptidases, and secondary metabolite gene clusters, we aimed to characterize both conserved and species-specific genomic features associated with biocontrol potential. This study provides a systematic framework for understanding the molecular basis of nematode parasitism and provides a genomic foundation for the rational improvement and application of highly efficient *P. lilacinum* -based biocontrol agents.

2. Results

2.1. Assembled Genomic Features and General Comparison

Comparative analysis of the genome assembly revealed substantial variation among four biocontrol fungi in genome size, GC content, and assembly contiguity (Table 1). Genome sizes ranged from 34 Mb in *A. niger* CBS 513.88 to 44.2 Mb in *P. chlamydosporia* 170. *A. niger* CBS 513.88 possessed the smallest genome (34 Mb), consisting of 19 scaffolds with a scaffold N50 of 2.5 Mb. In contrast, *P. lilacinum* PLFJ-1 exhibited a larger genome (38.5 Mb) but a more fragmented assembly (163 scaffolds), despite a comparable scaffold N50 (3.2 Mb). Notably, *T. harzianum* CBS 226.95 possessed a larger genome (41 Mb) with the highest number of scaffolds (532), suggesting lower assembly contiguity

 doi:10.20944/preprints202604.1074.v1

despite a relatively high contig N50 (360.6 kb). By contrast, *P. chlamydosporia* 170 possesses the largest genome (44.2 Mb), and exhibited the highest assembly quality. It contained only 49 scaffolds, with a scaffold N50 of 5.4 Mb and an L50 of 2, indicating the highest assembly contiguity among the four genomes.

GC content varied considerably, ranging from 47.5% in *T. harzianum* to 58.5% in *P. lilacinum*, which may reflect lineage-specific genomic features and evolutionary adaptations. Overall, these variations in genome size and assembly quality may impact the complexity of gene repertoires, particularly for gene families linked to parasitism and environmental adaptation.

Table 1. Genome assembly and annotation statistics.

Genome characteristics	*Aspergillus niger* CBS513.88	*Purpureocillium lilacinum* PLFJ-1	*Trichoderma harzianum* CBS 226.95	*Pochonia chlamydosporia* 170
Genome size	34 Mb	38.5 Mb	41 Mb	44.2 Mb
Total ungapped length	33.9 Mb	38.2 Mb	41 Mb	44.2 Mb
Number of scaffolds	19	163	532	49
Scaffold N50	2.5 Mb	3.2 Mb	2.4 Mb	5.4 Mb
Scaffold L50	6	5	7	4
Number of contigs	469	680	841	114
Contig N50	114 kb	150.1 kb	360.6 kb	2 Mb
Contig L50	96	76	38	8
GC percent	50.5	58.5	47.5	49.5
Genome coverage		152x	120x	211.0x
Assembly level	Scaffold	Scaffold	Scaffold	Chromosome
Number of chromosomes	8			7

2.2. Phylogenetic Relationships

The phylogenetic tree constructed using the Neighbor-Joining method with 1,000 bootstrap replicates revealed that the selected fungal species could be divided into two major, well-supported clades based on the target protein sequences (Figure 1). These clades broadly correspond to distinct ecological strategies, comprising an environmental saprophytic clade represented by the genus *Aspergillus*, and a parasitic or antagonistic clade represented by *Pochonia*, *Purpureocillium*, and *Trichoderma*. This divergence is supported by a bootstrap value of 100 at the root node, indicating that the two fungal groups have undergone pronounced phylogenetic differentiation over the course of long-term evolution.

Within the *Aspergillus* lineage, *A. flavus* and *A. oryzae* exhibit an extremely close phylogenetic relationship (bootstrap = 100), whereas the clustering of *A. fumigatus* and *A. niger* is supported by a comparatively low bootstrap value (bootstrap = 47), suggesting that the divergence within this lineage remains uncertain under the phylogenetic marker employed. This pattern may be attributable to frequent genomic rearrangements and extensive ecological diversification within the genus *Aspergillus*. Notably, *Pochonia chlamydosporia* and *P. globispora* form a strongly supported monophyletic clade, further corroborating the phylogenetic distinctiveness of the genus *Pochonia* as egg-parasitic nematode fungi. *Purpureocillium lilacinum* is closely related to this clade and exhibits a

 doi:10.20944/preprints202604.1074.v1

close evolutionary relationship (bootstrap = 98), suggesting shared evolutionary origins and potentially conserved parasitic mechanisms.

In addition, pronounced divergence was observed among species within the genus *Trichoderma*, with *T. reesei* and *T. virens* forming a stable clade, whereas *T. harzianum* occupies a more basal position, reflecting the diversity of antagonistic strategies and ecological adaptations within this genus. Overall, this phylogenetic framework provides an evolutionary context for interpreting differences in gene family composition and functional potential among the four fungi, particularly for gene families associated with host interaction and parasitism, such as proteases, chitinases, and secreted proteins.

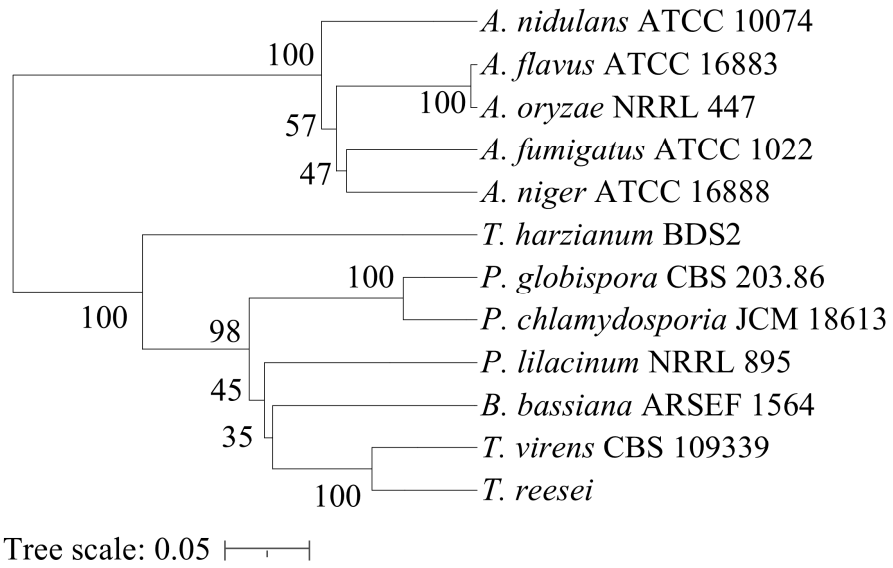

Figure 1. Phylogenetic tree constructed by the neighbor-joining algorithm among selected fungal species. Bootstrap values are shown on nodes as percentages of 1000 replicates. Only bootstrap values greater than 50% are indicated. The scale bar represents 0.05 substitutions per site.

2.3. Comparative Analysis of Orthogroups

Comparative analysis of orthologous gene families among the four fungal species identified a total of 19,838 orthogroups. Detailed category counts are provided in Supplementary Table S1. Of these, 5,257 orthogroups were shared across all four genomes, representing the conserved genetic core of these fungi (Figure 2). Among this shared core, numerous orthogroups were distributed among specific subsets of species. Notably, the three biocontrol-associated fungi (*P. lilacinum*, *T. harzianum*, and *P. chlamydosporia*) shared a higher number of orthogroups with each other than with *A. niger*, suggesting a closer functional and evolutionary relationship among these antagonistic or parasitic species. Species-specific orthogroups were also identified in each genome. Specifically, 4,082, 2,113, 2,651, and 2,820 unique orthogroups were identified in *A. niger*, *P. lilacinum*, *T. harzianum*, and *P. chlamydosporia*, respectively (Figure 2). These unique gene sets may contribute to ecological specialization, particularly traits associated with host interaction, environmental adaptation, and niche differentiation. In summary, UpSet analysis identified a conserved genomic foundation alongside variable, lineage-specific gene sets, highlighting significant differences in gene family composition among the four fungal species.

 doi:10.20944/preprints202604.1074.v1

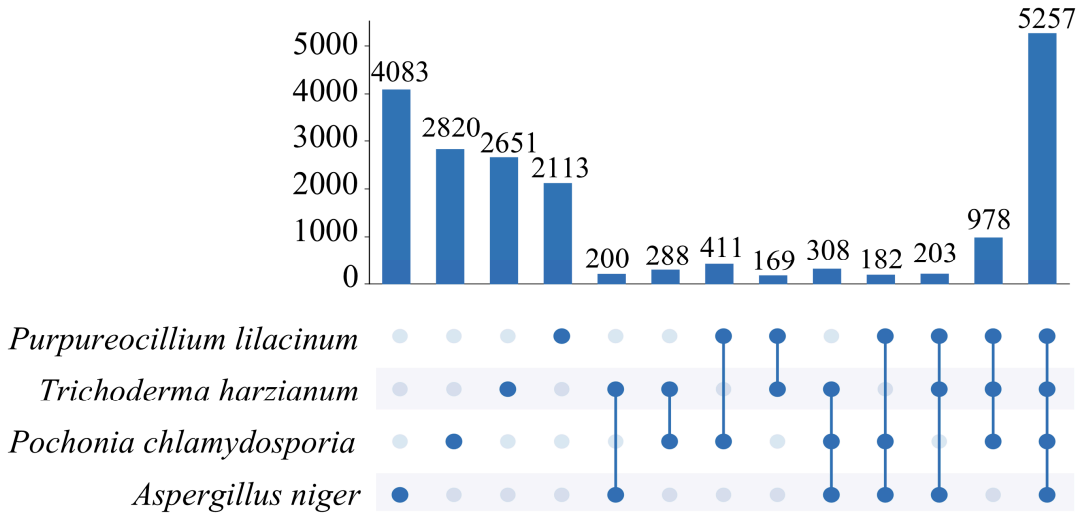

Figure 2. UpSet plot showing the shared and unique orthogroups among *P. lilacinum* PLFJ-1, *T. harzianum* CBS 226.95, *P. chlamydosporia* 170, and *A. niger* CBS 513.88. The vertical bars represent the number of orthogroups in each intersection, while the horizontal bars indicate the total orthogroups in each species.

2.4. Functional Annotation and Classification

Comparative analysis based on Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway annotations revealed pronounced differences in metabolic functional composition among the four fungal genomes (Figure 3A). Despite these differences, amino acid metabolism and carbohydrate metabolism were the most represented functional categories across all four fungi, indicating a high degree of conservation in core nutritional metabolic processes. Among the four species, *A. niger* and *T. harzianum* exhibited relatively higher numbers of annotated enzymes across most metabolic pathways, particularly in secondary metabolite biosynthesis, lipid metabolism, and xenobiotic biodegradation and metabolism. This pattern reflects their more complex metabolic networks and enhanced capacity for environmental adaptation. In contrast, *P. lilacinum* and *P. chlamydosporia* displayed relatively lower overall numbers of annotated metabolic pathways, although genes involved in glycan biosynthesis and amino acid metabolism remained well represented.

Comparative analysis of genome functional composition based on Eukaryotic Orthologous Groups (KOG) classification revealed a highly similar overall distribution pattern across major functional categories, with variation in the relative abundance of specific functional groups (Figure 3B). In terms of relative abundance of specific functional categories, the category “posttranslational modification, protein turnover, and chaperones” represented a substantial proportion in all species. “Signal transduction mechanisms” and “intracellular trafficking, secretion, and vesicular transport” were also consistently abundant across the four genomes. Differences among species were observed in several categories. The relative proportion of “cell wall/membrane/envelope biogenesis” variedwere present among the strains, with *P. chlamydosporia* and *P. lilacinum* exhibited higher representation in this category, whereas the category “defense mechanisms” showed relatively stable representation in *A. niger* and *T. harzianum.* In contrast, the categories “cell motility” ,”extracellular structures” , and “nuclear structure” were present at low proportions in all four species.

Gene Ontology (GO) functional annotation revealed differences in gene distribution among the four fungi across the three main categories: biological process, molecular function, and cellular component (Figure 3C). Within the “Biological Process” category, all four genomes were predominantly enriched in “metabolic process” (GO:0008152) and “cellular process” (GO:0009987). In addition, “regulation of biological process” (GO:0050789), “biological regulation” (GO:0065007), as well as “localization” (GO:0051179) and “establishment of localization” (GO:0051234) were also well represented, while “response to stimulus” (GO:0050896) was relatively less abundant with

 doi:10.20944/preprints202604.1074.v1

variation among species. Within the "Molecular Function" category, genes were predominantly enriched in "catalytic activity" (GO:0003824) and "binding" (GO:0005488), with smaller proportions assigned to "transcription regulator activity" (GO:0030528), "molecular transducer activity" (GO:0060089), and "transporter activity" (GO:0005215). Within the "Cellular Component" category, most genes were annotated to "cell" (GO:0005623) and "organelle" (GO:0043226), followed by "organelle part" (GO:0044422) and "macromolecular complex" (GO:0032991). By contrast, genes associated with the "extracellular region" (GO:0005576) and "membrane-enclosed lumen" (GO:0031974) were less abundant across all species.

Overall, the four biocontrol fungi exhibited a high degree of similarity at the level of GO, KOG and KEGG classification, with differences primarily reflected in the relative abundance of genes within specific functional categories.

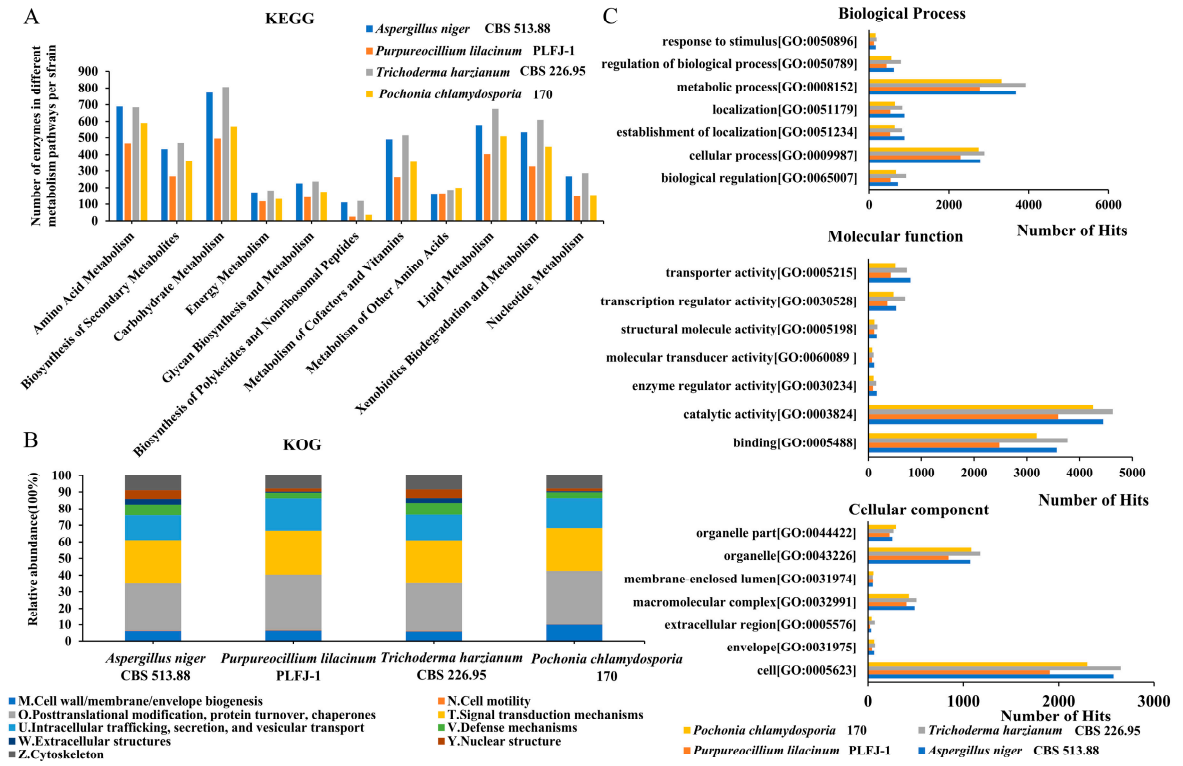

Figure 3. Functional classification of predicted proteins based on KEGG, KOG and GO annotation. (A) KEGG pathway categories. (B) KOG functional classification. (C) GO annotation across biological process, molecular function, and cellular component categories.

2.5. CAZymes, Peptidases and Phylogenetic Analysis of Key Hydrolytic Enzymes

To further explore the enzymatic mechanisms potentially involved in host interaction and biocontrol processes, carbohydrate-active enzymes (CAZymes) and peptidase families were analyzed across the four fungal genomes (Figure 4). Detailed category counts are provided in Supplementary Table S2. These enzymes were mainly assigned to the glycoside hydrolase (GH), glycosyltransferase (GT), carbohydrate esterase (CE), polysaccharide lyase (PL), and carbohydrate-binding module (CBM) families. A total of 2,054 CAZyme genes were identified. *T. harzianum* exhibited relatively higher numbers of GH and CBM family members, whereas *P. lilacinum* showed relative enrichment in GT and GH families (Figure 4A).

Peptidases were systematically identified and classified based on the MEROPS database (Figure 4B). Detailed category counts are provided in Supplementary Table S3. The results showed that all four genomes encode multiple peptidase-related genes, including serine peptidases, metallopeptidases, cysteine peptidases, and aspartic peptidases. Among all analyzed strains, serine

 doi:10.20944/preprints202604.1074.v1

peptidases and metallopeptidases constituted the predominant peptidase types. Notably, *P. lilacinum* and *P. chlamydosporia* harbored relatively higher numbers of genes in the S8 and S9 serine peptidase families. Specifically, *P. lilacinum* and *P. chlamydosporia* encoded 27 and 23 S8 serine peptidases, respectively, representing a marked expansion compared with *T. harzianum* and *A. niger*. In contrast, *T. harzianum* and *A. niger* exhibited higher gene abundances within metallopeptidase families. In addition, cysteine peptidases and aspartic peptidases were present in all four genomes, but their overall numbers were relatively low, and the interstrain differences were less pronounced than those observed for serine peptidases and metallopeptidases.

To investigate the evolutionary relationships of key hydrolytic enzymes associated with parasitism and host degradation, phylogenetic analyses were conducted for S8 serine peptidases and chitinases based on their amino acid sequences. The phylogenetic analysis of S8 serine peptidases showed that these proteins were grouped into several well-supported clades (Figure 4C). Members from *P. lilacinum* and *P. chlamydosporia* tended to cluster within certain clades and formed species-specific monophyletic subgroups, indicative of lineage-specific gene expansion. In contrast, S8 peptidases from *T. harzianum* and *A. niger* were distributed across multiple clades and frequently clustered with homologs from other species, indicating a more conserved evolutionary pattern. Phylogenetic analysis of chitinases revealed a different evolutionary trend (Figure 4D). Most chitinase proteins formed several conserved clades in which sequences from different fungal species were interspersed, reflecting a relatively high degree of evolutionary conservation. Nevertheless, some clades contained a higher proportion of chitinase members from *P. lilacinum* and *P. chlamydosporia*, suggesting moderate species-specific aggregation and limited lineage-specific expansion.

Taken together, the comparative analyses of CAZymes, peptidases, and the phylogenetic relationships of S8 serine peptidases and chitinases revealed both conserved and species-specific features among the four biocontrol fungi. These results provide insights into the potential enzymatic mechanisms underlying host degradation and the biocontrol capabilities of these fungi.

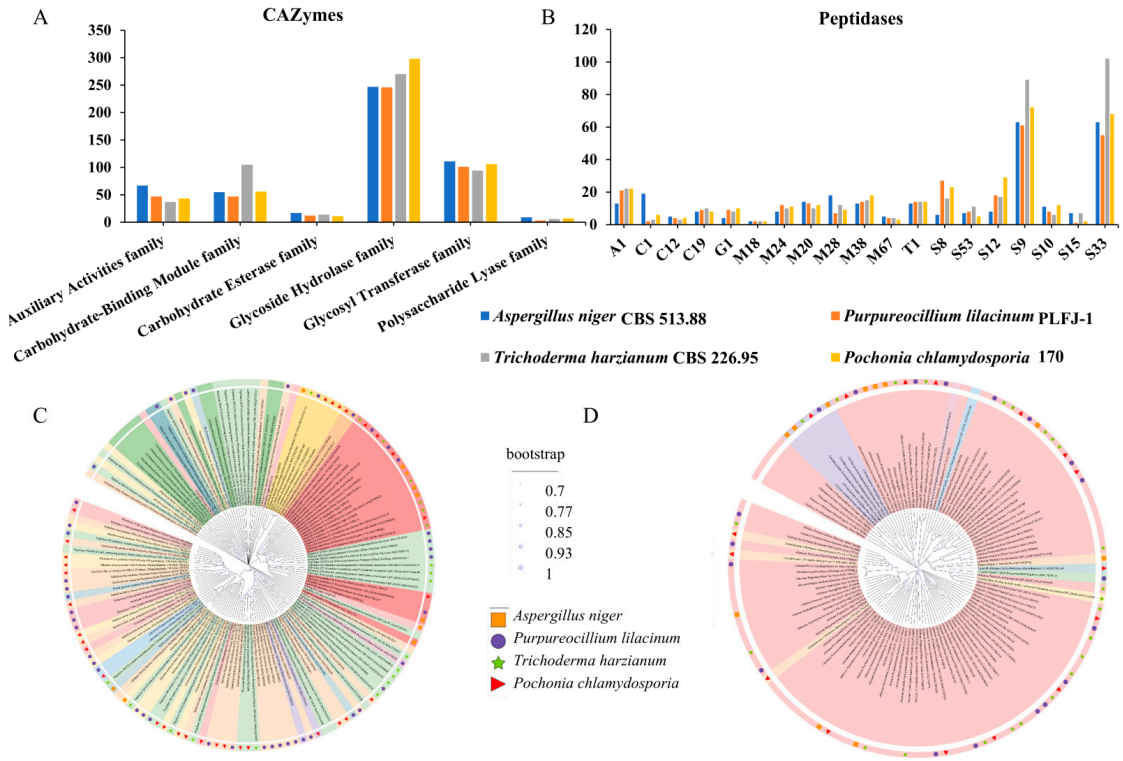

Figure 4. Comparative analysis of CAZymes and peptidases. (A) Bar plot showing the distribution of major CAZyme families. (B) Bar plot showing the distribution of peptidase families. (C) Phylogenetic tree of S8 serine proteases. (D) Phylogenetic tree of chitinases from different fungal species. Phylogenetic trees were constructed using the neighbor-joining method with 1000 bootstrap replicates; bootstrap values >50% are shown at nodes.

 doi:10.20944/preprints202604.1074.v1

2.6. Secondary Metabolite Biosynthetic Gene Clusters

Secondary metabolite biosynthetic gene clusters (BGCs) in the genomes of the four biocontrol fungi were systematically analyzed based on the predictions generated by the antiSMASH platform (Figure 5). Detailed category counts are provided in Supplementary Table S4. The results showed that all four genomes encoded multiple types of secondary metabolite biosynthetic gene clusters, mainly including polyketide synthase (PKS), nonribosomal peptide synthetase (NRPS), and PKS–NRPS hybrid clusters.

Comparative analysis revealed clear differences in both the number and composition of BGCs among the four fungi. A total of 77 (*A. niger*), 35 (*P. lilacinum*), 57 (*T. harzianum*), and 43 (*P. chlamydosporia*) BGCs were identified. Among the four genomes, clear variation was observed in the abundance of NRPS-like and PKS-like gene clusters, with *A. niger* exhibiting the highest numbers in both categories. In contrast, the secondary metabolite gene cluster profiles of the four fungi exhibited distinct distribution patterns, with PKS clusters representing the dominant category in several species. In addition to these major classes, other types of BGCs, including terpene and other hybrid clusters, were also identified across the four genomes, although their abundance varied among species. Overall, the diversity and composition of secondary metabolite gene clusters observed in the four fungal genomes may reflect differences in their metabolic capabilities and ecological adaptation strategies.

These findings highlight the distinct metabolic potentials and secondary metabolite profiles that may contribute to the species-specific biocontrol characteristics of the four fungi. Figure 6 summarizes the mechanistic framework of coordinated S8 proteases, CAZymes, and secondary metabolism underlying nematode biocontrol. This model illustrates the key functional modules associated with nematode antagonism, including the expansion and enrichment of S8 serine peptidases, differential composition of CAZyme families involved in cell wall and eggshell degradation, and species-specific distribution of secondary metabolite biosynthetic gene clusters. The coordinated action of hydrolytic enzymes and bioactive secondary metabolites contributes to fungal penetration, host tissue degradation, and nematocidal activity, representing a conserved yet diversified strategy for nematode biocontrol among these fungi.

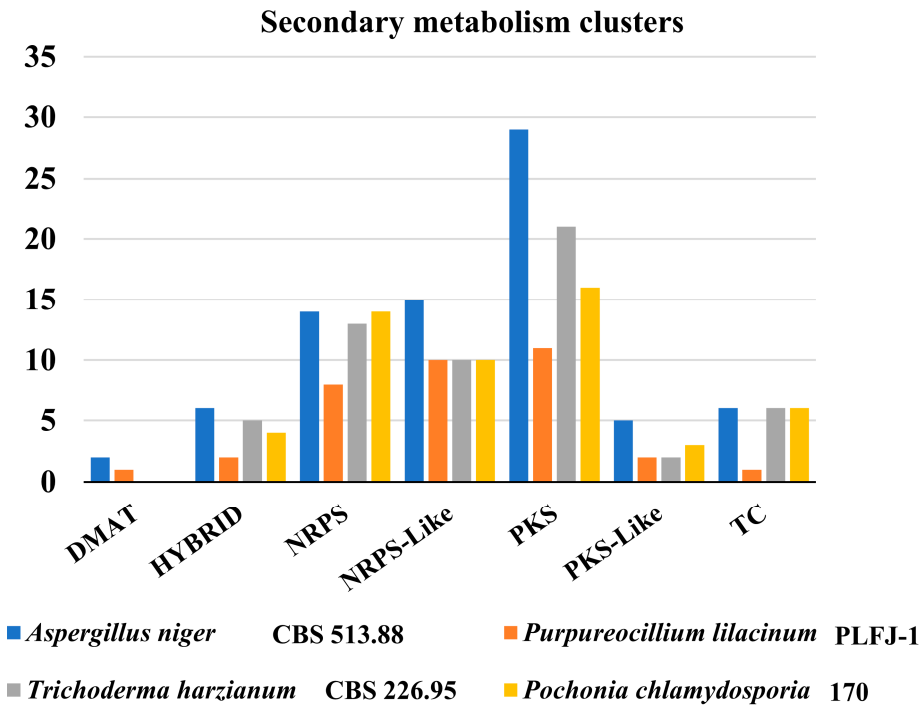

Figure 5. Distribution of secondary metabolite biosynthetic gene clusters in different fungal strains. Numbers of predicted gene clusters, including dimethylallyl tryptophan synthase (DMAT), hybrid, nonribosomal peptide

 doi:10.20944/preprints202604.1074.v1

synthetase (NRPS), NRPS-like, polyketide synthase (PKS), PKS-like, and terpene cyclase (TC) clusters, are shown for *A. niger* CBS 513.88, *P. lilacinum* PLFJ-1, *T. harzianum* CBS 226.95, and *P. chlamydosporia* 170.

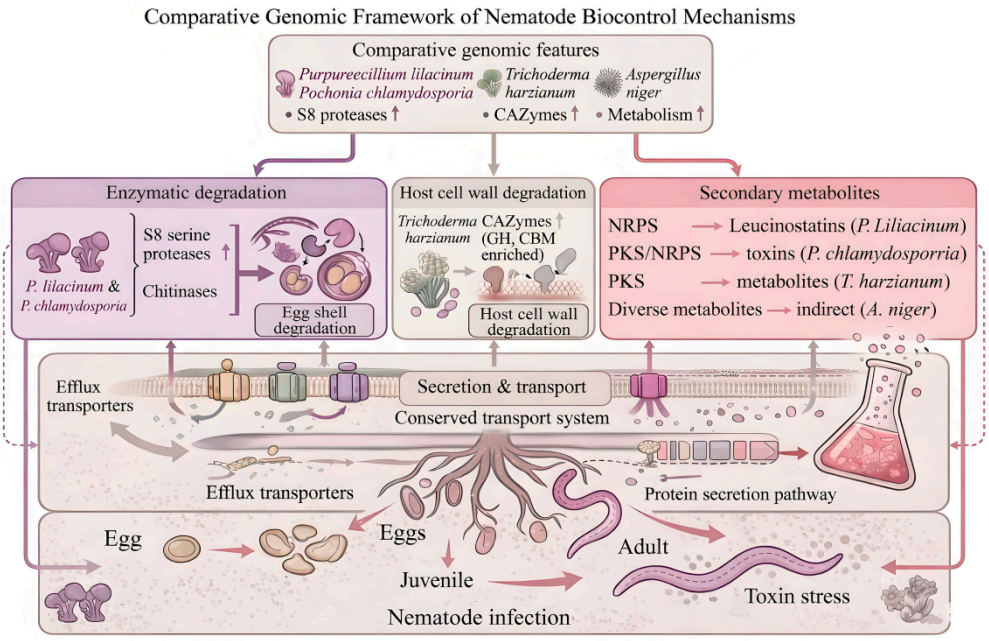

Figure 6. Mechanistic schematic illustrating the coordinated system of S8 proteases, CAZymes, and secondary metabolism underlying nematode biocontrol in *P. lilacinum*, *P. chlamydosporia*, *T. harzianum*, and *A. niger*. This model summarizes the synergistic actions of hydrolytic enzymes and secondary metabolites that mediate fungal penetration, host degradation, and nematocidal activity.

3. Discussion

Comparative genomic analysis revealed substantial variation in genome size, GC content, and gene composition among the four fungi, reflecting their distinct evolutionary trajectories and ecological strategies. The larger genome and higher assembly completeness observed in *P. chlamydosporia are* consistent with previous genomic studies showing that parasitic fungi tend to possess expanded gene repertoires supporting host recognition, penetration, and nutrient acquisition [27,31–33]. In contrast, *A. niger* possesses a relatively compact genome but retains extensive metabolic capacity, a feature widely recognized as characteristic of saprophytic fungi with strong environmental adaptability [23,24]. *P. lilacinum* and *T. harzianum* displayed intermediate genome sizes but differed in GC content and functional gene composition, suggesting that distinct selective pressures have shaped their genomic architectures toward different ecological and biocontrol strategies [25,26].

Phylogenetic analysis further demonstrated that nematode antagonism is not restricted to a single fungal clade, but has evolved independently across multiple fungal groups [27,28]. The close phylogenetic relationship between *P. chlamydosporia* and *P. lilacinum* is consistent with earlier reports suggesting partial conservation of parasitism-related mechanisms within Hypocreales [25,29], whereas *T. harzianum*, despite its phylogenetic distance, highlights the importance of functional gene content rather than phylogenetic proximity alone [30]. Similar patterns of convergent evolution have been documented in other nematode-trapping fungi, such as *Arthrobotrys* oligospora, where distinct trapping mechanisms evolved independently across lineages [27,29].

Functional annotation indicated that core metabolic pathways are conserved across the four fungi, whereas pathways associated with secondary metabolism and environmental interaction exhibit substantial variation [31,32]. The extensive metabolic networks observed in *A. niger* and *T. harzianum* are consistent with previous studies reporting that metabolic versatility enhances

 doi:10.20944/preprints202604.1074.v1

competitive fitness in heterogeneous soil and rhizosphere environments [23,33,34]. Conversely, *Purpureocillium lilacinum* and *Pochonia chlamydosporia* exhibited relatively streamlined metabolic profiles, which may reflect an evolutionary shift toward functional specialization associated with parasitic or host-adapted lifestyles [25,35,36]. Such metabolic specialization has been proposed as a strategy to allocate energetic resources preferentially toward infection-related processes rather than broad substrate utilization [35].

Secondary metabolite gene cluster analysis revealed marked differences in the abundance and composition of PKS, NRPS, and hybrid clusters. The enrichment of these clusters in *P. lilacinum* and *P. chlamydosporia* is consistent with previous genomic studies indicating that nematophagous fungi often rely on bioactive secondary metabolites to suppress or kill their hosts [25,31]. Notably, *P. lilacinum* is known to synthesize leucinostatins, a class of peptide antibiotics with demonstrated biological activity against multiple pathogens, highlighting the potential role in nematode suppression [25,37].

Among the hydrolytic systems, CAZyme profiling showed significant interspecific variation, particularly in families involved in polysaccharide degradation. The enrichment of GH and CBM families in *T. harzianum* is consistent with its well-documented capacity for plant and microbial cell wall degradation [37], whereas the relative abundance of GT and PL families in *P. lilacinum* may facilitate surface modification and host interaction during nematode infection [25]. Together, these findings are consistent with the hypothesis that secondary metabolites and CAZymes may act synergistically during nematode antagonism, but direct evidence from co-expression or gene knockout studies is needed to confirm this relationship [31,38].

Peptidases, particularly S8 serine proteases, play crucial roles in nutrient acquisition, extracellular protein degradation, and interactions with hosts or environmental substrates. The enrichment of serine peptidases, particularly members of S8 and S9 families in *P. lilacinum* and *P. chlamydosporia* are consistent with previous studies demonstrating the role of subtilisin-like peptidases in fungal parasitism of nematodes [34]. In contrast, *T. harzianum* and *A. niger* harbored higher numbers of genes belonging to metallopeptidase families, which are associated with extracellular protein degradation, nutrient acquisition, and biofilm formation, thereby facilitating resource acquisition and environmental adaptation in saprophytic or antagonistic fungi to obtain resources and adapt to complex environments [39]. This differential distribution suggests that distinct biocontrol fungi exhibit species-specific characteristics in the construction of their proteolytic systems and in their functional emphases. In general, the differences in peptidase family composition and abundance distribution among the biocontrol fungi reflect functional diversification in their modes of host interaction, nutrient acquisition strategies, and ecological adaptation pathways. In combination with the analyses of CAZymes and secondary metabolite gene clusters, it can be inferred that peptidases may act synergistically with polysaccharide-degrading enzymes and bioactive secondary metabolites, collectively forming an important molecular basis for fungal participation in biocontrol processes. From a genomic perspective, this study provides new theoretical insights into the action strategies of different biocontrol fungi.

Transporters and transcription factors provide a regulatory and functional framework supporting fungal biocontrol activity. Although no marked expansion of these components was observed, their conserved presence across all four fungi suggests a stable system for metabolite transport and gene regulation. In particular, Zn(2)-Cys(6) transcription factors are known to regulate secondary metabolism and enzyme expression, indicating their potential role in coordinating hydrolytic enzymes and metabolite production during host interaction[40,41].

Taken together, these comparative genomic analyses support a coordinated mechanism in which secondary metabolites, CAZyme repertoires, and enriched protease families act synergistically during nematode infection. Collectively, S8 proteases facilitate degradation of nematode cuticles and eggshell proteins, CAZymes contribute to polysaccharide breakdown, and secondary metabolites enhance nematocidal activity. These components function in a coordinated manner to support the biocontrol activity of *P. lilacinum* (Figure 6).

 doi:10.20944/preprints202604.1074.v1

4. Materials and Methods

4.1. Genome Data Collection

Genome sequences and corresponding annotation files of *P. lilacinum* PLFJ-1, *T. harzianum* CBS 226.95, *A. niger* CBS 513.88 and *P. chlamydosporia* 170 were obtained from publicly available databases, including NCBI Genome database and JGI MycoCosm platform. All genome datasets were processed under a unified framework to ensure comparability across species. Basic genome statistics, including genome size, GC content, scaffold number, and N50 values, were calculated based on the downloaded assemblies.

4.2. Orthologous Gene Family Analysis

Orthologous gene families were identified using OrthoFinder (v3.1.1) with default parameters. Protein sequences from all four species were used as input. The resulting orthogroups were used to identify shared (core) and species-specific gene families. The UpSet plot was generated using the Python library UpSetPlot to visualize intersections among orthogroups.

4.3. Phylogenetic Analysis

Genomes of twelve fungi associated with biological control or phylogenetically related taxa were retrieved from NCBI for comparative analysis, including *A. nidulans* ATCC 10074, *A. flavus* ATCC 16883, *A. oryzae* NRRL 447, *A. fumigatus* ATCC 1022, *A. niger* ATCC 16888, *T. harzianum* BDS2, *P. globispora* CBS 203.86, *P. chlamydosporia* 170, *P. lilacinum* NRRL 895, *B. bassiana* ARSEF 1564, *T. virens* CBS 109339 and *T. reesei*.

These species were selected to represent different ecological strategies relevant to nematode biological control. In particular, *P. chlamydosporia* and *P. lilacinum* are well-known nematophagous fungi capable of parasitizing nematode eggs, while *B. bassiana* represents an entomopathogenic fungus included for comparative analysis of host adaptation. Species of the genus *Trichoderma* (*T. harzianum*, *T. virens*, and *T. reesei*) were included as typical biocontrol fungi with strong capacities for secreting hydrolytic enzymes and secondary metabolites. Additionally, multiple *Aspergillus* species were incorporated as phylogenetically related reference taxa with well-annotated genomes, providing a robust evolutionary framework.

Phylogenetic analysis was performed based on conserved protein sequences. S8 serine proteases and chitinase protein sequences were extracted from each genome. Multiple sequence alignment and phylogenetic tree construction were conducted using the Neighbor-Joining method with 1,000 bootstrap replicates implemented in MEGA. The resulting trees were visualized using iTOL (Interactive Tree of Life).

4.4. Functional Annotation

Protein sequences were functionally annotated against multiple databases: KEGG annotation was performed using KAAS (KEGG Automatic Annotation Server), GO annotation was conducted using Blast2GO and KOG classification was assigned using eggNOG-mapper (v2.1). Functional categories were summarized and visualized using R (ggplot2).

4.5. Identification of CAZymes

Carbohydrate-active enzymes were identified using the dbCAN2 meta server, which integrates HMMER, DIAMOND, and Hotpep methods. Only CAZyme families supported by at least two methods were retained to improve annotation accuracy. Identified CAZymes were classified into GH, GT, CE, PL, and CBM families.

 doi:10.20944/preprints202604.1074.v1

4.6. Identification of Peptidases

Peptidases were annotated using the MEROPS database (https://www.ebi.ac.uk/merops/). Protein sequences were searched against the MEROPS database using BLASTP with an E-value threshold of 1×10^{-5}. Identified peptidases were classified into major categories, including serine, metallopeptidases, cysteine, and aspartic peptidases. S8 serine proteases were further extracted for phylogenetic analysis.

4.7. Secondary Metabolite Gene Cluster Prediction

Secondary metabolite biosynthetic gene clusters were predicted using antiSMASH (v6.0) with default settings. Predicted clusters were classified into PKS, NRPS, PKS-like, NRPS-like, hybrid, and terpene clusters. The distribution of BGC types was summarized and visualized.

4.8. Analysis of Transporters and Transcription Factors

Transporter proteins were annotated using the Transporter Classification Database (TCDB). Transcription factors were identified based on conserved domains using InterProScan (v5). TF families, including Zn(2)-Cys(6), bZIP, C2H2, and helix–loop–helix domains, were classified according to domain composition.

5. Conclusions

Comparative genomic analysis of four biocontrol fungi revealed conserved core metabolism alongside significant divergence in gene families related to host interaction. Variations in CAZymes, peptidases, and secondary metabolite biosynthetic gene clusters indicate distinct strategies for environmental adaptation and nematode antagonism. Phylogenetic and orthogroup analyses suggest that these traits have evolved independently across fungal lineages. Notably, the enrichment of S8 serine peptidases, together with differential CAZyme and secondary metabolite profiles, supports a coordinated mechanism for host infection and degradation. These findings provide a genomic basis for understanding fungal biocontrol and support the development of effective nematode control strategies.

Supplementary Materials: The following supporting information can be downloaded at the website of this paper posted on Preprints.org.

Author Contributions: X.Z., W.H. and R.S.B. conceived and designed the research. X.C., L.L., Z.Z. and M.C. conducted all experimental work, measurements, and analysis. X.C. and L.L. performed experimental analysis and wrote the manuscript. X.Z., W.H., J.L. and W.W. supervised and reviewed the manuscript. The final manuscript was reviewed and approved by all authors.

Funding: This work was supported by the National Natural Science Foundation of China (32400065), Doctoral Fund Project of University of Jinan (XBS2442), Youth Innovation Program Team in Universities of Shandong Province (2023KJ282), and the Project of Shandong Province Higher Educational Youth Innovation Science and Technology Program (2022KJ096).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding authors.

Conflicts of Interest: The authors declare no conflicts of interest.

References

 doi:10.20944/preprints202604.1074.v1

1. Yadav, H.; Roberts, P.A.; Lopez-Arredondo, D. Combating Root-Knot Nematodes (Meloidogyne Spp.): From Molecular Mechanisms to Resistant Crops. *Plants* **2025**, *14*, 1321.
2. Janati, S.; Houari, A.; Wifaya, A.; Essarioui, A.; Mimouni, A.; Hormatallah, A.; Sbaghi, M.; Dababat, A.A.; Mokrini, F. Occurrence of the Root-Knot Nematode Species in Vegetable Crops in Souss Region of Morocco. *Plant Pathol. J.* **2018**, *34*, 308–315.
3. Akhtar Hussain, M.; Parveen, G. Determining the Damage Threshold of Root-Knot Nematode, Meloidogyne Arenaria on Vigna Unguiculata (L.) Walp. *Rhizosphere* **2023**, *27*, 100714.
4. Wang, X.; Wang, J.; Duan, S.; Yan, X.; Wang, Y.; He, X.; Wu, W. Identification and Characterization of Root-Knot Nematodes Infecting Polygonatum Sibiricum and Peucedanum Praeruptorum in China. *Agronomy* **2024**, *14*, 782.
5. Zasada, I.A.; Ferris, H.; Elmore, C.L.; Roncoroni, J.A.; MacDonald, J.D.; Bolkan, L.R.; Yakabe, L.E. Field Application of Brassicaceous Amendments for Control of Soilborne Pests and Pathogens. *Plant Health Prog.* **2003**, *4*, 3.
6. Wram, C.L.; Zasada, I.A. Short-Term Effects of Sublethal Doses of Nematicides on *Meloidogyne Incognita*. *Phytopathology®* **2019**, *109*, 1605–1613.
7. Lian, X.; Liu, S.; Jiang, L.; Bai, X.; Wang, Y. Isolation and Characterization of Novel Biological Control Agent Clostridium Beijerinckii against Meloidogyne Incognita. *Biology* **2022**, *11*, 1724.
8. Win, P.P.; Kyi, P.P.; Maung, Z.T.Z.; Myint, Y.Y.; Cabasan, Ma.T.N.; De Waele, D. Host Status of Rotation Crops in Asian Rice-Based Cropping Systems to the Rice Root-Knot Nematode Meloidogyne Graminicola. *Trop. Plant Pathol.* **2016**, *41*, 312–319.
9. Timper, P.; Strickland, T.C.; Jagdale, G.B. Biological Suppression of the Root-Knot Nematode Meloidogyne Incognita Following Winter Cover Crops in Conservation Tillage Cotton. *Biol. Control* **2021**, *155*, 104525.
10. Kranti, K.V.V.S.; T. R. Kavitha; N.G. Ravichandra Technology Developed through Demonstration on the Management of Root Knot Nematode, Meloidogyne Incognita in Polyhouse Cultivated Cucumber by Soil Solarization. *Ecol. Environ. Conserv.* **2023**, *29*, 673–675.
11. Flores Francisco, B.G.; Ponce, I.M.; Plascencia Espinosa, M.Á.; Mendieta Moctezuma, A.; López Y López, V.E. Advances in the Biological Control of Phytoparasitic Nematodes via the Use of Nematophagous Fungi. *World J. Microbiol. Biotechnol.* **2021**, *37*, 180.
12. Degenkolb, T.; Vilcinskas, A. Metabolites from Nematophagous Fungi and Nematicidal Natural Products from Fungi as an Alternative for Biological Control. Part I: Metabolites from Nematophagous Ascomycetes. *Appl. Microbiol. Biotechnol.* **2016**, *100*, 3799–3812.
13. Rahman, M.U.; Chen, P.; Zhang, X.; Fan, B. Predacious Strategies of Nematophagous Fungi as Bio-Control Agents. *Agronomy* **2023**, *13*, 2685.
14. Liu, X.; Xiang, M.; Che, Y. The Living Strategy of Nematophagous Fungi. *Mycoscience* **2009**, *50*, 20–25.
15. El-Marzoky, A.M.; Elnahal, A.S.M.; Jghef, M.M.; Abourehab, M.A.S.; El-Tarabily, K.A.; Ali, M.A.M.S. Purpureocillium Lilacinum Strain AUMC 10620 as a Biocontrol Agent against the Citrus Nematode Tylenchulus Semipenetrans under Laboratory and Field Conditions. *Eur. J. Plant Pathol.* **2023**, *167*, 59–76.
16. Anastasiadis, I.A.; Giannakou, I.O.; Prophetou-Athanasiadou, D.A.; Gowen, S.R. The Combined Effect of the Application of a Biocontrol Agent Paecilomyces Lilacinus, with Various Practices for the Control of Root-Knot Nematodes. *Crop Prot.* **2008**, *27*, 352–361.
17. Singh, S.; Pandey, R.K.; Goswami, B.K. Bio-Control Activity of *Purpureocillium Lilacinum* Strains in Managing Root-Knot Disease of Tomato Caused by *Meloidogyne Incognita*. *Biocontrol Sci. Technol.* **2013**, *23*, 1469–1489.
18. Hajji, L.; Hlaoua, W.; Regaieg, H.; Horrigue-Raouani, N. Biocontrol Potential of Verticillium Leptobactrum and Purpureocillium Lilacinum Against Meloidogyne Javanica and Globodera Pallida on Potato (Solanum Tuberosum). *Am. J. Potato Res.* **2017**, *94*, 178–183.
19. Khan, A.; Williams, K.L.; Nevalainen, H.K.M. Infection of Plant-Parasitic Nematodes by Paecilomyces Lilacinus and Monacrosporium Lysipagum. *Biocontrol* **2006**, *51*, 659–678.
20. Bonants, P.J.M.; Fitters, P.F.L.; Thijs, H.; Belder, E.D.; Waalwijk, C.; Henfling, J.W.D.M. A Basic Serine Protease from Paecilomyces Lilacinus with Biological Activity against Meloidogyne Hapla Eggs. *Microbiology* **1995**, *141*, 775–784.

 doi:10.20944/preprints202604.1074.v1

21. Khan, A.; Williams, K.; Molloy, M.P.; Nevalainen, H. Purification and Characterization of a Serine Protease and Chitinases from Paecilomyces Lilacinus and Detection of Chitinase Activity on 2D Gels. *Protein Expr. Purif.* **2003**, *32*, 210–220.
22. Cavello, I.A.; Hours, R.A.; Cavalitto, S.F. Bioprocessing of "Hair Waste" by *Paecilomyces Lilacinus* as a Source of a Bleach-Stable, Alkaline, and Thermostable Keratinase with Potential Application as a Laundry Detergent Additive: Characterization and Wash Performance Analysis. *Biotechnol. Res. Int.* **2012**, *2012*, 1–12.
23. Pel, H.J.; De Winde, J.H.; Archer, D.B.; Dyer, P.S.; Hofmann, G.; Schaap, P.J.; Turner, G.; De Vries, R.P.; Albang, R.; Albermann, K.; et al. Genome Sequencing and Analysis of the Versatile Cell Factory Aspergillus Niger CBS 513.88. *Nat. Biotechnol.* **2007**, *25*, 221–231.
24. Cairns, T.C.; Nai, C.; Meyer, V. How a Fungus Shapes Biotechnology: 100 Years of Aspergillus Niger Research. *Fungal Biol. Biotechnol.* **2018**, *5*, 13.
25. Prasad, P.; Varshney, D.; Adholeya, A. Whole Genome Annotation and Comparative Genomic Analyses of Bio-Control Fungus Purpureocillium Lilacinum. *BMC Genomics* **2015**, *16*, 1004.
26. Xie, J.; Li, S.; Mo, C.; Xiao, X.; Peng, D.; Wang, G.; Xiao, Y. Genome and Transcriptome Sequences Reveal the Specific Parasitism of the Nematophagous Purpureocillium Lilacinum 36-1. *Front. Microbiol.* **2016**, *7*.
27. Meerupati, T.; Andersson, K.-M.; Friman, E.; Kumar, D.; Tunlid, A.; Ahrén, D. Genomic Mechanisms Accounting for the Adaptation to Parasitism in Nematode-Trapping Fungi. *PLoS Genet.* **2013**, *9*, e1003909.
28. Su, H.; Zhao, Y.; Zhou, J.; Feng, H.; Jiang, D.; Zhang, K.-Q.; Yang, J. Trapping Devices of Nematode-Trapping Fungi: Formation, Evolution, and Genomic Perspectives: Trapping Devices of Nematode-Trapping Fungi. *Biol. Rev.* **2017**, *92*, 357–368.
29. Yang, J.; Wang, L.; Ji, X.; Feng, Y.; Li, X.; Zou, C.; Xu, J.; Ren, Y.; Mi, Q.; Wu, J.; et al. Genomic and Proteomic Analyses of the Fungus Arthrobotrys Oligospora Provide Insights into Nematode-Trap Formation. *PLoS Pathog.* **2011**, *7*, e1002179.
30. Karlsson, M.; Durling, M.B.; Choi, J.; Kosawang, C.; Lackner, G.; Tzelepis, G.D.; Nygren, K.; Dubey, M.K.; Kamou, N.; Levasseur, A.; et al. Insights on the Evolution of Mycoparasitism from the Genome of Clonostachys Rosea. *Genome Biol. Evol.* **2015**, *7*, 465–480.
31. Keller, N.P.; Turner, G.; Bennett, J.W. Fungal Secondary Metabolism — from Biochemistry to Genomics. *Nat. Rev. Microbiol.* **2005**, *3*, 937–947.
32. De Vries, R.P.; Riley, R.; Wiebenga, A.; Aguilar-Osorio, G.; Amillis, S.; Uchima, C.A.; Anderluh, G.; Asadollahi, M.; Askin, M.; Barry, K.; et al. Comparative Genomics Reveals High Biological Diversity and Specific Adaptations in the Industrially and Medically Important Fungal Genus Aspergillus. *Genome Biol.* **2017**, *18*, 28.
33. Gibbons, J.G.; Rokas, A. The Function and Evolution of the Aspergillus Genome. *Trends Microbiol.* **2013**, *21*, 14–22.
34. Morton, O.; Hirsch, P.; Kerry, B. Infection of Plant-Parasitic Nematodes by Nematophagous Fungi – a Review of the Application of Molecular Biology to Understand Infection Processes and to Improve Biological Control. *Nematology* **2004**, *6*, 161–170.
35. Larriba, E.; Jaime, M.D.L.A.; Carbonell-Caballero, J.; Conesa, A.; Dopazo, J.; Nislow, C.; Martín-Nieto, J.; Lopez-Llorca, L.V. Sequencing and Functional Analysis of the Genome of a Nematode Egg-Parasitic Fungus, Pochonia Chlamydosporia. *Fungal Genet. Biol.* **2014**, *65*, 69–80.
36. Spatafora, J.W.; Chang, Y.; Benny, G.L.; Lazarus, K.; Smith, M.E.; Berbee, M.L.; Bonito, G.; Corradi, N.; Grigoriev, I.; Gryganskyi, A.; et al. A Phylum-Level Phylogenetic Classification of Zygomycete Fungi Based on Genome-Scale Data. *Mycologia* **2016**, *108*, 1028–1046.
37. Kubicek, C.P.; Herrera-Estrella, A.; Seidl-Seiboth, V.; Martinez, D.A.; Druzhinina, I.S.; Thon, M.; Zeilinger, S.; Casas-Flores, S.; Horwitz, B.A.; Mukherjee, P.K.; et al. Comparative Genome Sequence Analysis Underscores Mycoparasitism as the Ancestral Life Style of Trichoderma. *Genome Biol.* **2011**, *12*, R40.
38. Keller, N.P. Fungal Secondary Metabolism: Regulation, Function and Drug Discovery. *Nat. Rev. Microbiol.* **2019**, *17*, 167–180.
39. Yike, I. Fungal Proteases and Their Pathophysiological Effects. *Mycopathologia* **2011**, *171*, 299–323.

 doi:10.20944/preprints202604.1074.v1

40. Song, H.-J.; Li, X.-F.; Pei, X.-R.; Sun, Z.-B.; Pan, H.-X. Transcription Factors in Biocontrol Fungi. *J. Fungi* **2025**, *11*, 223.
41. Yin, W.; Keller, N.P. Transcriptional Regulatory Elements in Fungal Secondary Metabolism. *J. Microbiol.* **2011**, *49*, 329–339.