6-7-2022

Transcriptome Analysis of the Entomopathogenic Fungus Culicinomyces clavisporus

Dana E. Foresman
Nova Southeastern University

Thesis of
Dana E. Foresman

Submitted in Partial Fulfillment of the Requirements for the Degree of

Master of Science
Biological Sciences

Nova Southeastern University
Halmos College of Arts and Sciences

June 2022

Approved:
Thesis Committee

Committee Chair: Aurelien Tartar

Committee Member: Robert Smith

Committee Member: William Morgan

NOVA SOUTHEASTERN UNIVERSITY
HALMOS COLLEGE OF ARTS AND SCIENCES

Transcriptome Analysis of the Entomopathogenic Fungus *Culicinomyces clavisporus*

By
DANA FORESMAN

Submitted to the Faculty of
Halmos College of Arts and Sciences
in partial fulfilment of the requirements for
the degree of Master of Science with a specialty in:

Biological Sciences

Nova Southeastern University

June 2022

Acknowledgements

I would like to thank my major advisor, Dr. Aurelien Tartar, for all the guidance and support during my thesis experimentation, data analysis, and writing. Without his mentorship I would not be as prepared as I am to join the scientific community upon graduation. I thank my committee members, Dr. Aurelien Tartar, Dr. Robert Smith, and Dr. William Morgan, for providing me with support and expertise throughout my research and defense of my thesis. The Halmos College of Arts and Sciences Biology department has been a major support system for my development as a member of the scientific community, as the department supported my experience at the Fungal Genetics conference where I got to share my research with fellow scientists. Lastly, I would like to thank all my friends and family, those of whom encouraged and supported me while I completed my graduate education.

Table of Contents

Abstract

Culicinomyces clavisporus is an entomopathogenic fungus that can infect mosquito larvae, such as *Culex quinquefasciatus, Aedes aegypti, and Anopheles stephensi*. Whereas most fungal entomopathogens infect hosts through the cuticle, *C. clavisporus* initiates infection through ingestion. This suggests that the *C. clavisporus* genome may be mined for novel pathogenicity factors with the potential for vector control. To this end, a transcriptome analysis was initiated. The strain *C. clavisporus* ARSEF 582 was grown in modified PYG liquid cultures that was supplemented with whole, insect larvae (*Galleria mellonella*) to elicit the expression of genes involved in host-pathogen relationships. Total RNA samples were extracted and processed for cDNA library construction and Single Molecule Real Time (SMRT) sequencing (PacBio platform). A total of 3,512,145 reads were produced. Assembly of these reads was completed using CD-HIT-EST and revealed 8,266 unigenes. A phylogenetic analysis using actin gene sequences showed the *C. clavisporus* is closely related to both *D. coniospora* and *H. rhossiliensis*. The transcriptome annotation revealed 10 genes of interest to entomopathogenic fungi infection methods, including those potentially linked to the oral infection method. Of those 10 genes, 4 genes have functions related to cellular processes aiding in virulence, 4 genes linked to cuticle degradation, and 2 genes with potential links to the oral infection process. Overall, the number of unigenes identified from the transcriptomic analysis showed greater abundance of unigenes with possible link to oral infection than compared to cuticular degradation. This suggests that *C. clavisporus* utilizes oral infection as its main mode of infection.

Keywords

entomopathogen, transcriptomics, oral infection, virulence, pathogenicity

List of Figures and Tables

Introduction

Microbial Control of Insects

In nature, insect pathogens are microbes that harm insect hosts and serve an important role in regulating insect population numbers. This means insect pathogenic microbes can be and have been used in microbial control efforts. The term microbial control means the pathogen maintains the target (insect) host population density lower than it would be in their absence. These pathogens can include bacteria, fungi, and nematodes and can be applied like traditional chemical control agents. Fungi that are known insect pathogens are more commonly known as entomopathogenic fungi [11].

Entomopathogenic Fungi as Microbial Control Agents

The use of entomopathogenic fungi is far behind the advancements of other microbial control agents, such as bacteria and viruses, but there are prospects for this research. This is evident by the application of specific fungal strains on an operational level as an addition to the current chemical pesticides in use today [28]. For example, *Beauveria bassiana* was developed as a commercial mycoinsecticide by Mycotech. Through the production of aerial conidia, this fungal strain has been shown to be effective in the early application for both beetle and caterpillar pest control [41].

Historically, biocontrol using fungal agents has relied on the environmental release of various life stages of whole organisms. There are three methods of biocontrol when using whole microorganisms: inoculative release, augmentative application, and inundative application. Inoculative release involves small amounts of the fungus being applied early in the crop season directly on the crop. The expectation is that the fungus will establish in the environment, preventing the pest population from exceeding the economic threshold. Economic threshold is defined as the density of pest that still allows for economic return on the crop, taking into consideration the cost of treatment. Augmentative application involves increasing the presence of microorganisms already present in the environment. This method is used to prevent damage to the crop prior to the pest population growing past the threshold [9]. Inundative application involves applying a large amount of the fungus to the crop to control pests in the short term [41].

The type of control method chosen depends on the type of entomopathogenic fungi in use as well as the type of environment/crop the application is supposed to be deployed.

More recently, research on entomopathogenic fungi has included manipulation of several genes that have been identified as putative virulence factors [31]. This strategy has relied in part on the identification of the metabolites used by entomopathogenic fungi during host infection. Increased knowledge of the molecular basis of fungal-insect pathosystems has demonstrated that ingestion of fungal inoculate is not required for disease initiation in insect hosts. Pathogenesis is performed via attachment and germination through the external cuticle, which is where entomopathogenic fungi differ from other sources of microbial control models [28]. Spores or conidia are responsible for the invasion of the insect host body and circulatory system. The germination of the spore or conidia is possible because of a set of cascade reactions involving recognition and enzyme (e.g., chitinase) activation. Once the insect host has perished, the conidia can be dispersed either through wind or passively through infected cadavers [41]. It is widely believed that the genes coding for the sets of proteins and enzymes critical to host invasion can be expressed within the plant and used as a source of novel control strategies against insect disease.

Benefits of Transcriptomics

Transcriptomics allows for the understanding of gene function through the analysis of RNA sequences, and the technologies developed to study transcriptomes include RNA-seq and microarrays. It is important to consider how to assemble the transcriptome being investigated. If there is a transcriptome to use as a reference, then any method that provide short read sequences may be appropriate, but if there is no reference transcriptome, technologies that produce longer read sequences is necessary for correct reassembly [6]. An example of long read sequencing technology is PacBio's Single Molecule Real Time (SMRT) Iso-Seq sequencing. This can capture the full length of transcripts depending on the quality of the RNA and the generated cDNA sequences. Technologies that produce short read sequences require software to reconstruct the transcriptome, unlike when Iso-Seq sequencing is used [52]. The long read sequences produced by SMRT Iso-Seq sequencing consist of the whole gene sequence, getting rid of the need for a software to construct the full gene sequence.

Sequences generated during transcriptome analysis allow to catalog an organism's protein-coding potential or may indicate gene expression levels associated with specific environmental conditions. Difference in levels of gene expression can be seen through biotic and abiotic changes. Changes in the environment such as nutrient sources or host types can play a role in pathogenesis, invasion, and virulence of the entomopathogenic species [4]. This expression level analysis can be performed using the original read sequences provided through the RNA-seq files [52]. The sequences found upon analysis provide insight as to which genes are utilized during the pathogenicity process.

RNA-seq technologies have been utilized in many studies for the purpose of reconstructing the transcriptome of fungal pathogenic species in efforts to provide information on the molecular basis of the infection process and virulence. A study focused on the pathogenicity of the highly virulent *Curvularia lunata* strain CX-3, a fungus that targets maize crops in China. To investigate its pathogenicity, de novo sequencing using Illumina technologies provided transcriptome sequences for reassembly. Using *Magnaporthe grisea* as a reference, the CX-3 genome was constructed. The RNA extracted was reverse transcribed into cDNA, which contained unique gene tags. This cDNA was placed into a sequencing, and it was discovered that the virulence of *C. lunata* had several contributing factors, including genes involved in biosynthesis and other pathogenicity associated genes [17].

On the other hand, a study looking into annotation of the transcriptome of *Ascosphaera apis* used both PacBio long read and Illumina short read sequencing technologies. Using a reference genome of *A. apis*, 96% of the reads mapped compared to this reference. While these reads were successfully confirmed, many have various isoforms for the same gene. These novel isoforms were confirmed using gel electrophoresis, molecular cloning, and Sanger sequencing, and contributed to a reference full length transcriptome for future studies on *A. apis* [8].

Lastly, a study completed on *Lagenidium giganteum* which used next generation sequencing technologies (454 and PacBio) to obtain several transcriptomes. Major genes identified included glycoside hydrolase family 5 and family 20 (GH5 and GH20), which were shown to be underrepresented in non-entomopathogenic genomes. Comparative analysis also demonstrated that *L. giganteum* transcriptomes contained plant associated virulence factors [36,45].

Culicinomyces clavisporus

In the case of the *Culicinomyces* species, infection occurs mainly on aquatic larvae. This fungus is unique as it remains one of the few examples of fungal entomopathogens to initiate infection *per os*. After conidial ingestion, *Culicinomyces* is known to invade the insect host's digestive tract [43,44]. Only a small number of *Culicinomyces* isolates have been investigated, with little to no known variability among these isolates. In terms of host range, *Culicinomyces* species have shown effectiveness in the biological control of some species of mosquito larvae (aquatic stage). While instant death of insect host does not occur, there are lethal effects of the insect host later in the life cycle. In combination with the infection of *Culicinomyces*, the insect is more susceptible to chemical insecticides used on crops. However previous research on this fungus was only halted due to the slow growth of the isolates as well as high conidial inoculum needed for the previously mentioned infectivity.

This current study brings special attention to one of the best-known *Culicinomyces* species: *Culicinomyces clavisporus*. This entomopathogenic fungi is from the *Clavicipitaceeae* family and *Hypocreales* order. It was originally discovered in both Sydney, Australia as well as Chapel Hill, North Carolina. In both discoveries, *C. clavisporus* was isolated from marsh mosquito (anopheline) larvae when exposed to field collected water samples from the two locations. Since isolation, *C. clavisporus* has shown to be effective against the control of *Culex quinquefasciatus, Aedes aegypti, and Anopheles stephensi* [40].

Genes of Interest

Traditional virulence factors found during infection of the cuticle within include proteases, glycosidases, and lipases [2]. More specifically, cuticle degradation happens due to the activation and production of proteins from genes such as cytochrome P450 (CYP), chitinases, and subtilisin-like proteases. Once the cuticle has been breached, the entomopathogenic fungi produce secondary metabolites that aid in the continuation of the infection process. These secondary metabolites also contain antibacterial and antifungal functions to only allow this infection to kill the insect host [33]. The previous research into the transcriptome of entomopathogenic fungi that utilize cuticular infection routes when compared to the virulence factors and protein coding potential of an entomopathogenic species, such as *C. clavisporus*, would differ. The genes found in *C. clavisporus* and other entomopathogenic fungi that utilize

the oral infection route would provide novel genes and produced proteins that support the differing infection route.

Given that the oral infection method is not widely used among many entomopathogenic fungi, there is even less information on the genomes of these fungi that do utilize this method. To date, the whole genomes of both *B. bassiana* and *Metarhizium* spp. were sequenced to give insight into the molecular components involved in the oral infection method. Like entomopathogenic bacterium, theses fungi share homologous genes for the function of oral infection. Those sequenced genomes revealed a whole host of genes that are not commonly active during cuticular infection by entomopathogenic fungi, including heat liable bacterial-like toxins, cry-like delta enterotoxins, and bacterial-like zeta toxins. The greater abundance of these genes within the genome and transcriptome, including other novel oral infection related genes, is responsible for the increased oral toxicity that entomopathogenic fungi exhibit when involved in a route of infection other than the cuticle of the insect [33].

Objectives

Based on previous research pertaining to the infection process and virulence factors of entomopathogenic fungi, more specifically *C. clavisporus*, this current study will focus on (1) generating sequences representative of the *C. clavisporus* transcriptome for use as a reference to other *Culicinomyces* species, (2) identifying genes potentially utilized within the infection process of *C. clavisporus*, and (3) linking these genes with oral infection processes to further explain the unique infection nature of *C. clavisporus* as an entomopathogenic fungi. To address these objectives, a combination of RNA extraction, PacBio sequencing, and long read sequence identification and annotation were utilized.

Materials and Methods

Microbial Culture and DNA Extraction

The fungal strain of *Culicinomyces clavisporus* (ARSEF 582) was obtained from the US Department of Agriculture Agricultural Research Service Collection of Entomopathogenic Fungal Cultures (ARSEF, Ithaca, NY) and was grown on Peptone - Yeast Extract - Glucose (PYG). This PYG media contained peptone, yeast extract, distilled water, glucose, and agar [25].

Liquid PYG cultures (ca. 50 mL of broth in flasks) were inoculated using cut pieces of agar plates from which the organism was growing or spore transfer, then were incubated in the dark at 20°C. Fungal tissues were collected from liquid cultures through filtration, grounded in liquid nitrogen, and processed for DNA extraction using the Qiagen DNeasy Plant Mini Kit as per manufacturer's instructions.

Insect Experimentation, RNA Extraction, and Preliminary cDNA Synthesis

Modified PYG liquid cultures were made using DH_20, glucose, $CaC1_2$ $2H_20$, $MgCl_2$ $6H_20$, and soybean oil. Introduction of extracellular nutrient source (whole insect larvae) replaced traditional protein sources found in standard PYG media (Peptone, Yeast Extract) [25]. Late instar larvae from the giant wax moth, *Galleria mellonella*, were obtained commercially, flash frozen, and washed with 95% ethanol before being added to modified PYG liquid cultures. Cultures were then sterilized after extracellular nutrient source was added to the liquid media. Liquid culture was inoculated via spore pellet and allowed to grow for 14 days in the dark at 20°C. Similar to DNA extractions, total RNA extraction was performed on day 14 from the liquid culture growth of both the control and experimental conditions. After sample filtration and liquid nitrogen processing, the QIAGEN RNeasy Plant Mini Kit was used as per manufacturer's instructions. RNA samples were used for cDNA synthesis using the Invitrogen SuperScript III First-Strand Synthesis System for RT-PCR kit as per manufacturer's instructions.

Actin Gene Sequencing and Analysis of Concordance Between gDNA and cDNA

The universal actin primers ED30 (CTAGAAGCATTTGCGGTGGAC) and ED35 (CACGGYATYGTBACCAACTGGG) were diluted and used in conjunction with the extracted *C. clavisporus* gDNA. Polymerase Chain Reactions (PCRs) were performed using the following pattern repeated for 30 cycles: 95°C for 30 seconds, 50°C for 30 seconds, and 72°C for 1 minute. Gel electrophoresis with SYBR Safe was performed on PCR products to confirm presence of DNA. PCR products were gel extracted using the Zymoclean Gel DNA Recovery Kit (Zymo Research) as per manufacturer's instructions. Multiple gel extracted products were pooled together and purified using Zymo DNA Clean and Concentrator kit (Zymo Research). Final quality control gel electrophoresis was performed on purified Actin sample. Samples were sequenced commercially using Sanger sequencing (Psomagen). The resulting nucleotide sequence information was confirmed through homology searches and was used to design species

specific primers labeled CULIACT F & R. The species-specific primers were incorporated in the PCR reactions that use the *C. clavisporus* cDNA samples as templates, in an effort to obtain QC information on all RNA pools. Generated fragments were processed as described above, and the sequences obtained from gDNA, and cDNA samples were compared to each other to confirm similarity. The consensus cDNA sequence was input into NCBI's BLAST tool to find orthologous sequences [3]. A phylogenetic tree was generated using the Phylogeny.fr online program "One Click" mode analysis [13]. This program combines the steps of alignment, alignment refinement, phylogeny, and tree rendering.

cDNA Synthesis for NGS, qPCR and SMRTbell Sequencing

cDNA synthesis was performed using the LEXOGEN Teloprime Full Length cDNA Amplification kit as per manufacturer's instructions. The extracted RNA was used as template material. Following first strand synthesis, PCR reactions were performed for quality control purposes and incorporated the *Culicinomyces*-specific actin primers described above (CULIACT F & R). PCR products were purified and sequenced commercially, and resulting sequences were analyzed through homology searches. SYBR green-based qPCR reactions were performed on first strand cDNA samples to determine the optimal number of cycles, and these numbers were then used on traditional PCR reactions. Similar samples were pooled together. Additionally, size selection gel electrophoresis and extraction were performed on cDNA smears using original cDNA material. Size selection of 1.5-3 kb and 3-10kb were performed, and size selected fragments were purified, and quality tested for concentration and 260/280 ratio using the Nanodrop. Samples were shipped to the University of Florida Interdisciplinary Center for Biotechnology Research (ICBR) for library constructions and SMRT sequencing. Sequencing used the Sequel II system and resulted in the generation and parsing of high quality ("HiFi", or >Q20) reads.

Read Identification and Analysis

Reads generated from the SMRTbell sequencing were uploaded into Galaxy version 21.09.1 [1] The Fasta Statistics tool was used to assess the number of reads and their respective lengths. The Filter Sequences by Length tool was used to obtain manageable data file sizes for later applications. The size of these files was restricted to 500 Mb. Using the CD-HIT SUITE, CD-HIT-EST was utilized, keeping sequence identity parameters at 0.95 [23]. The Fasta files

were then merged using Galaxy's Fasta Merge Files and Filter Unique Sequences tool. CD-HIT-EST was used again on the merged file to produce a final dataset of *C. clavisporus* unigenes. This dataset was uploaded into the Blast2GO suite version 6.0.3 [20]. Homology searches using Blastx were used to annotate all reads, based on the analysis of 5 BLAST matches per read. A small section of identified reads were compared to a common reference sequence identified by the NCBI Blastx tool [3]. Common Reference sequence downloaded was the CDS region in a Fasta format. Once compared, reads were inputted through the reverse complement tool on the Sequence Manipulation Suite if necessary [42]. All reads were then aligned and trimmed to the reference sequence using the Multiple Sequence Alignment tool on EMBL-EBI [16]. Aligned sequences were then compiled for further analysis. Protein translation was performed using the NCBI ORF finder program. Protein sequence corresponding to known domains (as identified through the pfam search tool available at http://pfam.xfam.org/) were aligned to the orthologous reference sequence to create logos using Weblogo 3 [10].

Results

Microbial Growth Timeline

Cultures from liquid flasks (inoculated with agar pieces) were processed using centrifugation to collect fungal spores. These spores were used for inoculation of additional liquid flasks. Flask growth rates improved with this technique. *Figures 1A* and *B* represent growth 7 days after inoculation for both the control and experimental conditions. Microbial growth began to form in the surrounding liquid media in both conditions, but few larger growth hyphae were present in the control condition. *Figures 1C* and *D* show growth after 14 days. Microbial growth of control condition increased, demonstrated by comparison from *Figure 1A* to *C*. Increase in growth also occurred in the experimental condition (*Figure 1D*). Mycelium were observed to attach to larval nutrient sources between day 7 and day 14.

Taxonomic confirmation using the actin gene sequence fragment

Using universal primers ED35 and ED30, PCR products predicted to correspond to actin gene fragment were generated from *C. clavisporus* gDNA. These products were sequenced to obtain the sequence presented in *Figure 2*. Sequence analysis (homology searches) confirmed that the obtained sequence corresponded to actin and identified a potential intron. The generated

sequence allowed for the designing of two custom, species specific primers, CULIACTF and CULIACTR. The custom primers were used on cDNA samples and confirmed the intron and same gene sequence. The cDNA coding sequence was loaded into the NCBI Blast tool and identified the closest known relatives. The closest match to this cDNA sequence was the actin gene from *Drechmeria coniospora,* accession number XM_040802987.1 [50]. This match had an E value of 0.0, percent identity score of 93.14%, and query cover score of 99%.

Twenty (20) of the sequence Blast results were taken and utilized to construct a phylogenetic tree diagram based on the alignment and relatedness to the *C. clavisporus* actin gene sequence fragment, denoted as culi_actin on the tree (*Figure 3*). The *C. clavisporus* sequence was a part of a clade along with 2 other sequences from *D. coniospora* (XM_040802987.1) and *Hirsutella rhossiliensis* (XM_044861584.1). The bootstrap support value for the relatedness of *C. clavisporus* and *D. coniospora* sequence was 0.92. Overall, the clade containing *C. clavisporus*, *D. coniospora*, and *H. rhossiliensis* was supported by a bootstrap value was 0.92.

Summary Statistics

Total PacBio raw sequence data revealed 3,512,145 reads with an average read length of 1,732 base pair. The maximum length of the reads were 12,978 base pairs. After CD-HIT-EST analysis, 8,266 unigenes were identified with an average read length of 722 base pairs. The maximum length of those reads were 1200 base pairs. Using Blast2GO, out of the total number of unigenes, 5,700 reads (69%) resulted in Blast hits, leaving 2,566 reads that did not result in a similarity match.

Sequence Identification Analysis

Based on the Blast hits from the 5,700 unigenes, identification of 1,218 nonduplicated unigenes corresponded to genes from NCBI database. Table 1 shows the top 20 gene identifications with the most unigene hits. A total of 230 unigenes were shown to be homologous to Thioredoxin-like protein sequences, whereas the most common annotations corresponded to unigenes identified as ribosomal proteins (7 out of 20 of the top unigenes). Also, present in the list are unigenes matched to CyanoVirin-N Homology domains (59 total unigenes) and Zeta Toxins (52 total unigenes).

Of the 1,218 unique gene identification, 10 genes were putatively associated with the infection process of *C. clavisporus*. Table 2 identifies the function of each gene as well as the number of unigenes that matched to the gene identified. The cyanovirin-N homology domain had the largest number of unigenes at 59, with the lowest number of reads at 1 for both cytochrome p450 and dipeptidyl-peptidase.

Fifty nine (59) unigenes from the sequence data were identified as homologous to CyanoVirin-N Homology domain (CVNH) (*Table 3*). These sequences were compared to the cyanovirin-N sequence from *B. bassiana (D1-5)*, accession number KGQ04272.1, and an E value of 1E-20 were obtained for the 2 reads (124324128, 112985864). The sequence ID 124324128 was the overall better match with the percent identity score of 72.92% and the query cover score of 90%. These 2 reads align with the entirety of the reference sequence (*Figure 4A*). Sequences were further used to create the logo presented in *Figure 4B*, which highlights the similarity in the protein sequences corresponding to the CVNH domain (PF08881).

Three (3) unigenes from the sequence data were annotated as Cyclophilin (CYP). These reads (150078838, 14681277, 108201449) were compared to the Cyclophilin sequence from *Metarhizium guizhouense* (ARSEF 977), accession number KID83287.1 [22], and the read with the best match was sequence ID 150078838, with an E value of 2E-27, percent identity score of 45.39%, and query cover score of 94%. All of the unigenes align with the entirety of the reference sequence (*Figure 5*).

One (1) unigene from the sequence data was identified as homologous to Cytochrome P450.This read (63572254) was compared to the Cytochrome P450 sequence from *H. rhossiliensis*, accession number XP_044721909.1, with an E value of 1E-51, percent identity of 73.40%, and query cover of 92%. This read matches the reference sequence from bases 931-1,547 (*Figure 6*).

Two (2) unigenes from the sequence data were annotated as Dipeptidyl-peptidase 5 (DPP). These reads (2883929, 91556138) were compared to the partial Dipeptidyl-peptidase 5 sequence of *Tolypocladium capitatum*, accession number PNY23717.1 [39], with an E value of 1E-83. The sequence ID 2883929 was the best match, with a percent identity score of 79.12% and query cover score of 90%. All of the unigenes align with the entirety of the reference sequence (*Figure 7*).

One (1) unigene from the sequence data was identified as homologous to endochitinase A1. This read (44829103) was compared to the endochitinase A1 sequence from *Tolypocladium ophioglossoides* (*CBS 100239*), accession number KND89903.1 [38]. This read had an E value of 2E-114, percent identity score of 66.93%, and query cover score of 99%. This read matches the reference sequence from bases 1-782 (*Figure 8A*). Sequences were further used to create the logo presented in *Figure 8B*, which highlights the similarity in the protein sequences corresponding to the PF00704 protein domain. This domain is also references to as Glycoside Hydrolase family 18 (GH18), as defines by the CAZy database [14]. Enzymes of the GH18 family are mainly known as chitinases.

Fifteen (15) unigenes were annotated as non ribosomal peptide synthetases (NRPS) from the sequence data (*Table 4*). These unigenes were compared to the partial non ribosomal peptide synthetase sequence of *Trichoderma virens*, accession number AAX59995.1 [47]. The sequence ID 33883071 was the best match, with an E value of 1E-25, percent identity score of 49.04%, and query cover score of 90%. These unigenes match the reference sequence in the middle of the sequence (*Figure 9*).

Two (2) unigenes from the sequence data were identified as homologous to six-hairpin glycosidase. These reads (264389, 79429654) were compared to the six-hairpin glycosidase sequence from *M. guizhouense* (ARSEF 977), accession number KID92891.1 [22]. The sequence ID 264389 was the best match, with an E value of 1E-46, percent identity score of 85.23%, and query cover score of 98%. The 2 unigenes match the reference sequence from bases 1285-1551 (*Figure 10A*). Sequences were further used to create the logo presented in *Figure 10B*, which highlights the similarity in the protein sequences corresponding to the PF06824 protein domain. This domain is also referred to as Glycoside Hydrolase family 125 (GH125), as defined by the CAZy database [14]. Enzymes of the GH18 family are mainly known as mannosidases.

Three (3) unigenes were annotated as Salmonella Plasmid Virulence (SpvB) from the sequence data. These reads (17237673, 149227152, 165348163) were compared to SpvB domain containing protein from *Trematosphaeria pertusa*, accession number XP_033686269.1 [21]. The sequence ID 165348163 was the best match, with an E value of 1E-06, percent identity score of 52.83%, and query cover score of 100%. The 3 unigenes match the reference sequence. All the unigenes match the reference sequence from nucleotides 4,591 – 4,816 (*Figure 11*).

Three (3) of the unigenes from the sequence data were found to be homologous to Subtilisin-like protease. These reads (88343209, 123470957, 163512962) were compared to Subtilisin-like serine protease from *Thozetella sp.* (PMI_491), accession number KAH8892393.1. The sequence ID 88343209 was the best match, with an E value of 2E-05, percent identity score of 63.16%, and query cover score of 57%. These unigenes match the reference sequence from bases 488-654 (*Figure 12A*). Sequences were further used to create the logo presented in *Figure 12B,* which highlights the similarity in the protein sequences corresponding to the subtilase domain (PF00082).

Fifty two (52) unigenes were identified to be homologous to Zeta toxins from the sequence data (*Table 5*). These reads were compared to the Zeta toxin sequence from *Fusarium austroafricanum*, accession number KAF4448150.1 [26]. Two (2) reads (3736654, 69077145) had an E value of 1E-07, but the sequence ID 3736654 was the better match with a percent identity score of 67.44% and a query cover score of 69%. All of the unigenes align with the reference sequence in its entirety (*Figure 13*).

Discussion

C. clavisporus is an entomopathogenic fungi that utilizes a unique infection process as compared to other entomopathogenic fungi. Different from other entomopathogenic fungi, *C. clavisporus* infects their insect hosts via ingestion instead of attachment to the cuticle, which requires an altered set of genes engaged in this process. The analysis into the transcriptome of *C. clavisporus* has identified possible genes that are involved and aid in the disease process.

Phylogenetic Analysis

After phylogenic analysis of the actin gene fragment from *C. clavisporus*, the closest related organism was identified as *D. coniospora*, an endoparasitic nematophagous fungus (*Figure 3*). From past research, *D. coniospora* was known to follow a saprophytic lifestyle, but recent research has shed light on loss of genes associated with a saprophytic lifestyle and an increase in protein families necessary for an endoparasitic lifestyle. Much like the types of infection processes for entomopathogenic fungi, endoparasitic nematophagous fungi have three methods of infection: nematode-trapping fungi, female and egg parasites, and infection via conidia ingestion or spore adhesion. *D. coniospora* is a model organism for infection via spore

adhesion. Genome analysis performed on *D. coniospora* supported the claim that it has close relationships to insect pathogens such as *Metarhizium spp*., *B. bassiana*, and *H. minnesotensis* [50]. *B. bassiana* is an entomopathogenic fungi, similar to *C. clavisporus* and the gene sequence of six-hairpin glycosidase from our transcriptome analysis showed to best match an organism from *Metarhizium* spp. These two details provide support for a close phylogenetic relationship between *C. clavisporus* and *D. coniospora*. Phylogeny completed by Zhang et al. showed that *D. coniospora* was assigned to *Ophiocordyciptaceae*, which revealed shared lineage from an entomopathogenic ancestor. From this shared lineage, past research suggested that the infection process was inherited and adapted to its nematode host [50]. This explains why it was found to be *C. clavisporus'* closest relative from our phylogenetic analysis (*Figure 3*).

Outside of the closest relation, *C. clavisporus* is contained within a clade that included *H. rhossiliensis*. Past research has made the connection that *H. rhossiliensis* and *H. minnesotensis* are unique in their method to attack prey, nematodes specifically, and share the ability to produce spores to parasitize the nematode [29]. This makes *H. rhossiliensis* and *H. minnesotensis* closely related species, which mean *H. minnesotensis* can serve as a proxy to provide support to why *H. rhossiliensis* is a related species to *C. clavisporus*. Similar findings to Zhang et al. were found in the phylogenetic analysis of Lai et al. in that they support the divergence of endoparasitic species from entomopathogenic fungi. Unlike Zhang et al, Lai et al. did find evidence that the use of subtilisin genes were lost during this divergence [29,50]. Of the genes expanded upon in the genome, chitinases as well as NRPS genes were found in larger quantities than nematode trapping fungi and even in insect pathogens. This also supports the earlier hypothesis that the infection process inherited by entomopathogenic fungi was adapted to better fit the nematode host. From the evidence found in both Zhang et al. and Lai et al., our phylogenetic analysis of *C. clavisporus*' clade is supported.

Gene Association to Infection Method

C. clavisporus is known to utilize the route of oral infection to infect its insect hosts. From the transcriptome analysis, 10 genes were identified with specific interest in their possible roles that could aid in the infection process of *C. clavisporus* (Table 2). Based on lineage of all entomopathogenic fungi and their relatives, genes of importance are conserved within the genome such as endochitinase, NRPS, and six-hairpin glycosidase [7,12,33]. While these genes

may not abundantly active in all fungi during infection, there is still an advantage to conserving them within the genome. Genes associated with the cuticular infection route, such as endochitinase A1, subtilisin-like proteases, DPP, and six-hairpin glycosidases were identified within the *C. clavisporus* transcriptome yet function in one way or another to degrade the cuticle of the insect host (*Table 6*). The genes suggested to be associated with the oral infection route include zeta toxins and SpvB (*Table 7*) [33]. These two genes both involve an aspect of ingestion, mainly the pathogenic bacteria or fungi would release the product of these transcribed genes once ingested by the host. The remaining four genes of interest (CVNH, cyclophilins, cytochrome P450, and NRPS) play a role in pathogenicity, just in the context of impacting cellular processes related to virulence (*Table 8*) [7,27,34,48]. The continued conservation and utilization of these genes and their roles as virulence factors despite different methods of infection between entomopathogenic fungi show evidence of their relatedness as well as their general functionality as an entomopathogenic fungi. The initial infection and colonization of the insect host is only the first step, requiring the genes controlling cellular processes to continue the virulence of the entomopathogenic fungi until the death of the insect host ensues.

Overall, the genes of interest yielded 141 unigenes analyzed during this study of the *C. clavisporus* transcriptome. Of those 141, 6% or 8 unigenes were found to be associated with the cuticular infection route. The number of unigenes associated with the oral infection route was found to be 55 unigenes, or 39%. Lastly, there were 55%, or 78 unigenes, associated with regulating cellular functions related to virulence factors (*Figure 13*). *Figure 14A* shows a breakdown of the genes specifically associated with the possible infection methods utilized by entomopathogenic fungi, such as *C. clavisporus* and other related fungi. Out of the total 63 unigenes identified, 13% were associated with cuticular infection while 87% were associated with the oral infection route. *Figure 14B* shows the distribution of unigenes per gene in relation to infection method. While the diversity of genes associated with cuticle infection is greater, the sheer number of unigenes active within the *C. clavisporus* transcriptome analysis supports previous research conclusions that the preferred method of infection utilized by *C. clavisporus* is the oral infection method.

Genes Involved with Cellular Regulation Associated with Virulence

The first classification of genes from *Table 2* are those involved with the cellular processes that continue the virulence strength of the entomopathogenic fungi once infection has taken hold in the insect host. The first of which being CyanoVirin -N Homology (CVNH) domain, having had the most unigenes identified form the transcriptome analysis. Of the 59 unigenes, 2 of the reads were considered to best match the reference sequence from *B. bassiana*. *Figure 4* shows the MUSCLE alignment for sequence IDs 124324128 and 112985864 against the reference sequence. The CVNH domain has been found in many types of organisms, including symbiotic, saprotrophic, and pathogenic organisms. Potential functions for CVNHs within pathogenic fungi genomes include cell to cell recognition, more specifically the recognition of glycoproteins and surface-bound sugars. The recognition could also be used to identify metabolic signals in relation to the nutrient source status [27]. It is known that many fungal effectors are cysteine-rich, small, secreted proteins. From past research, the CVNH domain of the *Sclerotinia sclerotiorum* genome was analyzed for functionality. Overall, when the CVNH domain was silenced within *S. sclerotiorum*, the virulence and growth was significantly reduced. This indicates that the function of the CVNH domain is involved in the virulence of some fungal pathogens. These small effector proteins with a CVNH domain do not have the same impact in all cases, as there are more than 3 types of domain architecture that may alter the CVNH's impact on virulence [32]. From our transcriptome analysis, the continued identification of these unigenes, currently in *C. clavisporus*, only leads to the increased need to study the expression levels and domain architecture of these small effector proteins. This would be the main method to find out the specific function of the CVNH domain within the *C. clavisporus* transcriptome.

Along with genes involved in cell to cell recognition, genes that have roles in transcription regulation and the production of molecules are also of importance to the *C. clavisporus* infection process. The Nonribosomal peptide synthetase (NRPS) functions to biosynthesize small peptides without utilizing the ribosomal protein synthesis pathway, as previously stated. Past research focused on the deletion of different NRPS genes within *Cochliobolus heterostrophus*, which resulted in the decrease in virulence when NPS6 was deleted. The product produced from the transcription of the NRPS gene may have the impact

described on the pathogen's virulence. While the products of this gene are still unknown, it hypothesized that siderophores are produced, evidenced because most fungal siderophores are produced by NRPSs. Siderophores are used to acquire iron from hosts [35]. Recent research confirmed that this gene is conserved across fungi and their products have function regarding basic cellular needs, such as growth, pathogenesis, and reproduction [7]. From our transcriptome analysis, there were 15 unigenes identified, from *C. clavisporus* transcriptome, as homologous to the NRPS gene from *T. virens*. The best matching read was compared to the reference sequence and matched to a portion within the middle of the reference sequence (*Figure 9*). As demonstrated regarding other genes, the NRPS gene is suspected of going through horizonal transmission from bacteria genomes to fungi [7]. The NRPS gene was likely inherited and conserved within the transcriptome through close relation to other pathogenic organisms.

The regulation of environmental stress is also an important factor related to the virulence of entomopathogenic fungi. The cyclophilin (CYP) gene and genes with the CYP domain are responsible for transcription regulation, cell signaling, and other cellular processes [34]. The *C. clavisporus* transcriptome analysis revealed 3 unigenes identified as homologous to the cyclophilin gene from *M. guizhouense*. These unigenes (Sequence IDs 150078838, 14681277, 108201449) aligned completely with the reference sequence as seen in *Figure 5*. Although the understanding of the functionality of CYP is still the subject of investigation in filamentous fungi, studies have shown that genes with the CYP domain in *B. bassiana* has positively contributed to virulence during infection [34]. Both *B. bassiana* and *C. clavisporus* are entomopathogenic fungi belonging to the same order, *Hypocreales*. This would suggest that CYP could have the same or similar function during the *C. clavisporus* oral infection.

The intracellular functions that occur during entomopathogenic fungi infection rely on a series of molecules or proteins that work together for reactions to happen and products to be formed. Cytochrome P450 monooxygenases are a family of enzymes that functions closely with cyclophilins and regulate cellular processes. These enzymes are functionally versatile, aiding to their structurally diverse nature. They play a crucial role in pathogenic fungi given their function in producing secondary metabolites used during pathogenesis [48]. The reactions that cytochrome P450s mediate during pathogenesis occur because of their role as an electron donor, and ultimately have a major impact on the result of the cyclophilin reactions. Overall fungi

contain only a handful of cytochrome P450s, supported from the results of the transcriptome analysis [15]. The *C. clavisporus* transcriptome sequence revealed 1 unigene present during its infection process. *Figure 6* shows the alignment of this sequence to a reference sequence from *H. rhossiliensis*. As previously stated, it is known that *H. rhossiliensis* is *C. clavisporus'* closest known relative (*Figure 3*). The presence of the cytochrome P450 unigene and the cyclophilin unigenes suggest that reactions were taking place during infection to positively impact the pathogenesis process.

Genes with Possible Cuticular Infection Association

The second classification of genes found within the *C. clavisporus* transcriptome are those with functions involved with the cuticular infection method. These genes include endochitinase A1, subtilisin-like proteases, DPP, and six-hairpin glycosidases. The cuticular infection process involves attaching to the cuticle of the insect and killing the host. The composition of the cuticle is mainly made of chitin. The degradation of chitin is the function of endochitinase A1. From past research, fungi have been producing endochitinase to control more than just insect hosts, even plant pathogenic fungi. Recent research has shown that endochitinase has a crucial role in the action against the pathogen. The comparison of media type on endochitinase production and activity only furthered evidence of this anti-host enzyme [12]. *C. clavisporus* does not utilize the cuticular infection route, but our transcriptome analysis revealed endochitinase. Our transcriptome analysis identified 1 unigene that was identified as homologous to endochitinase A1. The reference sequence was from *T. ophioglossoides*, to which the read matched with first half of the reference sequence (*Figure 8*). While endochitinase is not directly involved in the oral infection route that *C. clavisporus* utilizes, having this gene conserved within the transcriptome may provide benefits currently unknown about its infection method.

Along with endochitinase, glycosidase, otherwise known as glycoside hydrolase (GH), functions to catalyze the hydrolysis of molecules on the insect cuticle, more specifically the glycosidic bonds on complex sugars present on the cuticle [33]. Entomopathogenic fungi possess a wide range of genes from the GH family, a small number of which are known as six-hairpin glycosidases. These six-hairpin glycosidases share a common structure consisting of 6 helical hairpins. This motif is found to resemble the bacterial and fungal α-L-rhamnosidase [49]. From the *C. clavisporus* transcriptome analysis, 2 unigenes were annotated as Six-hairpin

Glycosidases. These two reads were compared to a reference sequence from *M. guizhouense* and matched to the end of the reference sequence (*Figure 10*) Given the cuticular function of glycosidases, this gene is likely conserved within the genome of *C. clavisporus*, but not utilized during oral infection processes.

Subtilisin-like serine proteases are known to disrupt the physiological integrity of the host during infection. These proteases are used as a virulence factor for many pathogenic fungi, including those that utilize different infection methods and effects various host types. The function of subtilisin-like serine proteases involves the penetration and digestion of the cuticles of insects and nematodes. Using phylogenetic analysis, a recent study found clustering of these proteases from both nematode parasitic fungi and entomopathogenic fungi. Of the paralogous subtilisin-like serine proteases found in *Metarhizium anisopliae*, it was found that *PR1A* was a key virulence factor during cuticle degradation [30]. On the other hand, *B. bassiana* genes show that Subtilisin-like *PR1* proteases are rarely produced after initial infection occurs. From the transcriptome analysis of *C. clavisporus*, 3 unigenes were identified as homologous to subtilisin-like proteases. These reads were compared to a reference sequence from *Thozetella sp.* and corresponded to the middle of the sequence. Given the various levels of impact that *PR1* has on virulence regarding entomopathogenic fungi, a recent study focused on the enhancement of the *PR1* gene in combination with chitinase or a bacterial toxin gene [18]. In combination with the endochitinase A1 unigenes identified in the *C. clavisporus* transcriptome, the subtilisin-like proteases may have been functionally enhanced to play a significant role in the *C. clavisporus* infection process.

The cuticular infection route, like any other portion of the infection processes only works if genes work together to begin the initial infection. Dipeptidyl-peptidase (DPP) is a part of the serine protease family. Overall, these enzymes cleave dipeptides from proteins or peptides at the N-terminus [5]. Past research has shown evidence of DPPs possible involvement in pathogenesis from secreted DPP in yeast, *P. pastoris*. These secreted DPP enzymes were enzymatically active, meaning they play a role in digestion. The digestion in this case was keratin, but similar function could be evidenced in the *C. clavisporus* infection process from the presence of these DPPs in our transcriptome analysis [46]. Dipeptidyl-peptidase 5 was identified in 2 unigenes, best matching a DPP sequence from *T. capitatum*. The unigenes aligned completely with the

reference sequence as depicted in *Figure 7*. *Tolypocladium* species are known as "host jumpers" [39]. One type of host these fungi can inhabit are insects, which is what links *C. clavisporus* to *T. capitatum*. Added evidence is from when DPP is active within bacterium it has been proven to positively impact bacterial virulence using proteases [46]. *C. clavisporus* may utilize adapted forms of bacterial genes to infect insect hosts.

Proposed Gene Association with Oral Infectivity

The last gene classification is those suspected to be involved with the oral infection method. This method is utilized by entomopathogenic fungi that require the insect to ingest spores or other cellular molecules to initialize the fungal infection. The genes suspected to be involved include zeta toxins and the SpvB effector gene, both of which have a common origin, them being originally inherited from bacterial relatives from the entomopathogenic fungi lineage. First, zeta toxins found in transcriptome analyses of entomopathogenic fungi function similar to secondary metabolites [37]. From the *C. clavisporus* transcriptome analysis, 52 unigenes were found and identified as homologous to the zeta toxin gene from *F. austroafricanum*. These unigenes correspond with the entirety of the reference sequence (*Figure 12*). In *B. bassiana*, zeta toxins were identified in a genomic analysis. It was hypothesized that these zeta toxins could be involved in infection via ingestion of spores, or act as bactericidal agents [37]. Recent research has shed light on the utilization of zeta toxins, in that these genes were highly expressed during fungal infection of mosquitoes [2]. Given the knowledge that *C. clavisporus* is known to utilize oral infection methods compared to cuticular infection methods and its preferred host species are mosquitos (*C. quinquefasciatus, A. aegypti, and A. stephensi*), it is likely that zeta toxins are utilized during infection once fungal spores have been ingested by the host. The sequence abundance of this gene compared to genes more associated with cuticular infection provides support for the hypothesis that the main method of *C. clavisporus* infection through ingestion.

Next, the bacteria *Salmonella enterica* virulence is because of the SpvB effector gene, which is also the last of the genes of interest from our transcriptome analysis. This gene includes an operon attached to a large virulence plasmid. This effector gene functions to ADP-ribosylates and depolymerizes actin filaments inside the infected cells [19]. From this, recent studies have suggested that SpvB is released once the bacteria is inside the host [19]. The SpvB gene has been seen to be associated with fungal virulence as well [51]. From our analysis, 3 unigenes were

annotated and identified as SpvB from the *C. clavisporus* transcriptome and the reads were compared to the *T. pertusa* as a reference sequence. The adaptation of bacterial pathogenic genes to fungal hosts is very common, but the SpvB gene is more easily adapted to *C. clavisporus* due to the method of delivery. As stated, the bacterium must be inside the host cell before releasing the effector molecules. *C. clavisporus* spores must be ingested before the oral infectivity process can begin. This parallel suggests that the SpvB gene *C. clavisporus* utilizes functions similarly to the bacterial gene and suggests that the impact that SpvB has on bacterial virulence also occurs during the *C. clavisporus* oral infectivity process.

Conclusion

This study completed a comprehensive analysis of the transcriptome of *C. clavisporus* to provide evidence supporting the hypothesis that *C. clavisporus* utilizes oral infection routes as its main infection mode. A phylogenic tree analysis was completed using an actin gene sequence fragment from *C. clavisporus* and revealed its close relationships, *D. coniospora* and *H. rhossiliensis*. These organisms are parasitic to nematodes, not fungi. The transcriptome analysis of *C. clavisporus* identified 10 genes of interest, known to be utilized by other pathogenic organisms, that may shed light on the virulence and infection method of *C. clavisporus*. Genes found to regulate cellular functions related to virulence, such as CVNH, cyclophilins, cytochrome P450, and NRPS, were found in abundance within the transcriptome. While these are nonspecific to *C. clavisporus* as an entomopathogenic fungi, these virulence factors are shared among others from the same lineage. Genes such as endochitinase A1, subtilisin-like proteases, DPP, and six-hairpin glycosidases were suggested to be associated with cuticular infection routes, as they may function to degrade the cuticle of the insect host. The last of the genes of interest, zeta toxins and SpvB, were suggested to be associated with oral infection routes, which is the known infection route of *C. clavisporus* from past research. Importantly, unigenes identified as zeta toxins and SpvB were found to be more abundant, compared to the number of unigenes associated with cuticular infection genes. This evidence supports the hypothesis that *C. clavisporus* utilizes these genes during infection as opposed to genes involved in degrading the cuticle.

The results of this study would benefit from an in depth investigation into the expression of the gene products described above given the protein coding potential of the unigenes found

within our transcriptome analysis. Similar to many studies performed on entomopathogenic fungi, a study comparing the various *Culicinomyces* strains and their associated virulence factors would increase knowledge regarding all aspects of the oral infection route of other entomopathogenic fungi. The transcriptome sequences from this analysis may provide a basis for future studies on the *Culicinomyces spp*. regarding its virulence factors and their possible utilization in the bioinsecticide industry, like seen with other entomopathogenic fungi species such as *B. bassiana*. Of the genes of interest, the zeta toxin gene, the SpvB domain containing protein gene, and the CVNH domain are potential good candidate for further investigation aimed to develop a bioinsecticide from *C. clavisporus*.

References

1. Afgan, E., Baker, D., Batut, B., van den Beek, M., Bouvier, D., Čech, M., Chilton, J., Clements, D., Coraor, N., Grüning, B., Guerler, A., Hillman-Jackson, J., Jalili, J., Rasche, H., Soranzo, N., Goecks, J., Taylor, J., Nekrutenko, A., Blankenberg, D. (2018). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update, *Nucleic Acids Research*, 46(1) ,537–544.

2. Altimira, F., Arias-Aravena, M., Jian, L., Real, N., Correa, P., González, C., Godoy, S., Castro, J. F., Zamora, O., Vergara, C., Vitta, N., & Tapia, E. (2022). Genomic and Experimental Analysis of the Insecticidal Factors Secreted by the Entomopathogenic Fungus Beauveria pseudobassiana RGM 2184. *Journal of Fungi*, *8*(3), 253.

3. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990) "Basic local alignment search tool." J. *Mol. Biol*. 215:403-410.

4. Ball, B., Langille, M., & Geddes-McAlister, J. (2020). Fun (gi) omics: Advanced and Diverse Technologies to Explore Emerging Fungal Pathogens and Define Mechanisms of Antifungal Resistance. *Mbio*, *11*(5).

5. Beauvais, A., Monod, M., Wyniger, J., Debeaupuis, J. P., Grouzmann, E., Brakch, N., ... & Latge, J. (1997). Dipeptidyl-peptidase IV secreted by Aspergillus fumigatus, a fungus pathogenic to humans. *Infection and immunity*, *65*(8), 3042-3047.

6. Bhadauria, V., Popescu, L., Zhao, W.-S., & Peng, Y.-L. (2007). Fungal transcriptomics. *Microbiological research*, *162*(4), 285-298.

7. Bushley, K. E., & Turgeon, B. G. (2010). Phylogenomics reveals subfamilies of fungal nonribosomal peptide synthetases and their evolutionary relationships. *BMC Evolutionary Biology*, *10*(1), 26.

8. Chen, D., Du, Y., Fan, X., Zhu, Z., Jiang, H., Wang, J., Fan, Y., Chen, H., Zhou, D., & Xiong, C. (2020). Reconstruction and functional annotation of Ascosphaera apis full-length transcriptome utilizing PacBio long reads combined with Illumina short reads. *Journal of Invertebrate Pathology*, *176*, 107475.

9. Cook, J., Bruckart, W., Coulson, J., Goettel, M., Humber, R., Lumsden, R., Maddox, J., McManus, M., Moore, L., Meyer, S., Quimby, P., Stack, J., Vaughn, J. (1996). Safety of Microorganisms Intended for Pest and Plant Disease Control: A Framework for Scientific Evaluation, Biological Control, 7(3), 333-351.

10. Crooks GE, Hon G, Chandonia JM, Brenner SE (2004). WebLogo: A sequence logo generator, Genome Research, 14:1188-1190.

11. Daly, H. V., Daly, P. E. H. V., Purcell, A. P. E. A. H., Doyen, J. T., Purcell, A. H., Ehrlich, P. R., Daly, B. B., & Ross, E. S. (1998). Introduction to Insect Biology and Diversity. *Oxford University Press.*

12. Deng, S., Lorito, M., Penttilä, M., & Harman, G. E. (2007). Overexpression of an endochitinase gene (ThEn-42) in Trichoderma atroviride for increased production of antifungal enzymes and enhanced antagonist action against pathogenic fungi. *Applied biochemistry and biotechnology*, *142*(1), 81–94.

13. Dereeper A.*, Guignon V.*, Blanc G., Audic S., Buffet S., Chevenet F., Dufayard J.F., Guindon S., Lefort V., Lescot M., Claverie J.M., Gascuel O. *Phylogeny.fr: robust phylogenetic analysis for the non-specialist.* Nucleic Acids Res. 2008 Jul 1;36(Web Server issue): W465-9.

14. Drula, E., Garron, M.L., Dogan, S., Lombard, V., Henrissat, B., Terrapon, N. (2022).The carbohydrate-active enzyme database: functions and literature, Nucleic Acids Research, Volume 50, Issue D1, Pages D571–D577.

15. Durairaj, P., Hur, J. S., & Yun, H. (2016). Versatile biocatalysis of fungal cytochrome P450 monooxygenases. *Microbial cell factories*, *15*(1), 1-16.

16. Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 32(5):1792-1797.

17. Gao, S., Li, Y., Gao, J., Suo, Y., Fu, K., Li, Y., & Chen, J. (2014). Genome sequence and virulence variation-related transcriptome profiles of *Curvularia lunata*, an important maize pathogenic fungus. *BMC genomics*, *15*(1), 1-18.

18. Gao, B. J., Mou, Y. N., Tong, S. M., Ying, S. H., & Feng, M. G. (2020). Subtilisin-like Pr1 proteases marking the evolution of pathogenicity in a wide-spectrum insect-pathogenic fungus. *Virulence*, *11*(1), 365–380.

19. Gotoh, H., Okada, N., Kim, Y. G., Shiraishi, K., Hirami, N., Haneda, T., Kurita, A., Kikuchi, Y., & Danbara, H. (2003). Extracellular secretion of the virulence plasmid-encoded ADP-ribosyltransferase SpvB in Salmonella. *Microbial pathogenesis*, *34*(5), 227–238.

20. Götz S., Garcia-Gomez JM., Terol J., Williams TD., Nagaraj SH., Nueda MJ., Robles M., Talon M., Dopazo J. and Conesa A. (2008). High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic acids research*, 36(10), 3420-35.

21. Haridas, S., Albert, R., Binder, M., Bloem, J., LaButti, K., Salamov, A., Andreopoulos, B., Baker, S., Barry, K., Bills, G., Bluhm, B. H., Cannon, C., Castanera, R., Culley, D., Daum, C., Ezra, D., González, J. B., Henrissat, B., Kuo, A., & Grigoriev, I. (2020). 101 Dothideomycetes genomes: A test case for predicting lifestyles and emergence of pathogens. *Studies in Mycology, 96*.

22. Hu, X., Xiao, G., Zheng, P., Shang, Y., Su, Y., Zhang, X., Liu, X., Zhan, S., Leger, R. J. S., & Wang, C. (2014). Trajectory and genomic determinants of fungal-pathogen speciation and host adaptation. *Proceedings of the National Academy of Sciences*, *111*(47), 16796-16801.

23. Huang, Y., Niu, B., Gao, Y., Fu, L., Li, W. CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics,* 2010(26): 680-682.

24. Karthika, A., Ramachandran, B., Chitra, J., Prabhu, D., Rajamanikandan, S., Veerapandiyan, M., & Jeyakanthan, J. (2021). Molecular dynamics simulation of Toxin-Antitoxin (TA) system in Acinetobacter baumannii to explore the novel mechanism for inhibition of cell wall biosynthesis: Zeta Toxin as an effective therapeutic target. *Journal of cellular biochemistry*, *122*(12), 1832–1847.

25. Kerwin, J. L., & Petersen, E. E. (1997). Fungi: oomycetes and chytridiomycetes. In *Manual of techniques in insect pathology* (pp. 251-268). Academic Press.

26. Kim, HS., Lohmar, J.M., Busman, M. *et al.* Correction to: Identification and distribution of gene clusters required for synthesis of sphingolipid metabolism inhibitors in diverse species of the filamentous fungus *Fusarium*. *BMC Genomics* **21,** 712 (2020).

27. Koharudin, L. M., Viscomi, A. R., Jee, J. G., Ottonello, S., & Gronenborn, A. M. (2008). The evolutionarily conserved family of cyanovirin-N homologs: structures and carbohydrate specificity. *Structure (London, England: 1993)*, *16*(4), 570–584.

28. Lacey, L., & Goettel, M. (1995). Current developments in microbial control of insect pests and prospects for the early 21st century. *Entomophaga*, *40*(1), 3-27.

29. Lai, Y., Liu, K., Zhang, X., Zhang, X., Li, K., Wang, N., Shu, C., Wu, Y., Wang, C., Bushley, K. E., Xiang, M., & Liu, X. (2014). Comparative genomics and transcriptomics analyses reveal divergent lifestyle features of nematode endoparasitic fungus Hirsutella minnesotensis. *Genome biology and evolution*, *6*(11), 3077–3093.

30. Li, J., Yu, L., Yang, J., Dong, L., Tian, B., Yu, Z., ... & Zhang, K. (2010). New insights into the evolution of subtilisin-like serine protease genes in Pezizomycotina. *BMC evolutionary biology*, *10*(1), 1-14.

31. Lovett, B., & St. Leger, R. J. (2018). Genetically engineering better fungal biopesticides. *Pest management science*, *74*(4), 781-789.

32. Lyu, X., Shen, C., Fu, Y. *et al.* Comparative genomic and transcriptional analyses of the carbohydrate-active enzymes and secretomes of phytopathogenic fungi reveal their significant roles during infection and development. *Sci Rep* **5,** 15565 (2015).

33. Mannino, M. C., Huarte-Bonnet, C., Davyt-Colo, B., & Pedrini, N. (2019). Is the insect cuticle the only entry gate for fungal infection? Insights into alternative modes of action of entomopathogenic fungi. *Journal of fungi*, *5*(2), 33.

34. Mo, C., Xie, C., Wang, G., Liu, J., Hao, Q., Xiao, X., & Xiao, Y. (2019). Genome-wide identification and characterization of the cyclophilin gene family in the nematophagous fungus Purpureocillium lilacinum. *International journal of molecular sciences*, *20*(12), 2978.

35. Oide, S., Moeder, W., Krasnoff, S., Gibson, D., Haas, H., Yoshioka, K., & Turgeon, B. G. (2006). NPS6, encoding a nonribosomal peptide synthetase involved in siderophore-mediated iron metabolism, is a conserved virulence determinant of plant pathogenic ascomycetes. *The Plant cell*, *18*(10), 2836–2853.

36. Olivera, I. E., Fins, K. C., Rodriguez, S. A., Abiff, S. K., Tartar, J. L., & Tartar, A. (2016). Glycoside hydrolases family 20 (GH20) represent putative virulence factors that are shared by animal pathogenic oomycetes but are absent in phytopathogens. *BMC microbiology*, *16*(1), 1-11.

37. Ortiz-Urquiza A. (2021). The Split Personality of Beauveria bassiana: Understanding the Molecular Basis of Fungal Parasitism and Mutualism. *mSystems*, *6*(4), e0076621.

38. Quandt, C. A., Bushley, K. E., & Spatafora, J. W. (2015). The genome of the truffle-parasite Tolypocladium ophioglossoides and the evolution of antifungal peptaibiotics. *BMC Genomics*, *16*(1), 553.

39. Quandt, C. A., Patterson, W., & Spatafora, J. W. (2018). Harnessing the power of phylogenomics to disentangle the directionality and signatures of interkingdom host jumping in the parasitic fungal genus Tolypocladium. Mycologia, 110(1), 104–117.

40. Rodrigues, J., Luz, C., & Humber, R. A. (2017). New insights into the in vitro development and virulence of Culicinomyces spp. as fungal pathogens of Aedes aegypti. *Journal of invertebrate pathology*, *146*, 7-13.

41. Shah, P., & Pell, J. (2003). Entomopathogenic fungi as biological control agents. *Applied microbiology and biotechnology*, *61*(5), 413-423.

42. Stothard P (2000) The Sequence Manipulation Suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *Biotechniques* 28:1102-1104.

43. Sweeney, A. W. (1979). Infection of mosquito larvae by Culicinomyces sp. through anal papillae. *Journal of Invertebrate Pathology*, *33*(2), 249-251.

44. Sweeney, A. W. (1975). The Mode of Infection of the Insect Pathogenic Fungus Culicinomyces in Larvae of the Mosquito Culex Fatigans. *Australian Journal of Zoology*, *23*(1), 49-57.

45. Velasquez, P. F. Q., Abiff, S. K., Fins, K. C., Conway, Q. B., Salazar, N. C., Delgado, A. P., Dawes, J. K., Douma, L. G., & Tartar, A. (2014). Transcriptome analysis of the entomopathogenic oomycete Lagenidium giganteum reveals putative virulence factors. *Applied and environmental microbiology*, *80*(20), 6427-6436.

46. Vermout, S., Baldo, A., Tabart, J., Losson, B., & Mignon, B. (2008). Secreted dipeptidyl peptidases as potential virulence factors for Microsporum canis. *FEMS Immunology & Medical Microbiology*, *54*(3), 299-308.

47. Wei, X., Yang, F., & Straney, D. C. (2005). Multiple non-ribosomal peptide synthetase genes determine peptaibol synthesis in Trichoderma virens. *Canadian journal of microbiology*, *51*(5), 423-429.

48. Xiao, G., Ying, SH., Zheng, P. *et al.* Genomic perspectives on the evolution of fungal entomopathogenicity in *Beauveria bassiana*. *Sci Rep* **2,** 483 (2012).

49. Yadav, V., Yadav, D., & Yadav, K. D. S. (2010). In silico analysis of α-L-rhamnosidase protein sequences from different source organisms. *Onl. J. Bioinform*, *11*(2), 293-301.

50. Zhang, L., Zhou, Z., Guo, Q., Fokkens, L., Miskei, M., Pócsi, I., Zhang, W., Chen, M., Wang, L., Sun, Y., Donzelli, B. G., Gibson, D. M., Nelson, D. R., Luo, J. G., Rep, M., Liu, H., Yang, S., Wang, J., Krasnoff, S. B., Xu, Y., … Lin, M. (2016). Insights into Adaptations to a Near-Obligate Nematode Endoparasitic Lifestyle from the Finished Genome of Drechmeria coniospora. *Scientific reports*, *6*, 23122.

51. Zhang, Z., Lu, Y., Xu, W. *et al.* (2020). Influence of genetic diversity of seventeen *Beauveria bassiana* isolates from different hosts on virulence by comparative genomics. *BMC Genomics* **21,** 451.

52. Zhao, L., Zhang, H., Kohnen, M. V., Prasad, K. V., Gu, L., & Reddy, A. S. (2019). Analysis of transcriptome and epitranscriptome in plants using PacBio Iso-Seq and nanopore-based direct RNA sequencing. *Frontiers in genetics, 10*, 253.

Appendix

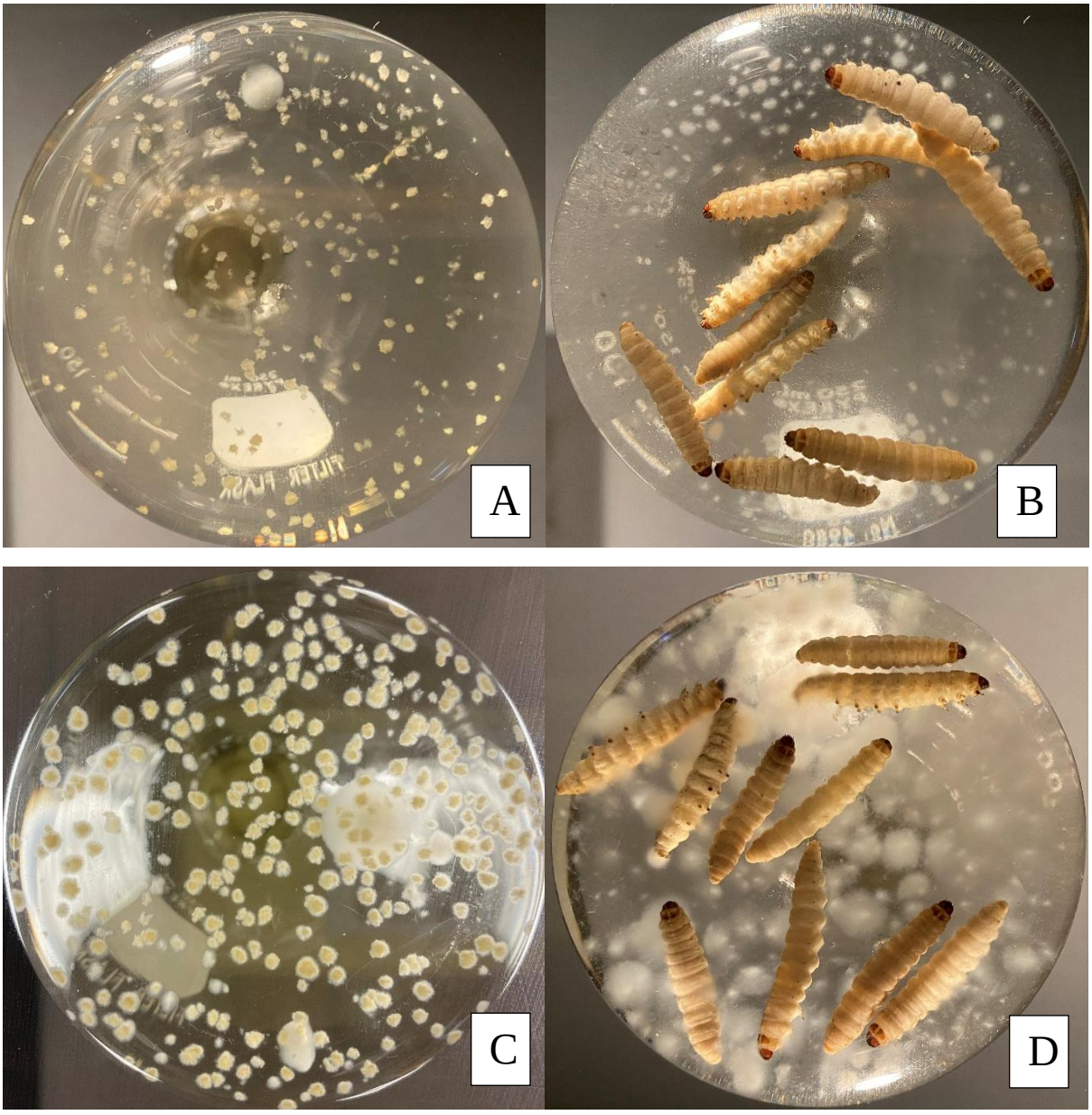


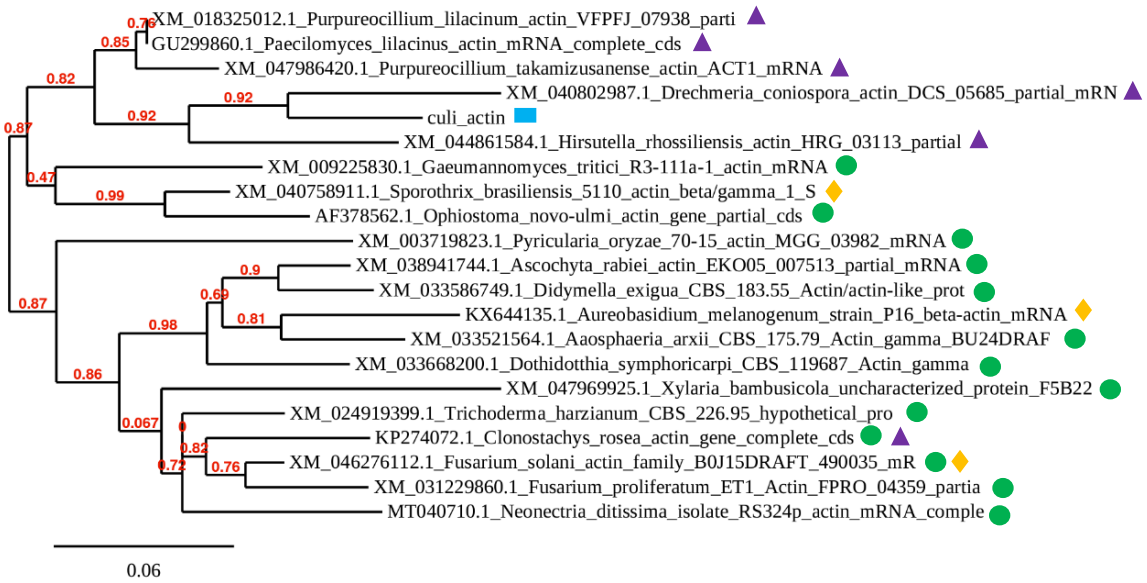
Figure 1: Growth of *C. clavisporus* within control and experimental liquid PYG flasks over the course of 14 days

(A) represents the control liquid flask inoculated by spore transfer after 7 days of growth. (B) represents the experimental liquid flask inoculated by spore transfer after 7 days of growth. (C) represents the growth of *C. clavisporus* after 14 days in the control PYG flask. (D) represents the growth of *C. clavisporus* after14 days in the experimental PYG flask.

CACACACCTTCTACAACGAGCTGCGTGTGGCCCCCGAGGAGCACCCCGTCCTGCTCACCGAGGC
CCCCATCAACCCCAAGTCAAACCGTGAGAAGATGACGCAGATCGTCTTTGAGACTTTCAACGCC
CCCGCCTTCTACGTCTCCATCCAGGCCGTCCTGTCGCTCTATGCCTCTGGCCGCACCACGGGTA
TTGTTCTTGACTCTGGTGACGGCGTCACCCACGTTGTCCCCATCTACGAGGGTTTCGCTCTTCC
CCACGCCATTGCTCGCGTTGACATGGCTGGCCGTGACCTCACCGACTACCTTATGAAGATTCTG
GCCGAGCGCGGGTACACCTTCTCCACCACTGCCGAGCGAGAAATCGTCCGAGACATCAAGGAGA
AGCTCTGCTACGTCGCCCTCGACTTTGAGCAGGAGATCCAGACGGCCGCCCAGAGCTCCAGTCT
GGAGAAGTCGTACGAGCTTCCCGACGGTCAGGTCATCACCATCGGCAACGAGCGATTCCGCGCC
CCCGAGGCTCTGTTCCAGCCCTCTGTCCTGGGTCTTGAGAGCGGCGGCATCCACGTCACGACGT
ACAACTCCATCATCAAGTGCGATGTGGATGTCCGAAAGGATCTTTACGGCAACATTGTCATG**GT**
GAGCATCTCGGCGAGGAAGGGAAATCAGCAACGGCTAGGCTAACAAGTGGCAGTCTGGTGGTAC
CACCATGTACCCCGGTCTCTCGGACCGCATGCAGAAGGAGATTACTTCGCTCGCCCCGTCGTCC
ATGAAGATCAAGATCATTGCTCCTCCCGAGCGAAAGTACTCGGTCTGGATCGGTGGTTCCATCC
TCGCCTCGCTGTCCACCTTCCAGCAGATGTGGATCTCGAAGCAGGAGTACGACGAGAGCGGCCC
CTCGATTGTCCACCGCAA

Figure 2: Sequence read obtained for the *C. clavisporus* actin gene fragment

Intron identified (in gray) and custom primers, CULIACT F (in green) and CULIACT R (in blue).



Legend

Type of Pathogen	Symbol
Plant Pathogen	●
Insect Pathogen	■
Nematode Pathogen	▲
Mammal Pathogen	♦

Figure 3: Phylogenetic tree analysis of *C. clavisporus* cDNA sequence

The *C. clavisporus* sequence, denoted by culi_actin, was found to be most closely related to the actin gene sequence from *D. coniospora*, with a 0.92 bootstrap support value. The overall clade in which *C. clavisporus* is within contains the actin gene sequence from *H. rhossiliensis*, and this relationship is supported with a 0.92 bootstrap value.

```
KGQ04272.1      ATGTCCTTTCACACGAGCAGCTCCAACATTCGTGTCGATGATGGCCACATCCTCCGCGCC
124324128       ATGTCTTTTCATGAATCTTCTTCCAACATTCGTGTCGATGATGGACACATTCTCCGGGCT
112985864       ATGTCTTTTCATGAATCTTCTTCCAACATTCGTGTCGATGATGGACACATTCTCCGGGCT
                ***** *****          ********************** ***** ***** **

KGQ04272.1      ACTCTCCGTAACGACAACGGCGAGGAGGTCGACAGCGAGCTCGACCTGAACAGCTGCCTT
124324128       TCGGTCCAAAACGGAAGCGGTGAGACGGTTGACAGCGAGGTTGATCTGAATCGGTGCCTC
112985864       TCGGTCCAAAACGGAAGCGGTGAGACGGTTGACAGCGAGGTTGATCTGAATCGGTGCCTC
                 *  ***  ****  * *** ***  *** ********* * ** *****  * *****

KGQ04272.1      AGCAACAGCAACGGGCGCTTTGCAT-GGGATGGCGCGAATTTCTCCGAGTCCGCCGAAAA
124324128       GGAAACGACCACGGGCGATTCGTATCCCGGTGGTGCAGACTTTTCCCACTCTGCGGAAGG
112985864       GGAAACGACCACGGGCGATTCGTAT-GGGGGGGTGAAGACTTTTCCCACTCTGCGGAAGG
                 * ***  * ******* ** * **   *  ** *   * ** *** * ** ** ***

KGQ04272.1      CATC-TCCTTCACTCTCGAGGGCGATGCTCAGGTCCCGATTCTTCGCGCCCGCCTGCGTG
124324128       CGTC-TCCTTTGCAATCGAGGG---TGCAGGCGTACCCATCTTGCGCGCCCGTCT-CGCG
112985864       CGTCTTCCTTTGCAATCGAGGG---TGCAGGCGTACCCATCTTGCGCGCCCGTCT-CGCG
                * ** *****  *  *******   ***    ** ** **  * ******** ** ** *

KGQ04272.1      CATC----TGACGAATCCTACGACGCCGATGTCAACCTTGCCGAGCGTATCGCCAACAAC
124324128       AATGGG--TGGCGGCCACGTTGATGCCGACGTCAACCTCAGCGAGCGCATTGGCAACAAC
112985864       AATGGGGATGGCGGCCACGTTGATGCCGACGCCAACCTCAGCGAGCGCATTGGCAACAAC
                 **     ** **    *   ** ***** * ******   ****** ** * *******

KGQ04272.1      AATGGCAGCCTGTCCTTTGGTAGGTTTCTCAATTTCCAGCTAGAGTATCGTGCTGGA---
124324128       AATGGCAGCTTG----------------------------------------------
112985864       AATGGCAGCTTGGT-----GTACGATTGTCGATTGCGGCAGAGCACACCACCCTACAATA
                ********* **

KGQ04272.1      --AGCGACCAGCTAACAGTTCATGCAGTTTGA
124324128       --------------------------------
112985864       ATAAAAGTCCAATAACAAAGTAGTGAATTCA-
```

A

B

Figure 4: Read alignments of CVNH domain

The reference sequence used for CVNH was from *B. bassiana* (Accession number KGQ04272.1) (A) MUSCLE alignment of the 2 reads, the better overall match (Sequence ID 124324128) highlighted in yellow, shown ID 124324128 matched the entirety of its sequence to the reference sequence. The alignment of sequence ID 112985864 match extended to the end of the reference sequence. (B) Sequence logo highlighting the conserved protein domain corresponding to CVNH.

```
KID83287.1      ATGGGCAA---GAAGGTCTTCTTCGACATTACCTGGGAGGG--TCCTGTCATGCAGGGCG
14681277        ATAGACGA---AAAAGTCTTCTTCGACCTCACGTGGGAGGGG-CCCCGTCCTTCAGAACG
150078838       ATGGCCAACTCCAGAGTCTTCTTCGACCTCACGTGGGAGGGGCCCCCGTCCTTCAGAACG
108201449       ATGGCCAACTCCAGAGTCTTCTTCGACCTCACGTGGGAGGG--CCCCGTCCTTCAGAACG
                ** * * *    *  ************ * ** ********   ** *** * ***  **

KID83287.1      GCAAGCCTACCAATACTGTCAAAGAGCAGTCTGGTCGCATCAGCTTCAACCTATTTGACG
14681277        GCAAGATCACGTCTACCGTCAAAGAGCAGCAGGGTCGCATCACCTTCAACCTCTTCGACA
150078838       GCAAGATCACGTCTACCGTCAAAGAGCAGCAGGGTCGCATCACCTTCAACCTCTTCGACA
108201449       GCAAGCCCCCCCCCCCCGTCAAAGAGCAGCAGGGTCGCATCACCTTCAACCTCTTCGACA
                *****    *     * ************   ********** ********* ** ***

KID83287.1      ACGTCGTTCCCAAGACCGCCGAGAACTTCCGCG-CTCTTTGCACTGGTGAAAAGGGTTTT
14681277        ACATCGTTCCCAAGACTG----------------------GCACCGGCGAGAAAGGGTTC
150078838       ACATCGTTCCCAAGACTGCCGCAAACTTCCGCGCCCCTCTGCACCGGCGAGAAAGGGTTC
108201449       ACATCGTTCCCAAGACTGCCGCAAACTTCCGCG-CCCTCTGCACCGGCGAGAAAGGGTTC
                ** ************* *                      **** ** ** ** ** **

KID83287.1      GGTTACTCTGGATCGTCTTTCCACCGCATCATCCCCG-ACTTTATGCTTCAGGGTGGTGA
14681277        GGCTACAAGGGCTCGTCCTTTCACCGCGTCATCCCGA-ACTTTATGCTCCAGGGCGGTGA
150078838       GGCTACAAGGGCTCGTCCTTTCACCGCGTCATCCCCGAACTTTATGCTCCAGGGCGGTGA
108201449       GGCTACAAGGGCTCGTCCTTTCACCGCGTCATCCCGA-ACTTTATGCTCCAGGGCGGTGA
                ** ***   ** ***** ** ****** *******   ********** ***** *****

KID83287.1      CTTCACACGTGGCAATGGCACCGGTGGCAAGTCCATCTATGGTGACAAA-TTCGCCGACG
14681277        TTTCACCCGTGGTAACGGCACTGGCGGTAAGTCCATATACGGCGAAAAG-TTTGCCGACG
150078838       TTTCACCCGTGGTAACGGCACTGGCGGTAAGTCCATATACGGCGAAAAG-TTTGCCGACG
108201449       TTTCACCCGTGGTAACGGCACTGGCGGTAAGTCCATATACGGCGAAAAAGTTTGCCGACG
                 ***** ***** ** ***** ** ** ******** ** ** ** **  ** *******

KID83287.1      AGAACTTCAAGCTGAATCACGACCGCCCTGGACTGTTGTCCATGGCCAACGCCGG-CCCC
14681277        AGAACTTCAAGGAGAAGCACACCAAGCCCGGCCTGCTGTCCATGGCCAACGCCGG-CCCC
150078838       AGAACTTCAAGGAGAAGCACACCAAGCCCGGCCTGCTGTCCATGGCCAACGCCGGCCCCC
108201449       AGAACTTCAAGGAGAAGCACACCAAGCCCGGCCTGCTGTCCATGGCCAACGCCGGCCCCC
                ***********  *** ***  *   ** ** *** ****************** ****

KID83287.1      AACACAAATGGCTCCCAGTTCTTCGTCACCACTGTCGTTACCTCATGGCTCAATGGACGG
14681277        AACACCAACGGCTCGCAGTTCTTCGTCACAACCGTGGTCACGGAGTGGCTAGACGGCAAG
150078838       AACACCAACGGCTCGCAGTTCTTCGTCACAACCGTGGTCACGGAGTGGCTAGACGGCAAG
108201449       AACACCAACGGCTCGCAGTTCTTCGTCACAACCGTGGTCACGGAGTGGCTAGACGGCAAG
                ***** ** ***** ************** ** ** ** **    *****  * **   *

KID83287.1      CACGTCGTCTTCGGCGAAGTCGCTGACGAGGAGTCCATGA-ACATTGTCAAGGCCCTCGA
14681277        CATGTCGTCTTTGGCGAGGTTGCCGAGAGGACTTCGATGC-ATGTCGTCAAGTCGGTCGA
150078838       CATGTCGTCTTTGGCGAGGTTGCCGACGAGGCTTCGATGCGATGTCGTCAAGTCGGTCGA
108201449       CATGTCGTCTTTGGCGAGGTTGCCGACGAGGCTTCGATGC-ATGTCGTCAAGTCGGTCGA
                ** ******** ***** ** ** **   *   ** ***  *  * ****** *  ****

KID83287.1      GGCCACTGGTTCCCAGAGCGGTACTGTCAAGTACAGCAAACGCCCCACCATCGTCAA-GT
14681277        GGCCACGGGCTCCGACTCTGGCAAAATCAAGTACAGCAAGCCCCCGACCATTGTTGA-GT
150078838       GGCCACGGGCTCCGACTCTGGCAAAATCAAGTACAGCAAGCCCCCGACCATTGTTGACGT
108201449       GGCCACGGGCTCCGACTCTGGCAAAATCAAGTACAGCAAGCCCCCGACCATTGTTGA-GT
                ****** ** *** *    ** *   ************* * *** ***** **  * **

KID83287.1      CTGGTGAGCTGTAA
14681277        CTTGTAAGCTGCCA
150078838       CTTGTAAGCTGCCA
108201449       CTTGTAAGCTGCCA
                ** ** *****  *
```

Figure 5: MUSCLE alignment of cyclophilin unigenes

Of the 3 reads identified as Cyclophilin, sequence ID 150078838 (highlighted in yellow) was found to be the best match to the reference sequence from *M. guizhouense* (Accession number KID83287.1).

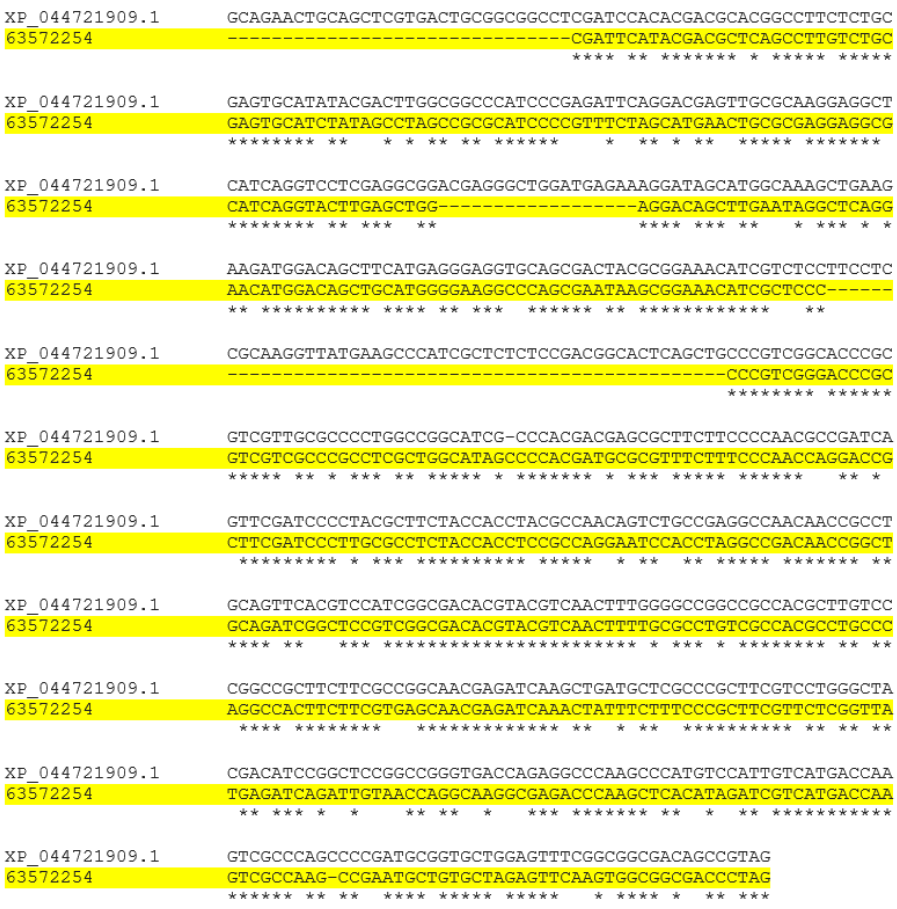


Figure 6: MUSCLE alignment of cytochrome P450 unigene

The read found from the C. clavisporus transcriptome analysis (Sequence ID 63572254) was compared to a reference sequence from *H rhossiliensis,* (Accession number XP_044721909.1). The read matched the reference sequence from nucleotide 931 -1,549.

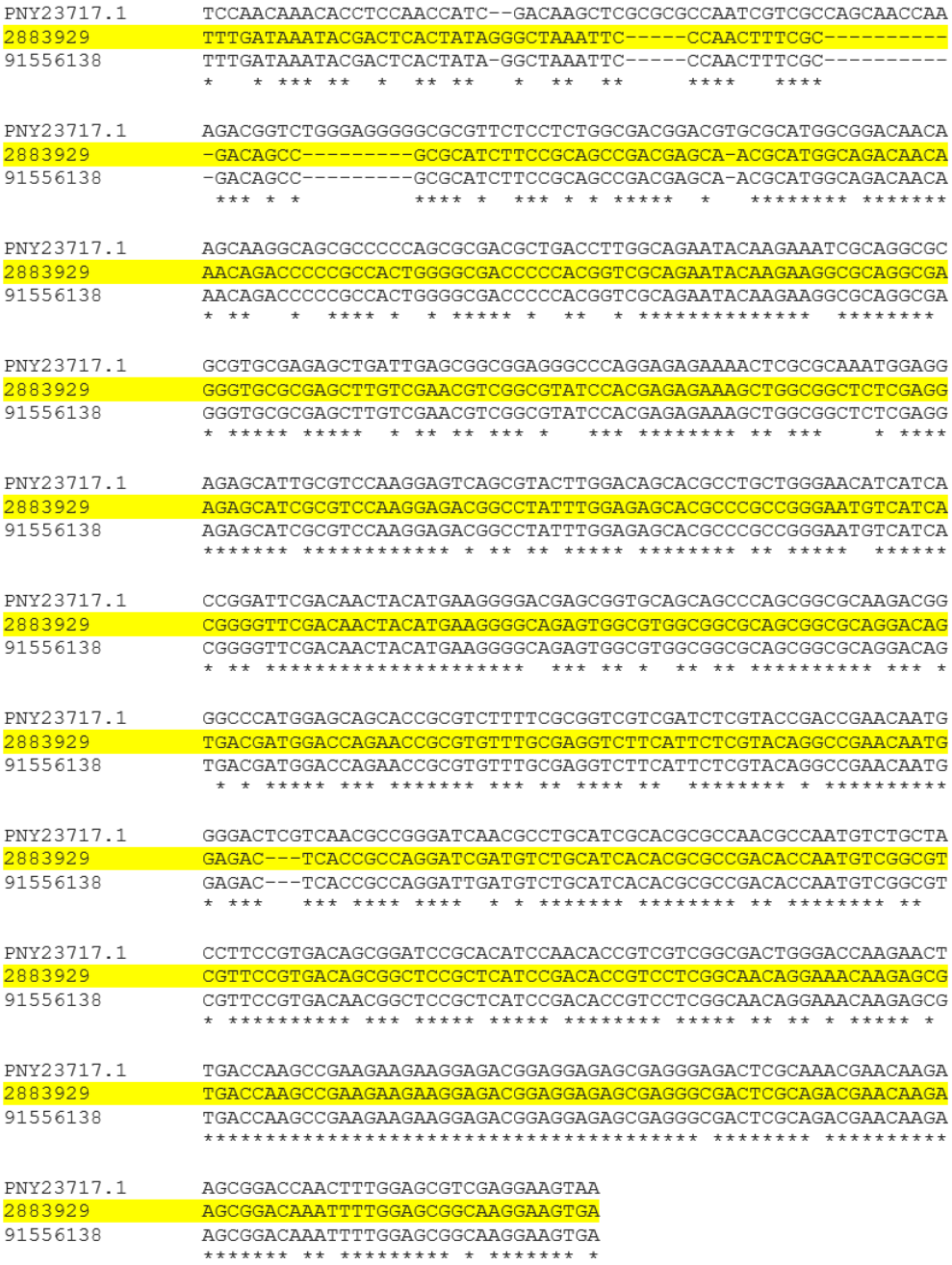


Figure 7: MUSCLE alignment of dipeptidyl-peptidase 5 reads

Of the 2 reads found homologous to Dipeptidyl-peptidase 5, Sequence ID 2883929 (highlighted in yellow) best matched the reference sequence from *T. capitatum* (Accession number PNY23717.1).

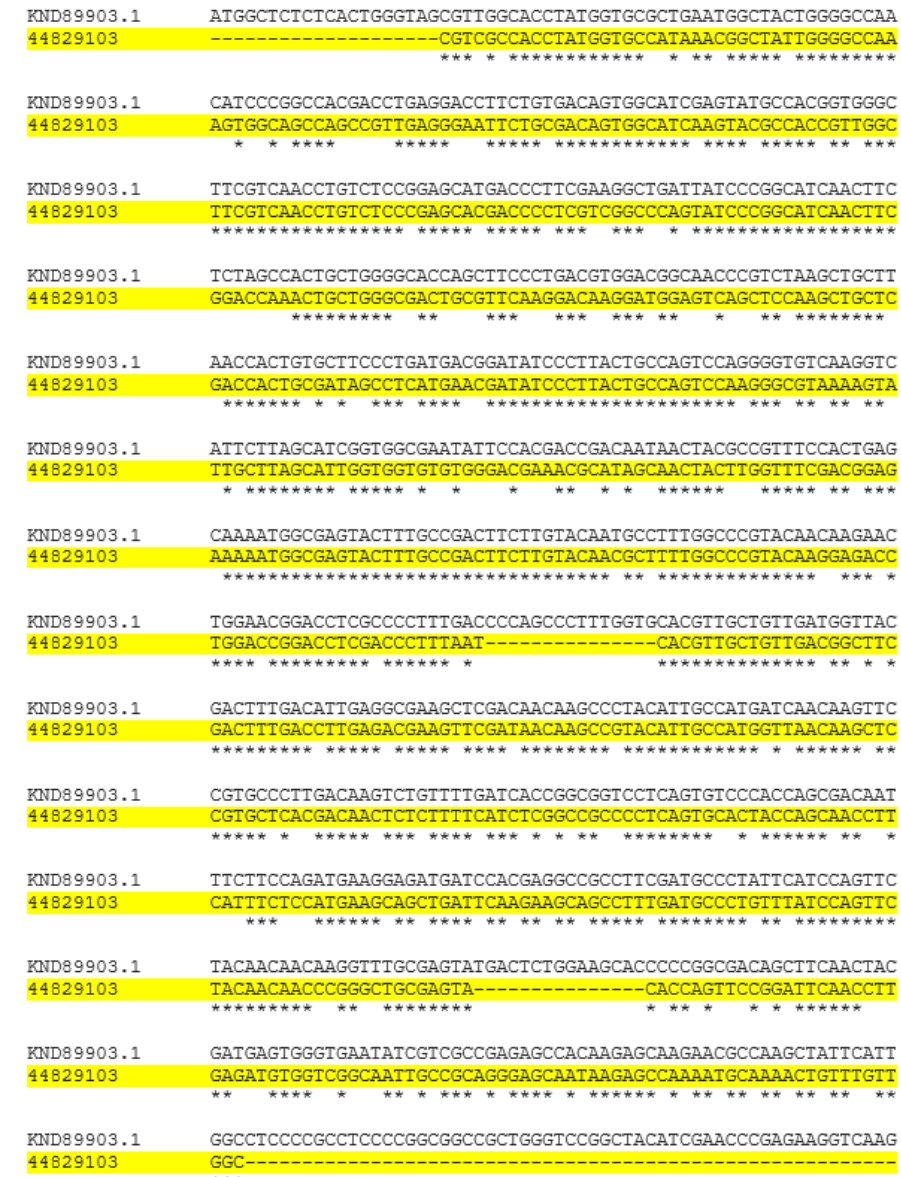


Figure 8: Alignment of reads for endochitinase A1

The reference sequence used for the endochitinase A1 sequence was from *T. ophioglossoides* (Accession number KND89903.1). (A) Sequence ID 44829103 (highlighted in yellow). This read corresponded with the reference sequence from nucleotides 1 – 783. (B) Sequence logo highlighting the conserved protein domain corresponding to endochitinase A1.

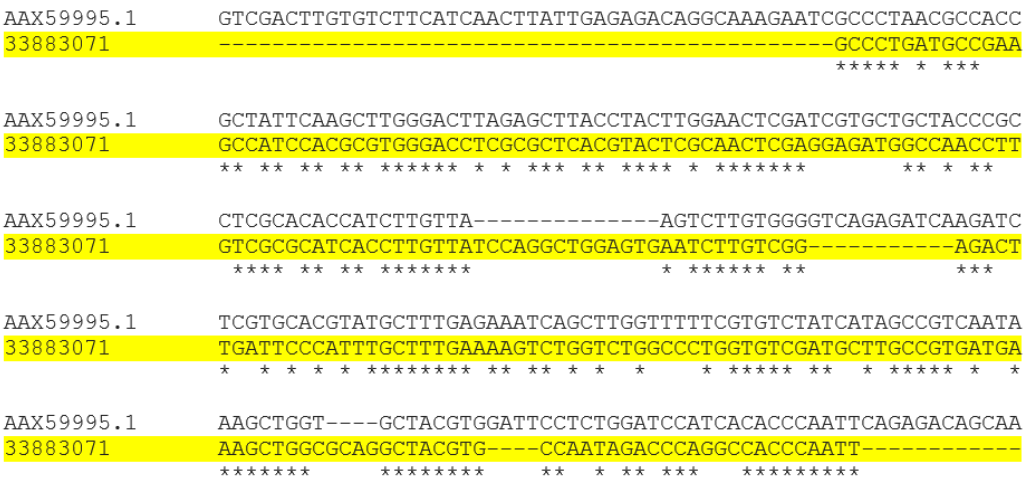
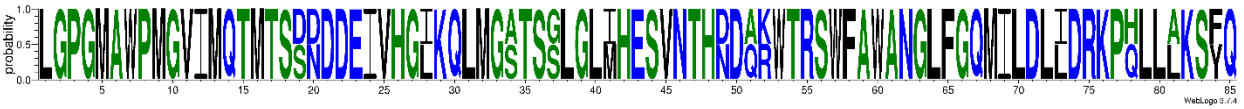


Figure 9: MUSCLE alignment of non ribosomal peptide synthetases (NRPS)

Sequence ID 33883071 (highlighted in yellow) was the best match to the reference sequence from *T. virens* (Accession number AAX59995.1). This read matched the reference sequence from nucleotides 2,327 – 2,568.

```
KID92891.1      GGACCAGTTCTTAATTCAACTGGTGGCCCGCACCTGGGTCCCGGAATGGCATGGCCGATG
264389          ------------------------GGGCCTCACCTTGGCCCTGGAATGGCTTGGCCCATG
79429654        -----------------------------------CTTGGCCCTGGAATGGCTTGGCCCATG
                                                   ** ** ** ******** ***** ***

KID92891.1      GGCGTTATCATGCAAACCATGACTTCGAGTGACGACGACGAGATTGTGCATGGTATTAAG
264389          GGCGTCATCATGCAGACAATGACGTCCGACAACGATGACGAGATTGTACACGGCCTCAAG
79429654        GGCGTCATC---CAGACAATGACGTCCGACAACGATGACGAGATTGTACACGGCCTCAAG
                ***** ***   ** ** ***** **     **** ********** ** **  * ***

KID92891.1      CAGCTAATGGGGGCGACAAGTGGGCTTGGGCTCATCCATGAGTCTGTTAATACGCATGAT
264389          CAGCTCATGGGTTCCACCAGCAGTCTTGGGCTCATGCACGAGTCGGTGAATACCCATAAT
79429654        CAGCTCATGGGTTCCACCAGCAGTCTTGGGCTCATGCACGAGTCGGTGAATACCCATAAT
                ***** *****  * ** **  * ********** ** ***** ** ***** *** **

KID92891.1      GATCAGAGGTGGACTCGGTCTTGGTTTGCCTGGGCGAACGGTCTCTTTGGGCAAATGATT
264389          GATGCAAAATGGACTAGATCCTGGTTTGCCTGGGCAAACGGTTTGTTCGGGCAAATGATT
79429654        GATGCAAAATGGACTA----CTGGTTTGCCTGGGCAAACGGTTTGTTCGGGCA-------
                ***   *  ******      ************** ****** * ** *****

KID92891.1      CTGGACCTTTTGGACAGAAAGCCACATCTACTCGCGAAAAGCTACCAGTAG
264389          CTGGATCTCATTGACAGAAAGCCCCAGCTTCTCTTGAAGAGTTTCCAGAAT
79429654        ----------------------------------------------------
```

A



B

Figure 10: Read alignments for six-hairpin glycosidase unigenes

The reference sequence used for six-hairpin glycosidase was from *M. guizhouense* (Accession number KID92891.1). (A) MUSCLE alignment for the reads compared to the reference sequence. The read Sequence ID 264389 (highlighted in yellow) was the best match to the reference sequence. These reads corresponded with the reference sequence from around nucleotide 1,285 – 1,551. (B) Sequence logo highlighting the conserved protein domain corresponding to six-hairpin glycosidase.

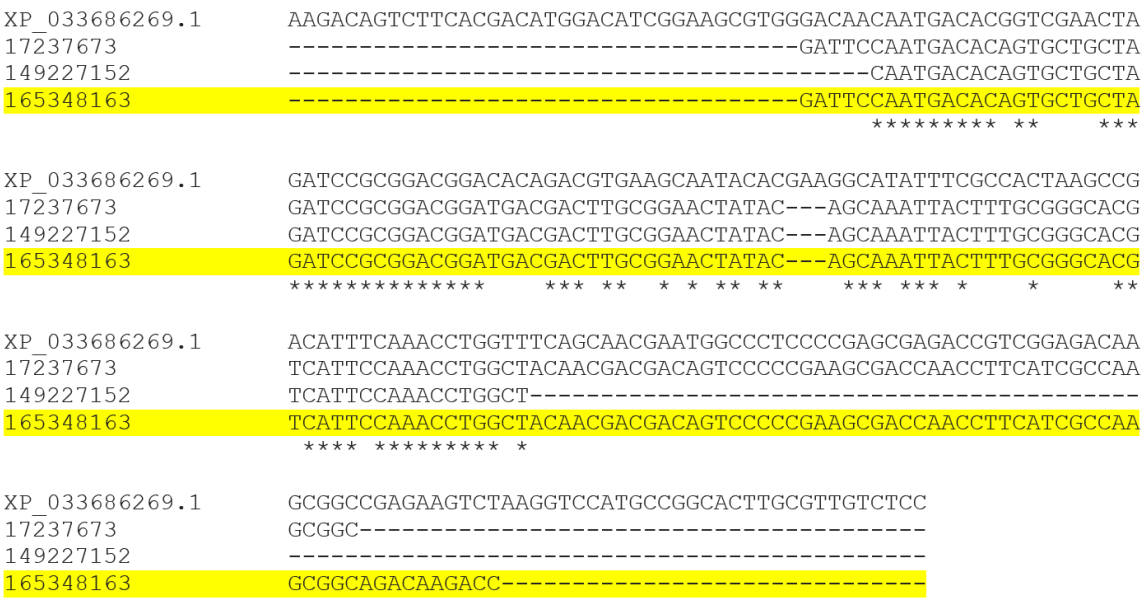
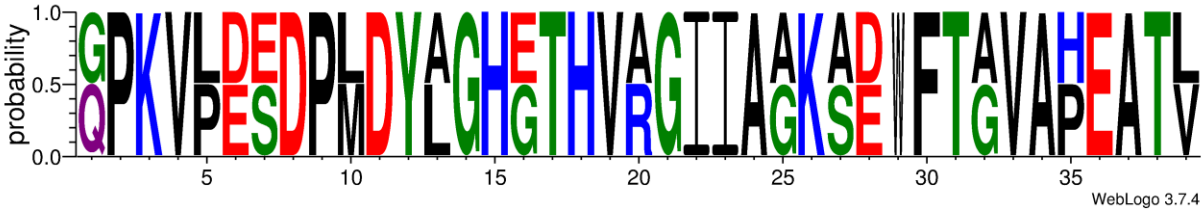


Figure 11: MUSCLE alignment of SpvB domain containing protein genes

Sequence ID 165348163 (highlighted in yellow) was the best match to the reference sequence from *T. pertusa* (accession number XP_033686269.1). This read matched the reference sequence from nucleotides 4,591 – 4,816.

```
KAH8892393.1        CAGTCCTGGCCGCAGGGCCCTAAAGTCCCTGACGAGGACCCTCTGGACTACCTTGGTCAC
88343209            -----------GCAG---CCTAAAGTTCTTGAAAGTGACCCAATGGACTACGCGGGCCAC
123470957           -----------GCAG---CCTAAAGTTCTTGAAAGTGACCCAATGGACTACGCGGGCCAC
163512962           -----------GCAG---CCTAAAGTTCTTGAAAGTGACCCAATGGACTACGCGGGCCAC
                               ****   ******** * ***    *****  ********   ** ***

KAH8892393.1        GGTACGCACGTGGCGGGCATCATTGCTGGCAAGTCGGACTGGTTCACCGGAGTGGCCCCT
88343209            GAAACTCATGTTCGAGGCATAATTGCTGCTAAGGCAGAGTAGTTCA--------------
123470957           GAAACTCATGTTCGAGGCATAATTGCTGCTAAGGCAGAGTAGTTCA--------------
163512962           GAAACTCATGTTCGAGGCATAATTGCTGCTAAGGCAGAGTAGTTCA--------------
                    *  ** ** **    ***** *******  *** * ** * *****
```

A



B

Figure 12: Read alignments of subtilisin-like proteases unigenes

The reference sequence was from *Thozetella sp.* (Accession number KAH8892393.1). (A) The sequence ID 88343209 (highlighted in yellow) was the best match to the reference sequence. These reads are homologous to the reference sequence from nucleotides 552 – 646. (B) Sequence logo highlighting the conserved subtilisin-like protease domain.

Figure 13: MUSCLE alignment of the zeta toxin unigenes

Of the 2 best matches were compared to the reference sequence of *F. austroafricanum* (Accession number KAF4448150.1). The better overall match was sequence ID 3736654, but all of the reads match the entirety of the reference sequence.

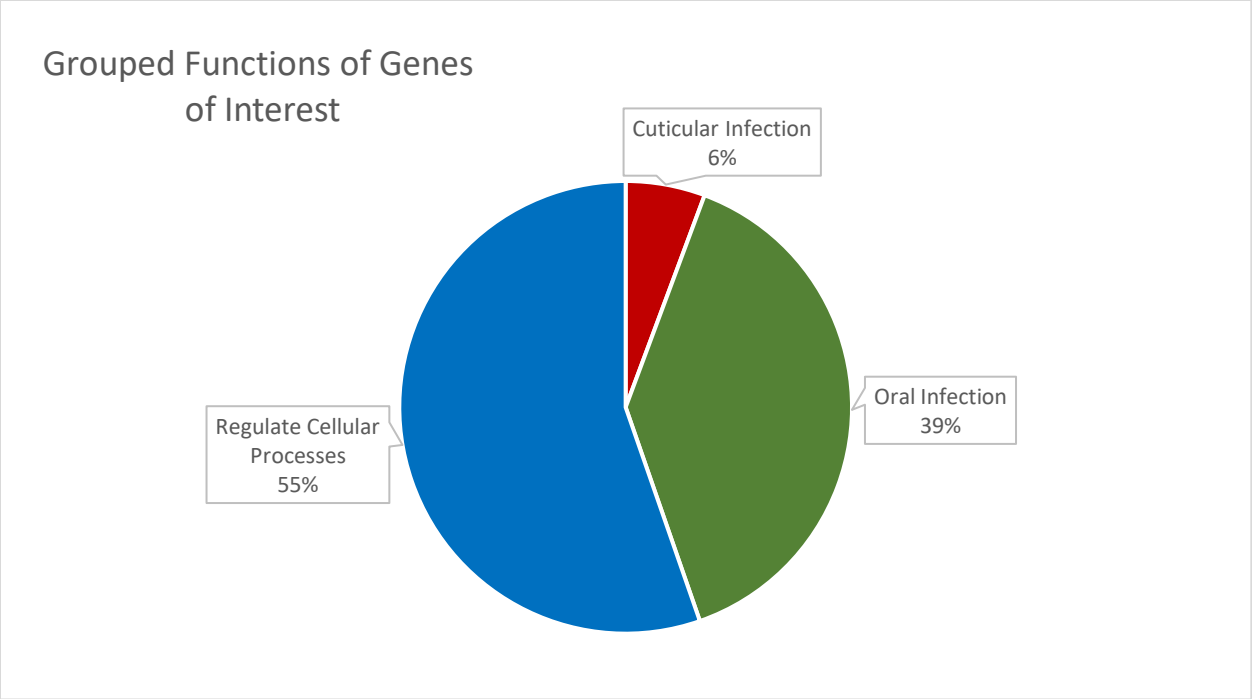
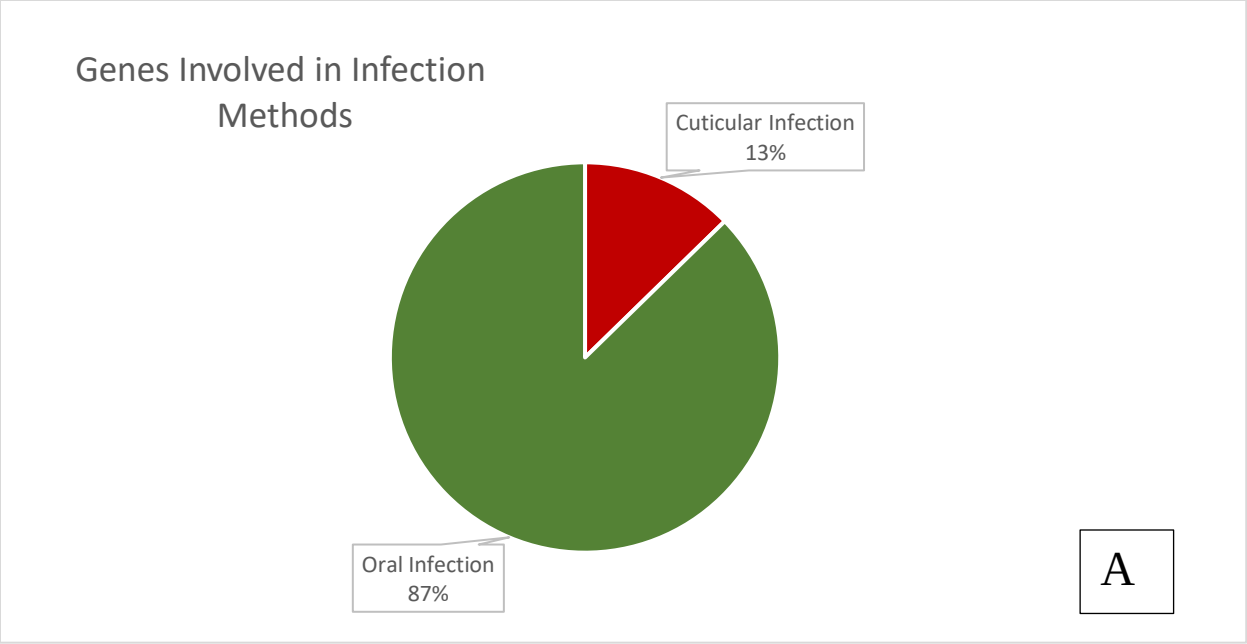


Figure 14: Hypothetical role in pathogenicity for the genes of interest

Of the 141 unigenes identified as having a special interest in the transcriptome analysis, 55% of the unigenes regulate punitive cellular processes associated with virulence, 39% of the unigenes are hypothesized to be involved in oral infection methods used by fungi, and 6% of the unigenes may be involved in cuticular infection methods utilized by fungi.



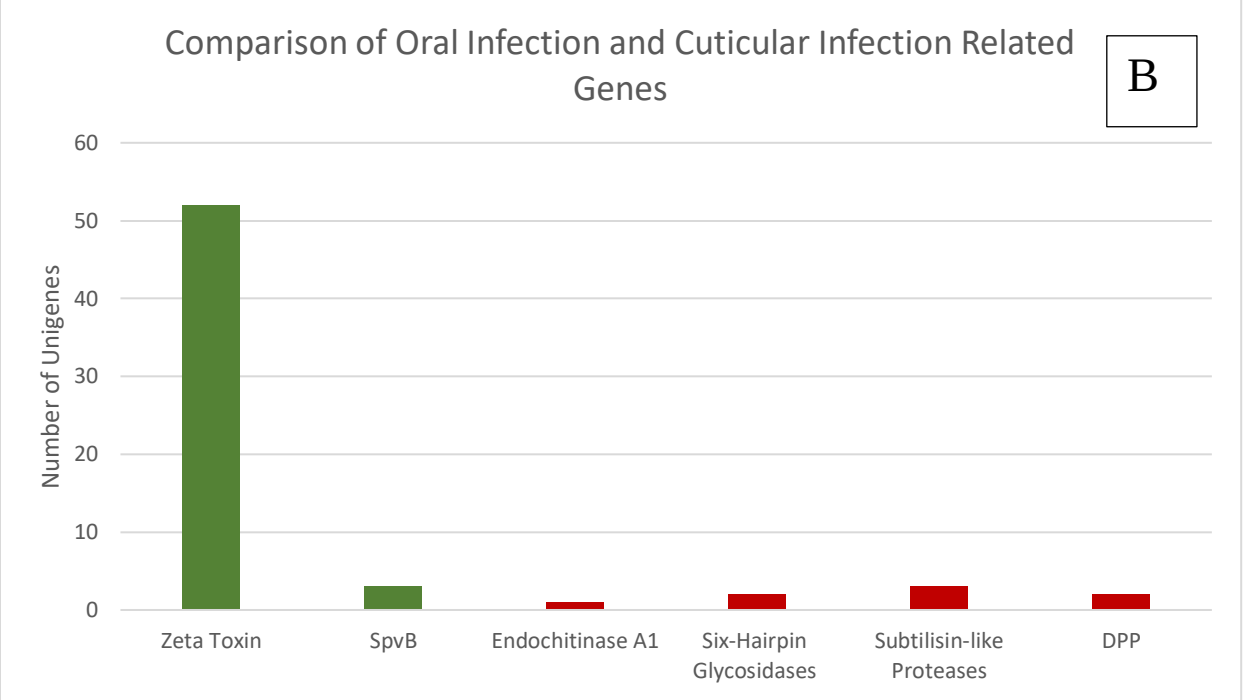


Figure 15: Comparison of genes utilized during cuticular infection and oral infection methods

(A) Graphical representation of gene distribution according to their punitive function and roles in pathogenesis. Overall, 87% of the unigenes are utilized during oral infection compared to 13% of the unigenes involved with cuticular infection methods. (B) Graphical representation of the number of unigenes associated with each gene of interest identified. Genes related to oral infection are shown in green, with an overall total of 55 unigenes and those related to cuticular infection are shown in red, with a total of 8 unigenes.

Table 1: Top 20 Unigene Identification from the *C. clavisporus* Transcriptome Analysis

Gene Name	Number of Unigenes
Thioredoxin-like Protein	230
Ribosomal Protein L22e	146
Glutathione S-transferase/chloride Channel	129
Cytochrome C	117
Actin Lateral Binding Protein	113
60S Ribosomal Protein L30	95
40S Ribosomal Protein S15	95
HSP20	94
Histone H2B	90
60S Ribosomal Protein L12	86
Peptidyl-prolyl Cis-trans Isomerase	85
40S Ribosomal Protein S23	74
Serine/threonine Protein Kinase	73
60S Ribosomal Protein L32	73
60S Ribosomal Protein L25	72
30 kDa Heat Shock Protein	65
Mitochondrial Import Receptor Subunit or Translocase Domain-containing Protein	61
CyanoVirin-N Homology domain (CVNH)	59
Hsp20/Alpha Crystallin Family Protein	56
Zeta Toxin	52

Table 2: Genes of Interest During *C. clavisporus* Infection Process

Gene Name	Function	Number of Unigenes	Best Unigene Match Sequence ID	E value	Percent Identity	Query Cover
CyanoVirin-N Homology domain (CVNH)	Involved in cell to cell recognition of glycoproteins and surface-bound sugars	59	124324128	1E-20	72.92%	90%
Cyclophilin (CYP)	Ubiquitous proteins that regulate cellular processes and transcription	3	150078838	2E-27	45.39%	94%
Cytochrome P450 (CYP)	Ubiquitous enzymes involved in the biosynthesis of secondary metabolites	1	63572254	1E-51	73.40%	92%
Dipeptidyl-peptidase 5 (DPP)	Enzymes that cleave dipeptides from proteins or peptides at the N-terminus	2	2883929	1E-83	79.12%	90%
Endochitinase A1	Enzymes for degrading a chitin based cell wall	1	44829103	2E-114	66.93%	99%
Non Ribosomal Peptide Synthetases (NRPS)	Catalyze biosynthesis of small bioactive peptides via a thiotemplate mechanism independent of ribosomal	15	33883071	1E-25	49.04%	90%

	protein synthesis mechanism					
Six-hairpin Glycosidase	Hydrolyses glycosidic bonds from the surface of the insect cuticle	2	264389	1E-46	85.23%	98%
Salmonella Plasmid Virulence (SpvB) Effector Protein	ADP-ribosylates and depolymerizes actin filaments inside the infected cells	3	165348163	1E-06	52.83%	100%
Subtilisin-like Proteases	Cuticle degrading proteases	3	88343209	2E-05	63.16%	57%
Zeta Toxin	Bacterial-like toxin suggesting the utilization of the bacterial toxin-antitoxin system to control cell stasis and death	52	3736654	1E-07	67.44%	69%

Table 3: CVNH Domain Read IDs

2295400	15073603	36833553	38470332	41288402	47712873	77006774	90768312
104858240	121964046	146931997	148573007	174459221	4786481	6228436	12716697
20776881	26085741	55968933	59246551	84149118	89260563	99223288	99616196
102433398	109250537	116130395	124324128	137625757	139461297	171902671	173212181
5767319	13765374	37291556	60229161	1706824	5900200	6293792	34210825
38144054	43452407	60032837	70714911	79954949	111739476	112985864	120521017
120783644	126615665	126748729	152633654	162595642	164824679	165478727	7340440
25560571	48235878	54395903					

Table 4: NRPS Read IDs

23725094	26674914	41224798	75038872	90506775	91424427	112462058	124193314
138150923	140050709	143458430	146211733	6555154	29559446	33883071	

Table 5: Zeta Toxin Read IDs

2294989	3736654	8522053	9702046	11666106	12912351	26149732	32179372
34472567	36242782	41091090	41747804	42533422	52300372	57737217	61801175
65339514	68880865	69077145	69534067	72483447	72877344	79628487	80415470
80937770	83429112	88868421	90899324	96666138	111609503	112658200	112920406
118490514	125108379	131794397	134217822	141036027	146080105	147456124	165609896
174784982	176163134	179308178	180355321	180355552	22676594	24709847	32704549
47710631	55903009	56820540	57674361				

Table 6: Genes of Interest Involved with Cuticular Infection

Gene Name	Unigenes Identified
Endochitinase A1	1
Six-Hairpin Glycosidase	2
Subtilisin-like Proteases	3
DPP	3

Table 7: Genes of Interest Involved in Oral Infection

Gene Name	Unigenes Identified
Zeta Toxin	52
SpvB	3

Table 8: Genes of Interest Involved with Cellular Processes Affecting Virulence

Gene Name	Unigenes Identified
CVNH	59
Cyclophilins	3
Cytochrome P450	1
NRPS	15