Preprints.org

Article

Not peer-reviewed version

Genome Sequence, Comparative Genome Analysis and Expression Profiling of Chitinase GH18 Gene Family in Cordyceps javanica Bd01

Tao Zhu , Mehboob Hussain , Jingyi Ning , Xiao Chen , Chunlan Shi , Dewei Yang , Xi Gao , Guoxing Wu *

Posted Date: 28 January 2025

doi: 10.20944/preprints202501.2089.v1

Keywords: Cordyceps javanica; GH18; phylogenetic; conservatism

Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 January 2025 doi:10.20944/preprints202501.2089.v1

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Genome Sequence, Comparative Genome Analysis and Expression Profiling of Chitinase GH18 Gene Family in *Cordyceps javanica* Bd01

Tao Zhu, Mehboob Hussain, Jingyi Ning, Xiao Chen, Chunlan Shi, Dewei Yang, Xi Gao and Guoxing Wu *

State Key Laboratory for Conservation and Utilization of Bio-Resources in Yunnan, Yunnan Agricultural University, Kunming 650201, China.

* Correspondence: chonchon@ynau.edu.cn; Tel.: +8613608849697

Abstract: The fungus *Cordyceps javanica* is known for entomopathogenicity and effective in control of various arthropods. Here, we aimed to reveal chitinase GH18 gene family expansion through high put sequencing of genome of *C. javanica* strain Bd01 isolated from *Xylotrechus quadripes* larvae. The genome was 34 Mb in size with 9,590 protein-coding genes. By comparative genome analysis, it was found that family GH18 of chitinase genes was expanded in *C. javanica* Bd01. Phylogenetic analysis of twenty-seven GH18 genes with those of 4 other species showed that genes were clustered into three groups based on their conserved domains. The GH18 genes clustered in same group had the same protein motifs and orthologs. The molecular mass of GH18 genes varied from 14.03 kDa - 81.41 kDa, and theoretical pI from 4.40 - 7.92. Most of the chitinases were extracellular, hydrophilic, thermostable and negatively charged with good in vivo half-life. Furthermore, a three-dimensional model of GH18 was constructed using SWISS-MODEL server. These findings provide comprehensive insights into pathogenicity-related genes, their role and conservatism.

Keywords: *Cordyceps javanica*; GH18; phylogenetic; conservatism

1. Introduction

Entomopathogenic fungi (EPF) have a broad host range and represents one of the largest groups of entomopathogenic microorganisms. They are also among the earliest entomopathogenic microbes applied in pest control. The fungus *Cordyceps javanica* (Formerly known as *Isaria javanica*) belongs to the family *Cordycipitaceae* (*Ascomycota*: *Sordariomycetes*: *Hypocreales*) [1]. According to the existing reports, it has shown a great biocontrol potential; it can control 11 species across 10 genera of insects. Pathogenicity tests have demonstrated that this fungus exhibits strong lethal effects on various insect species across different orders. It has shown great toxicity against several insect pest of the order Hemiptera and induced 100% mortality in *Diaphorina citri* [2], the *Bemisia tabaci* [3]and *Bemisia argentifolii* [4], as well as aphids such as the *Hyalopterus pruni* and the *Aphis pomi* [5], and the *Empoasca vitis* [6]. Additionally, it is effective against the *Thrips palmi* (Thysanoptera) [7] and red imported fire ant *Solenopsis invicta* (*Hymenoptera*) [8]. In the case of the order Isoptera, it is also lethal to the *Coptotermes gestroi* [9]. Furthermore, the fungus has demonstrated good insecticidal effects on several Lepidopteron species, including the *Mamestra brassicae* [10], and the *Lymantria dispar* [11].

Entomopathogenic fungi initiates the fungal infection through insect cuticles adhesion by secreting various enzymes, mainly chitinases, proteases, and lipases. These enzymes are crucial determinants of pathogenicity to the insect hosts [12–16]. The chitinases of pathogenic fungi play important roles not only in spore germination, septum formation, cell division, and morphogenesis but also in host interactions [17–20]. Based on amino acid sequence similarity, chitinases are classified into two families: glycoside hydrolase 18 and 19 [21]. The GH18 chitinase gene family is widely distributed among bacteria, fungi, viruses, animals, and higher plants [22–25]. GH18 chitinase play a

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 January 2025 doi:10.20944/preprints202501.2089.v1

role in deglycosylation of glycoproteins [26,27]. Until now, 75 chitinases of entomopathogenic fungi have been characterized from *M. anisopliae*, *B. bassiana* and *I. fumosorosea.* The chi2 and chi3, identified from *M. anisopliae*, were involved in the pathogenicity of the fungus [28,29]. Overproduction of Bbchit1 found in *B. bassiana* enhanced the virulence of *B. bassiana* for aphids [30]. *The* Ifchit1 of *I. fumosorosea* was involved in fungal growth and virulence functions, while the Ifu-chit2 may play a role in the early stage of *pathogenesis* [31–33]. However, little is known about chitinases and other entomopathogenic-related genes in the fungus *C. javanica*, and there is no information available in the literature about the structural and functional properties of *C. javanica* chitinases.

In this study, we aimed to reveal the chitinase GH18 gene family evolution based on sequencing and assembling the genome of *C. javanica* Bd01 isolated from the larvae of *Xylotrechus quadripes*. Furthermore, we performed a comparative genome analysis of *C. javanica* Bd01 genome with other available genomes of entomopathogens. Moreover, the phylogenetic relationships, structural domains, motifs, and protein models of the *C. javanica* Bd01 GH18 chitinase were analyzed. This would provide additional information on the genomic background of entomopathogenic fungi, facilitating a deeper understanding of their biological functions. These findings would serve as a reference for future protein manipulation and can be used for further studies on their mechanisms of action.

2. Results

2.1. Genome Sequencing, Assembly and Annotation

Complete genome sequencing of *C. javanica* Bd01 using Novaseq 6000 platforms (Illumina, USA) revealed a total number of 78.1 million clean reads (~229.6 × coverage). The assembled genome represents 98.28% completeness (based on BUSCO analysis) with a contig N50 length of 5.4 Mb. The resulting genome assembly was based on 8 contigs with a 53.18% of GC content, 9590 protein-coding genes, 43 rRNA genes, and 131 tRNA genes. The repetitive sequence 0.34Mb, represents 1.04% of the total genome (Figure 1; Table 1).

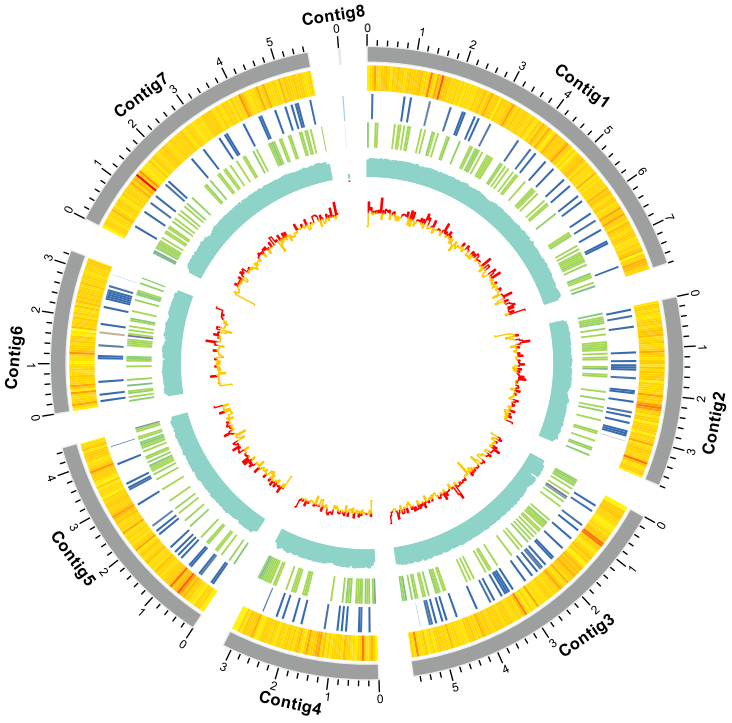


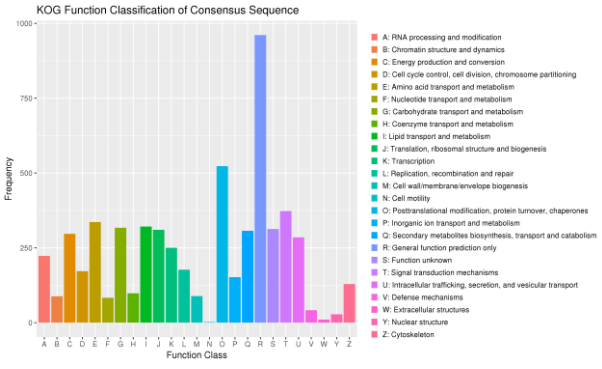
Figure 1. Circos map of the 8 contigs for *Cordyceps javanica* strain Bd01. **Note:** From inside to outside: Gene Density, Non-coding RNA Density, Repetitive Sequence Coverage, GC Percentage, GC Skew.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 January 2025 doi:10.20944/preprints202501.2089.v1

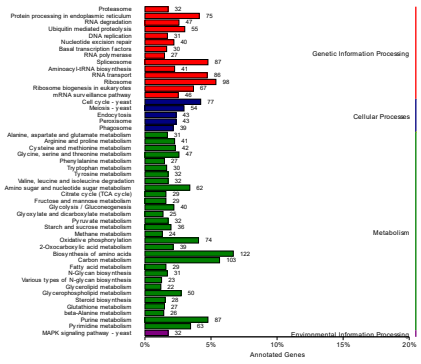
Table 1. Genome features of *Cordyceps javanica* Bd01.

Sample	*Cordyceps javanica* Bd01
Coverage (fold)	229.59X
No. of all sequence	698,137
Bases in all sequence (bp)	7,808,156,751
Largest length (bp)	193,316
N50 Len (bp)	18,161
N90 Len (bp)	4,902
G+C content	53.18
No. of all contigs	8
Contig Length (bp)	34,008,118
No. of large contigs (> 1000 bp)	8
Bases in large contigs (bp)	34,008,118
Contig N50 (bp)	5,679,179
Contig N90 (bp)	3,285,737
Gene num	9,590
Gene total length (bp)	17,091,976
Gene average length (bp)	1,782.27
ExonLen (bp)	15,026,841
AveExonLen (bp)	568.12
ExonNum	26,450
AveExonNum	2.76
rRNA	43
tRNA	131

Pfam domains were assigned to 9,590 proteins based on InterProscan program. Subsequently, a total of 5,236 proteins were assigned to the eukaryotic orthologous groups (KOG) databases (Supplementary Table S1; Figure 2A). The abundance of metabolism was about 50%, which was the highest in these four categories. 3111 genes (32.7%) were annotated using KEGG map, and KEGG annotations contained four major pathways including metabolism (41.2%), genetic information processing (24.5%), cellular processes (8.2%), environmental information processing (1%) (Supplementary Table S2; Figure 2B). Gene ontology (GO) terms were divided into three major function categories: molecular functions (51.1%), cellular components (30.6%) and biological processes (18.3%). The 6,990 genes (73.7%) were annotated and assigned to the three categories, 35 genes were classified into the secondary metabolite biosynthetic process (Supplementary Table S3; Figure 2C), and 1096 genes were unknown function in KOG, GO, and KEGG.

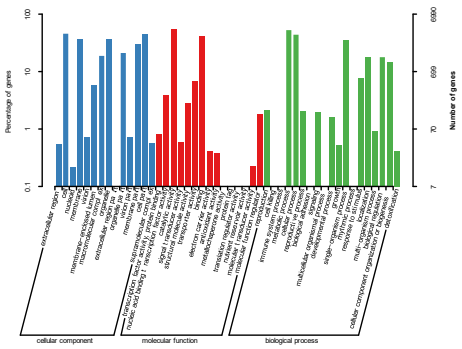


A



B

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 January 2025 doi:10.20944/preprints202501.2089.v1

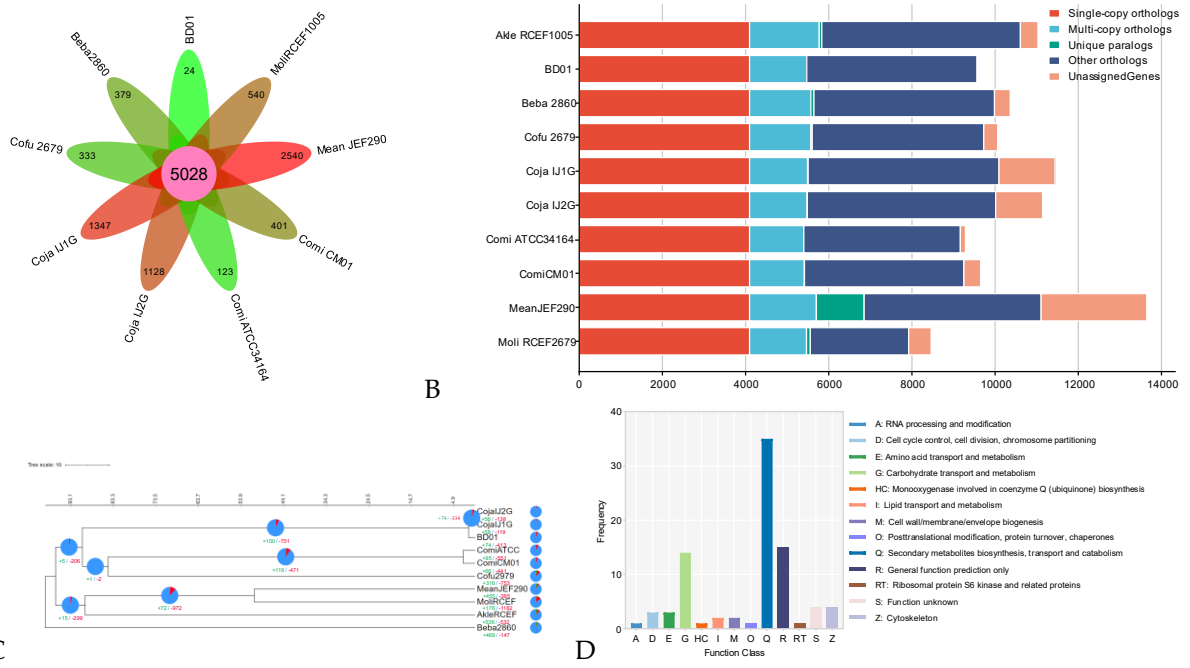


C

Figure 2. (A) KOG functional annotation classification of *Cordyceps javanica* Bd01; (B) Classification statistics of KEGG annotations of *Cordyceps javanica* Bd01; (C) Statistics of GO annotation classification of *Cordyceps javanica* Bd01. **Note:** The horizontal coordinate is the classification content of KOG, and the vertical coordinate is the number of genes. In different functional categories, the proportion of genes reflects the metabolic or physiological bias in the corresponding period and environment, which can be scientifically explained in combination with the distribution of research objects in each functional category.

2.2. Comparative Genomic Analysis

To evaluate the genome evolution across *C. javanica* Bd01 and its related entomopathogenic fungal strains and to identify factors underlying lineage diversification, we evaluated changes associated with fungal evolution in relation to Expansion and contraction of gene families, as well as gene duplications and deletions that occurred within the gene families. By OrthoFinder analysis, among the ten fungal genomes (*C. javanica* Bd01, *C. javanica* IJ1, *C. javanica* IJ2, *C. fumosorosea* ARSEF2679, *C. militaris* CM01, *C. militaris* ATCC34164, *Akanthomyces lecanii* RCEF1005, *B. bassiana* ARSEF2860, *M. libera* RCEF2490 and *M. anisopliae* JEF290), a total of 11,781 orthologous groups were constructed, and 42.68% (5028) of which were shared among the ten strains (Figure 3A, 3B), indicating their close evolutionary relationships.



Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 January 2025 doi:10.20944/preprints202501.2089.v1

Figure 3. (A,B) Comparative analysis of homologous genomes; (C) Phylogenetic tree of *Cordyceps javanica* Bd01 with different strains; (D) KOG functional annotation classification of expansion gene in *Cordyceps javanica* Bd01. **Note:** The green and red numbers in the evolutionary tree represent the number of expansion genes and contraction genes at each node, respectively.

Then, based on the amino acid sequences of 4,096 single-copy orthologous groups, we obtained the maximum likelihood phylogenetic tree of the ten fungal strains. It was shown that *C. fumosorosea* and *C. militaris* formed a branch, independent from the *C. javanica* branch (Figure 3C; Supplementary Table S4), indicating that the three species have a closer relationship with each other than with other species. The number of gene family expansion and contraction was mapped to the phylogenetic tree (Figure 3C). 74 and 413 gene families were identified as expansion and contraction in the fungus *C. javanica* Bd01, respectively (Figure 3C). To investigate the function of gene family expansion in *C. javanica* Bd01 during the evolutionary process, the present study was conducted to annotate the KOG function of the gene family expanded at the differentiation node of *C. javanica* Bd01 and other related species of the genus *Cordyceps*. Based on KOG annotation (Figure 3D), in addition to the unknown functions, the gene families for expansion in *C. javanica* Bd01 had the highest proportion of gene families annotated to the ABC transporter. Using CAZy database models, the number of genes for glycoside hydrolases in the *C. javanica* Bd01 genome was markedly greater than that in other entomopathogenic fungal genomes (Supplementary Table S5), especially in family GH18 and family CBM50.

2.3. Genome-Wide Identification and Characterization of GH18 Genes in Cordyceps javanica Bd01

A total of 27 putative GH18 gene sequences were obtained from the genome of *C. javanica* Bd01 after an HMM of CAZy, InterPro, and SwissProt databases. All candidate GH18 sequences were further analyzed using the CDD and SMART databases to confirm the presence of conserved domains. Twenty-seven putative GH18 genes were obtained after eliminating short length (100 bp) and low identity sequences (Supplementary Table S6).

2.4. Conserved Motif and Chromosomal Location of GH18 Genes in Cordyceps javanica Bd01

In addition to the predictions of subcellular localizations and the analysis of trans-membrane domains, the evaluation of conservative motifs is one of the important means for the functional prediction of GH18 proteins. In this study, a sum of 15 conserved motifs (motifs 1–15) were predicted using MEME online software (Figure 5A). Several motifs were found in most GH18 proteins, such as motifs 1, 2, 3, 7, and 10. Additionally, Cordyceps0G049700.1 contained motifs 1–15 and motif 3 existed in all GH18 members, whereas Cordyceps0G070700.1 only included one member (Figure 5A).

Most of GH18 proteins within the same clan exhibited similar motif components while a high discrepancy was observed among different clans, indicating that the GH18 members within the same family may undertake semblable functions, and that some motifs may play a vital role in the family-specific functions. The chromosomal location of each GH18 gene in the *C. javanica* Bd01 genome is shown in Figure 5B. A total of 27 GH18 are unevenly distributed on 7 contigs. Out of the 27 GH18, seven were mapped onto contig 1, followed by 6 on contig 4, whereas only 1 GH18 were located on contig 3. The GH18 with relatively high densities were detected on the top and middle arms of the contig.

2.5. Phylogenetic Analysis of GH18 Family Proteins

To predict the functions and better understand the evolutionary relationships of GH18 chitinase proteins among different species, a phylogenetic tree was constructed using the 27 GH18 chitinase protein sequences of *C. javanica* Bd01 and 63 GH18 protein sequences of 4 other fungal species (*C. sinensis*, *C. militaris*, *C. cicadae*, and *B. bassiana*). The phylogenetic tree clustered the 27 GH18 chitinase protein sequences of *C. javanica* Bd01 into three major groups (A, B, and C) (Figure 5). This result was consistent with that using the conserved domains (figure 4A).

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 January 2025 doi:10.20944/preprints202501.2089.v1

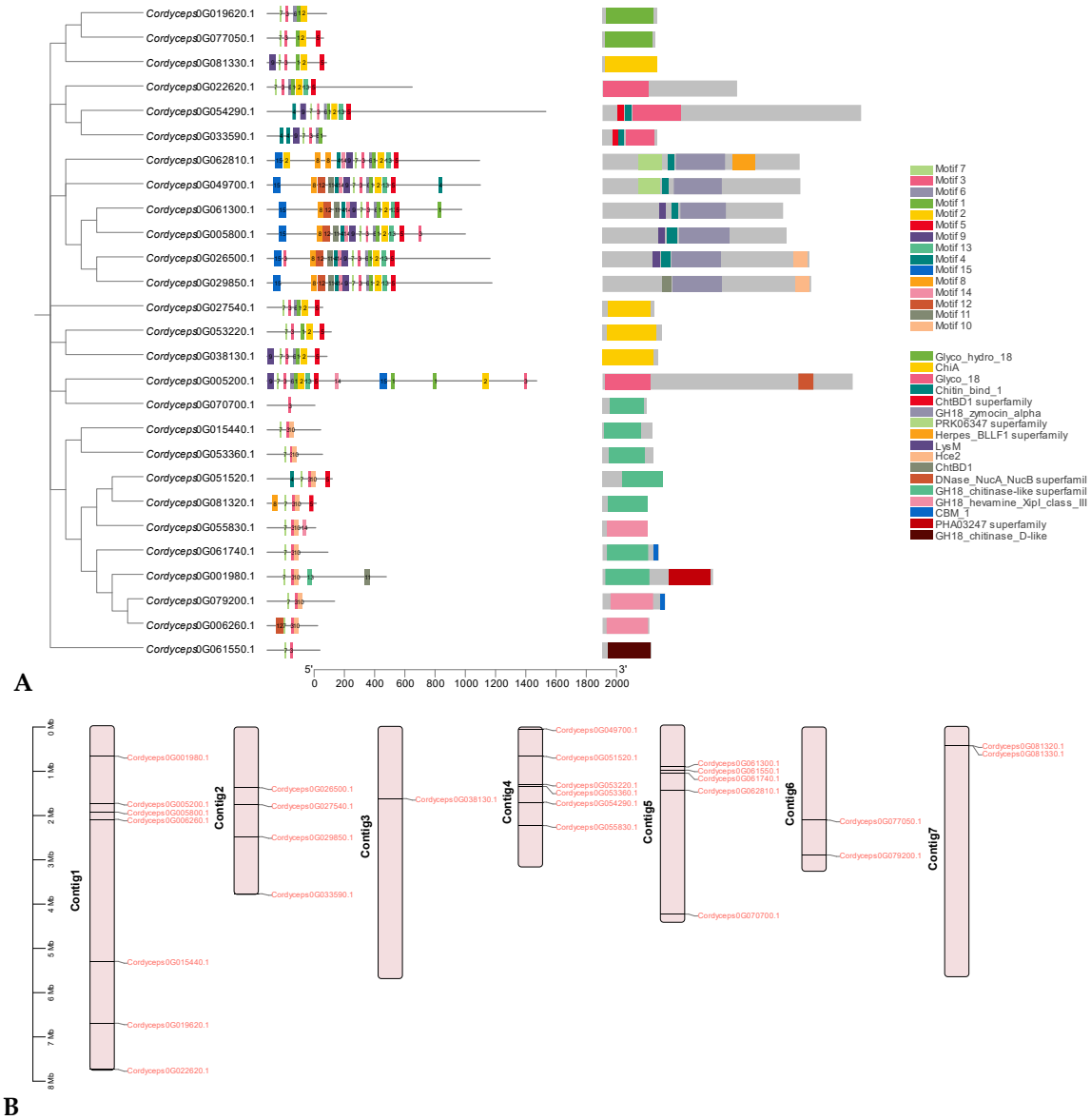


Figure 4. Conserved motif (A) and chromosomal location (B) of GH18 genes in *Cordyceps javanica* Bd01.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 January 2025 doi:10.20944/preprints202501.2089.v1

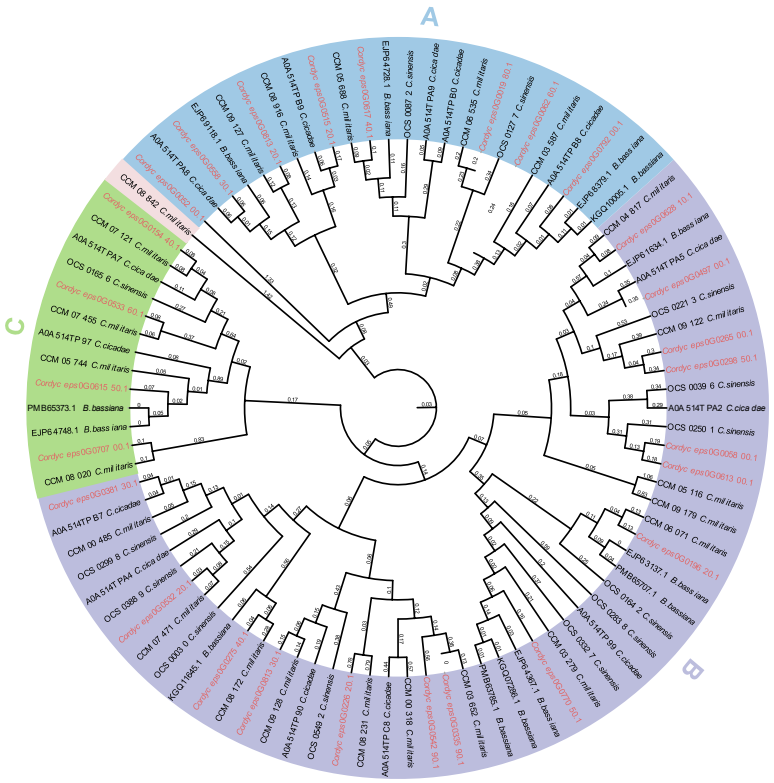


Figure 5. Phylogenetic relationship of GH18 from *Cordyceps javanica* Bd01 and four other fungal species. **Note:** Maximum likelihood trees were constructed based on an alignment of GH18 family catalytic domain amino acid sequences. Percent bootstrap support from 1,000 iterations. *Cordyceps javanica* Bd01 GH18 genes are marked as red typeface(s).

2.6. Characterisation of Physicochemical Properties

The predicted physicochemical properties as computed by the ProtParam tool showed that the chitinases under investigation from *C. javanica* Bd01 had a varied number of amino acids (315 - 1842) and molecular weights (14.03 kDa – 81.41 kDa) (Table 2). The theoretical pI of all the chitinases ranged from acidic pH (4.40) to slightly alkaline pH (7.92). Analyses of the sequences showed that the number of negatively charged amino acid residues (aspartate and glutamate) was higher than the number of positive residues (arginine and lysine) in all the strains, except *Cordyceps*0G055830.1 in *C. javanica* Bd01 (Table 2). The instability index, the measure of the in vivo half-life of a protein, of most of the analysed chitinases were under 40. GRAVY values for all the chitinase were predicted to be negative (Table 2). This variable is used for the prediction of the hydrophobicity or hydrophilicity of a protein over the entire amino acid sequence. The aliphatic index of the chitinases in this study ranged between 59.48 to 89.41. Aliphatic index of 30% chitinases was above 75, while another 40% chitinases were between 65-75 and the rest 30% was around 64. Due to the wide variation in the aliphatic index of chitinases from the same source, *C. javanica* Bd01, it is predicted that the enzymes had varying temperature optima and thermostability. The amino acid composition of chosen chitinases from *C. javanica* Bd01 was evaluated and profiled as depicted in Figure 6A. The results substantiated that alanine and glycine were the most widely distributed amino acids in the selected chitinases. These protein sequences were also rich with amino acids such as leucine, serine, threonine, aspartic acid, valine, proline and isoleucine. Secondary structures of proteins are governed by the inclination of amino acids to form either helix or sheet form.

Table 2. Physico-chemical properties of the GH18 gene family in *Cordyceps javanica* Bd01.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 January 2025 doi:10.20944/preprints202501.2089.v1

Gene name	Number of amino acids	Molecular weight (kDa)	Total number of negatively charged residues (Asp+Glu)	Total number of positively charged residues (Arg+Lys)	pI	Instability index	Aliphatic Index	GRAVY	subcellular localization
*Cordyceps*0G001980.1	787	81.41	49	47	7.92	39.67	63.52	-0.206	extr
*Cordyceps*0G005200.1	1782	19.91	249	206	5.33	40.35	62.25	-0.604	nucl
*Cordyceps*0G005800.1	1310	14.35	157	126	5.55	44.52	76.86	-0.268	plas
*Cordyceps*0G006260.1	333	35.07	29	17	4.52	33.17	68.68	-0.186	extr
*Cordyceps*0G015440.1	353	38.95	50	31	4.73	43.44	89.41	-0.194	mito
*Cordyceps*0G019620.1	392	43.82	57	43	5.03	48.63	78.65	-0.329	mito
*Cordyceps*0G022620.1	958	10.54	107	92	5.58	45.09	64.73	-0.421	extr
*Cordyceps*0G026500.1	1473	15.9	180	148	5.39	28.36	63.65	-0.443	extr
*Cordyceps*0G027540.1	367	39.06	33	28	6.11	46.74	83.32	-0.126	extr
*Cordyceps*0G029850.1	1486	16.06	182	143	5.2	30.46	67.27	-0.412	extr
*Cordyceps*0G033590.1	389	42.6	36	32	6.2	33.37	64.73	-0.456	extr
*Cordyceps*0G038130.1	395	44.06	54	33	4.72	37.83	67.72	-0.508	extr
*Cordyceps*0G049700.1	1409	15.1	138	125	5.83	43.01	65.44	-0.376	extr
*Cordyceps*0G051520.1	430	45.22	45	36	5.44	30.03	73.35	-0.231	extr
*Cordyceps*0G053220.1	423	46.19	39	38	6.53	26.44	73.22	-0.319	extr
*Cordyceps*0G053360.1	365	39.84	31	25	5.74	38.65	80.47	-0.147	extr
*Cordyceps*0G054290.1	1842	20.43	251	204	5.63	36.42	72.49	-0.461	extr
*Cordyceps*0G055830.1	320	34.08	29	30	7.53	24.01	76.34	-0.218	extr
*Cordyceps*0G061300.1	1286	14.03	150	127	5.68	43.02	73.79	-0.265	extr
*Cordyceps*0G061550.1	348	36.71	24	23	6.12	31.95	84.45	0.009	extr
*Cordyceps*0G061740.1	401	42.25	28	24	4.94	36.66	59.48	-0.183	extr
*Cordyceps*0G062810.1	1404	15.14	140	104	5.3	41.7	70.89	-0.236	mito

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 January 2025 doi:10.20944/preprints202501.2089.v1

*Cordyceps*0G070700.1	315	34.54	35	32	5.69	28.53	78.44	-0.320	extr
*Cordyceps*0G077050.1	372	41.41	52	47	5.63	33.53	64.87	-0.483	extr
*Cordyceps*0G079200.1	445	46.94	24	23	6.29	41.27	61.21	-0.290	extr
*Cordyceps*0G081320.1	324	33.98	31	19	4.4	37.51	80.46	-0.108	extr
*Cordyceps*0G081330.1	392	43.13	34	31	5.5	35.96	73.67	-0.246	extr

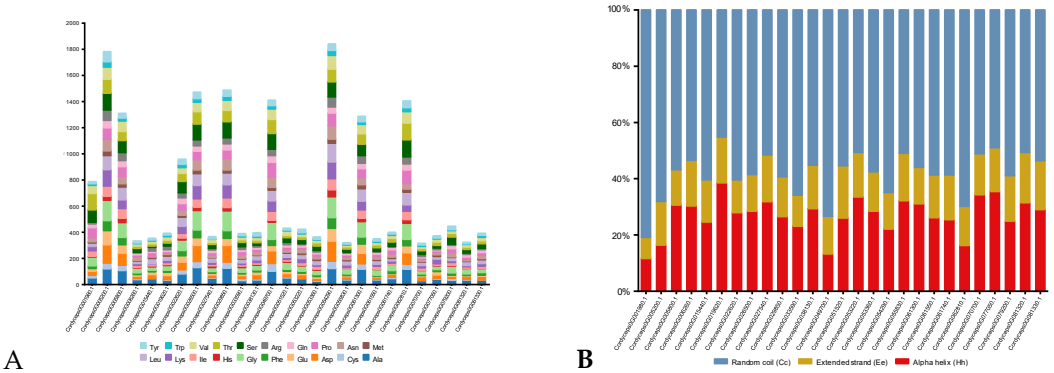


Figure 6. (A) Amino acid composition of GH18 genes from *Cordyceps javanica* Bd01; (B) Percentage of secondary structure elements in *Cordyceps javanica* Bd01 GH18 genes.

2.7. Secondary and Tertiary Structure Analysis

The secondary structures of all the analysed chitinases are dominated by a random coil region as this conformation has the highest mean value of approximately 58.3% (Figure 6B). This was followed by the alpha helix region (27%) and Extended strand (14.7%).

The tertiary structure of proteins is critical in distinguishing the function of proteins, their interactions with other compounds such as ligands, other proteins, and nucleotides as well as to understand the phenotypical effects of mutations. The Swiss-Model server was used to predict the three-dimensional structure of *C. javanica* Bd01 GH18 chitinases based on the known crystal structures of homologous proteins. For the 3D structure prediction, Most of the GH18 protein sequences have very similar three-dimensional structures, and their structural models are shown in Figure 7, with a higher degree of similarity among members within subfamilies than among subfamilies.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 January 2025 doi:10.20944/preprints202501.2089.v1

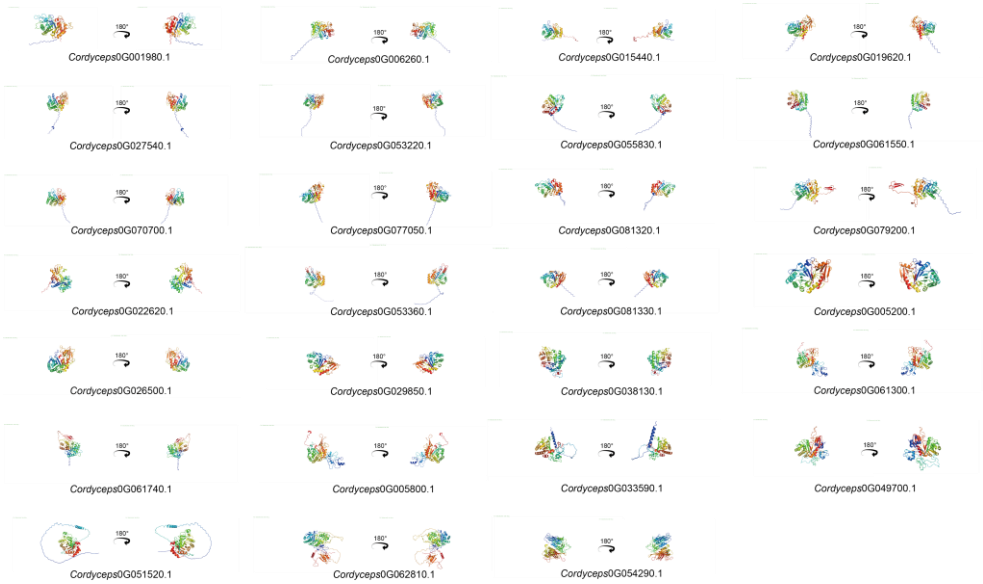


Figure 7. 3D structure of *Cordyceps javanica* Bd01 GH18 genes. **Note:** The left and right plots represent symmetrical structures rotated 180° along the y-axis.

2.8. Analysis of cis-Regulatory Elements in the GH18 Gene Family of Cordyceps javanica Bd01

To further investigate the cis-acting elements in the promoter region of the GH18 gene, the 2000 bp region upstream of the transcription start site of the GH18 gene family was analyzed using the PlantCARE online tool. In terms of element types, hormone responsiveness was the most diverse and was found in all GH18 gene promoter regions. Regarding the distribution of individual elements, 90% of the family members contained CGTCA motif, G-BOX and ABRE elements in the promoter region. It is suggested that these genes may bind to various transcription factors and be involved in the regulation of pathogenicity, and some of the elements in the promoter region of the GH18 gene are shown in the Figure 8.

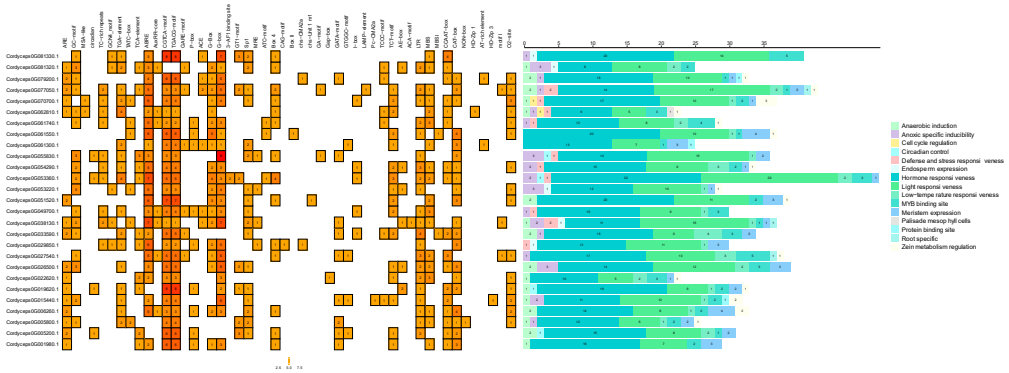


Figure 8. Diagram of cis-acting elements of the promoter sequence of the GH18 genes in *Cordyceps javanica* Bd01.

3. Discussion

Complete genome sequencing of *C. javanica* Bd01 revealed a total number of 78.1 million clean reads (~229.6 × coverage) and BUSCO analysis represents 98.28% completeness with a contig N50 length of 5.4 Mb. The resulting genome assembly was based on 8 contigs with a 53.18% of GC content, 9,590 protein-coding genes, 43 rRNA genes, and 131 tRNA genes. The repetitive sequence 0.34Mb, which represents 1.04% of the total genome. These findings are in line with Yu et al. [34]; their findings revealed a genome size of 47,239,278 bp, comprising 27 contigs, with a GC content of 51.16%

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 January 2025 doi:10.20944/preprints202501.2089.v1

from sequencing of *Fusarium solani* KMZW-1. The genome completeness was reported as 97.93% (BUSCO analysis). Moreover, the DFVF sequence identifier was Fusarium 0G092560.1, and AntiSMASH analysis identified 35 gene clusters associated with secondary metabolite biosynthesis. Similarly, the genome of *C. cicadae* strain CCAD02 (a sexual type of *I. cicadae*) was also sequenced using asexual fruiting bodies [35]. Both of them have close genome size (33.8~33.9 M).

Delineating gene family emergence and extinction within phylogenetically related organisms can identify molecular determinants that underlie species adaptation and evolution [36]. By the comparative genomic analysis, we found that the Chitinases (GH18) is significantly expanded in genes in the genomes of *C. javanica* Bd01. The GH18 chitinases are diverse multigene families in a wide range of organisms and are known to play essential roles in biological processes like growth, nutrient acquisition, interspecific interactions, pathogenesis, and defence [37,38]. Interestingly, chitinases from *B. bassiana* and other entomopathogenic fungi have been identified as important virulence factors in their pathogenicity towards arthropods and nematodes [39]. Many fungi feed on chitin or chitin-containing organisms. Subgroup B chitinases appear mainly to be involved in nutritional functions and more aggressive functions such as invasion and pathogenesis [40]. Most studied subgroup B chitinases in mycoparasitic and entomopathogenic fungi are inducible by nutritional stimuli including chitin or host-specific carbon sources. The expression of subgroup B chitinases, such as Chit33 in *Trichoderma harzianum*, Ech30 in *T. atroviride* and Chi2 in *M. anisopliae* was upregulated during starvation but repressed by glucose or other easily metabolizable carbon sources [40]. Subgroup B chitinases also play a role in insect infection. Knockout of Chi2 from *M. anisopliae* resulted in decreased virulence toward insects, while Chi2 overexpression constructs showed higher efficiency in host killing [41]. The other two subgroups (subgroup A and C) of fungal chitinases are also related to pathogenesis [42–45]. The compartmentation and localisation of different proteins, including enzymes have been well linked to their general biological functions and the reactions they catalyse specifically. In *C. cicadae* genome, Lu and colleagues identified 135 carbohydrate-degrading enzymes, including 16 GH18 proteins [35]. This observation is in accordance with the previous reports that entomopathogenic fungi had more GH18 chitinases than plant pathogenic fungi and mammalian pathogenic fungi [46]. We identified 27 GH18 chitinase genes and putative genes related to the pathogenicity. Furthermore, it is believed that information on the subcellular localisation may provide useful insights into the specific enzymatic pathways of proteins and serve as a guide in subsequent wet-lab experiments.

Thus, in the present study, a variety of readily available computational tools were employed to gain insights into the overall physical parameters; primary, secondary and tertiary structures; functional analysis, domains and motifs, and protein model of *C. javanica* Bd01 chitinase. However, the function of GH18 in the pathogenesis of *C. javanica* Bd01 is unknown and needs to be further investigated, and we are only making predictions based on the existing conditions.

The molecular weight of the enzyme is a critical characteristic for the selection of appropriate purification protocol. The molecular weight range observed in this study suggested the use of gel exclusion chromatography with Sephadex G-100 SF having a fractionation range of 2-120 kDa, best suited for the purification of chitinases under investigation. Orthology analysis revealed that all the *C. javanica* Bd01 GH18 genes were homologous to the GH18 genes of the 4 species used in the phylogenetic tree. This suggests that *C. javanica* Bd01 GH18 genes and those of the other 4 species evolved from a common ancestor. This indicates the divergence in the relative amount of basic/acidic amino acids in their sequences. Hence isoelectric focusing mediated purification should be tailor-made for the individual chitinases as they are expected to precipitate in buffers with different pH values. However, the results showed that most of the *C. javanica* Bd01 GH18 chitinases contain more acidic amino acid residues than basic residues. Results of pI analysis showed that most of the chitinases are acidic enzymes, which is in agreement with previous studies [47,48].This characteristic can also be used for the purification of the said chitinases by employing anion exchange chromatography with weak and strong anion exchangers such as DEAE-sepharose and Qsepharose. These results are in correlation with the earlier reports where chitinases from *B. bassiana* has been

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 January 2025 doi:10.20944/preprints202501.2089.v1

purified with DEAE-cellulose and Mono Q column [49,50]. Recent findings have suggested the effects of nascent peptide charge on protein translation efficiency and protein expression [51]. Hence, it can be speculated that most of the chitinases in this study, being more acidic will be polysome-translated proteins. Proteins with instability index below 40 were predicted to possess in vivo half-life of above 16 hours, whereas, those with instability index above 40 have in vivo half-life of below 5 hours [52]. Therefore, it could be inferred that besides ten chitinases, all the other *C. javanica* Bd01 chitinases analysed, possess remarkable stability in vivo. It is posited that the ten proteins might be products of gene duplications in the organism as strain possess other chitinases which were predicted to be relatively stable [53]. The results from this prediction could specifically indicate the significant thermodynamic and kinetic stabilities of the chitinases which are believed to be necessary for their natural function as entomopathogenic enzymes as well for their potential applications in industrial processes [54]. Alanine, leucine, aspartic acid and proline have a high propensity to form helix confirmation while valine, isoleucine, threonine, glycine, and serine favours the formation of sheet confirmation [53]. The dominance of random coils and alpha-helix regions among the secondary structures indicate the stability of the enzymes as well as a significant amount of conservation within chitinases GH18 from *C. javanica* Bd01 [55].

In order to verify the accuracy of genome sequencing and to further analyse the function of the GH18 gene family of *C. javanica* Bd01, 27 typical genes were cloned in this study. It was found that all five GH18 genes had α helix structures, β folding structures, and irregular curling structures. The proportion of irregular curling to amino acids of each gene was the largest, while the proportion of α helix and β folding structure to amino acids of each gene was relatively small. It is speculated that different RNA splicing methods participate in the expression of chitinase genes, and more chitinase or other proteins can be expressed in the different infection processes and in different host types, allowing the pathogen to expand its host range.

In filamentous fungus, only the TATA box, CCAAT motifs, and CT rich sections of core promoter elements have been functionally studied. Other elements, such as Inr, DPE, MTE, BRE, and CpG islands, mostly present in higher eukaryotes, are not characterized in filamentous fungi. Further, multiple copies of a cis-acting region were added to *Aspergillus* to enhance its promoter activity [56]. The majority of fungal promoters, viz., gpdA, oliC, trpC, citA, and agaA, include a region rich in pyrimidines (CT) upstream of the TSP. The presence of these CT-rich sequences was more evident in highly expressed genes (especially housekeeping 400 Page 6 of 24 World Journal of Microbiology and Biotechnology (2024) 40:400 genes) and in genes missing the TATAAA and CCAAT motifs, indicating a potential role for these sequences as promoter elements in filamentous fungi [57]. G-boxes are a class of cis-acting elements, including the ACGT family, that are capable of responding to a variety of environmental cues, including ABA, light, UV, injury, and pathogen signaling [58,59].

In summary, a total of 27 GH18 genes were identified in *C. javanica* Bd01, with variation in gene structure, protein length, and physicochemical properties. The lower GRAVY score and higher aliphatic index substantiated the hydrophilicity and thermostability of the chitinases. Collectively from the phylogeny and structure analysis, it was observed that the twenty-seven sequences are highly conserved and might also share a similar mechanism for chitinase activity. approach afforded some useful information about the chitinases from *C. javanica*, which could be helpful for isolation and characterisation of the enzyme in vitro and to understand the possible structure and functions of unknown proteins and hence could be subsequently exploited for various applications.

4. Materials and Methods

4.1. Culture of C. javanica Bd01

The strain *C. javanica* Bd01 was isolated from the dead larvae of *Xylotrechus quadripes* and the fungal isolate belongs to the collection of the Institute of Microbiology, Chinese Academy of Sciences, Conservation No. CGMCC23078. The isolate was identified morphologically and by means of the nucleotide sequence of the ITS1-5.8S rRNA-ITS4 region. The internal transcribed spacer sequence of

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 January 2025 doi:10.20944/preprints202501.2089.v1

C. javanica Bd01 was deposited in the GenBank nucleotide sequence database under the accession number MZ831846 [60]. Routinely maintained on sabouraud dextrose agar medium with yeast extract (SDAY) [61]. Bd01 was grown on SDAY at 27°C temperature, 70% ± 5 relative humidity and 12:12 (dark: light) photoperiod for 14 days.

4.2. DNA Extraction and Sequencing

DNA extraction was carried out on pure cultures of C. javanica Bd01 grown on SDAY. The experimental procedure was performed according to the standard protocol provided by Oxford Nanopore Technologies (ONT, Shanghai, China), including sample quality testing, library construction, library quality testing, and library sequencing.

Genomic DNA was extracted using the Omega Fungal DNA Kit D3390-02 according to the manufacturer's instructions (Illumina, USA). Purified genomic DNA was quantified by TBS-380 fluorometer (Turner BioSystems Inc., Sunnyvale, CA). High quality DNA (OD260/280=1.8~2.0, > 15ug) was used to do further research.

Genome was sequenced using a combination of PacBio Sequel Single Molecule Real Time (SMRT) and Illumina sequencing platforms. The Illumina data was used to evaluate the complexity of the genome. For Illumina sequencing, at least 5μg genomic DNA was used for each strain in the sequencing library construction. DNA samples were sheared into 400-500 bp fragments using a Covaris M220 Focused Acoustic Shearer following the manufacture's protocol. Illumina sequencing libraries were prepared from the sheared fragments using the NEXTflex™ Rapid DNA-Seq Kit. Furthermore, in sum 5' prime ends were first end-repaired and phosphorylated. After that, the 3' ends were A-tailed and ligated to sequencing adapters. The third step was to enrich the adapters-ligated products using PCR (Polymerase chain reaction). The prepared libraries were then used for paired-end Illumina sequencing (2 × 150 bp) on an Illumina HiSeq X Ten machine. For Pacific Biosciences sequencing, an aliquot of 8 μg DNA was spun in a Covaris g-TUBE (Covaris, MA) at 6,000 RPM for 60 seconds using an Eppendorf 5424 centrifuge (Eppendorf, NY). DNA fragments were then purified, end-repaired and ligated with SMRTbell sequencing adapters following manufacturer's recommendations (Pacific Biosciences, CA). Resulting sequencing library were purified three times using 0.45 x volumes of Agencourt AMPure XP beads (Beckman Coulter Genomics, MA) following the manufacturer's recommendations. Next, a ~10kb insert library was prepared and sequenced on one SMRT cell using standard methods.

4.3. Genome Assembly and Annotation

The genome sequence was assembled using both the PacBio reads and Illumina reads. The original image data was then transferred into sequence data via base calling, which is defined as raw data or raw reads and saved as FASTQ files. A statistic of quality information was applied for the quality trimming, by which the low-quality data can be removed to form the clean data. The reads then were assembled into contigs using CANU. Finally error correction of the PacBio assembly results was performed using the Illumina reads.

The predicted protein were blast (e-value: 1e-5) against Nr [62], Swiss-Prot [63], TrEMBL [63], KEGG [64], KOG [65]. Blast2go [66] was used for GO [67] annotation. Hmmer [68] was used for Pfam [69] annotation.

4.4. Orthologous and Phylogenomic Analysis

We used OrthoFinder version 2.5.5 [70] to analyze gene families of ten entomopathogens, including *C. javanica* Bd01, *C. javanica* IJ1, *C. javanica* IJ2, *Cordyceps fumosorosea* ARSEF2679, *Cordyceps militaris* CM01, *C. militaris* ATCC34164, *Akanthomyces lecanii* RCEF1005, *Beauveria bassiana* ARSEF2860, *Moelleriella libera* RCEF2490 and *Metarhizium anisopliae* JEF290. For each of the 4,096 single-copy gene families, alignment of amino acid sequences from six strains was performed by MUSCLE version 5.0 [71]. Sequences from all single-copy gene families were concatenated by an in-

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 January 2025 doi:10.20944/preprints202501.2089.v1

house Perl script. Then, the poor alignment regions were removed using Gblocks version 0.91b [72]. The phylogenetic trees were constructed based on the alignment using the Neighbor-Joining method (1,000 repeats) with the parameters of the Jones-Taylor-Thornton model, uniform rates among sites, and partial deletion of gaps was done in MEGA version 11.0.13 [73].

4.5. Identification and Characterization of GH18 Gene Family in C. javanica Bd01

The annotated protein sequences of *C. javanica* Bd01 were used as queries for a hidden Markov model (HMM) searched against the SwissProt [63], InterPro [74], and carbohydrate-active enzymes databases (CAZy) [75] using HMMER database. The retrieved sequences were searched against the SMART [76] and the National Center for Biotechnology Information (NCBI) Conserved Domain Search Service tool [77] to confirm the conserved domains for the GH18 gene family. The number of amino acids, theoretical molecular weight (MW), instability index, aliphatic index, grand average of hydropathicity (GRAVY) and isoelectric point (PI) of the GH18 proteins were predicted using ExPasy-ProtParam tool (http://web.expasy.org/protparam/) [78]. The subcellular localization was predicted using WoLF PSORT (https://wolfpsort.hgc.jp/) [79] web server.

4.6. Conserved domain, Conserved Motif Analyses, and Chromosomal Location

The conserved domain structures were collected by aligning protein domain and protein family models constructed as multiple sequence using the NCBI Batch CD-Search [80] Server. The conserved motifs of the genes were predicted using MEME version 5.5.7 [81] with default parameters. The secondary structure and tertiary structure of the chitinase GH18 gene family were predicted using SOPMA version 2.16.0 [82]. server and SWISS-MODEL version 1.0.2 [83] workspace, respectively.

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

Author Contributions: T. Z. analyzed the data and prepared the manuscript. M. H., J. N., X. C., C. S., D. Y. and X. G. assisted in data interpretation, and helped in writing. G.W. supervised the project, secured funding, and provided overall guidance and mentorship. All authors have read and agreed to the manuscript.

Funding: This research was financially supported by the Reserve Talents for Yunnan Young and Middle-aged Academic and Technical Leaders, China [No. 202105AC160037 and 202205AC160077].

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The original contributions presented in this study are included in the article and Supplementary Materials.

Acknowledgments: Whole-genome sequencing was performed at the Biomarker & Tech Co., Ltd. (Beijing, China). We would like to express our gratitude to Professor Jia-Ying Zhu, College of Forestry, Southwest Forestry University, China, for commenting on an earlier draft of the manuscript.

Conflicts of Interest: The authors declare no competing interests.

References

1. Luangsa-Ard, J.J.; Hywel-Jones, N.L.; Manoch, L.; Samson, R.A. On the Relationships of *Paecilomyces* Sect. *Isarioidea* Species1. *Mycol. Res.* **2005**, *109*, 581–589, doi:10.1017/S0953756205002741.
2. Gallou, A.; Serna-Domínguez, M.G.; Berlanga-Padilla, A.M.; Ayala-Zermeño, M.A.; Mellín-Rosas, M.A.; Montesinos-Matías, R.; Arredondo-Bernal, H.C. Species Clarification of Isaria Isolates Used as Biocontrol Agents against Diaphorina Citri (Hemiptera: Liviidae) in Mexico. *Fungal Biol.* **2016**, *120*, 414–423, doi:10.1016/j.funbio.2015.11.009.
3. Xie, L.; Han, J.H.; Kim, J.J.; Lee, S.Y. Effects of Culture Conditions on Conidial Production of the Sweet Potato Whitefly Pathogenic Fungus *Isaria Javanica*. *Mycoscience* **2016**, *57*, 64–70, doi:10.1016/j.myc.2015.09.002.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 January 2025 doi:10.20944/preprints202501.2089.v1

4. Pan, H.R.; Lin, Y.; Lv, B.K. Evaluation of *Isaria javanica* WH-EP-1 for the control of *Bemisiaargentifolii biotype* B. *Compendium of Research on Agricultural Improvement in Hualien District* **2019**, 89–103.
5. Hasan, W.A.; Assaf, L.H.; Abdullah, S.K. Occurrence of Entomopathogenic and Other Opportunistic Fungi in Soil Collected from Insect Hibernation Sites and Evaluation of Their Entomopathogenic Potential. *Bull. Iraq Nat. Hist. Mus. (P-ISSN: 1017-8678 E-ISSN: 2311-9799)* **2012**, *12*, 19–27.
6. Min, C. Wettable Powder Development of Isaria Javanica for Control of the Lesser Green Leafhopper,Empoasca Vitis. *J. Biol. Control* **2014**.
7. Park, S.E.; Kim, J.C.; Lee, S.J.; Lee, M.R.; Kim, S.; Li, D.; Baek, S.; Han, J.H.; Kim, J.J.; Koo, K.B.; et al. Solid Cultures of Thrips-Pathogenic Fungi *Isaria Javanica* Strains for Enhanced Conidial Productivity and Thermotolerance. *J. Asia-Pac. Entomol.* **2018**, *21*, 1102–1109, doi:10.1016/j.aspen.2018.08.005.
8. Hu, Q.; Liu, S.; Yin, F.; Cai, S.; Zhong, G.; Ren, S. Diversity and Virulence of Soil-Dwelling Fungi *Isaria* Spp. and *Paecilomyces* Spp. against *Solenopsis Invicta* (Hymenoptera: Formicidae). *Biocontrol Science and Technology* **2011**, *21*, 225–234, doi:10.1080/09583157.2010.515733.
9. Lopes, R.S.; Svedese, V.M.; Portela, A.P. a. S.; Albuquerque, A.C.; Luna-Alves Lima, E.A. Virulência e Aspectos Biológicos de Isaria Javanica (Frieder & Bally) Samson & Hywell-Jones Sobre Coptotermes Gestroi (Wasmann) (Isoptera: Rhinotermitidae). *Arq, Inst, Biol, (Online)* **2011**, 565–572.
10. Cabanillas, H.E.; Jones, W.A. Effects of Temperature and Culture Media on Vegetative Growth of an Entomopathogenic Fungus Isaria Sp. (Hypocreales: Clavicipitaceae) Naturally Affecting the Whitefly, Bemisia Tabaci in Texas. *Mycopathologia* **2009**, *167*, 263–271, doi:10.1007/s11046-008-9176-2.
11. Shimazu, M.; Takatsuka, J. Isaria Javanica (Anamorphic Cordycipitaceae) Isolated from Gypsy Moth Larvae, Lymantria Dispar (Lepidoptera: Lymantriidae), in Japan. *Appl. Entomol. Zool. - APPL ENTOMOL ZOOL* **2010**, *45*, 497–504, doi:10.1303/aez.2010.497.
12. Ortiz-Urquiza, A.; Keyhani, N.O. Action on the Surface: Entomopathogenic Fungi versus the Insect Cuticle. *Insects* **2013**, *4*, 357–374, doi:10.3390/insects4030357.
13. Huang, Z.; Hao, Y.; Gao, T.; Huang, Y.; Ren, S.; Keyhani, N.O. The Ifchit1 Chitinase Gene Acts as a Critical Virulence Factor in the Insect Pathogenic Fungus Isaria Fumosorosea. *Appl. Microbiol. Biotechnol.* **2016**, *100*, 5491–5503, doi:10.1007/s00253-016-7308-z.
14. Ortiz-Urquiza, A.; Keyhani, N.O. Molecular Genetics of Beauveria Bassiana Infection of Insects. *Adv Genet* **2016**, *94*, 165–249, doi:10.1016/bs.adgen.2015.11.003.
15. Valero-Jiménez, C.A.; Wiegers, H.; Zwaan, B.J.; Koenraadt, C.J.M.; van Kan, J.A.L. Genes Involved in Virulence of the Entomopathogenic Fungus Beauveria Bassiana. *J. Invertebr. Pathol.* **2016**, *133*, 41–49, doi:10.1016/j.jip.2015.11.011.
16. Wang, C.; Wang, S. Insect Pathogenic Fungi: Genomics, Molecular Interactions, and Genetic Improvements. *Annu. Rev. Entomol.* **2017**, *62*, 73–90, doi:10.1146/annurev-ento-031616-035509.
17. Elad, Y.; Chet, I.; Henis, Y. Degradation of Plant Pathogenic Fungi by Trichoderma Harzianum. *Can. J. Microbiol.* **1982**, *28*, 719–725, doi:10.1139/m82-110.
18. Inbar, J.; Chet, I. The Role of Recognition in the Induction of Specific Chitinases during Mycoparasitism by Trichoderma Harzianum. *Microbiology* **1995**, *141*, 2823–2829, doi:10.1099/13500872-141-11-2823.
19. Adams, D.J. Fungal Cell Wall Chitinases and Glucanases. *Microbiology* **2004**, *150*, 2029–2035, doi:10.1099/mic.0.26980-0.
20. Gonfa, T.G.; Negessa, A.K.; Bulto, A.O. Isolation, Screening, and Identification of Chitinase-Producing Bacterial Strains from Riverbank Soils at Ambo, Western Ethiopia. *Heliyon* **2023**, *9*, e21643, doi:10.1016/j.heliyon.2023.e21643.
21. Henrissat, B.; Bairoch, A. New Families in the Classification of Glycosyl Hydrolases Based on Amino Acid Sequence Similarities. *Biochem J* **1993**, *293 (Pt 3)*, 781–788, doi:10.1042/bj2930781.
22. Kawase, T.; Saito, A.; Sato, T.; Kanai, R.; Fujii, T.; Nikaidou, N.; Miyashita, K.; Watanabe, T. Distribution and Phylogenetic Analysis of Family 19 Chitinases in Actinobacteria. *Appl. Environ. Microbiol.* **2004**, *70*, 1135–1144, doi:10.1128/AEM.70.2.1135-1144.2004.
23. Seidl, V. Chitinases of Filamentous Fungi: A Large Group of Diverse Proteins with Multiple Physiological Functions. *Fungal Biol. Rev.* **2008**, *22*, 36–42, doi:10.1016/j.fbr.2008.03.002.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 January 2025 doi:10.20944/preprints202501.2089.v1

24. Hartl, L.; Zach, S.; Seidl-Seiboth, V. Fungal Chitinases: Diversity, Mechanistic Properties and Biotechnological Potential. *Appl. Microbiol. Biotechnol.* **2012**, *93*, 533–543, doi:10.1007/s00253-011-3723-3.
25. Adrangi, S.; Faramarzi, M.A. From Bacteria to Human: A Journey into the World of Chitinases. *Biotechnol. Adv.* **2013**, *31*, 1786–1795, doi:10.1016/j.biotechadv.2013.09.012.
26. Stals, I.; Samyn, B.; Sergeant, K.; White, T.; Hoorelbeke, K.; Coorevits, A.; Devreese, B.; Claeyssens, M.; Piens, K. Identification of a Gene Coding for a Deglycosylating Enzyme in Hypocrea Jecorina. *FEMS Microbiol Lett* **2010**, *303*, 9–17, doi:10.1111/j.1574-6968.2009.01849.x.
27. Tzelepis, G.; Hosomi, A.; Hossain, T.J.; Hirayama, H.; Dubey, M.; Jensen, D.F.; Suzuki, T.; Karlsson, M. Endo-β-N-Acetylglucosamidases (ENGases) in the Fungus Trichoderma Atroviride: Possible Involvement of the Filamentous Fungi-Specific Cytosolic ENGase in the ERAD Process. *Biochem. Biophys. Res. Commun.* **2014**, *449*, 256–261, doi:10.1016/j.bbrc.2014.05.017.
28. Boldo, J.T.; Junges, A.; do Amaral, K.B.; Staats, C.C.; Vainstein, M.H.; Schrank, A. Endochitinase CHI2 of the Biocontrol Fungus Metarhizium Anisopliae Affects Its Virulence toward the Cotton Stainer Bug Dysdercus Peruvianus. *Curr. Genet.* **2009**, *55*, 551–560, doi:10.1007/s00294-009-0267-5.
29. Staats, C.C.; Kmetzsch, L.; Lubeck, I.; Junges, A.; Vainstein, M.H.; Schrank, A. Metarhizium Anisopliae Chitinase CHIT30 Is Involved in Heat-Shock Stress and Contributes to Virulence against Dysdercus Peruvianus. *Fungal Biol.* **2013**, *117*, 137–144, doi:10.1016/j.funbio.2012.12.006.
30. Fang, W.; Leng, B.; Xiao, Y.; Jin, K.; Ma, J.; Fan, Y.; Feng, J.; Yang, X.; Zhang, Y.; Pei, Y. Cloning of Beauveria Bassiana Chitinase Gene Bbchit1 and Its Application to Improve Fungal Strain Virulence. *Appl. Environ. Microbiol.* **2005**, *71*, 363–370, doi:10.1128/AEM.71.1.363-370.2005.
31. Meng, H.; Wang, Z.; Meng, X.; Xie, L.; Huang, B. Cloning and Expression Analysis of the Chitinase Gene Ifu-Chit2 from Isaria Fumosorosea. *Genet. Mol. Biol.* **2015**, *38*, 381–389, doi:10.1590/S1415-475738320150003.
32. Huang, Z.; Hao, Y.; Gao, T.; Huang, Y.; Ren, S.; Keyhani, N.O. The Ifchit1 Chitinase Gene Acts as a Critical Virulence Factor in the Insect Pathogenic Fungus Isaria Fumosorosea. *Appl. Microbiol. Biotechnol.* **2016**, *100*, 5491–5503, doi:10.1007/s00253-016-7308-z.
33. Wang, C.; Gao, T.; Huang, Y.; Huang, Z. Effect of Ifchit1 Gene of Isaria Fumosorosea on Mortality, Oviposition and Oxidase Activities of Bemisia Tabaci. *Biocontrol Sci. Technol.* **2017**.
34. Yu, J.; Hussain, M.; Wu, M.; Shi, C.; Li, S.; Ji, Y.; Hussain, S.; Qin, D.; Xiao, C.; Wu, G. Whole-Genome Sequencing of the Entomopathogenic Fungus Fusarium Solani KMZW-1 and Its Efficacy against Bactrocera Dorsalis. *Curr. Issues Mol. Biol.* **2024**, *46*, 11593–11612, doi:10.3390/cimb46100688.
35. Lu, Y.; Luo, F.; Cen, K.; Xiao, G.; Yin, Y.; Li, C.; Li, Z.; Zhan, S.; Zhang, H.; Wang, C. Omics Data Reveal the Unusual Asexual-Fruiting Nature and Secondary Metabolic Potentials of the Medicinal Fungus Cordyceps Cicadae. *BMC Genomics* **2017**, *18*, 668, doi:10.1186/s12864-017-4060-4.
36. Mitreva, M.; Jasmer, D.P.; Zarlenga, D.S.; Wang, Z.; Abubucker, S.; Martin, J.; Taylor, C.M.; Yin, Y.; Fulton, L.; Minx, P.; et al. The Draft Genome of the Parasitic Nematode Trichinella Spiralis. *Nat. Genet.* **2011**, *43*, 228–235, doi:10.1038/ng.769.
37. Gruber, S.; Seidl-Seiboth, V. Self versus Non-Self: Fungal Cell Wall Degradation in Trichoderma. *Microbiology* **2012**, *158*, 26–34, doi:10.1099/mic.0.052613-0.
38. Nagpure, A.; Choudhary, B.; Gupta, R.K. Chitinases: In Agriculture and Human Healthcare. *Crit. Rev. Biotechnol.* **2014**, *34*, 215–232, doi:10.3109/07388551.2013.790874.
39. Berini, F.; Katz, C.; Gruzdev, N.; Casartelli, M.; Tettamanti, G.; Marinelli, F. Microbial and Viral Chitinases: Attractive Biopesticides for Integrated Pest Management. *Biotechnol. Adv.* **2018**, *36*, 818–838, doi:10.1016/j.biotechadv.2018.01.002.
40. Hartl, L.; Zach, S.; Seidl-Seiboth, V. Fungal Chitinases: Diversity, Mechanistic Properties and Biotechnological Potential. *Appl. Microbiol. Biotechnol.* **2012**, *93*, 533–543, doi:10.1007/s00253-011-3723-3.
41. Boldo, J.T.; Junges, A.; do Amaral, K.B.; Staats, C.C.; Vainstein, M.H.; Schrank, A. Endochitinase CHI2 of the Biocontrol Fungus Metarhizium Anisopliae Affects Its Virulence toward the Cotton Stainer Bug Dysdercus Peruvianus. *Curr Genet* **2009**, *55*, 551–560, doi:10.1007/s00294-009-0267-5.
42. Carsolio, C.; Benhamou, N.; Haran, S.; Cortés, C.; Gutiérrez, A.; Chet, I.; Herrera-Estrella, A. Role of the Trichoderma Harzianum Endochitinase Gene, Ech42, in Mycoparasitism. *Appl Environ Microbiol* **1999**, *65*, 929–935, doi:10.1128/AEM.65.3.929-935.1999.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 January 2025 doi:10.20944/preprints202501.2089.v1

43. Woo, S.L.; Donzelli, B.; Scala, F.; Mach, R.; Harman, G.E.; Kubicek, C.P.; Del Sorbo, G.; Lorito, M. Disruption of the Ech42 (Endochitinase-Encoding) Gene Affects Biocontrol Activity in Trichoderma Harzianum P1. *Mol. Plant-Microbe Interact.* **1999**, *12*, 419–429, doi:10.1094/MPMI.1999.12.5.419.
44. Gruber, S.; Vaaje-Kolstad, G.; Matarese, F.; López-Mondéjar, R.; Kubicek, C.P.; Seidl-Seiboth, V. Analysis of Subgroup C of Fungal Chitinases Containing Chitin-Binding and LysM Modules in the Mycoparasite Trichoderma Atroviride. *Glycobiology* **2011**, *21*, 122–133, doi:10.1093/glycob/cwq142.
45. Gruber, S.; Kubicek, C.P.; Seidl-Seiboth, V. Differential Regulation of Orthologous Chitinase Genes in Mycoparasitic Trichoderma Species. *Appl. Environ. Microbiol.* **2011**, *77*, 7217–7226, doi:10.1128/AEM.06027-11.
46. Shang, Y.; Xiao, G.; Zheng, P.; Cen, K.; Zhan, S.; Wang, C. Divergent and Convergent Evolution of Fungal Pathogenicity. *Genome Biol. Evol.* **2016**, *8*, 1374–1387, doi:10.1093/gbe/evw082.
47. Seidl, V.; Huemer, B.; Seiboth, B.; Kubicek, C.P. A Complete Survey of Trichoderma Chitinases Reveals Three Distinct Subgroups of Family 18 Chitinases. *FEBS J.* **2005**, *272*, 5923–5939, doi:10.1111/j.1742-4658.2005.04994.x.
48. Junges, Â.; Boldo, J.T.; Souza, B.K.; Guedes, R.L.M.; Sbaraini, N.; Kmetzsch, L.; Thompson, C.E.; Staats, C.C.; de Almeida, L.G.P.; de Vasconcelos, A.T.R.; et al. Genomic Analyses and Transcriptional Profiles of the Glycoside Hydrolase Family 18 Genes of the Entomopathogenic Fungus Metarhizium Anisopliae. *PLOS One* **2014**, *9*, e107864, doi:10.1371/journal.pone.0107864.
49. Havukkala, I.; Mitamura, C.; Hara, S.; Hirayae, K.; Nishizawa, Y.; Hibi, T. Induction and Purification of *Beauveria Bassiana* Chitinolytic Enzymes. *Journal of Invertebrate Pathology* **1993**, *61*, 97–102, doi:10.1006/jipa.1993.1017.
50. Fang, W.; Leng, B.; Xiao, Y.; Jin, K.; Ma, J.; Fan, Y.; Feng, J.; Yang, X.; Zhang, Y.; Pei, Y. Cloning of Beauveria Bassiana Chitinase Gene Bbchit1 and Its Application to Improve Fungal Strain Virulence. *Appl Environ Microbiol* **2005**, *71*, 363–370, doi:10.1128/AEM.71.1.363-370.2005.
51. Requião, R.D.; Fernandes, L.; de Souza, H.J.A.; Rossetto, S.; Domitrovic, T.; Palhano, F.L. Protein Charge Distribution in Proteomes and Its Impact on Translation. *PLOS Comput. Biol.* **2017**, *13*, e1005549, doi:10.1371/journal.pcbi.1005549.
52. Gasteiger, E.; Hoogland, C.; Gattiker, A.; Duvaud, S.; Wilkins, M.R.; Appel, R.D.; Bairoch, A. Protein Identification and Analysis Tools on the ExPASy Server. In *The Proteomics Protocols Handbook*; Walker, J.M., Ed.; Humana Press: Totowa, NJ, 2005; pp. 571–607 ISBN 978-1-59259-890-8.
53. Idicula-Thomas, S.; Balaji, P.V. Understanding the Relationship between the Primary Structure of Proteins and Their Amyloidogenic Propensity: Clues from Inclusion Body Formation. *Protein eng. des. sel.: PEDS* **2005**, *18*, 175–180, doi:10.1093/protein/gzi022.
54. Gohel, V.; Naseby, D.C. Thermalstabilization of Chitinolytic Enzymes of *Pantoea Dispersa*. *Biochem. Eng. J.* **2007**, *35*, 150–157, doi:10.1016/j.bej.2007.01.009.
55. Gouripur, G.C.; Kaliwal, R.B.; Kaliwal, B.B. In Silico Characterization of Beta-Galactosidase Using Computational Tools. *J. Bioinform. Seq. Anal.* **2016**, *8*, 1–11, doi:10.5897/JBSA2015.0101.
56. Minetoki, T.; Tsuboi, H.; Koda, A. Development of High Expression System with the Improved Promoter Using the Cis-Acting Element in Aspergillus Species.; November 1 2003.
57. Sakekar, A.A.; Gaikwad, S.R.; Punekar, N.S. Protein Expression and Secretion by Filamentous Fungi. *J. Biosci.* **2021**, *46*, 5, doi:10.1007/s12038-020-00120-8.
58. Kim, S.R.; Choi, J.L.; Costa, M.A.; An, G. Identification of G-Box Sequence as an Essential Element for Methyl Jasmonate Response of Potato Proteinase Inhibitor II Promoter. *Plant Physiol.* **1992**, *99*, 627–631, doi:10.1104/pp.99.2.627.
59. Menkens, A.E.; Schindler, U.; Cashmore, A.R. The G-Box: A Ubiquitous Regulatory DNA Element in Plants Bound by the GBF Family of bZIP Proteins. *Trends Biochem. Sci.* **1995**, *20*, 506–510, doi:10.1016/s0968-0004(00)89118-5.
60. Liu, Q.J. et al. Identification, culture, and pathogenicity of an entomopathogenic fungus from the larvae of *Xylotrechus quadripes*. *Science of Western Forestry* **2022**, 89–96, doi:10.16473/j.cnki.xblykx1972.2022.05.013.
61. Gallou, A.; Serna-Domínguez, M.G.; Berlanga-Padilla, A.M.; Ayala-Zermeño, M.A.; Mellín-Rosas, M.A.; Montesinos-Matías, R.; Arredondo-Bernal, H.C. Species Clarification of Isaria Isolates Used as Biocontrol

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 January 2025 doi:10.20944/preprints202501.2089.v1

Agents against Diaphorina Citri (Hemiptera: Liviidae) in Mexico. *Fungal Biol.* **2016**, *120*, 414–423, doi:10.1016/j.funbio.2015.11.009.

62. Fu-chu, H. Integrated Nr Database in Protein Annotation System and Its Localization. *Comput. Eng.* **2006**.
63. Boeckmann, B.; Bairoch, A.; Apweiler, R.; Blatter, M.-C.; Estreicher, A.; Gasteiger, E.; Martin, M.J.; Michoud, K.; O'Donovan, C.; Phan, I.; et al. The SWISS-PROT Protein Knowledgebase and Its Supplement TrEMBL in 2003. *Nucleic Acids Res.* **2003**, *31*, 365–370, doi:10.1093/nar/gkg095.
64. Kanehisa, M.; Goto, S.; Kawashima, S.; Okuno, Y.; Hattori, M. The KEGG Resource for Deciphering the Genome. *Nucleic Acids Res.* **2004**, *32*, D277-280, doi:10.1093/nar/gkh063.
65. Tatusov, R.L.; Galperin, M.Y.; Natale, D.A.; Koonin, E.V. The COG Database: A Tool for Genome-Scale Analysis of Protein Functions and Evolution. *Nucleic Acids Res.* **2000**, *28*, 33–36, doi:10.1093/nar/28.1.33.
66. Conesa, A.; Götz, S.; García-Gómez, J.M.; Terol, J.; Talón, M.; Robles, M. Blast2GO: A Universal Tool for Annotation, Visualization and Analysis in Functional Genomics Research. *Bioinform. (Oxf. Engl.)* **2005**, *21*, 3674–3676, doi:10.1093/bioinformatics/bti610.
67. Ashburner, M.; Ball, C.A.; Blake, J.A.; Botstein, D.; Butler, H.; Cherry, J.M.; Davis, A.P.; Dolinski, K.; Dwight, S.S.; Eppig, J.T.; et al. Gene Ontology: Tool for the Unification of Biology. The Gene Ontology Consortium. *Nat. Genet.* **2000**, *25*, 25–29, doi:10.1038/75556.
68. Eddy, S.R. Profile Hidden Markov Models. *Bioinform. (Oxf. Engl.)* **1998**, *14*, 755–763, doi:10.1093/bioinformatics/14.9.755.
69. Finn, R.D.; Coggill, P.; Eberhardt, R.Y.; Eddy, S.R.; Mistry, J.; Mitchell, A.L.; Potter, S.C.; Punta, M.; Qureshi, M.; Sangrador-Vegas, A.; et al. The Pfam Protein Families Database: Towards a More Sustainable Future. *Nucleic Acids Res* **2016**, *44*, D279-285, doi:10.1093/nar/gkv1344.
70. Emms, D.M.; Kelly, S. OrthoFinder: Phylogenetic Orthology Inference for Comparative Genomics. *Genome Biol.* **2019**, *20*, 238, doi:10.1186/s13059-019-1832-y.
71. Edgar, R.C. Muscle5: High-Accuracy Alignment Ensembles Enable Unbiased Assessments of Sequence Homology and Phylogeny. *Nat. Commun.* **2022**, *13*, 6968, doi:10.1038/s41467-022-34630-w.
72. Castresana, J. Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis. *Mol. Biol. Evol.* **2000**, *17*, 540–552, doi:10.1093/oxfordjournals.molbev.a026334.
73. Tamura, K.; Stecher, G.; Kumar, S. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Mol. Biol. Evol.* **2021**, *38*, 3022–3027, doi:10.1093/molbev/msab120.
74. Blum, M.; Andreeva, A.; Florentino, L.C.; Chuguransky, S.R.; Grego, T.; Hobbs, E.; Pinto, B.L.; Orr, A.; Paysan-Lafosse, T.; Ponamareva, I.; et al. InterPro: The Protein Sequence Classification Resource in 2025. *Nucleic Acids Research* **2025**, *53*, D444–D456, doi:10.1093/nar/gkae1082.
75. Lombard, V.; Golaconda Ramulu, H.; Drula, E.; Coutinho, P.M.; Henrissat, B. The Carbohydrate-Active Enzymes Database (CAZy) in 2013. *Nucleic Acids Res.* **2014**, *42*, D490–D495, doi:10.1093/nar/gkt1178.
76. Letunic, I.; Khedkar, S.; Bork, P. SMART: Recent Updates, New Developments and Status in 2020. *Nucleic Acids Res.* **2021**, *49*, D458–D460, doi:10.1093/nar/gkaa937.
77. Lu, S.; Wang, J.; Chitsaz, F.; Derbyshire, M.K.; Geer, R.C.; Gonzales, N.R.; Gwadz, M.; Hurwitz, D.I.; Marchler, G.H.; Song, J.S.; et al. CDD/SPARCLE: The Conserved Domain Database in 2020. *Nucleic Acids Res.* **2020**, *48*, D265–D268, doi:10.1093/nar/gkz991.
78. *The Proteomics Protocols Handbook*; Walker, J.M., Ed.; Humana Press: Totowa, NJ, 2005; ISBN 978-1-58829-343-5.
79. Horton, P.; Park, K.-J.; Obayashi, T.; Fujita, N.; Harada, H.; Adams-Collier, C.J.; Nakai, K. WoLF PSORT: Protein Localization Predictor. *Nucleic Acids Res.* **2007**, *35*, W585-587, doi:10.1093/nar/gkm259.
80. Wang, J.; Chitsaz, F.; Derbyshire, M.K.; Gonzales, N.R.; Gwadz, M.; Lu, S.; Marchler, G.H.; Song, J.S.; Thanki, N.; Yamashita, R.A.; et al. The Conserved Domain Database in 2023. *Nucleic Acids Res.* **2023**, *51*, D384–D388, doi:10.1093/nar/gkac1096.
81. Bailey, T.L.; Johnson, J.; Grant, C.E.; Noble, W.S. The MEME Suite. *Nucleic Acids Res.* **2015**, *43*, W39-49, doi:10.1093/nar/gkv416.
82. Paysan-Lafosse, T.; Andreeva, A.; Blum, M.; Chuguransky, S.R.; Grego, T.; Pinto, B.L.; Salazar, G.A.; Bileschi, M.L.; Llinares-López, F.; Meng-Papaxanthos, L.; et al. The Pfam Protein Families Database: Embracing AI/ML. *Nucleic Acids Res* **2024**, *53*, D523-534, doi:10.1093/nar/gkae997.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 January 2025 doi:10.20944/preprints202501.2089.v1

83. Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F.T.; de Beer, T.A.P.; Rempfer, C.; Bordoli, L.; et al. SWISS-MODEL: Homology Modelling of Protein Structures and Complexes. *Nucleic Acids Res.* **2018**, *46*, W296–W303, doi:10.1093/nar/gky427.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.