Proteomics Profiling and Functional Characterization of *Caenorhabditis elegans* Excreted/Secreted Proteins

Thesis by

Wen Chen

In Partial Fulfillment of the Requirements for

the Degree of

Doctor of Philosophy in Biology

CALIFORNIA INSTITUTE OF TECHNOLOGY

Pasadena, California

2019

(Defended May 13th, 2019)

Wen Chen
ORCID: 0000-0001-8056-5711

ACKNOWLEDGEMENTS

I can't believe I am finally facing this special moment, the finish line of a long marathon and the most important transition of my life. This is not the end of my scientific career but is definitely a new beginning of being on longer a student. I have enjoyed studying at Caltech all the time because of the intelligent people here. I am also extremely lucky to get to know some of the most amazing people in the world, who greatly supported me through the graduate school.

First and foremost, I would thank my advisor Dr. Paul Sternberg. His insights, critical way of thinking and passion toward science have guided me through the project. His patience and encouragement helped me to finally push myself to the finishing line. I know Paul is the best mentor that one can have. And I think I still need to learn from him for a lifetime. I also would like to thank Dr. Animesh Ray. I really benefited a lot from his intelligence and advices on several side projects. Animesh is also very nice and generous to give me suggestions on all levels of science, from the big picture to the detailed presentation skills.

I owe thanks to my committee members: Professor Douglas Rees, Sarkis Mazmanian, and Dianne Newman. I really appreciate all the efforts they made to save time from busy schedule and provide insightful suggestions and critical comments on my project. Without their care and understanding, I would even have failed to find a thesis defense date at the last minute.

I want to thank the great colleagues in the Sternberg lab. Pei-Yin Shih and Cynthia Chai tolerated all types of stupid questions from me with great patience. They are my

number one life savers in lab. James Lee helped me a lot on the expression profiles clustering and comparative proteomics. I want to thank David Angeles for the *C. elegans* food preference project we worked on together and his model building is the key feature of it. I want to thank the experienced postdocs, Chun-Hao Chen, Mengyi Cao, Han Wang, and Hillel Schwartz. They are always there to consult with any type of questions, from knowledge of nematodes to advices on careers. I would also thank following labmates, Sandy Wang, Porfirio Quintero, Ravi Nath, Jonathan Liu, Katie Brugman, Ping Hsueh, Daniel Leighton, Sarah Cohen, and Liz Holman, for helpful discussions and intelligent input. I need thank Heenam Park for help in building CRISPR lines, Mandy Tan for preparing plates, Barbara Perry for the strain requests, and the WormBase people. I need thank Sarah Torres for putting in the orders and for help with a lot of administrative stuff. They are at the same time great friends who make my life in the lab so pleasant. I want to thank Pei for being a great workout buddy and for introducing me to the world of keeping fit. I need to thank Cynthia for her love and care in organizing so many events in lab. I can't forget how touched I was when she planned my birthday party in lab. Mengyi and Chun-Hao are so nice to listen to my troubles and are so understanding and supportive. I want to thank Katie for all the movies she took me to. I also want to thank my labmates for the food explorations as well.

I owe thanks to the Proteome Exploration Laboratory people. I want to thank Dr. Annie Moradian, Dr. Michael Sweredoski, Dr Sonja Hess, and Roxana Eggleston-Rangel for their help in proteomic experiments and the data analysis. I also need thank Fan Gao for bioinformatic consulting.

I have also met a lot of great friends outside the lab. First, I want to give special

thanks to Lu Pan. Meeting you is the best thing that happened in my life. I know I can find you to calm me down at any time with any kind of trouble. I feel so safe and open in front of you because I know you will not laugh at me no matter how stupid or how wrong I am. I will always cherish the memories of the food adventures, coffee breaks, and trips that we did together. I also want to thank Cheng Li, Tianhao Le, and Ellen Han for pleasant chatting and doing a lot of fun stuff together. I need thank Jin Yang and Lu Xu for organizing the hiking events. I owe thanks to my great former roommates, Grace Chow, Jennifer Maqueda, and Natalia Luna. I am very grateful to Xun Wang and Can Li for helping me with experiments and for food adventures. I want to thank Sally Zhang and Yingxin Deng, for the knitting and coffee time we spent together. There are also many other friends I owe thanks to: Zhihong Tan, Zihan Liu, Yuting Liu, Yutao Qi, Qi Hu, Qiong Zhang, Xu Yi, and Jinyi Fan.

I owe a lot of thanks to Jun Chen, my boyfriend and best roommate. Your self-discipline and hard-working nature always encourage me to keep up. I am very grateful to have you with me along this journey.

Finally, I would like to thank my family for their love and support. At this moment, I miss my grandparents and parents so much. I know I made my parents worry a lot when I was in a low mood, especially considering I am so far away from home. But now I wish my parents could join me in celebrating the biggest achievement and that they will be proud of me.

ABSTRACT

Excretory-secretory products (ESPs) are first characterized and defined in parasitic nematode proteomics studies as the combination of various biomolecules that are continuously excreted or secreted into the environment throughout the whole life cycle. ESPs are particularly interesting to many scientists as anti-parasitic vaccine candidates and as promising drug targets since large portions of ESPs are active enzymes that potentially function directly at the parasite-host or worm-environment interfaces. However, majority of the parasites lack whole genome sequence knowledge and genome-editing tools. Thus, the number of ESPs identified is limited and many functions of ES proteins cannot be elucidated. Therefore, we use the most studied nematode, *Caenorhabditis elegans*, as the model to characterize the composition of excreted/secreted proteins with the help of nano-liquid chromatography coupled with tandem mass spectrometry (nanoLC-MS/MS). We characterized more than 509 excreted/secreted proteins with mix-staged worms, including many metalloproteases, cysteine proteases, and lysozymes. Proteases and proteases inhibitors are a major group in *C. elegans* ESPs. We performed stable isotope dimethyl labeling quantitative proteomics and compared *C. elegans* ESPs on different bacteria diets. Lysozymes are not only enriched in *C. elegans* ESPs but are also up-regulated in response to pathogen and bacteria.

Comparative studies of expression profiles of developmental life stages and pathogen infections elucidate the dynamics in regulating ESP components. We successfully identified stage-specific ESP groups associated with L1, L3, adult, L2 dauer, and post-dauer. We demonstrated that proteases activities are down regulated by increased protease inhibitor expressions, while during dauer exit proteases expressions are increased. The

comparison between dauer excretome/secretome and RNA-seq dauer expression profiles revealed 91 ESP encoding genes that are highly expressed in dauers. We performed dauer formation assay to these dauer-associated gene mutants. The great prediction rate confirmed that our comparative method is the simplest way to quickly pick out candidates for functional assays. Similarly, we employed this comparative method to pathogen-induced transcriptomes. We reported a group of genes that are associated with *Serratia marcescens* infection and a group of bacterial pathogens responding genes. We confirmed the roles of *C. elegans* ESPs in immuoregulation by infection assays with various pathogens. Lysosomes and cysteine protease inhibitor are among the most important genes in innate immune response pathway of *C. elegans* defending pathogen infection.

The recent discovery of a *C. elegans* sibling species, *Caenorhabditis inopinata,* allows the deeply comparative study for evolutional interpretation. The excretome/secretome of *C. inopinata* has not been characterized. We took advantage of the sensitive and high-throughput technique of nanoscale liquid chromatography coupled to tandem mass spectrometry (nano LC-MS/MS) to directly characterize the protein components of *C. inopinata* excretome/secretome. Functional annotations reveal several protein families, including C-type lectins, Cathepsin Z, Cathepsin B family, transthyretin, and saposin-like families, suggesting ESPs play critical roles in regulating innate immune response. We compared *C. inopinata* excretome/secretome with *C. elegans*. The structures are highly conserved across species, suggesting the sibling species share common mechanism to respond to environmental stimuli.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS AND/OR TABLES

Figure *Page*

Table *Page*

Appendix Table *Page*

Chapter 1

INTRODUCTION

1.1 Thesis overview

Excretory/Secretory (ES) Products and the research history in parasites

In the nematode studies, excretory/secretory products describe the combination of a wide range of biomolecule that are continuously excreted or secreted out, including surface antigens, non-immunogenic glycolipids, glycans, bioactive lipids, heat shock proteins, detoxifying enzymes, surface proteases, and other metabolic products [1]–[4].

Excretory/secretory products in parasites have been studied for several decades now [5]. Parasitic nematode chronic infection to host requires successful invasion of host barrier, migrating through the tissue to a suitable niche and bypassing the host immune system to survive and reproduce [6]–[8]. Thus, the mechanism of how parasites modify host immune system has been studied intensively. Excretory/secretory products, as the active biomolecules functioning directly at the interaction surface of host and parasites, are at the center of the study. The investigation of excretory/secretory products will benefit the therapy of parasitic nematodes in many aspects. The antiallergic and anti-inflammatory effectors in excretory/secretory products would be great anti-parasite vaccine candidates. A lot of excretory/secretory products are enzymatically active proteins, providing the possibility of intervention with small drug molecules [9]. The featuring molecules in excretory/secretory products can also serve as biomarkers for early diagnosis of parasite infection [10].

The majority of the excretory/secretory products are proteins, which will be referred to as excretory/secretory proteins (ESPs). Most works that have been done with excretory/secretory products so far focus on the protein components only. The

excretory/secretory proteins have been characterized in the following parasites: *Ascaris suum[11], [12], Brugia malayi[6], [11], [13] , Dirofilaria immitis[10], Meloidogyne incognita[9], Bursaphelenchus xylophilus[14], Heligomosomoides polygyrus[15], Ostertagia ostertagi[16], Ancylostoma caninum[17], [18], Strongyloides ratti*[19]–[21], *Teladorsagia circumrcincta[22][23], Trichinella pseudospiralis[24], Trichinella spiralis[24], Haemonchus contortus[25], [26]* and *Nippostrongylus brassilliensis[27]*. However, in many of these nematodes, the numbers of identified ESPs are limited by several reasons. First, the main methods to characterize excretory/secretory proteins are functional assays like enzyme activity assay, prediction of secretion by bioinformatics, 2D SDS-PAGE, prediction transcriptome and proteomics [20], [26]–[28]. The numbers of identified excretory/secretory proteins range from 2 to 1500 and are limited by the characterizing method. Among these, proteomic is the most sensitive method to detect excretory/secretory proteins directly. High throughput proteomics also allow large scale protein characterization [29]. Second, genomes of many parasites have not yet been fully sequenced. This hinders the annotation of gene models and therefore limits the number of identified excretory/secretory proteins. In fact, proteins missing annotations or hypothetical proteins are very common in parasitic excretomes/secretomes.

ESPs identified in parasitic nematodes and their immune regulation functions

ESPs are active directly at parasite/host interface and regulate host immunity at every step from initial recognition to downstream effectors [30]. The successful invasion, migration, and reproduction of parasites requires continuous suppression of host immune system by ESPs [30]–[32]. All these facts suggest that ESPs are potential anti-parasitic

vaccine targets to solve the emerging anthelmintic resistance [1]. Several antigens from ESPs are under clinical trials [33].

Several groups of excreted/secreted proteins have been characterized and their immune regulation roles have been studies intensively and their putative or partially understood immune regulation mechanisms are listed [11], [14], [15], [23].

Proteases, including cysteine-, aspartyl-, serine-, and metalloproteinases, may be involved in degradation of host tissues or bacteria food, helping host invasion and/or nutrition uptake [11].

Proteases inhibitors, may function by inhibiting host proteolytic enzyme activity. Known protease inhibitors in parasitic ESPs include cystatins (cysteine protease inhibitor) and serpins (serine protease inhibitor). Proteases inhibitors could also regulate host immune system by blocking antigen processing. In *B. malayi*, CPI-2 inhibits asparaginyl endopeptidase (AEP) and therefore inhibits antigen processing by human B cells [34]. Proteases inhibitors could also regulate cytokine levels and T cell proliferation [11].

Antioxidants could detoxify reactive oxygen species produced by host phagocytes and maintain homeostasis of nematodes. Observed antioxidants in ESPs include thioredoxin peroxidases, peroxiredoxins, catalases, superoxide dismutases (SODs), and glutathione peroxidase [30].

C-type lectins and galectins have putative roles in many immune responses including antigen uptake and presentation, cell adhesion, and T cell polarization [30], [35] [36].

Cytokine homologues are another group of conserved proteins in excretome/secretome. One most studied cytokine homologoue is *B. malayi* macrophage migration inhibitory facter (MIF). Bm-MIF mimics mammalian cytokines and could interact with human MIF

receptor and thus suppress pro-inflammatory activation in host [31], [37], [38], [39].

Although tranthyretin-like proteins are common in nematodes secretome, their functional role in regulating immune system remains elusive [11].

Heat shock proteins are another common group in secretome. One hypothesis is that heat shock proteins are released in the stress response pathway. There are also reports showing parasite secreted HSPs are antigens [11], [40].

Lipid binding proteins in ESPs are hypothesized to be involved in membrane trafficking , signaling pathways, and interaction with innate pattern-recognition receptors, including nematode polyprotein allergens (NPA) and fatty acid and retinol-binding (FAR) proteins [30]. FAR protein could bind to small lipids like vitamin A and interfere functions like macrophages [41].

Some ESPs could also be post-translationally modified, forming antigenic glycoconjugates. One example is a leucine aminopeptidse from *Acanthocheilonema viteae*, ES-62, heavily conjugating with an unusual phosphorylcholine moiety. ES-62 was shown to induce Th2 anti-inflammmatory response and suppress Th1 immune response, altering cytokine levels and T and B cell proliferation [30], [42].

Caenorhabditis elegans **as model to study excreted/secreted proteins**

Caenorhabditis elegans has been intensively studied as a model organism since 1963, leading to numerous breakthroughs in elucidating pathways and mechanisms like apoptosis, aging and metabolism [43]–[45]. *C. elegans* has a short life cycle of 3 days. It is a small organism with only 959 somatic cells in hermaphrodites and 1033 in males. The cell lineage and neuron connectome were fully mapped. The whole genome of *C. elegans* has

been fully sequenced for two decades now and benefits from laboratory work on genes and function characterizations. Multiple genetic manipulation tools are available including RNA interference, CRISPR/Cas9 gene editing, and overexpression. Numerous functional and behavioral assays have already been established, making *C. elegans* a great model to study many phenomena and diseases including Alzheimer's disease, autism, cancer, and aging [46]–[53].

However, the study of excreted/secreted proteins in *C. elegans* lags far behind compared with the works in parasitic nematodes. Researches underestimated the broad presence and complicated composition of excreted/secreted proteins in free-living nematodes. Thus, in Chapter 2, we reported proteomic characterization of *C. elegans* excretome/secretome and for the first time profiled the protein components systematically. We identified 509 proteins, and functions for the majority of the ESPs remain obscure. This work could serve as a great platform for annotating parasitic nematodes excreted/secreted proteins with no fully sequenced genomes or with few function annotations. Benefiting from the mutant's availabilities and well-established functional characterizing techniques and assays, studying functions of *C. elegans* ESPs would provide better knowledge and prediction of their parasite homologues.

C. elegans, as a free-living nematode, undergoes four larval stages after hatch and reach reproductive stage in about 3 days under favorable condition [54]. Under unfavorable conditions, *C. elegans* would enter an alternative development cycle, forming non-aging dauer larvae that can survive for months [55]. Dauer has distinct behaviors like nictation to attach to other animals and disperse [56]. This phoresy phenomena is hypothesized as the pre-adapting step toward parasitism [7], [57], [58]. Indeed, dauer share a lot of common

features with parasite infective larvae, including slim bodies, a constricted esophagus, and a closed mouth [57]. Dafachronic acid (DA) is required in both dauer and parasites for dauer/infective larvae formation. This shared mechanism strongly supports that parasitism may be acquired from dauer to infective larvae evolution [59]–[61]. ESPs functions are the key to the parasite's infections. Thus, studying the evolution of ESPs between free-living nematodes and parasites could provide clues for how parasitism evolved. In Chapter 4, we report the excretome/secretome of *C. elegans* sibling species, *C. inopinata*, to contribute to the collection of nematode excretomes/secretomes [62].

C. elegans could also be used as a great model to study the developmental regulations of ESPs composition [50], [63]. There are no known reports on whether specific groups of ESPs are up or down regulated within a certain developmental stage. Whether they are actively involved in the dauer decision making is unknown. In Chapter 3, we reported the first attempts to identify stage-specific ESPs in *C. elegans*.

C. elegans encounters a wide variety of bacteria and fungi in nature [64]. Among these, several bacteria and fungi have been proved to be pathogenic to *C. elegans*, causing shortened life span or non-lethal disease [65]–[68]. *C. elegans* ESPs are the major players in immunoregulation. RNAi experiments targeting several ESPs including *lys-1, lys-8, cpi-1, cpi-2, spp-5,* and *asp-3* showed increased pathogen susceptibility or altered life span, supporting the hypothesis that one major role of *C. elegans* ESPs is to defend pathogen infection [50], [69]–[71]. However, the mechanisms and corresponding ESPs of *C. elegans* responding to different pathogens are largely unknown. In Chapter 3, we identify pathogen-specific ESPs in order to elucidate the host-pathogen interaction and to identify host pathways exploited by different pathogens.

1.2 Summary

The excretome/secretome of the free-living model animal, *C. elegans*, had not been noticed and characterized. This field had long been neglected and scientists underestimated the complexity and evolutional importance of *C. elegans* ESPs.

Thus, in Chapter 2, with the sensitive and high throughput proteomic method, we for the first time systematically profile the protein components of *C. elegans* excretome/secretome. Functional annotation of ESPs revealed the presentence of several important protein families, including proteases, proteases inhibitors, lectins, lysozymes, superoxide dismutase, and peroxidases. We predicted functions for ESPs, including bacteria/pathogen defense, immune system regulation, response to stimuli, and nutrition uptake. We also for the first time demonstrate that the proteases and proteases inhibitors in ESPs are enzymatically active *in vivo*.

Quantitative proteomic method was employed to profile excretomes/secretomes under different temperatures and bacterial diet exposure. This is the first attempt to study the dynamic regulation of *C. elegans* excretome/secretome. We show under different temperatures that the main structures of the excretome/secretome remain stable. Upon exposure to pathogenic bacteria, lysozymes are highly up-regulated to defend against toxic effects.

In Chapter 3, we take a further step to push the comparative method to a broader application, taking advantage of the deeply sequenced *C. elegans* transcriptomes of different developmental stages. Clustering revealed the stage-specific ESPs. Dauer-specfic ESPs are of particular interest due to the similarities between dauer and parasite infective

larvae. To provide a cross-reference to our prediction method, we directly characterized dauer ESPs with Mass-Spec. The most interesting observation is that proteases inhibitors are up-regulated in dauer formation and proteases are up-regulated in dauer exit, suggesting that protease activity level maybe be related with dauer decision making.

Pathogen-specific ESPs are revealed by clustering the pathogen-exposed transcriptomes. We validated our prediction by multiple infection assays. The results showed our method is the quickest way to select out candidate genes for functional assays.

In Chapter 4, we characterized the protein components of the *C. elegans* sibling species, *C. inopinata*. This provides a great example of how *C. elegans* ESPs facilitate the annotation to the gene model of a new species. The comparisons between species showed protein families are conserved in excretome/secretome. Thus, the knowledge of ESPs in *C. elegans* can be used to annotate the poorly understood genome or excretome/secretome in parasites.

In summary, the characterization of *C. elegans* ESPs is the first study towards free-living nematodes excretome/secretome. How *C. elegans* responds to environmental changes is unclear and a large portion of identified ESPs has unknown functions. Our work provides a solid basis to study the mechanisms and signaling pathways of how *C. elegans* maintains homeostasis. Also, our work would be useful to help annotate poorly understood parasitic excretome/secretome based on homologies. It also enables the comparison with available parasites excretomes/secretomes, providing clues for how parasitism evolves.

1.3 References

[1] J. Borloo *et al.*, "In-Depth Proteomic and Glycomic Analysis of the Adult-Stage *Cooperia oncophora* Excretome/Secretome," *J. Proteome Res.*, vol. 12, no. 9, pp. 3900–3911, Sep. 2013.

[2] H. Soblik *et al.*, "Life Cycle Stage-resolved Proteomic Analysis of the Excretome/Secretome from Strongyloides ratti—Identification of Stage-specific Proteases," *Mol. Cell. Proteomics MCP*, vol. 10, no. 12, Dec. 2011.

[3] L.-P. Leroux *et al.*, "Analysis of the Trichuris suis excretory/secretory proteins as a function of life cycle stage and their immunomodulatory properties," *Sci. Rep.*, vol. 8, no. 1, p. 15921, Oct. 2018.

[4] A. Mahmuda *et al.*, "Proteomic identification of larval excretory / secretory diagnostic biomarkers for early hyperinfection in strongyloidiasis," 2017.

[5] D. Ditgen, E. M. Anandarajah, K. A. Meissner, N. Brattig, C. Wrenger, and E. Liebau, "Harnessing the Helminth Secretome for Therapeutic Immunomodulators," *BioMed Research International*, 2014.

[6] S. Bennuru, R. Semnani, Z. Meng, J. M. C. Ribeiro, T. D. Veenstra, and T. B. Nutman, "Brugia malayi excreted/secreted proteins at the host/parasite interface: stage- and gender-specific proteomic profiling," *PLoS Negl. Trop. Dis.*, vol. 3, no. 4, p. e410, 2009.

[7] C. Dieterich and R. J. Sommer, "How to become a parasite - lessons from the genomes of nematodes," *Trends Genet. TIG*, vol. 25, no. 5, pp. 203–209, May 2009.

[8] H. Soblik *et al.*, "Life Cycle Stage-resolved Proteomic Analysis of the Excretome/Secretome from Strongyloides ratti—Identification of Stage-specific Proteases," *Mol. Cell. Proteomics MCP*, vol. 10, no. 12, Dec. 2011.

[9] S. Bellafiore, Z. Shen, M.-N. Rosso, P. Abad, P. Shih, and S. P. Briggs, "Direct identification of the Meloidogyne incognita secretome reveals proteins with host cell reprogramming potential," *PLoS Pathog.*, vol. 4, no. 10, p. e1000192, Oct. 2008.

[10] J. Geary *et al.*, "First analysis of the secretome of the canine heartworm, Dirofilaria immitis," *Parasit. Vectors*, vol. 5, no. 1, p. 140, Jul. 2012.

[11] J. F. Chehayeb, A. P. Robertson, R. J. Martin, and T. G. Geary, "Proteomic Analysis of Adult Ascaris suum Fluid Compartments and Secretory Products," *PLoS Negl. Trop. Dis.*, vol. 8, no. 6, p. e2939, Jun. 2014.

[12] A. R. Jex *et al.*, "*Ascaris suum* draft genome," *Nature*, vol. 479, no. 7374, pp. 529–533, Nov. 2011.

[13] "Stage- and Gender-Specific Proteomic Analysis of Brugia malayi Excretory-Secretory Products."

14] R. Shinya, H. Morisaka, T. Kikuchi, Y. Takeuchi, M. Ueda, and K. Futai, "Secretome Analysis of the Pine Wood Nematode Bursaphelenchus xylophilus Reveals the Tangled Roots of Parasitism and Its Potential for Molecular Mimicry," *PLoS ONE*, vol. 8, no. 6, Jun. 2013.

[15] J. P. Hewitson *et al.*, "Secretion of Protective Antigens by Tissue-Stage Nematode Larvae Revealed by Proteomic Analysis and Vaccination-Induced Sterile Immunity," *PLoS Pathog.*, vol. 9, no. 8, p. e1003492, Aug. 2013.

[16]H. Saverwyns *et al.*, "Identification and characterization of a novel specific secreted protein family for selected members of the subfamily Ostertagiinae (Nematoda)," *Parasitology*, vol. 135, no. Pt 1, pp. 63–70, Jan. 2008.

[17]J. Mulvenna, B. Hamilton, S. H. Nagaraj, D. Smyth, A. Loukas, and J. J. Gorman, "Proteomics analysis of the excretory/secretory component of the blood-feeding stage of the hookworm, Ancylostoma caninum," *Mol. Cell. Proteomics MCP*, vol. 8, no. 1, pp. 109–121, Jan. 2009.

[18]E. M. Schwarz, Y. Hu, I. Antoshechkin, M. M. Miller, P. W. Sternberg, and R. V. Aroian, "The genome and transcriptome of the zoonotic hookworm *Ancylostoma ceylanicum* identify infection-specific gene families," *Nat. Genet.*, vol. 47, no. 4, pp. 416–422, Apr. 2015.

[19]H. Soblik *et al.*, "Life cycle stage-resolved proteomic analysis of the excretome/secretome from Strongyloides ratti--identification of stage-specific proteases," *Mol. Cell. Proteomics MCP*, vol. 10, no. 12, p. M111.010157, Dec. 2011.

[20]G. Garg and S. Ranganathan, "In silico secretome analysis approach for next generation sequencing transcriptomic data," *BMC Genomics*, vol. 12, no. Suppl 3, p. S14, Nov. 2011.

[21]V. L. Hunt *et al.*, "The genomic basis of parasitism in the *Strongyloides* clade of nematodes," *Nat. Genet.*, vol. 48, no. 3, pp. 299–307, Mar. 2016.

[22]H. Craig, J. M. Wastling, and D. P. Knox, "A preliminary proteomic survey of the in vitro excretory/secretory products of fourth-stage larval and adult Teladorsagia circumcincta," *Parasitology*, vol. 132, no. Pt 4, pp. 535–543, Apr. 2006.

[23]S. K. Smith *et al.*, "Proteomic analysis of excretory/secretory products released by Teladorsagia circumcincta larvae early post-infection," *Parasite Immunol.*, vol. 31, no. 1, pp. 10–19, Jan. 2009.

[24]M. W. Robinson, R. Greig, K. A. Beattie, D. J. Lamont, and B. Connolly, "Comparative analysis of the excretory-secretory proteome of the muscle larva of Trichinella pseudospiralis and Trichinella spiralis," *Int. J. Parasitol.*, vol. 37, no. 2, pp. 139–148, Feb. 2007.

[25]J. A. Gadahi *et al.*, "Proteomic Analysis of the Excretory and Secretory Proteins of Haemonchus contortus (HcESP) Binding to Goat PBMCs In Vivo Revealed Stage-Specific Binding Profiles," *PLoS ONE*, vol. 11, no. 7, Jul. 2016.

[26]R. Laing *et al.*, "The genome and transcriptome of Haemonchus contortus, a key model parasite for drug and vaccine discovery," *Genome Biol.*, vol. 14, no. 8, p. R88, Aug. 2013.

[27]Y. M. Harcus *et al.*, "Signal sequence analysis of expressed sequence tags from the nematode Nippostrongylus brasiliensis and the evolution of secreted proteins in parasites," *Genome Biol.*, vol. 5, no. 6, p. R39, 2004.

[28]S. Bennuru, R. Semnani, Z. Meng, J. M. C. Ribeiro, T. D. Veenstra, and T. B. Nutman, "Brugia malayi Excreted/Secreted Proteins at the Host/Parasite Interface: Stage- and Gender-Specific Proteomic Profiling," *PLoS Negl. Trop. Dis.*, vol. 3, no. 4, Apr. 2009.

[29]"High-Throughput Proteomics | Annual Review of Analytical Chemistry."

[30]J. P. Hewitson, J. R. Grainger, and R. M. Maizels, "Helminth immunoregulation: the role of parasite secreted proteins in modulating host immunity," *Mol. Biochem. Parasitol.*, vol. 167, no. 1, pp. 1–11, Sep. 2009.

[31]R. T. Semnani, M. Law, J. Kubofcik, and T. B. Nutman, “Filaria-Induced Immune Evasion: Suppression by the Infective Stage of Brugia malayi at the Earliest Host-Parasite Interface,” *J. Immunol.*, vol. 172, no. 10, pp. 6229–6238, May 2004.

[32]J. Correale and M. Farez, “Association between parasite infection and immune responses in multiple sclerosis,” *Ann. Neurol.*, vol. 61, no. 2, pp. 97–108, 2007.

[33]J. M. Hawdon, B. F. Jones, D. R. Hoffman, and P. J. Hotez, “Cloning and Characterization of Ancylostoma-secreted Protein A NOVEL PROTEIN ASSOCIATED WITH THE TRANSITION TO PARASITISM BY INFECTIVE HOOKWORM LARVAE,” *J. Biol. Chem.*, vol. 271, no. 12, pp. 6672–6678, Mar. 1996.

[34]J. Murray, B. Manoury, A. Balic, C. Watts, and R. M. Maizels, “Bm-CPI-2, a cystatin from Brugia malayi nematode parasites, differs from Caenorhabditis elegans cystatins in a specific site mediating inhibition of the antigen-processing enzyme AEP,” *Mol. Biochem. Parasitol.*, vol. 139, no. 2, pp. 197–203, Feb. 2005.

[35]A. Loukas and R. M. Maizels, “Helminth C-type Lectins and Host–Parasite Interactions,” *Parasitol. Today*, vol. 16, no. 8, pp. 333–339, Aug. 2000.

[36]D. Ditgen *et al.*, “Comparative characterization of two galectins excreted-secreted from intestine-dwelling parasitic versus free-living females of the soil-transmitted nematode Strongyloides,” *Mol. Biochem. Parasitol.*, vol. 225, pp. 73–83, Oct. 2018.

[37]X. Zang *et al.*, “Homologues of human macrophage migration inhibitory factor from a parasitic nematode. Gene cloning, protein activity, and crystal structure,” *J. Biol. Chem.*, vol. 277, no. 46, pp. 44261–44267, Nov. 2002.

[38]Y. Cho *et al.*, "Structural and Functional Characterization of a Secreted Hookworm Macrophage Migration Inhibitory Factor (MIF) That Interacts with the Human MIF Receptor CD74," *J. Biol. Chem.*, vol. 282, no. 32, pp. 23447–23456, Aug. 2007.

[39]R. M. Maizels, M. L. Blaxter, and A. L. Scott, "Immunological genomics of Brugia malayi: filarial genes implicated in immune evasion and protective immunity," *Parasite Immunol.*, vol. 23, no. 7, pp. 327–344, 2001.

[40]B. Maresca and G. S. Kobayashi, "Hsp70 in parasites: as an inducible protective protein and as an antigen," *Experientia*, vol. 50, no. 11, pp. 1067–1074, Nov. 1994.

[41]C. B. Stephensen, "Vitamin A, infection, and immune function," *Annu. Rev. Nutr.*, vol. 21, pp. 167–192, 2001.

[42]W. Harnett and M. M. Harnett, "Filarial nematode secreted product ES-62 is an anti-inflammatory agent: therapeutic potential of small molecule derivatives and ES-62 peptide mimetics," *Clin. Exp. Pharmacol. Physiol.*, vol. 33, no. 5–6, pp. 511–518, Jun. 2006.

[43]L. C. Clark and J. Hodgkin, "Commensals, probiotics and pathogens in the Caenorhabditis elegans model," *Cell. Microbiol.*, vol. 16, no. 1, pp. 27–38, 2014.

[44]M. J. Gravato-Nobre and J. Hodgkin, "Caenorhabditis elegans as a model for innate immunity to pathogens," *Cell. Microbiol.*, vol. 7, no. 6, pp. 741–751, 2005.

[45]T. Kaletta and M. O. Hengartner, "Finding function in novel targets: C. elegans as a model organism," *Nat. Rev. Drug Discov.*, vol. 5, no. 5, p. 387, May 2006.

[46]A. G. Alexander, V. Marfil, and C. Li, "Use of Caenorhabditis elegans as a model to study Alzheimer's disease and other neurodegenerative diseases," *Front. Genet.*, vol. 5, p. 279, 2014.

[47]E. F. Griffin, K. A. Caldwell, and G. A. Caldwell, "Genetic and Pharmacological Discovery for Alzheimer's Disease Using Caenorhabditis elegans," *ACS Chem. Neurosci.*, vol. 8, no. 12, pp. 2596–2606, 20 2017.

[48]Y. Wu and Y. Luo, "Transgenic C. elegans as a model in Alzheimer's research," *Curr. Alzheimer Res.*, vol. 2, no. 1, pp. 37–45, Jan. 2005.

[49]N. V. Kirienko, K. Mani, and D. S. Fay, "Cancer models in C. elegans," *Dev. Dyn. Off. Publ. Am. Assoc. Anat.*, vol. 239, no. 5, pp. 1413–1448, May 2010.

[50]C. L. Kurz and M.-W. Tan, "Regulation of aging and innate immunity in C. elegans," *Aging Cell*, vol. 3, no. 4, pp. 185–193, 2004.

[51]C. L. Kurz and M.-W. Tan, "Regulation of aging and innate immunity in C. elegans: Aging and innate immunity in C. elegans, C. L. Kurz and M.-W. Tan," *Aging Cell*, vol. 3, no. 4, pp. 185–193, Jul. 2004.

[52]K. Schmeisser and J. A. Parker, "Worms on the spectrum - C. elegans models in autism research," *Exp. Neurol.*, vol. 299, no. Pt A, pp. 199–206, 2018.

[53]W.-R. Wong *et al.*, "Autism-associated missense genetic variants impact locomotion and neurodevelopment in Caenorhabditis elegans," *Hum. Mol. Genet.*

[54]T. Stiernagle, "Maintenance of C. elegans," *WormBook*, 2006.

[55]X. Karp, "Working with dauer larvae," *WormBook*, pp. 1–19, Aug. 2018.

[56]H. Lee *et al.*, "Nictation, a dispersal behavior of the nematode *Caenorhabditis elegans*, is regulated by IL2 neurons," *Nat. Neurosci.*, vol. 15, no. 1, pp. 107–112, Jan. 2012.

[57]M. Crook, “The dauer hypothesis and the evolution of parasitism: 20 years on and still going strong,” *Int. J. Parasitol.*, vol. 44, no. 1, pp. 1–8, Jan. 2014.

[58]W. Wong, “How to Evolve into a Parasite,” *Sci. Signal.*, vol. 2, no. 54, pp. ec26–ec26, Jan. 2009.

[59]H. Aguilaniu, P. Fabrizio, and M. Witting, “The Role of Dafachronic Acid Signaling in Development and Longevity in Caenorhabditis elegans: Digging Deeper Using Cutting-Edge Analytical Chemistry,” *Front. Endocrinol.*, vol. 7, Feb. 2016.

[60]K. J. Dumas *et al.*, “Functional divergence of dafachronic acid pathways in the control of C. elegans development and lifespan,” *Dev. Biol.*, vol. 340, no. 2, pp. 605–612, Apr. 2010.

[61]T.-M. Li *et al.*, “No Significant Increase in the Δ4- and Δ7-Dafachronic Acid Concentration in the Long-Lived glp-1 Mutant, nor in the Mutants Defective in Dauer Formation,” *G3 Genes Genomes Genet.*, vol. 5, no. 7, pp. 1473–1479, Jul. 2015.

[62]N. Kanzaki *et al.*, “Biology and genome of a newly discovered sibling species of Caenorhabditis elegans,” *Nat. Commun.*, vol. 9, no. 1, p. 3216, 10 2018.

[63]D. L. Riddle, T. Blumenthal, B. J. Meyer, and J. R. Priess, *Regulation of Translation During Development*. Cold Spring Harbor Laboratory Press, 1997.

[64]E. A. Abada, H. Sung, M. Dwivedi, B.-J. Park, S.-K. Lee, and J. Ahnn, “C. elegans behavior of preference choice on bacterial food,” *Mol. Cells*, vol. 28, no. 3, pp. 209–213, Sep. 2009.

[65]J. Hodgkin, P. E. Kuwabara, and B. Corneliussen, “A novel bacterial pathogen, Microbacterium nematophilum, induces morphological change in the nematode C. elegans,” *Curr. Biol.*, vol. 10, no. 24, pp. 1615–1618, Dec. 2000.

[66]J. Höflich *et al.*, "Loss of srf-3-encoded nucleotide sugar transporter activity in Caenorhabditis elegans alters surface antigenicity and prevents bacterial adherence," *J. Biol. Chem.*, vol. 279, no. 29, pp. 30440–30448, Jul. 2004.

[67]C. Darby, "Interactions with microbial pathogens," *WormBook*, 2005.

[68]H. R. Nicholas and J. Hodgkin, "The ERK MAP Kinase Cascade Mediates Tail Swelling and a Protective Response to Rectal Infection in C. elegans," *Curr. Biol.*, vol. 14, no. 14, pp. 1256–1261, Jul. 2004.

[69]S. Miyata, J. Begun, E. R. Troemel, and F. M. Ausubel, "DAF-16-Dependent Suppression of Immunity During Reproduction in Caenorhabditis elegans," *Genetics*, vol. 178, no. 2, pp. 903–918, Feb. 2008.

[70]M. E. Yanos, C. F. Bennett, and M. Kaeberlein, "Genome-Wide RNAi Longevity Screens in Caenorhabditis elegans," *Curr. Genomics*, vol. 13, no. 7, pp. 508–518, Nov. 2012.

[71]A. V. Samuelson, R. R. Klimczak, D. B. Thompson, C. E. Carr, and G. Ruvkun, "Identification of Caenorhabditis elegans genes regulating longevity using enhanced RNAi-sensitive strains," *Cold Spring Harb. Symp. Quant. Biol.*, vol. 72, pp. 489–497, 2007.

Chapter 2

PROTEOMICS PROFILING AND FUNCTIONAL CHARACTERIZATION OF *CAENORHABDITIS ELEGANS* EXCRETOME/SECRETOME

2.1 Abstract

Excretory-secretory proteins (ESPs) are first characterized and defined in parasitic nematode proteomics studies as the combination of various biomolecules that are continuously excreted or secreted into the environment throughout the whole life cycle. ESPs are particularly interesting to many scientists as anti-parasitic vaccine candidates and as promising drug targets since large portions of ESPs are active enzymes that potentially function directly at the parasite-host or worm-environment interfaces. ESPs are also reported to play pivotal roles in many critical pathways, regulating nematode survival, reproduction, food processing, and innate immune response. However, proteomics studies using parasitic nematodes as models are limited due to lack of whole genome sequence knowledge and lack of genome-editing tools. Thus, the number of ESPs identified is limited and many functions of ES proteins are elusive. Therefore, we use the most studied nematode, *Caenorhabditis elegans*, as the model to characterize the composition of excreted/secreted proteins with the help of nano-liquid chromatography coupled with tandem mass spectrometry (nanoLC-MS/MS). In summary, we characterized more than 509 excreted/secreted proteins with mix-staged worms, including many metalloproteases, cysteine proteases, and lysozymes. With gene ontology analysis, many proteins are annotated to play roles in defending bacterial infection and regulating pathogen susceptibility. Proteases and proteases inhibitors are a major group in *C. elegans* ESPs. We performed stable isotope dimethyl labeling quantitative proteomics and compared *Caenorhabditis elegans* excretomes/secretomes on different bacteria diets. Lysozymes are

not only enriched in *C. elegans* ESPs but are also up-regulated in response to pathogens and bacteria.

2.2 Introduction

Many parasitic worms are associated with severe diseases, infecting people, animals, and plants and causing heavy economic burdens around the world [1]–[3]. For example, *Echinococcus granulosus* is responsible for Cystic hydatid disease (CHD) [4], [5]. Filarial parasites are the cause of Lymphatic filariasis (LF) [6]. Pine wood nematode *Bursaphelenchus xylophilus* is a notorious forest pest, causing pine wilt disease [7]. Data has shown that 1334 million people are at risk of parasitic nematodes infection and 3 billion dollars is spent each year on parasitic worm-related diseases treatment or prevention [2]. Excretory-secretory (ES) products are a wide range of biomolecules, primarily proteins, which are continually excreted and secreted from nematodes through all life stages [1], [8]–[10]. ESPs of parasitic worms have attracted the attention of many researchers for decades as pathogenicity factor candidates. Upon infection, parasitic nematodes need to penetrate tissue barriers and migrate through the host tissue to a suitable niche while evading host immune defenses [1]. ESPs are believed to mediate this complex process and function by regulating the host immune system, allowing parasites to successfully invade and reproduce inside or transfer between hosts. Complex components of ESPs, including hundreds of different proteins, have multiple functions regarding modulation of signaling pathways, nutrient transport and/or uptake, digestion, blood coagulation, and so on [2], [7], [11]–[14].

All these facts address the importance and urgent requirements for more knowledge of these physiologically and clinically important biomolecules. Many parasitic nematodes even develop comprehensive and highly specific-host-dependent life stages during the long

history of adaptation and co-evolution [11]. Differences in protein excretion/secretion among nematodes may relate to the ecological niche of each parasite and the pathological changes that they induce. A better understanding of how these specific interactions take place and how they achieve precise control and biological function clearly requires more study of the nematodes excretome/secretome. The excresome/secretome is important for cell communication, cell adhesion, and interaction with the environment. During development, secreted proteins are essential for cell fate specification and cell migration [12], [14], [15].

ESPs are active directly at the interface between nematodes/environment and parasite/host. Many proteases are included and play a pivotal role with their catalytic activities. Proteases in ESPs are great candidates for potential new drug targets for intervention and vaccine design, especially considering the ongoing resistance problem with current anti-parasite drugs [16]–[20] .

Recently, with the benefit of high-throughput proteomic technology coupled with developing genome and transcriptome sequencing, several proteomic studies have been done in different parasite-animal models, providing rich information on parasitism in different species [11], [21]–[24]. The ESPs profiles have been identified from *Brugia malayi*, the canine hookworm *Ancylostoma caninum*, the plant parasitic nematode *Meloidogyne incognita,* and from *Strongyloides ratti* [11], [25], [26][26], [27]. However, not all of these nematode datasets were analyzed against complete genomes (or transcriptomes). Some of the compilations are less completely assigned than others, which hinders the comparison between nematodes species. Additionally, only a limited number of secreted proteins have been cloned and characterized. Limited knowledge regarding the

molecules involved in the pathology as well as lack of techniques for functional analysis in parasitic nematodes greatly restricts the ability to elucidate the molecular mechanisms of pathogenicity [26].

Thus, excretory-secretory products of *C. elegans* were completely profiled using proteomic analysis and would provide insights to the molecular basis of how *C. elegans* adapt to changes in the environment.

2.3 Results

Characterization of *C. elegans* Excreted/Secreted Proteins (ESPs) using nanoLC-MS/MS

The complete content of *C. elegans* ESPs remains unknown to date. Proteomic analysis, at present, is the most powerful and successful approach in directly identifying a large set of secreted proteins. A large-scale identification of secreted proteins using nano-liquid chromatography coupled with tandem mass spectrometry (nanoLC-MS/MS) analysis was used in this study, characterizing the ESP profile of *C. elegans* with good quality.

C. elegans wild type N2 strain was cultured in liquid culture using *E. coli* HB101 as food source. Initial concentration of 3 worms/μl was used in a 250 ml S Medium system and worms were grown at 20°C for 7 days. Mix-staged nematodes were then harvested by centrifuge and washed intensively six times in M9 buffer. Worms were left in 1 ml M9 buffer for incubation at 20°C for another 4 hours. Supernatant was collected and filtered by a 0.22μm syringe filter unit and treated as the unconcentrated *C. elegans* Excreted/Secreted Proteins. *C. elegans* ESPs sample was further digested by lysine/trypsin and subjected to nano LC-MS/MS for peptide characterization (**Figure 2.1**).

The number of *C. elegans* proteins identified in one single test is fairly sensitive to the quality of prepared excretome/secretome sample (low background contaminating peptides noise, high reading depth), ranging from 395 to 637 secreted/excreted proteins. As shown in Figure 2.2, 228 proteins were identified in four biological replicates and 605 proteins were shown up in at least 2 out of 4 replicates. We further excluded proteins that were identified less than 2 times by MS/MS, reducing the set to 509 proteins. We treated this

subset of 509 proteins as a confident profile of *C. elegans* ESPs and majority of proteins were also identified with two or more specific peptides (**Figure 2.2**). The detailed composition of ESP is provided in a concise form in Appendix Table 2.1, which provided details for mass-spec readout of these 509 *C. elegans* secreted/excreted proteins (**Appendix Table 2.1**). This confident subset of 509 proteins was further used in subsequent functional annotations.

These numbers of ESPs identified in Mass-Spec are comparable with earlier parasites secretome data: a proteomic study with *Strongyloides ratti* identified 586 proteins in total and 852 proteins were characterized in *Brugia malayi* ESPs [25], [26]. With continually refining and experimentally verification after publication of the first draft of *Caenorhabditis elegans* genome, 20,242 protein-coding genes are now included in the release 210 of WormBase (http://www.wormbase.org) [28] . Our data covered about 2.5% (509 out of 20,242) of the whole worm protein-coding genes.

Identified *C. elegans* ESPs were annotated to be secreted

We first searched the whole list of *C. elegans* ESPs against UniprotKB for secretion information. 25 proteins that were manually annotated to be secreted were listed (**Table 2.1**). As we expected, aspartic proteases like *asp-6* and *asp-3*, zinc metalloproteinase were important proteinases that may mediate critical substrate processing for nutrient uptake. Transthyretic-like protein family members were reported to be key extracellular proteins with immunomodulatory potential [2]. This knowledge strongly supports that the identified *C. elegans* ESPs are collected from secretion with biological meanings.

The prediction of a signal peptide is typically used as a first filter to identify secreted proteins [29]. Thus, we examined the presence of signal peptide and location of signal peptide cleavage sites in *C. elegans* ESPs using SignalP5.0 [30]. 368 out of 509 proteins, which represented 72.3% of the whole list of *C. elegans* ESPs, were predicted to contain a signal peptide. This suggested that the majority of identified proteins were very possibly secreted. The proportion of the identified proteins bearing a secretion signal was higher than the percentage of signal peptide containing proteins in reported literature on other nematodes: *B. malayi* (55%) and *B. xylophilus* (41%) [7], [25].

The remaining 27.7% proteins may have unknown secretory signals or lack of a classical signal peptide, or may be secreted through non-classical secretory pathways. In parasitic nematodes, thioredoxins and macrophage migration inhibitory factor homolog in *B. malayi* were proven to be exported despite lacking a signal peptide [31], [32] . Helminth parasites were shown to produce exosome, carrying immunoglobulins and metabolic enzymes [33]–[36]. Exosome proteome of *E. caproni* even explained 54% of the characterized secretome [37]. Thus, non-canonical secretion mechanism might be a common feature in nematodes [37].

Next, since *C. elegans* genes have been intensively studied for several decades, we did a further search in WormBase, collecting all expression data (**Appendix Table 2.2**). We then examined the cell/tissue enrichment to find which anatomical parts are statistically over-represented (**Figure 2.3**). As shown in Figure 2.3, the intestine is the most enriched tissue source of *C. elegans* ESPs, suggesting a large portion of ESPs may regulate food processing and nutrient uptake. *C. elegans,* as a free-living nematode, mainly interacts with

bacteria. The intestine is the major location to interact with outer environment, including processes related to food digestion and defending against harmful bacteria or pathogens.

Several genes even have published fluorescent protein reporter lines of expression patterns. The *C. elegans* ESPs are expressed in intestine, excretory cell and gland cell, amphid sheath cell, hypodermis and uterine epithelial cell and more [38], [39]. The *C. elegans* excretory system in *C. elegans* contains four cells: pore cell, duct cell, canal cell, and a fused pair of gland cells [40]. The excreted/secreted materials pass gland and canal cells and are deposited outside through duct cell [41]. The published expression patterns once more confirmed that the proteins identified in proteomics of *C. elegans* ESPs were secreted/excreted.

Cellular components contaminations were inevitable yet controlled at reasonable level in *C. elegans* ESPs

Major contamination of the results comes from the *E. coli* food that was used during the culture of worms. In preliminary Mass-Spec runs, 70% of identified proteins were of *E. coli* origin. To solve this problem, liquid culture of *C. elegans* was harvested after 7 days when no obvious food trace could be observed. Worms were washed intensively 6 times to further remove bacteria attached to worm surface or defecated from worm body. After optimization of the whole workflow, each Mass-Spec run only contained less than 10 proteins of *E. coli* origin.

Another major source of contamination came from the lysate of dead worms. A drop of liquid culture was examined carefully under a compound microscope before harvesting the nematodes. Only cultures with living worms and without other fungal or bacterial

contamination were used in subsequent incubation and Mass-Spec sample preparation. Enzyme digestion (like trypsin) and MS sample preparation steps may also have introduced contaminating proteins, yet these proteins can be easily identified and excluded from the final annotation of the results.

As shown in Appendix Table 2.1, some cellular proteins, for example, ribosomal proteins and actins, are included in *C. elegans* ESPs. These may at first sight be considered as contaminates from worm lysate. However, we carefully checked the viability of worms before sample collection and minimized the effects of worm lysate. The parasitic nematodes excretome/secretome also have lots of ribosomal proteins and actins [7], [11], [25]. The control done with whole worm lysate confirmed that ribosomal proteins and actins are *bona fide* ESPs [7].

Brief overview of *C. elegans* Excretome/Secretome structure shows several protein families are main players

The most abundant proteins identified in *C. elegans* ESPs under the standard culture condition were listed (**Table 2.2**). *C. elegans* excretome/secretome shared many proteins in common with other nematodes secretome, including lysozyme family members, small heat shock proteins, aspartic proteases, cystatins (cysteine protease inhibitors), serine protease inhibitors (serpins), lectins, and transthyretin family proteins [7], [25]. The common features among species indicated that the conservation and potential evolution relationships during parasitism acquisition could be elucidated through comparing *C. elegans* excretome/secretome data with ESPs from other nematodes.

Gene Ontology analysis

Another main purpose of ESPs study in *C. elegans* is to utilize the excretome/secretome data collected in the Mass-Spec experiment and to transform it into rich information of clues and guides for following functional characterization for individual protein entry. One obvious and critical challenge is to bring order into this overwhelming amount of data. Gene Ontology has had a great success in developing precise terminology that can be used across many different species, providing structured means to describe the biological processes, cell components, and molecular functions of gene products. The precise role of ES proteins from parasitic nematodes in mediating cellular processes is largely unknown due to the lack of knowledge in function annotation. However, the best currently available genome annotation with *C. elegans* has a great advantage in explaining functions of excretory/secretory proteins.

We conducted Gene Ontology (GO) term analysis to the list of 509 excreted/secreted proteins, using the whole worm genome data as the reference background. Thus, we could examine the relative abundance of secreted proteins and detect whether the secreted proteins are enriched in the excretome/secretome result. A considerable amount of software for ontological analysis of gene lists has been published over the past 5 years, each of them having advantages and drawbacks and each approaching data or vocabularies in a slightly different manner. Gene Ontology Consortium (powered by PANTHER, http://geneontology.org) [42]–[46] and the Database for Annotation, Visualization and Integrated Discovery (DAVID, http://david.abcc.ncifcrf.gov) [47], [48] are among some of the most well-known tools. We used both tools here to examine our Mass-Spec data, which gave back similar annotations. The GO analysis results shown below are generated with the

former tool. The 509 proteins with gene id information were submitted under GO analysis, with 508 proteins successfully annotated with 1 unmapped WormBase gene id. Also, the top 50 most abundant proteins identified in the preliminary result were submitted to similar analysis as well to assure the validation of this method.

PANTHER overrepresentation test was performed to GO terms of cellular component annotation with Fisher's exact test and corrected by calculating false discovery rate [43]. Over or under represented GO terms of cellular component were listed (**Table 2.3**). The enriched GO terms are associated with secretion or suggest proteins are presenting at surface, as most enriched cellular components are found to be vesicle lumen (GO:0031983), membrane raft (GO:0045121), vacuole (GO:0005773), lysosome (GO:0005764) or extracellular region (GO:0005576) **(Figure 2.4)**. Genes with GO terms associated with some level of secretion explained almost half of the whole gene list, strongly supporting the conclusion that the proteins characterized in Nano LC-MS/MS are from secretion.

A similar test was performed with biological process GO terms. Clearly, several major subgroups can be identified from the biological process GO term list. The first group contains peptidoglycan metabolic process (GO:0000270), peptidoglycan catabolic process (GO:0009253), glycosaminoglycan catabolic process (GO:0006027), cell wall macromolecule catabolic process (GO:0016998), glycosphingolipid catabolic process (GO:0046479), glycolipid catabolic process (GO:0019377), chitin metabolic process (GO:0006030), glucose metabolic process (GO:0006006), aminoglycan catabolic process (GO:0006026). These GO terms indicate *C. elegans* ESPs may involve in regulating the surface glycan modification, surface galectin presentation, and surface coat synthesis.

Surface glycan and galectins are key molecules in interacting with binding and invasion of environmental bacteria and pathogens. The second group contains gland morphogenesis (GO:0022612), gland development (GO:0048732), and pharyngeal gland morphogenesis (GO:1905905). This finding is consistent with our hypothesis that the *C. elegans* excretory system and head sensory neurons are continuously secreting proteins to facilitate food uptake and chemical sensing, therefore modifying nematodes' behavior according to environmental signals. The third group of biological process GO terms includes defense response to Gram-positive bacterium (GO:0050830), response to superoxide (GO:0000303), innate immune response (GO:0045087), and defense response to Gram-negative bacterium (GO:0050829). This is consistent with our expectation that excretory-secretory proteins are capable of regulating immune responses and helping nutrition uptake or removing pathogens/cell debris, probably through lysosome and phagolysosome. The fourth group includes response to oxygen radical (GO:0000305), response to external biotic stimulus (GO:0043207), response to other organism (GO:0051707), and response to biotic stimulus (GO:0009607). This subset of ESPs again confirmed the critical role of ESPs in response to stimuli, thus adapting to the changing environment. Moreover, proteolysis (GO:0006508) and regulation of proteolysis (GO:0030162) are also well represented in *C. elegans* ES products, supporting the earlier assertion that many of the ES proteins are enzymatically active, playing important roles in regulating biological activity and metabolism. On the other side, GO terms depleted in Mass-Spec result support the conclusion that the data collected were truly from secretion once more. The underrepresented biological process GO terms mainly contain nucleic acid metabolic process (GO:0090304), regulation of transcription, DNA-templated (GO:0006355), and

regulation of RNA biosynthetic process (GO:2001141). These are mainly terms associated with proteins that function inside the nuclei **(Figure 2.5 and Table 2.4)**.

GO terms enriched in molecular function **(Figure 2.6 and Table 2.5)** are mainly lysozymes and proteases and proteases inhibitors, including galactosidase activity (GO:0015925), serine-type exopeptidase activity (GO:0070008), lysozyme activity (GO:0003796), peptidoglycan muralytic activity (GO:0061783), threonine-type peptidase activity (GO:0070003), nutrient reservoir activity (GO:0045735), aspartic-type endopeptidase activity (GO:0004190), cysteine-type endopeptidase activity (GO:0004197), serine hydrolase activity (GO:0017171), endopeptidase inhibitor activity (GO:0004866), and serine-type endopeptidase inhibitor activity (GO:0004867). GO terms depleted in Mass-Spec result included nucleic acid binding (GO:0003676), DNA-binding transcription factor activity (GO:0003700), ion channel activity (GO:0005216), and transcription regulator activity (GO:0140110). Proteins related with these activities are usually believed to function in cellular or nuclear parts, with less possibility to be found in excretome/secretome in *C. elegans.*

KEGG and reactome pathway analysis

KEGG and reactome pathway analysis were performed to identify pathways that are over-represented in *C. elegans* ESPs **(Figure 2.7 and Figure 2.8)** [49]–[52] . Innate immune system pathway and protein degradation pathway are the most over-represented pathway, suggesting one major role of *C. elegans* ESPs is to continuously fight with the changing surroundings, specifically defending infection of bacteria pathogens.

Pfam domain search and InterProScan sequence search

Each protein sequence in *C. elegans* excretome/secretome data was also searched against Pfam and InterPro database **(Figure 2.9 and Figure 2.10) [53], [54]**. Each protein sequence was also BLASTed against each other, and the BLAST plot was made based on the similarity between sequences **(Figure 2.11)**. Key features and enriched protein domains can provide information to the ESPs functions. We could also gain better understanding of the structure of *C. elegans* ESPs by knowing what the major protein families in the result are. C-type lectins, aspartic peptidases (peptidase A1), serine peptideses (peptidase S10), cysteine proteases (peptidase C1), and transthyretins are the major protein families in C. elegans ESPs. Protein-protein interactions were also plotted by STRING **(Figure 2.12)** [55]. Cathepsin Z family and Cathepsin B family members are the cores to interact with other ESPs.

Proteases and proteases inhibitors in ESPs are enzymatically active

Peptidases contribute to host specificity, host range, and virulence [7], [25].To detect putative proteases (also termed peptidase or proteinases) and proteases inhibitors, a MEROPS batch BLAST search (http://merops.sanger.ac.uk) was performed as an additional approach besides GO terms [56]. A total of 509 secreted protein sequences were subjected to the MEROPS BLAST search and classified into detailed MEROPS proteases or proteases inhibitor families (E-value cutoff of 1e-4). This approach predicted 88 putative proteases and 87 proteases inhibitors, contributing to 34.4% of annotated genes in list. The percentage of peptidases in the *B. xylophilus* secretome was 10.6%, whereas in M. incognita and *B. malayi* they were 6.4% and 5.2%, respectively [7], [25].

Carbohydrate-active enzymes (CAZymes) are key regulators to metabolism of glycoconjugates, oligosaccharides, and polysaccharides. The size and diversity of CAZymes can provide clues to nematode nutritional strategy and host specificity. dbCAN (automated CAZyme annotation) database was searched to identify CAZymes in *C. elegans* ESPs [57]. With HMMER3 search, 17 glycoside hydrolases (GH) were identified in *C. elegans* ESPs, belonging to 11 GH families **(Table 2.6)**.

We directly examined *in vivo* enzyme activities of candidate *C. elegans* proteases and proteases inhibitors using utse (uterine seam cell) development model. Utse (uterine seam cell) has proven to be an excellent system to study cell outgrowth defects. During the L4 stage, utse undergoes cell outgrowth and nuclear migration. The cell body grows bidirectionally along the anterior-posterior axis and nuclei segregate into two groups, migrating and settling at the edges of the utse cell body. Both arms will reach final shape at L4 lethargus stage [58]. Utse is very sensitive to levels and activities of a series of complex regulators, including two astacin metalloproteases, *nas-21* and *nas-22*. Astacin family members are highly conserved in functions while *nas-20* and *nas-31* were detected in excretome/secretome. Additionally, one speculated ESPs function is to regulate tissue migration during host invasion, sharing a common feature with cell outgrowth and migration during *in vivo* development and with metastatic cancer spreading tumors from one tissue to another. Thus, utse would serve as a great platform to observe the effects of proteases or protease inhibitors. Generally, using RNA interference (RNAi) to target proteases might lead to the failure of cell outgrowth in utse while RNAi targeting protease inhibitors would cause overgrowth of cell body [58].

We screened 44 RNAi strains targeting selected genes from ESPs data with a putative protease or protease inhibitor function. Uste cell body was labeled with mCHERRY and nuclei of utse were labeled with GFP to allow tracking of the morphology of the utse cell. Defects in utse development, including shortened outgrowth of utse arms, incorrect migration of cell nuclei, were scored and recorded **(Figure 2.13 and Table 2.7)**. Among these genes, *clec-15, ttr-16, ttr-17,* and *mig-6* resulted in a pretty high rate of deficiencies of animals with a ratio of 50%, 60%, 40%, and 60%, respectively. Another 12 strains have around 20% defect rates. Indeed, we observed the failure of development in utse arms with proteins that have protease activities (example of *cpr-1*). Also, arms with abnormal shapes were observed with RNAi result targeting protease inhibitors. We calculated the p-values in comparison to the empty vector (RNAi) using Fisher's exact test. 7 genes (8 stains, both *mig-6* strains passed $p < 0.05$) in total are within the $p<0.05$ threshold (*clec-15, cpi-1, ttr-16, ttr-17, cpi-2, asp-6, mig-6*). These genes passed FDR test at 0.05 as well. (The positive rate is 8/44 = 0.0182, giving a new alpha equal to 0.009.)

Thus, proteases and proteases inhibitors contribute to a big portion in *C. elegans* ESPs. The RNAi targeting proteases and proteases inhibitors lead to cell outgrowth deficiency, directly confirming proteases and proteases inhibitors are enzymatically active *in vivo*. This fact makes studying proteases and proteases inhibitors even more interesting since it provides the possibility to interfere with the enzyme activities using small drug molecules. In other parasitic nematodes, several proteases were also proven to be active and disruption of protease activity would cause failure in parasites infection.

Culturing worms at different temperatures does not change the majority of *C. elegans* ESPs

We established the standard protocol to perform proteomics study of *C. elegans* ESPs. By changing the culturing conditions, we could detect *C. elegans* ESPs using same technique but expect the composition of new ESPs was modified according to the cultural condition. Therefore, comparison of ESPs under different conditions would allow quick selection of candidate genes for subsequent large-scale reverse genetic experiments. Here, we employed stable isotope dimethyl labeling quantitative proteomics since it is the most quick, easy, and affordable comparative Mass-Spec method.

The standard protocol for proteomics characterization of *C. elegans* ESPs was performed to a mix-staged population of nematodes culturing at 20°C on an *Escherichia coli* HB101 diet.

15°C - 25°C temperature range is the physiological condition for *C. elegans* growth. Temperature beyond this range would have harmful effect to nematode development [59]. *C. elegans* N2 strains were cultured and harvested worms were incubated at 15°C, 20°C and 25°C respectively. Scatter plots of pair-wise proteomics results showed that the majority of ESPs components intensities remain stable between comparisons of 15°C to 20°C and 20°C to 25°C **(Figure 2.14 and Figure 2.15)**.

Lysozyme expressions are up-regulated in *C. elegans* Excretome/secretome upon exposure to pathogens and different bacteria diets

Strikingly, almost all known *C. elegans* behaviors are affected by food [59]. Literature has shown that different food diets greatly affect and shape the metabolism of this free-

living soil nematode [59]–[62]. Additionally, more than a third of the putative secreted proteins are up-regulated upon exposure to pathogens, indicating that a substantial fraction may have a role in immune response [63]. We also performed phenotype enrichment analysis to *C. elegans* ESPs **(Figure 2.16)**. The top three categories are all very interesting. The top 1 enriched phenotype is "avoids bacteria lawn", indicating the roles of immunoregulators in response to bacteria and pathogens. The second most enriched phenotype is "dauer constitutive". Dauer stage is usually considered to be equivalent to the parasite infective larvae. Infective juveniles are the most active stage for excretory/secretory proteins to function, mediating successfully invasion and migration into host. The *C. elegans* ESPs associated with dauer phenotype would lead to interesting comparison between dauer and infective larvae. The third most enriched phenotype is "molt variant", indicating ESPs may play pivotal roles in molting process. ESPs are supposed to function at critical transition of life styles, including molting process. Earlier ES results in *Brugia malayi* showed molting larvae continuously secreting immune system regulators, including leucyl aminopeptidase (ES-62) and galectins [25]. All these evidences lead to the conclusion that the composition of ESPs is actively regulated and is adapted to the changing environment.

Thus, in order to explore the metabolism changes adapting to different bacteria diets, we cultured the *C. elegans* on the following bacteria: *E. coli* HB101*, E. coli* OP50, *Bacillus subtilis, Bacillus mycoides, Bacillus simplex, Bacillus megaterium, Microbacterium nematophilum* CBX102.

The differences between two *E. coli* diets, HB101 and OP50, were not very significant, while the majority of protein components were detected at similar level **(Figure 2.17)**.

Bacillus bacteria are common soil bacterium. While *Bacillus subtilis* has a similar nutrition content as *E. coli, Bacillus simplex* and *Bacillus megaterium* are known poor quality food to support nematode growth [59], [64]. *E. coli* is not a soil bacterium that *C. elegans* would encounter in the natural habitat. Thus, excretome/secretome profiling of *C. elegans* on soil bacteria may be a better representation of the wild status of *C. elegans* metabolism.

However, the qualities of proteomics on bacteria diets were not comparable to ESPs feeding on *E. coli*. The technical difficulties mainly come from the interference of contamination proteins. First, bacteria strains are much stickier than *E. coli* and were hard to fully be depleted and removed simply through wash steps in sample preparation. *Microbacterium nematophilum* is even a pathogen to *C. elegans*, forming bumps and attaching to the anal part of worms [65]–[68]. Thus, the number of contamination proteins of bacteria source increased significantly in the final proteomic results. Second, *C. elegans* cultured on bacteria food had a higher death rate and released cellular proteins to the culturing system. Thus, the number of non-secreted proteins was increased. The comparisons were further hindered by the complexity of excretome/secretome as well as the great variations among replicates.

We were able to detect several proteins that are highly up-regulated when challenging with poor-quality food or pathogen **(Table 2.8)**. Among these proteins, lysozymes were up-regulated by the most. *ilys-2* has a 188-fold change with feeding with

Bacillus subtilis and a 1155-fold change when feeding with *Bacillus mycoides*. To validate the transcriptional level change of lysozymes, we performed quantitative Real Time quantitative PCR (RT-qPCR) to *ilys-2* and *ilys-3*. Indeed, *ilys-3* was up-regulated by 4.45 fold when feeding on *Bacillus subtilis* and up-regulated by 3.92 fold when feeding on *Bacillus mycoides*. *ilys-2* was up-regulated by 23.88 fold when feeding on *Bacillus subtilis* and up-regulated by 35.29 fold when feeding on *Bacillus mycoides* **(Figure 2.18)**. This result strongly supports that *C. elegans* ESPs composition is under regulation and is adapted to the environment. Lysozymes as the main antimicrobial molecules are up regulated in response to defending bacteria and pathogens.

2.4 Discussion

Several proteomics studies of excreted/secreted proteins in parasitic nematodes have been published, resolving proteins from a couple hundreds to even one thousand. Several secreted proteins, like ES-62, have been proven to play pivotal roles for the infective larvae to invade host by simple infection assays [25]. Proteases from cathepsin B and Z families have been shown to be enzymatically active *in vitro* [69], [70] . All the successes in studying the parasites ESPs greatly contribute to the identification of new anti-parasite vaccines or provide candidates to small drug molecule designs. However, the successes are limited at multiple aspects. First, only a couple of parasitic nematodes have a fully sequenced and annotated genome. The excretome/secretome of many parasites, for example, bovine lungworm *Dictyocaulus viviparous*, were identified by searching against expressed sequence tag (EST). This limits the number of excreted/secreted proteins that can be identified and also introduce errors in annotating the proteins. Second, the annotation level of parasites genome is very limited. It is really hard to find annotations beyond sequences and there are very little reported expression patterns and phenotype information. In fact, many parasites excretome/secretome were annotated based on *C. elegans* BLAST homologs. Third, the functional characterizations of parasite ESPs are limited. Many techniques, including genome-editing tools including CRISPR, gene overexpression by microinjection, gene expression knockdown by RNAi, and gene knockout, are not available in parasitic nematodes studies. In proteomics papers of parasitic ESPs, many can only identify less than 100 proteins with very few annotations to the proteins [16], [71]. The

functional assays are hard to develop as well due to the difficulty to generate homozygous gene mutants. This is harmful to elucidate excreted/secreted protein functions.

C. elegans, one of the free-living nematodes, has been the first and most studied nematode for several decades. However, no one has reported the proteomics profiling of excreted/secreted proteins in *C. elegans*. People have underestimated the importance and complexity of *C. elegans* ESPs all the time. Our proteomics result is the first report to fully characterize the protein components of *C. elegans* excretome/secretome. 509 proteins were identified through nanoLC-MS/MS. This number is astonishing at first glance, since no one has expected a free-living nematode would have such a complicated excretome/secretome composition. When we think about the fact that even free-living nematodes are facing the continuously changing environments and need to respond to stimuli actively to achieve successful survival in the niche, this number is reasonable. Also, the number of identified *C. elegans* ESPs is comparable to the parasite ESPs, indicating the ESPs across nematodes may be evolutionally conserved. Our proteomic result provides the basis to allow comparison between *C. elegans* and parasitics nematodes. The knowledge learned in *C. elegans* ESPs could be transferred to predict protein function and help functional assay design and drug design in the future.

Thus, we can take advantage of the thoroughly annotated *C. elegans* genome and available Gene Ontology tools to find the major protein families in ESPs. We first prove that the majority of proteins detected in proteomics are from secretion. Most ESPs are expressed in the intestine and may be related to food digestion and stimuli response. Our result stressed the importance of lysozymes, lectins, proteases, and protease inhibitors. Lysozymes and lectins function as the immune system regulators. They are the key players

in facilitating defending bacteria/pathogens and nutrition uptake. Protease and protease inhibitors are a big group in *C. elegans* ESPs **(Figure 2.19)**. The abundance and varieties of protease species allow the processing of a wide range of substrates, ensuring the successful survival in a complex niche. The enzyme activity gains particular interest of many scientists since it opens the possibly to interrupt enzymatically activity by small molecules. Indeed, we proved proteases and proteases inhibitors are active *in vivo*. With the advanced gene manipulation techniques and the abundant mutant libraries in *C. elegans*, many functional assays can be designed and tested in the future.

After the successful establishment of standard protocol of characterizing *C. elegans* ESPs, we performed the stable isotope dimethyl labeling proteomics to allow comparison of ESPs collected under different conditions. The composition of ESPs is fast changing and highly dynamic. It is under accurate regulation and is responding to the stimuli and adapting to changing environment all the time. We have shown the major protein families remain stable when increasing the culturing temperature although there is a slightly tendency to increase expressions of genes related to defense pathway. When challenging the nematodes with different bacteria and pathogens, the lysozymes are highly expressed. We further validated this result by RT-qPCR and transgenetic promoter reporter lines. The comparative proteomics greatly narrow down the candidate gene sets to be tested in functional assays.

Other useful comparative datasets could be generated by using *C. elegans* of different stages or by separating hermaphrodites and males of *C. elegans*. This comparison would give information of ESPs associated with different life styles and genders. Since parasites have more complicated lifestyles and require transfer between host species, stage-specific

ESPs will provide insights towards the mechanisms of nematodes life stage transition. The ESPs associated with specific genders may facilitate the study of how males and hermaphrodites recognize and attract to each other in nature.

Using *C. elegans* mutants can generate useful comparative datasets as well. For example, since we have shown *C. elegans* ESPs include glycan-processing enzymes and lectins, srf mutants may be a good candidate to explore the composition changes due to the modified surface reactivity to antibody and lectins [72], [73]. Ivermectin disrupts *Brugia malayi* excretory-secretory apparatus function [74]. We could also detect whether it can disrupt *C. elegans* excretory-secretory system with proteomics. This work will allow future screening of small molecules that can interrupt excretome/secretome using our standard protocol as reference. These small molecules can be potential anthelmintic drugs.

This comparative proteomic method is limited by the detection power of Mass-Spec and the accuracy for quantitation. Using RNA-seq expression datasets would overcome this drawback of proteomics. We can benefit directly from the collection of hundreds of RNAseq expression datasets that are available in *C. elegans* without the efforts and money to repeat all conditions using Mass-Spec.

In summary, we established the standard profile of *C. elegans* ESPs and successfully identified 509 excreted/secreted proteins. Lysozymes, lectins, proteases, proteases inhibitors are the major protein groups in *C. elegans* excretome/secretome. *C. elegans* ESPs involve in many important pathways, including bacteria/pathogen defense, immune system regulation, nutrition uptake, and response to stimuli. Comparative proteomics allow quick selection of gene candidates for functional characterization. Lysozymes are up-regulated in

expression when exposing *C. elegans* to bacteria, consistent with their immuno-regulation roles.

2.5 Materials and Methods

Excreted/Secreted Proteins Sample preparation

C. elegans wild type N2 strain was cultured in large quantities in liquid culture using *E. coli* HB101 as food source [75]. Initial concentration of 3 worms/µl was used to a 250 ml S Medium system and worms were grown at 20°C for 7 days while shaking at 200 rpm. A drop of worm culture was examined under a compound microscope. Cultures with contaminating bacteria or fungal and cultures with large amounts of dead worms would be discarded. Mix-staged nematodes were then harvested by centrifuge and washed intensively six times in M9 buffer. Worms were left in 1 ml M9 buffer for incubation at 20°C for another 4 hours. Supernatant was collected and filtered by a 0.22µm syringe filter unit and treated as the unconcentrated *C. elegans* Excreted/Secreted Proteins.

nanoLC-MS/MS and data process

All excreted/secreted protein samples were digested by LysC and Trypsin enzymes after the reduction and alkylation of Cysteines. In-solution digested *C. elegans* ESPs samples were chemically labeled using dimethyl labeling strategy. Briefly digested tryptic peptides are tagged on primary amines (N-terminus and amino group of lysine) using a mixture of cyanoborohydride and formaldehyde in their unlabeled and stable isotope-labeled forms. In this study samples were labeled with regular formaldehyde and cyanoborohydride for the light label (+28) and deuterated formaldehyde and cyanoborohydride to generates a mass increase of +32 for the intermediate label. Then the samples were mixed desalted and subjected to LC-MS/MS analysis on a nanoflow LC

system, EASY-nLC II, (Thermo Fisher Scientific) coupled to a LTQ Orbitrap mass spectrometer (Thermo Fisher Scientific, Bremen, Germany) equipped with a Nanospray Flex ion source (Thermo Fisher Scientific).

For the EASY-nLC II system, solvent A consisted of 97.8% H2O, 2% ACN, and 0.2% formic acid and solvent B consisted of 19.8% H2O, 80% ACN, and 0.2% formic acid. Samples were directly loaded onto a 16-cm analytical HPLC column (75 mm ID) packed in-house with ReproSil-Pur C18AQ 3 um resin (120A° pore size, Dr. Maisch, Ammerbuch, Germany). The column was heated to 45° C. The peptides were separated with a 120 min gradient at a flow rate of 350 nL/min. The gradient was as follows: 2–30% Solvent B (110 min), 30–100% B (1 min), and 100% B (9 min). Eluted peptides were then ionized using a standard coated silica tip (New Objective, Woburn, MA) as an electrospray emitter and introduced into the mass spectrometer. The LTQ Orbitrap was operated in a data-dependent mode, automatically alternating between a full-scan (m/z 300-1700) in the Orbitrap and subsequent MS/MS scans of the 15 most abundant peaks in the linear ion trap (Top15 method). Data acquisition was controlled by Xcalibur 2.0.7 and Tune 2.4 software (Thermo Fisher Scientific).

Raw data was analyzed using MaxQuant (v. 1.5.3.30) [76], [77]. Spectra were searched against *C. elegans* entries in UniProt as well as a contaminant database containing common proteins like trypsin and keratins. Precursor mass tolerance was 4.5 ppm after recalibration and fragment tolerance was 0.5 Da. Carbamidomethylation of cysteine was specified as a fixed modification and protein N-terminal acetylation and oxidation of methionine were specified as variable modifications. Trypsin was the specified digestion enzyme and up to two missed cleavages were allowed. Score were thresholded so as to achieve a 1% FDR at

the PSM, peptide, and protein levels as estimated by a decoy database. Match-between-runs, iBAQ, and LFQ quantitation were enabled.

Bacteria strains

E. coli HB101*, E. coli* OP50, *Bacillus subtilis, Bacillus mycoides* were cultured in LB at 37°C. *Bacillus simplex, Bacillus megaterium, Microbacterium nematophilum* CBX102 were obtained from Caenorhabditis Genetics Center (CGC) and cultured at 37°C.

Stable isotope dimethyl labeling quantitative proteomics

After enzyme digestion of excreted/secreted protein samples, stable isotope dimethyl labeling protocol was followed as previously described to allow quantitative comparison for two samples [78]. The rest of the procedures were the same with regular nanoLC-MS/MS.

BLAST

Each protein sequence in the ESPs based on *C. elegans* genome release (WS 271) was blasted against NCBI using BLASTP search with the following parameters: BLAST expectation value (e-value) 1.0E-3, number of blast hits 20, HSP length cutoff 33[79] .

Bioinformatics

Gene Ontology analysis was performed with Gene Ontology Consortium powered by PANTHER[43]–[45]. WormBase enrichment analysis tools were used to find enriched tissue, phenotype, and GO terms [80], [81]. The protein sequences were further searched

against Pfam search (version 32.0) with HMMER3 [54], SignalP (version 5.0) [29], InterProScan [53], KEGG [49]–[51], STRING [55], and MEROPS [56]for proteases and proteases inhibitors, dbCAN2 [57] for automated Carbonhydrate-active enzyme annotation. Blast2GO [82], [83]was used to add more GO and InterProScan results .

BLAST map

The *C. elegans* ESPs BLAST map was created similarly to [84]. All-to-all BLAST [85]was performed with the 509 *C. elegans* ESPs identified in this work, with a threshold of E-value < 0.1. The map was created using igraph package in R [86]. Edges between nodes (proteins) were plotted as percent identity scores, scaled between 0 and 1, in a force-directed graph. Annotations for protein classes were pulled from WormBase [28] and Uniprot [87].

RNA interference

RNAi was performed by feeding nematodes dsRNA-producing bacteria using standard procedures with modifications to utse system [58]. PS6640 *qyIs[Cdh-3 mk62-63::membrane cherry]; unc-119(ed4); kuIs29[unc-119(+) + egl-13::GFP(pWH17)]* was generated in Sternberg lab. Animals were scored at young adult or L4 lethargus stage. Image was taken with Zeiss LSM 710 Inverted confocal microscope with a ×100 Plan-APOCHROMAT objective and ZEN acquisition software.

RT-qPCR

Quantitative PCR was performed as previously described using *pmp-3* as a reference gene [88]. *C. elegans* wilde type N2 worms were cultured on *E. coli* OP50*, Bacillus subtilis* and *Bacillus mycoides* lawn for 6 days. Worms were washed off plates with M9 and washed for another 5 times. The concentrations of worms were counted under compound microscope. Equal numbers of worms were used to RNA extraction steps. RNAs were prepared by Trizol extraction and purified by RNAeasy kit. Reverse transcription was carried out using SuperScript III First-Strand Synthesis System for RT-PCR from Invitrogen. 8 μL of RNA was mixed with 1 μL of $(dT)_{20}$ and 1 μL of 10mM dNTP and the mixture was incubated at 65°C for 5 min then left on ice for 10 min. 2 μL of 10X RT buffer, 4 μL of 25mM $MgCl_2$, 2 μL of 0.1M DTT, 1 μL of RNaseOUT and 1 μL of SuperScript III RT was added to 10 μL of mixture from the first step. The new mixture was incubated at 50°C for 50 min, followed by 85°C for 5 min. 1 μL of RNase H was added to the system and new mixture was incubated for 20 min at 37°C. qPCR was performed using the Roche LightCycler 480 SYBR Green I Master in the LightCycler 480 System. The cDNA was diluted by12.5 fold. 5 μL of diluted cDNA was mixed with 3 μL of PCR grade water, 2 μL of 5 μM primers, 10 μL of 2X master mix. Each qPCR reaction was performed with three technical replicates and three biological replicates. Crossing point-PCR-cycle (Cp) averages were computed for each group of three technical replicates; these values were then subtracted from the respective average Cp value of the reference gene.

Primers used in RT-qPCR:

pmp-3

5'-primer: GTTCCCGTGTTCATCACTCAT

3'-primer: ACACCGTCGAGAAGCTGTAGA

ilys-2

5’-primer: CTATTGCGGTCGCCTACGC

3’-primer: GAACATCCGCAGCAGCTGTG

ilys-3

5’-primer: GACTATTGCGGTCGCCTACG

3’-primer: GCACAGCTAAGATCATTCGCG

2.6 Figures

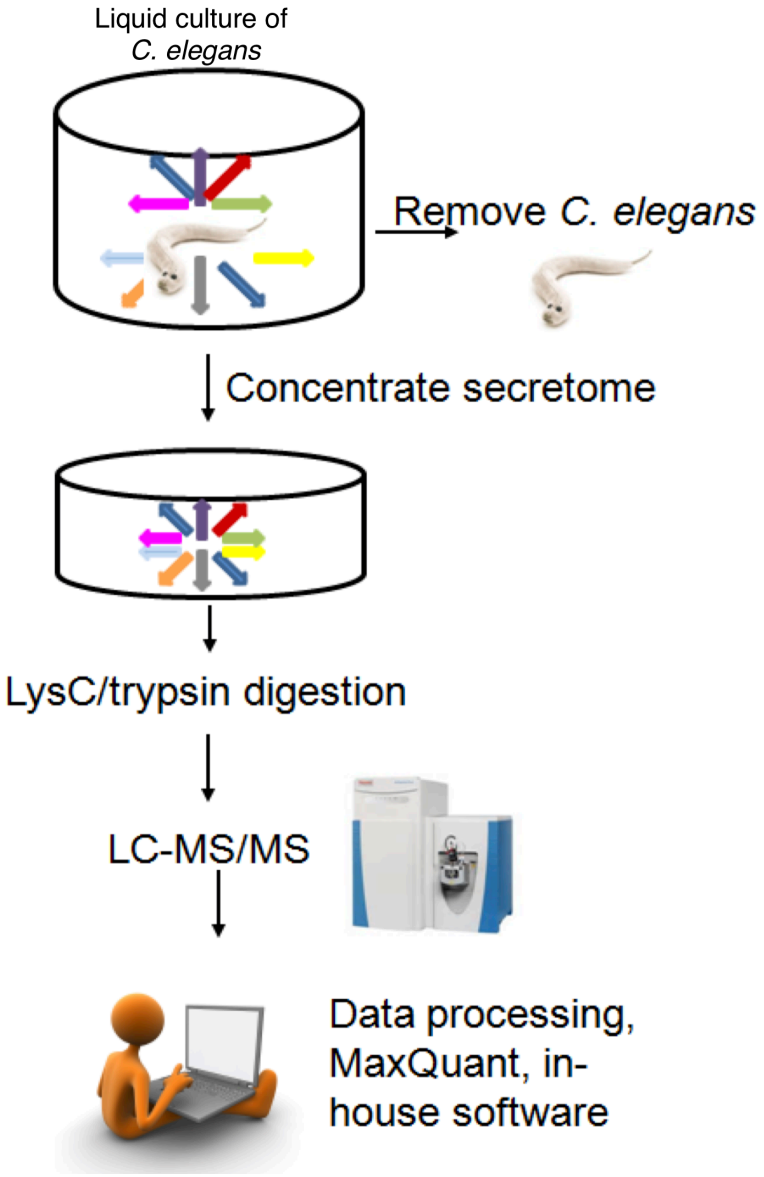


Figure 2.1 Workflow of proteomics characterization of *C. elegans* ESPs. large quantities of *C. elegans* were collected and incubated in buffer, followed by enzyme digestion and LC-MS/MS and then MaxQuant processing to identify peptides.

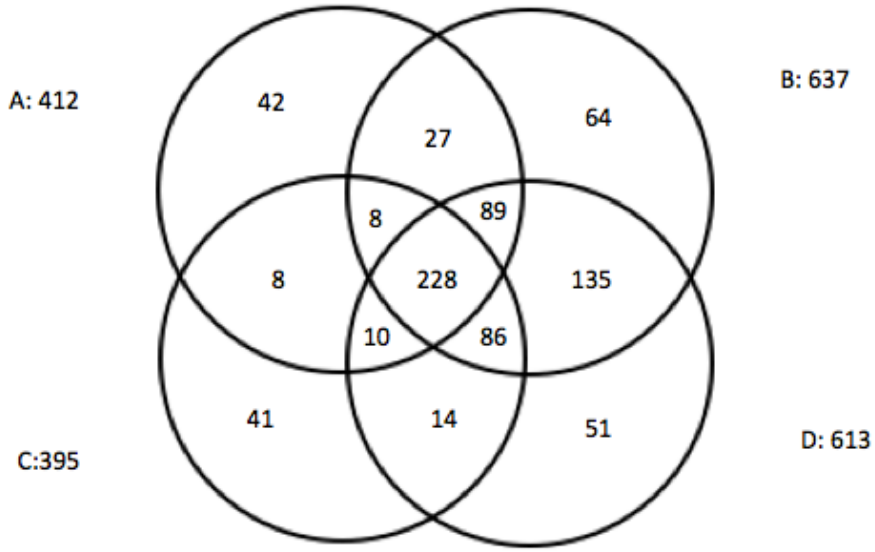


Figure 2.2 Venn diagram showing overlap of identified *C. elegans* ESPs from four biological replicates of nano LC-MS/MS. Four biological replicates were labeled with A, B, C, and D and numbers represent the count of proteins.

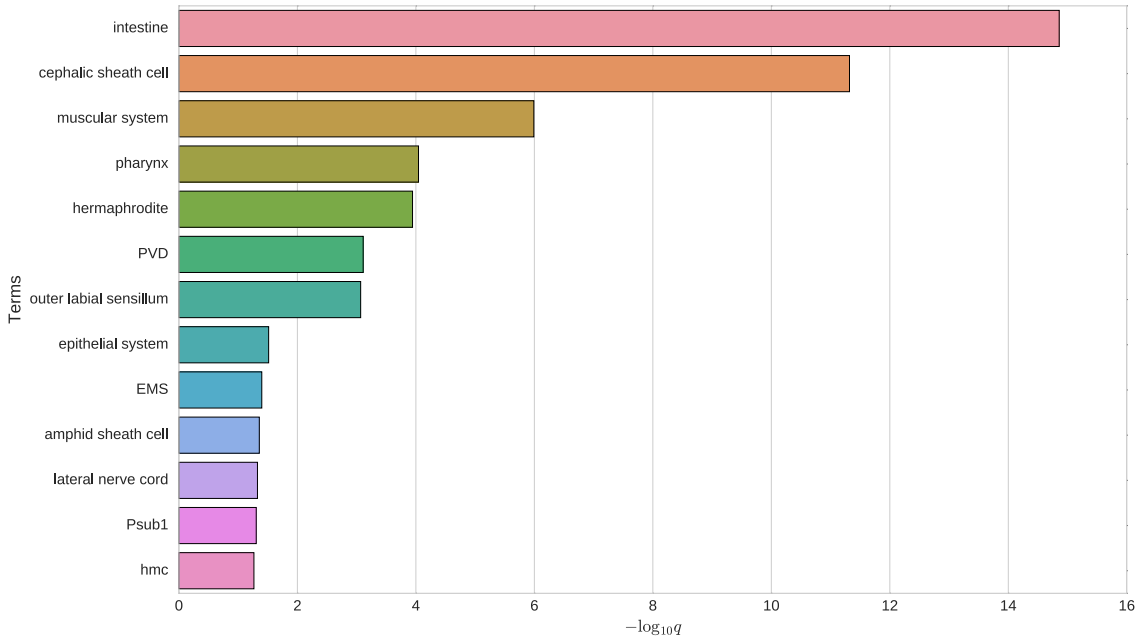


Figure 2.3 Tissue enrichment Analysis of *C. elegans* ESPs. Tissue enrichment analysis tool from WormBase was used to find out over-represented anatomical tissues.

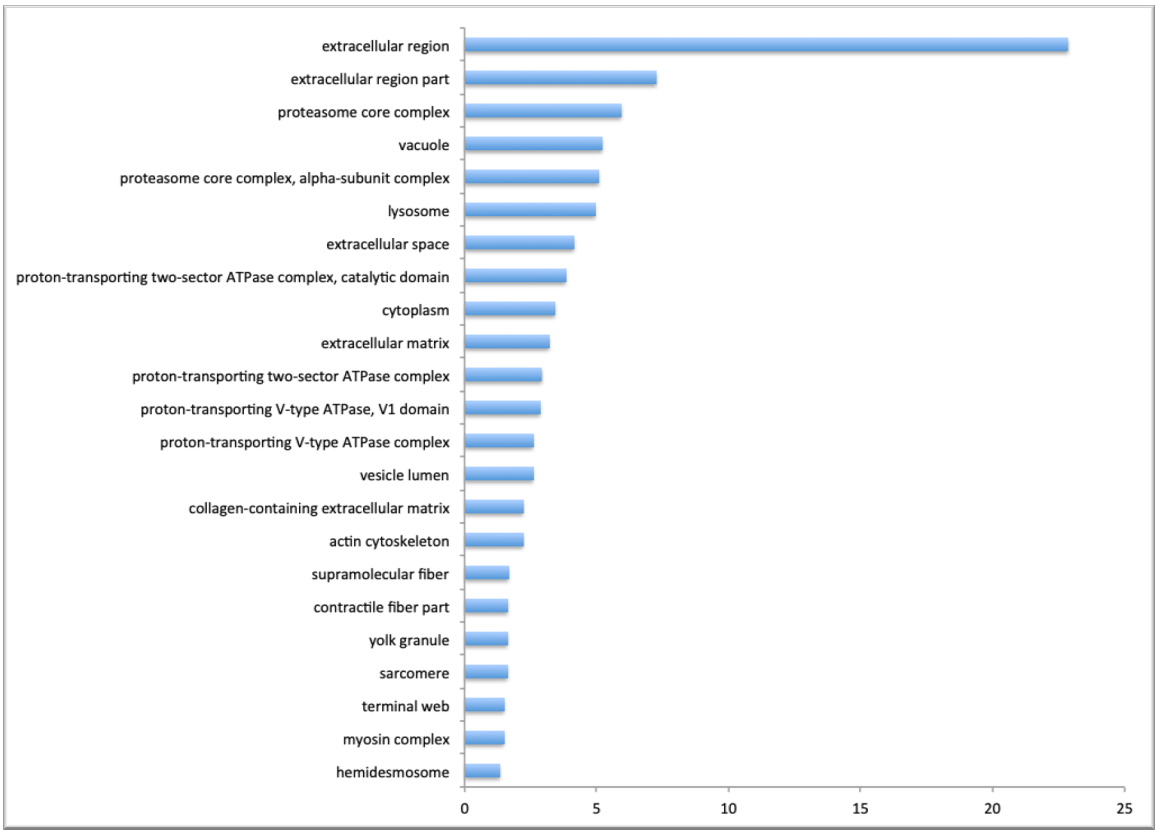


Figure 2.4 Gene Ontology enrichment analysis result on cellular component Enrichment analysis was used with a cut-off of FDR $p<0.05$. GO terms were ranked by –log (FDR). FDR, false discovery rate.

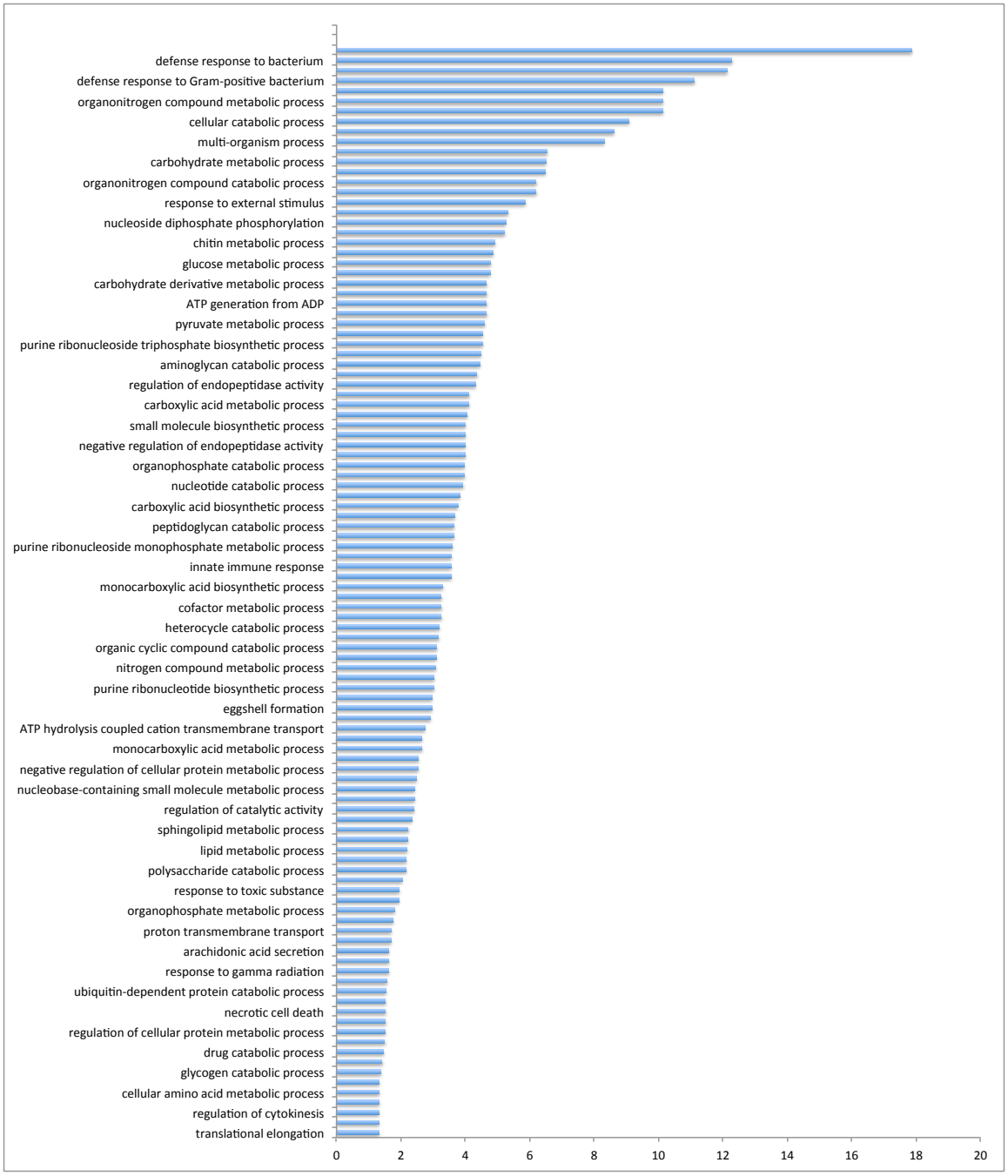


Figure 2.5 Gene Ontology enrichment analysis result on biological process. Enrichment analysis was used with a cut-off of FDR $p<0.05$. GO terms were ranked by –log (FDR). FDR, false discovery rate.

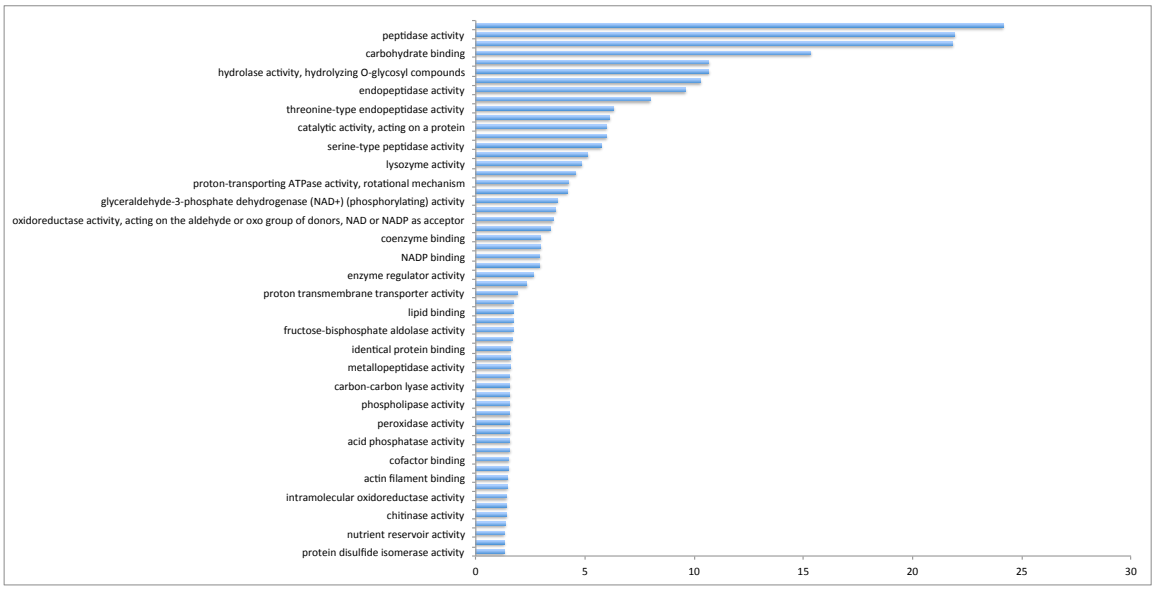


Figure 2.6 Gene Ontology enrichment analysis result on molecular function. Enrichment analysis was used with a cut-off of FDR $p<0.05$. GO terms were ranked by –log (FDR). FDR, false discovery rate.

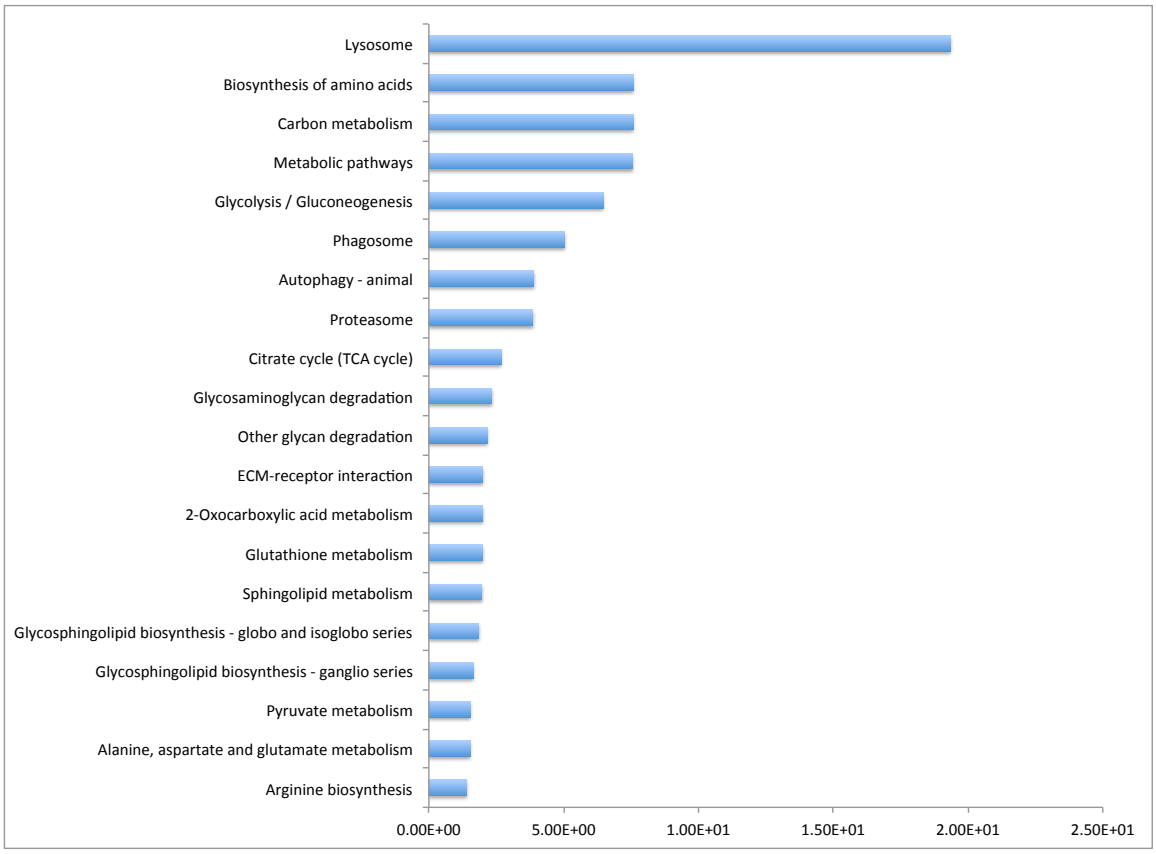


Figure 2.7 Top 20 enriched KEGG pathways in *C. elegans* ESPs. Enrichment analysis was used with a cut-off of FDR $p<0.05$. Pathways were ranked by –log (FDR). FDR, false discovery rate.

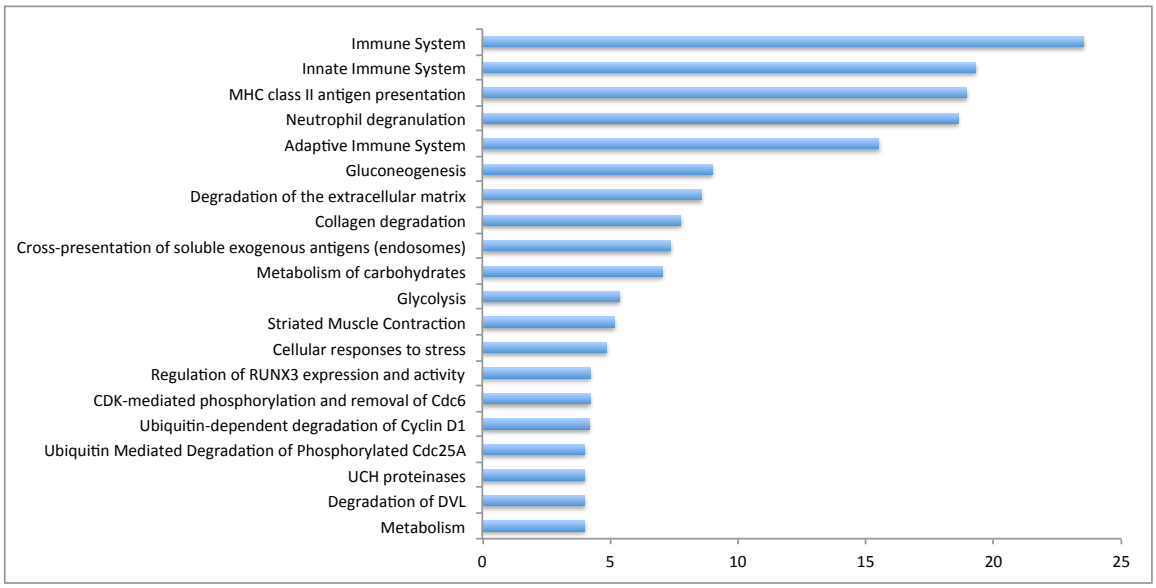


Figure 2.8 Top 20 enriched reactome pathways in *C. elegans* ESPs. enrichment analysis was used with a cut-off of FDR $p<0.05$. Pathways were ranked by –log (FDR). FDR, false discovery rate.

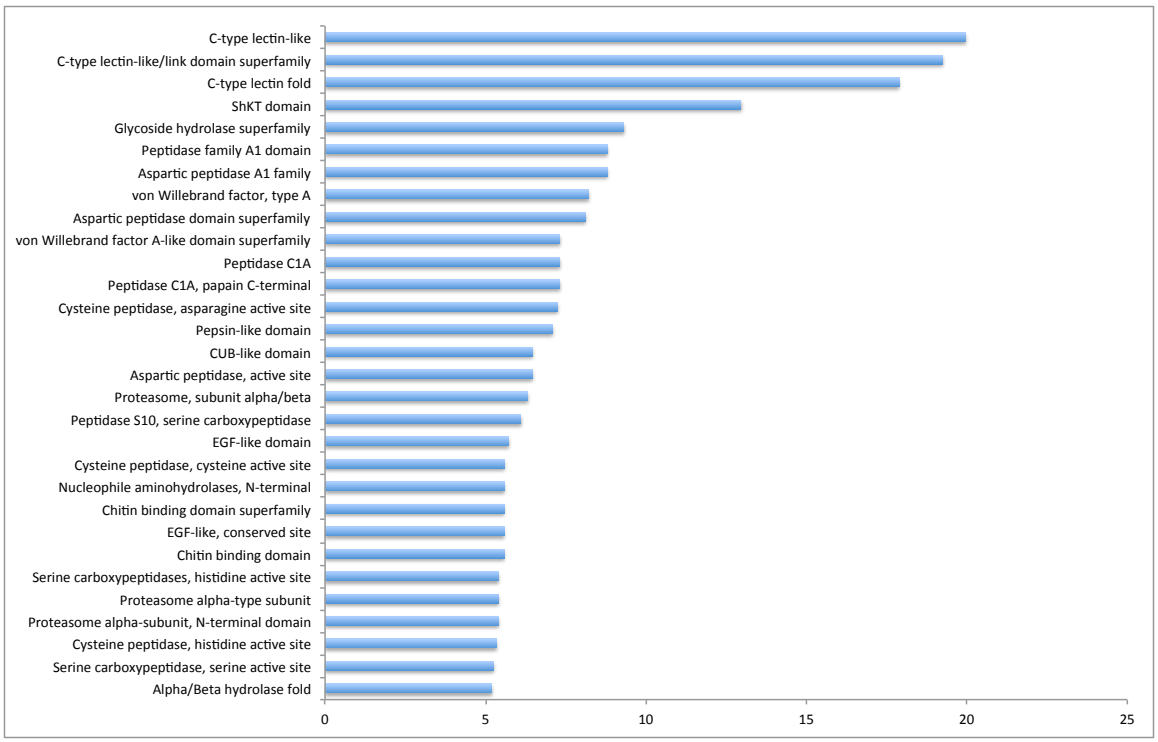


Figure 2.9 Top 30 enriched protein domain structures after searching against InterPro enrichment analysis was used with a cut-off of FDR $p<0.05$. Pathways were ranked by –log (FDR). FDR, false discovery rate.

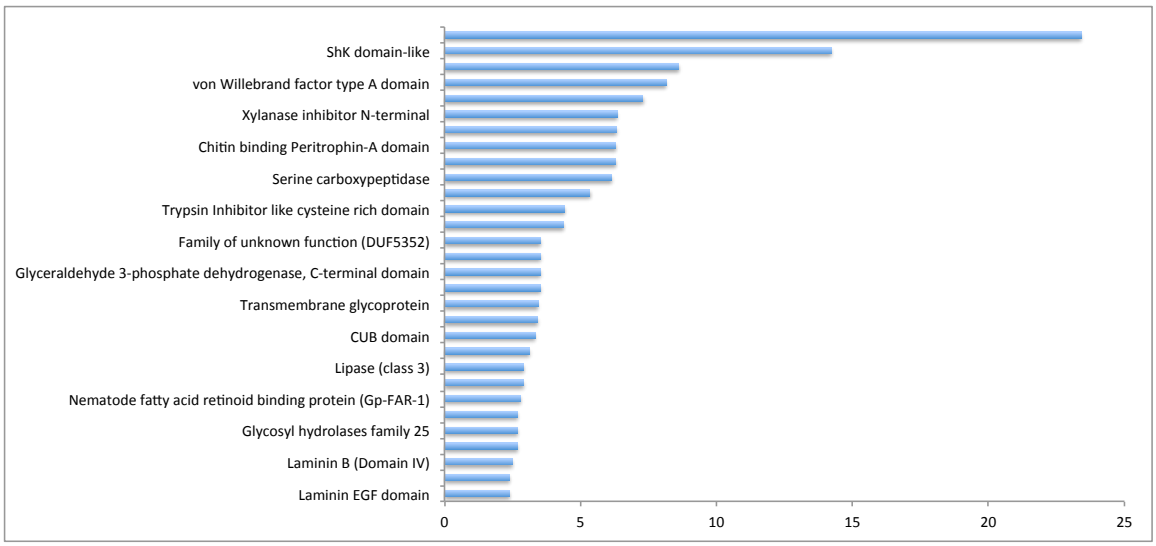


Figure 2.10 Top 15 enriched protein domain structures after searching against Pfam enrichment analysis was used with a cut-off of FDR $p<0.05$. Pathways were ranked by –log (FDR). FDR, false discovery rate.

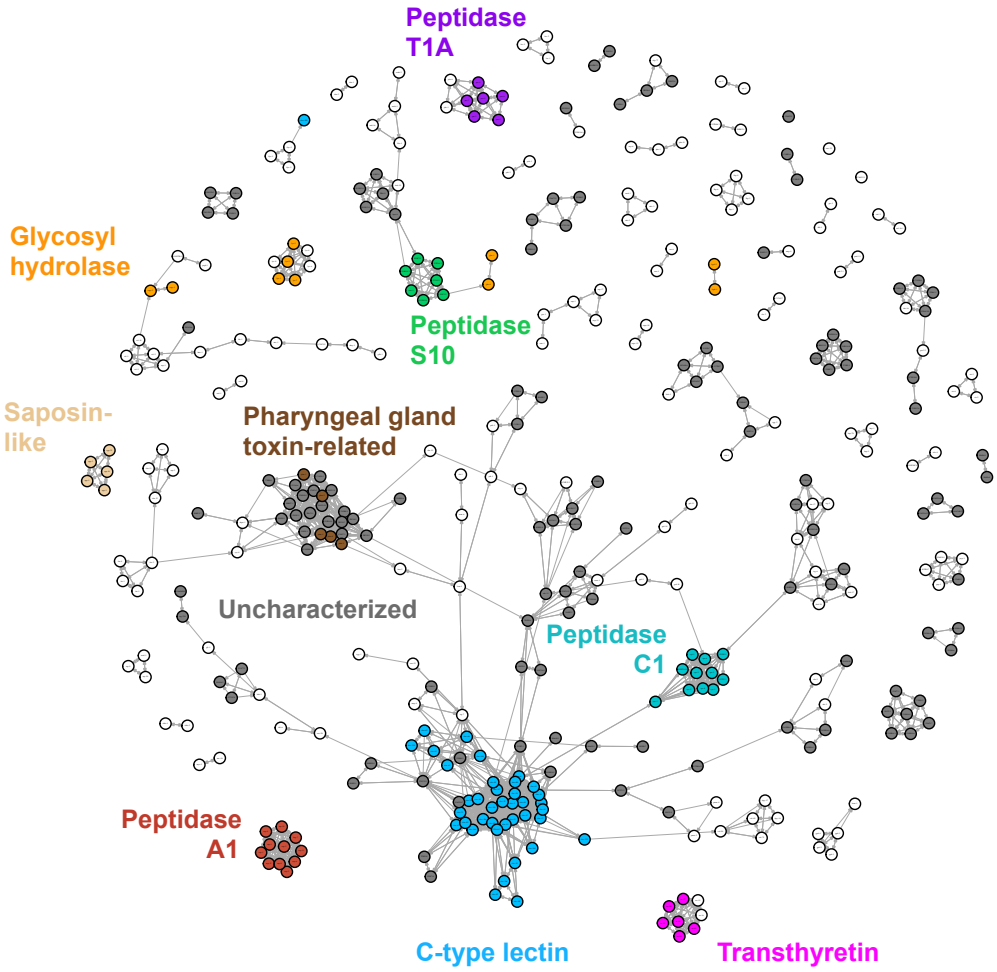


Figure 2.11 BLAST map of *C. elegans* ESPs. Edges indicate similarity by BLAST, with an E-value < 0.1. Color-coded are the ten protein classes with the highest numbers among the ESPs.

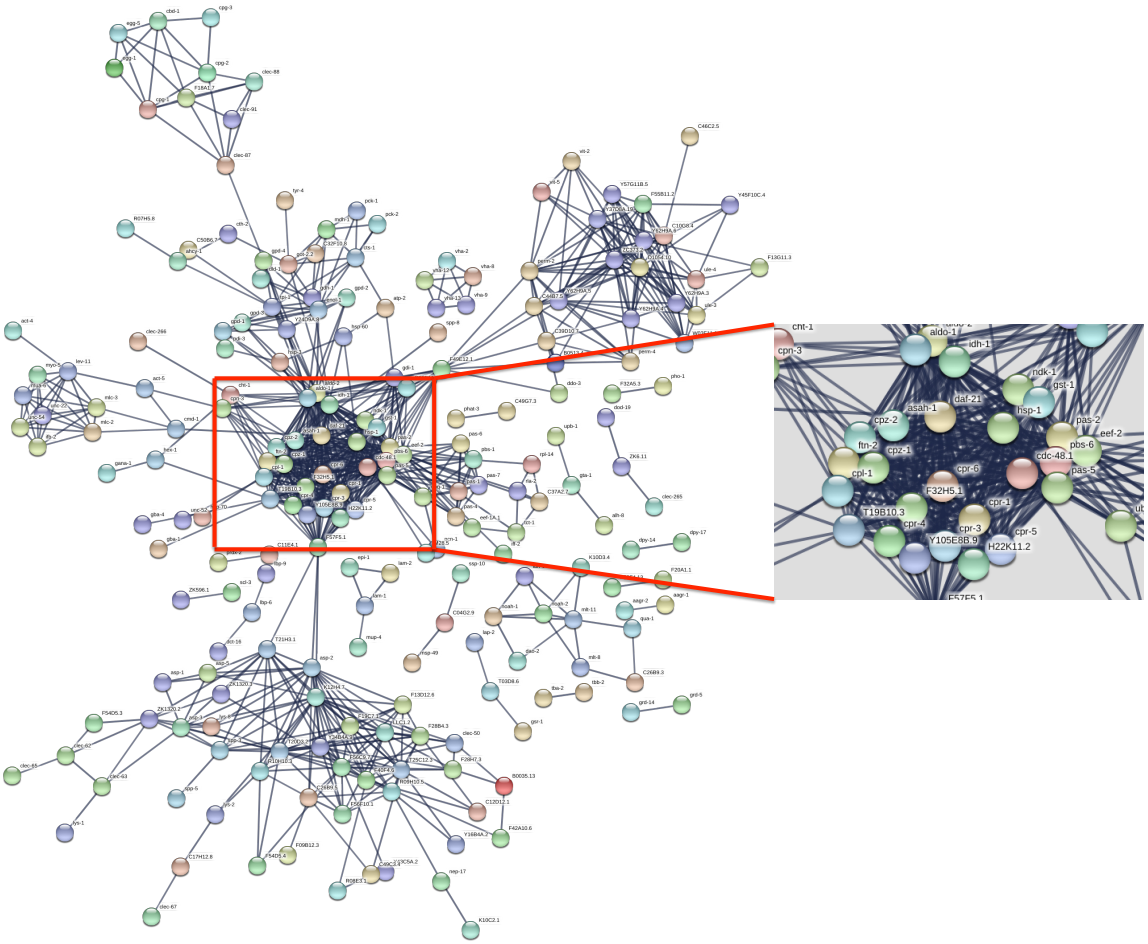


Figure 2.12 Protein-Protein interaction network of *C. elegans* ESPs visualized by STRING. Color saturation of the edges represents the confidence score of a functional association. The section in the middle was enlarged to the right, showing several proteins from Cathepsin Z family and Cathepsin B family, which are at the core position to interact with other proteins.

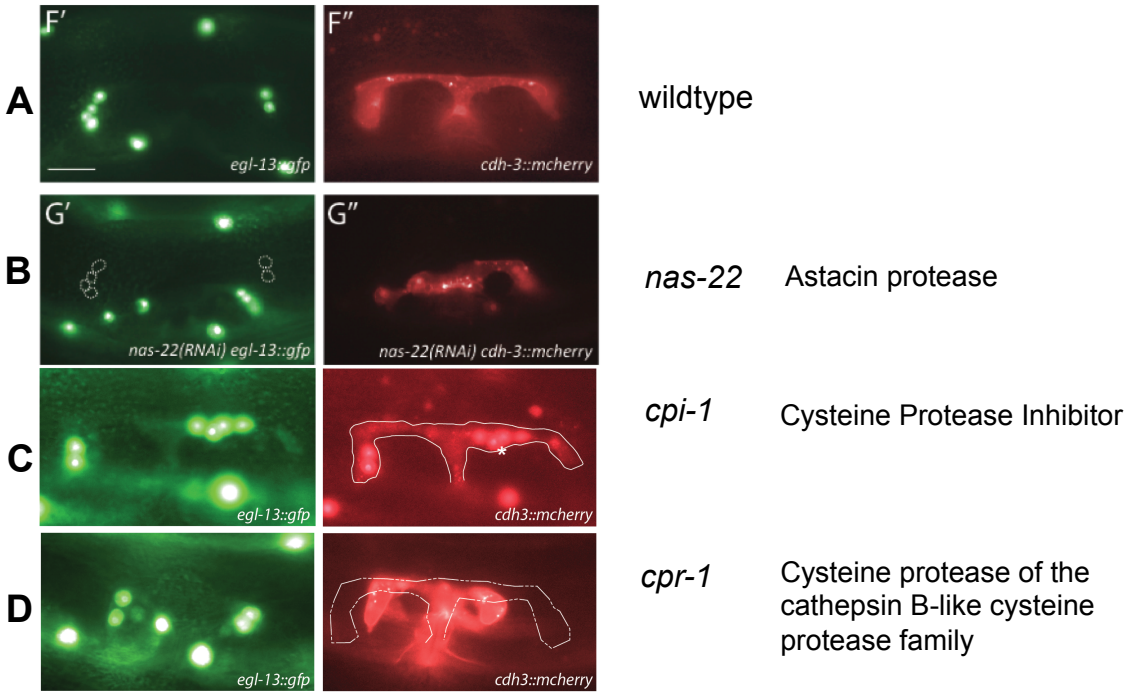


Figure 2.13 Defective in utse arm outgrowth and nuclei localization when using RNAi to knockdown expression of proteases and proteases inhibitors. (A) Left: the nuclei localization in wild type (N2) L4 lethargus worms. Right: utse cell body marked by *cdh-3::mcherry*. (B) Knocking down expression of *nas-22* gene by RNAi. Left: dotted circles indicated the nuclei postion in wild type. Right: disrupted cell body of utse. (C) Knocking down expression of *cpi-1* gene by RNAi. Left: nuclei migration was disrupted in utse. Right: cell outgrowth looked similar to wild type. (D) Knocking down expression of *cpr-1* gene by RNAi. Left: nuclei migration was disrupted in utse. Right: disrupted cell body of utse. Panel A and B are modified from [58].

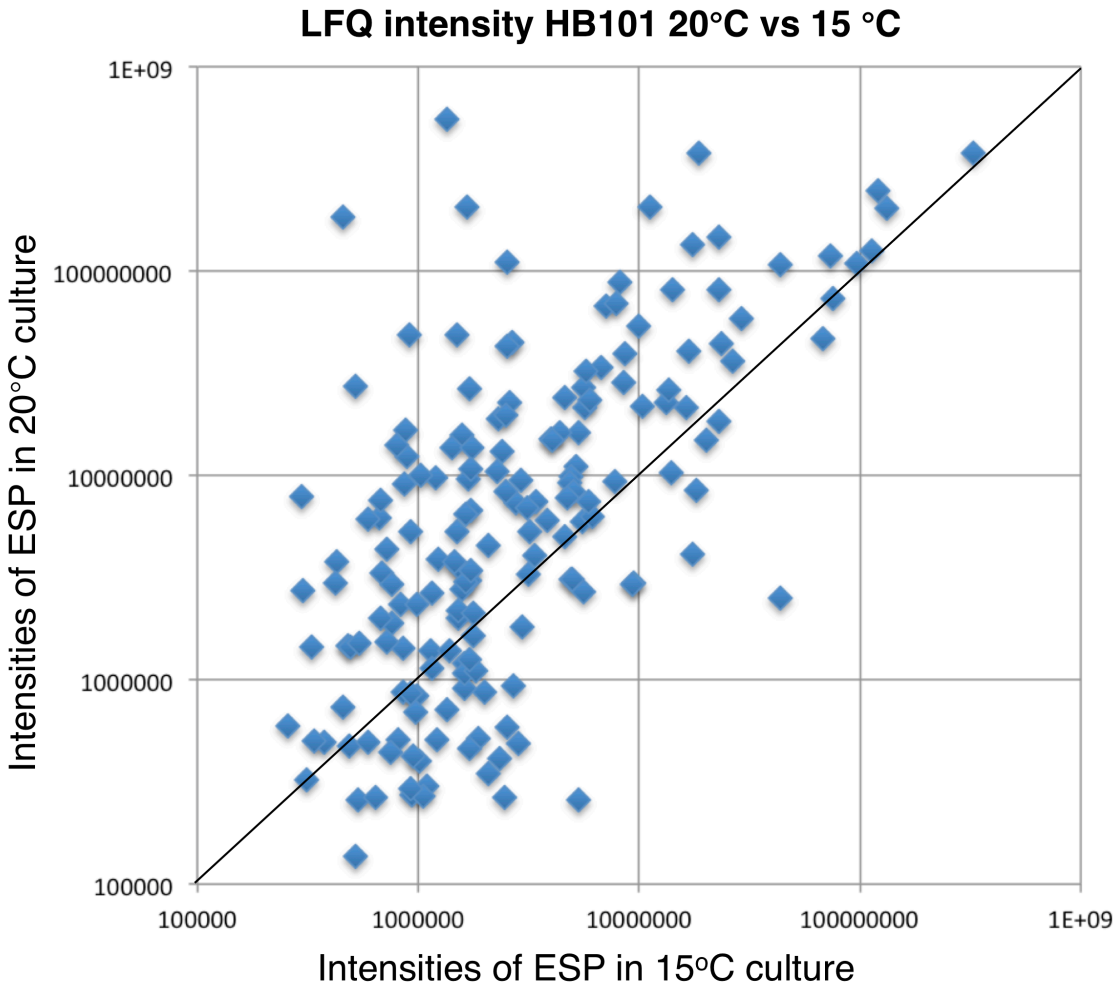


Figure 2.14 Intensity scatter plot of *C. elegans* ESPs from nematodes cultured at 20°C over *C. elegans* ESPs from nematodes cultured at 15°C. Each dot represented one excreted/secreted protein.

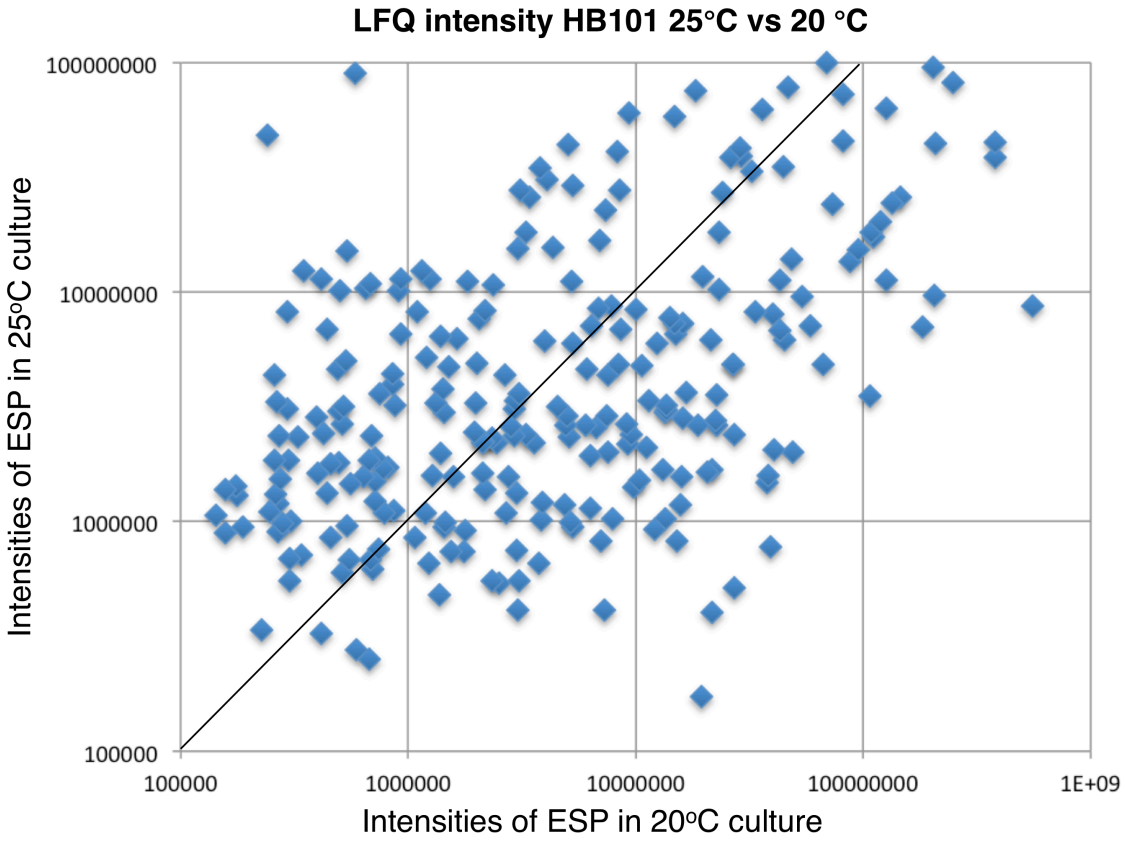


Figure 2.15 Intensity scatter plot of *C. elegans* ESPs from nematodes cultured at 25°C over *C. elegans* ESPs from nematodes cultured at 20°C. Each dot represented one excreted/secreted protein.

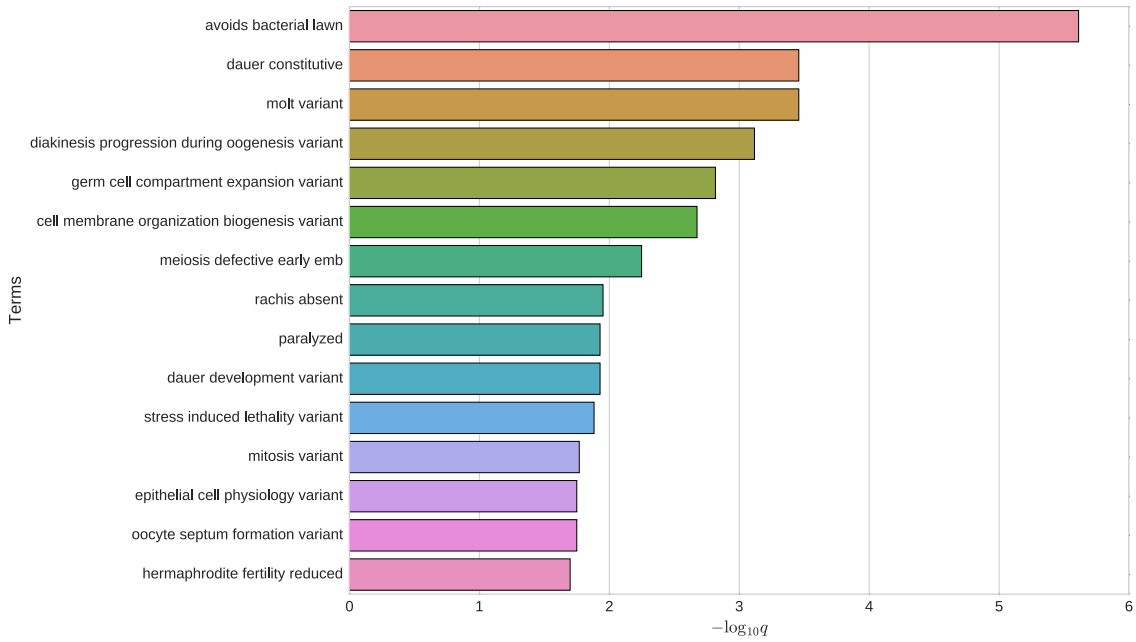


Figure 2.16 Phenotype enrichment Analysis of *C. elegans* ESPs. Phenotype enrichment analysis tool from WormBase was used to find out over-represented gene associated phenotypes.

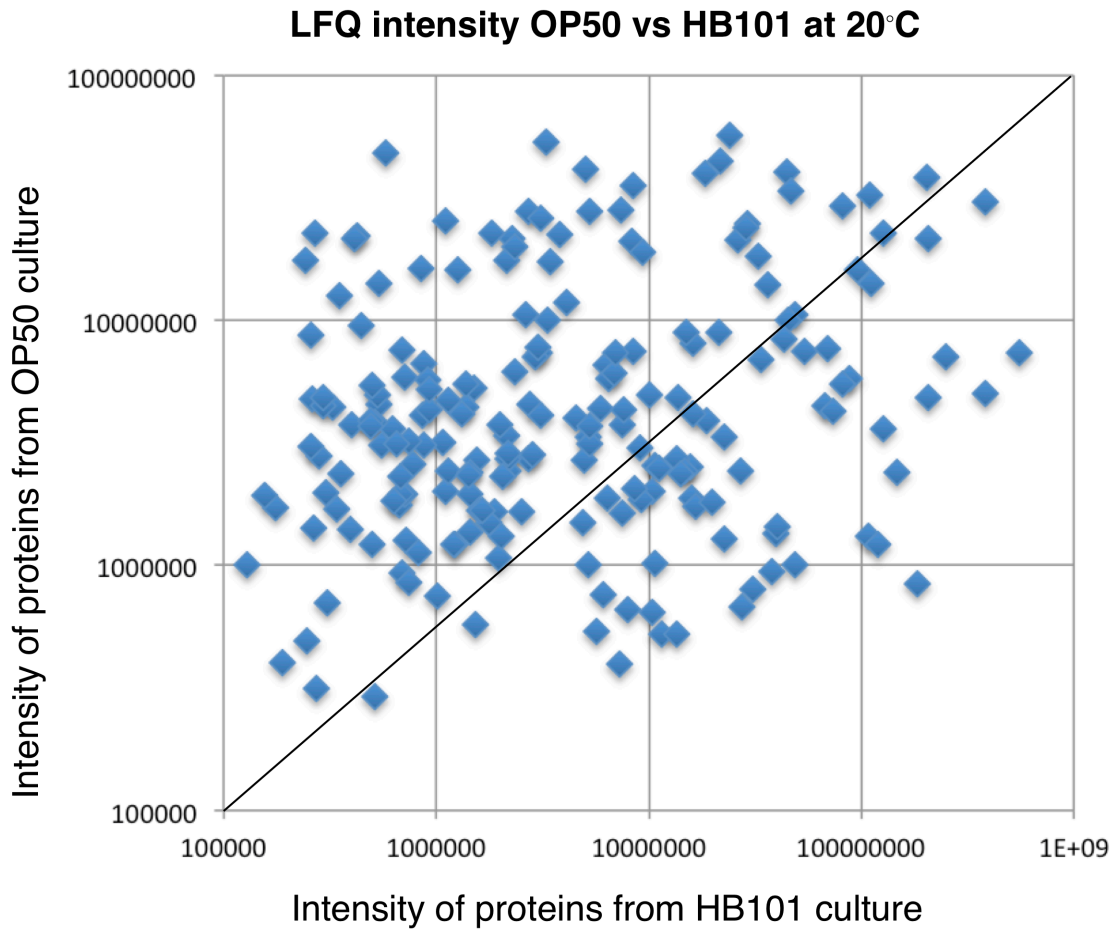


Figure 2.17 Intensity scatter plot of *C. elegans* ESPs from nematodes cultured with *E. coli* OP50 diet over ESPs with *E. coli* HB101 diet. Each dot represented one excreted/secreted protein.

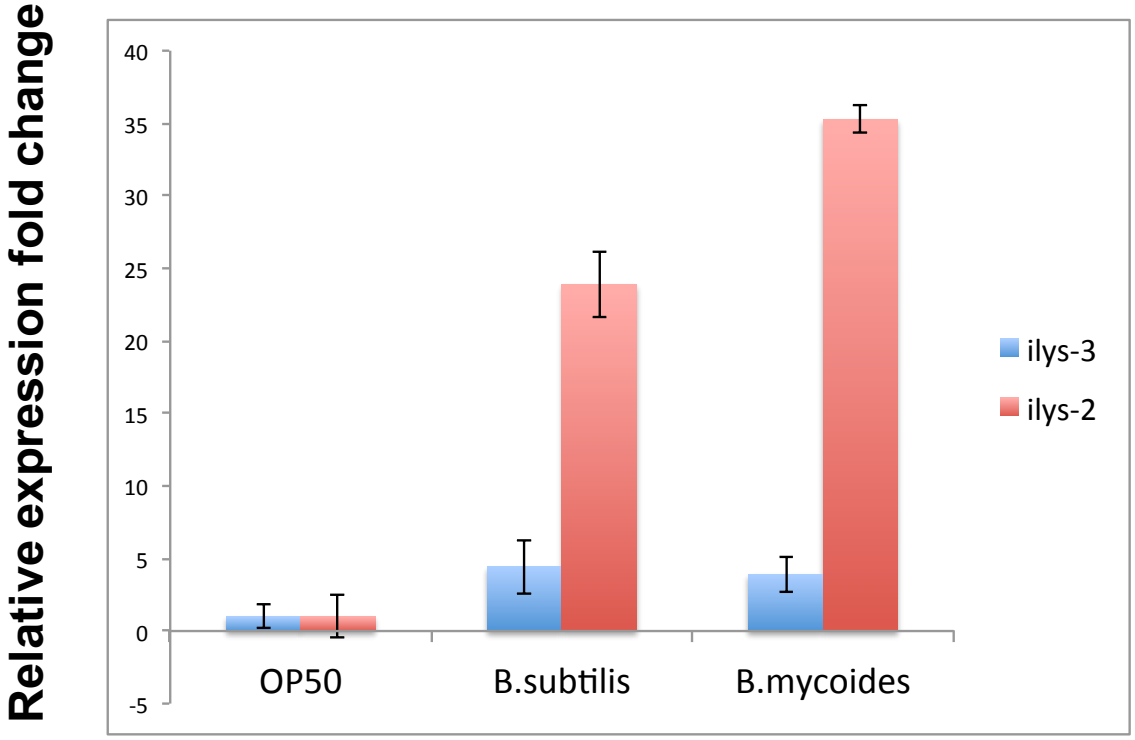


Figure 2.18 qRT-PCR results showed expression level of *ilys-2* and *ilys-3* were increased when challenging *C. elegans* with *Bacillus subtilis* and *Bacillus mycoides.* *OP50, E. coli* OP50.

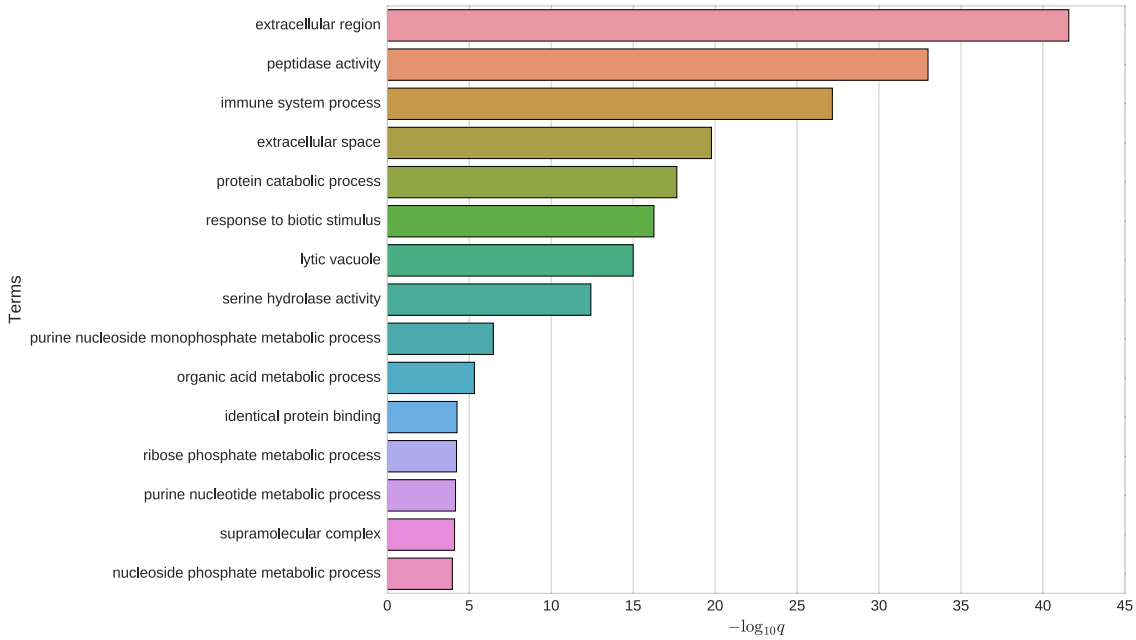


Figure 2.19 GO enrichment Analysis of *C. elegans* ESPs. GO enrichment analysis tool from WormBase was used to find over-represented anatomical GO terms.

2.7 Tables

Protein ID	Gene names	Protein names
O01530	asp-6	Aspartic protease 6
O62053	C08F11.11	UPF0375 protein C08F11.11
O45944	Y45F10C.4	UPF0375 protein Y45F10C.4
P34383	far-2	Fatty-acid and retinol-binding protein 2
O76840-3	mig-6	Papilin
Q21059	hch-1	Zinc metalloproteinase nas-34
Q9XWV2	Y37D8A.2	Putative phospholipase B-like 1
P55956	asp-3	Aspartic protease 3
Q9U3R0	C08F11.12	UPF0375 protein C08F11.12
Q9XWC2	Y73F4A.1	DOMON domain-containing protein Y73F4A.1
Q9U256	Y52B11A.8	Phospholipase A2-like protein Y52B11A.8
P18947	vit-4	Vitellogenin-4
Q18594	C44B7.5	Uncharacterized protein C44B7.5
O17861	F37H8.5	GILT-like protein F37H8.5
O62146	F09B12.3	Putative phospholipase B-like 2
H9G352	unc-52	Basement membrane proteoglycan
P55955	ttr-16	Transthyretin-like protein 16
Q22285	ttr-46	Transthyretin-like protein 46
Q22396	nas-20	Zinc metalloproteinase nas-20
Q9BL07	Y54F10AM.8	Putative phospholipase B-like 3
O77469-3	fbl-1C	Fibulin-1
Q7JLI1-2	nas-31	Zinc metalloproteinase nas-31
G8JY38	vit-2	Vitellogenin-2
P34714	ost-1	SPARC
Q03610	ZC84.1	Uncharacterized protein ZC84.1

Table 2.1 Identified *C. elegans* ESPs with reported evidence of secretion from Uniprot. Only 25 identified ESP from *C. elegans* were reported to be secreted proteins.

WormBase ID	Protein names	Gene names	Short annotation
WBGene00000219	Aspartic protease 6	asp-6	Aspartic protease 6
WBGene00017691		ilys-5	Invertebrate LYSozyme;Invertebrate LYSozyme
WBGene00000214		asp-1	ASpartyl Protease
WBGene00008358		nspc-18	Nematode Specific Peptide family, group C
WBGene00000535		cpi-1	Cystatin
WBGene00004990		spp-5	SaPosin-like Protein family
WBGene00016781			Uncharacterized protein
WBGene00015392		nspc-10	Nematode Specific Peptide family, group C
WBGene00000784	Cathepsin B-like cysteine proteinase 4	cpr-4	Cathepsin B-like cysteine proteinase 4
WBGene00009639		nspc-15	Nematode Specific Peptide family, group C
WBGene00004999		spp-14	SaPosin-like Protein family
WBGene00008572	Uncharacterized protein F08B12.4	F08B12.4	Uncharacterized protein F08B12.4
WBGene00007458	UPF0375 protein C08F11.11	C08F11.11	UPF0375 protein C08F11.11
WBGene00010204			Uncharacterized protein
WBGene00013867			Uncharacterized protein
WBGene00000218		asp-5	ASpartyl Protease
WBGene00008841			Uncharacterized protein
WBGene00004987		spp-2	SaPosin-like Protein family
WBGene00017127			Uncharacterized protein
WBGene00017881		asp-13	ASpartyl Protease
WBGene00016670		ilys-3	Invertebrate LYSozyme
WBGene00009895		scl-2	SCP-Like extracellular protein
WBGene00019682	Putative serine protease K12H4.7	K12H4.7	Putative serine protease K12H4.7
WBGene00000216	Aspartic protease 3	asp-3	Aspartic protease 3
WBGene00003096		lys-7	LYSozyme

Table 2.2 Top 25 most abundant proteins identified in *C. elegans* excretome/secretome. Several protein families, including aspartic and serine proteases, lysozymes, nematode specific peptide family, and saposin-like protein family, were the most abundant proteins in *C. elegans* ESPs.

GO cellular component complete	Caenorhabditis elegans - REFLIST (19921)	Ce_ESP(508)	Ce_ESP (expected)	Ce_ESP (over/ under)	Ce_ESP (fold Enrichment)	Ce_ESP(raw P-value)	Ce_ESP (FDR)
vesicle lumen (GO:0031983)	3	3	0.08	+	39.21	2.89E-04	9.18E-03
proteasome core complex, alpha-subunit complex (GO:0019773)	7	6	0.18	+	33.61	3.40E-07	2.56E-05
proteasome core complex (GO:0005839)	14	8	0.36	+	22.41	3.26E-08	2.62E-06
membrane microdomain (GO:0098857)	47	24	1.2	+	20.02	3.27E-21	6.59E-19
membrane region (GO:0098589)	47	24	1.2	+	20.02	3.27E-21	5.64E-19
membrane raft (GO:0045121)	47	24	1.2	+	20.02	3.27E-21	4.94E-19
proton-transporting V-type ATPase, V1 domain (GO:0033180)	8	4	0.2	+	19.61	1.60E-04	6.02E-03
proton-transporting two-sector ATPase complex, catalytic domain (GO: 0033178)	15	7	0.38	+	18.3	6.96E-07	4.67E-05
proton-transporting V-type ATPase complex (GO:0033176)	19	5	0.48	+	10.32	2.68E-04	8.74E-03
lysosome (GO:0005764)	99	22	2.52	+	8.71	2.54E-13	2.55E-11
lytic vacuole (GO:0000323)	100	22	2.55	+	8.63	3.03E-13	2.81E-11
proton-transporting two-sector ATPase complex (GO:0016469)	38	8	0.97	+	8.26	1.57E-05	9.49E-04
proteasome complex (GO:0000502)	42	8	1.07	+	7.47	2.97E-05	1.70E-03
vacuole (GO:0005773)	150	28	3.83	+	7.32	6.47E-15	7.10E-13
endopeptidase complex (GO:1905369)	44	8	1.12	+	7.13	3.98E-05	2.09E-03
collagen-containing extracellular matrix (GO:0062023)	46	7	1.17	+	5.97	3.23E-04	9.75E-03
extracellular region (GO:0005576)	575	83	14.66	+	5.66	3.31E-35	1.33E-32
peptidase complex (GO:1905368)	56	8	1.43	+	5.6	1.80E-04	6.60E-03
extracellular region part (GO:0044421)	345	48	8.8	+	5.46	6.06E-20	8.13E-18
extracellular space (GO:0005615)	306	41	7.8	+	5.25	1.14E-16	1.38E-14
extracellular matrix (GO:0031012)	95	11	2.42	+	4.54	6.68E-05	2.88E-03
myofibril (GO:0030016)	97	11	2.47	+	4.45	7.92E-05	3.30E-03
whole membrane (GO:0098805)	274	29	6.99	+	4.15	6.36E-10	5.48E-08
sarcomere (GO:0030017)	89	9	2.27	+	3.97	7.51E-04	2.11E-02
contractile fiber (GO:0043292)	160	13	4.08	+	3.19	4.04E-04	1.19E-02
supramolecular fiber (GO:0099512)	304	20	7.75	+	2.58	1.93E-04	6.84E-03
supramolecular polymer (GO:0099081)	306	20	7.8	+	2.56	2.09E-04	7.21E-03
supramolecular complex (GO:0099080)	311	20	7.93	+	2.52	2.55E-04	8.55E-03
cytoplasmic part (GO:0044444)	2572	92	65.59	+	1.4	1.04E-03	2.84E-02
Unclassified (UNCLASSIFIED)	8437	262	215.15	+	1.22	4.15E-05	2.09E-03
cellular_component (GO:0005575)	11484	246	292.85	-	0.84	4.15E-05	2.00E-03
nucleus (GO:0005634)	2472	28	63.04	-	0.44	4.46E-07	3.16E-05
plasma membrane (GO:0005886)	1350	14	34.43	-	0.41	9.68E-05	3.77E-03
membrane (GO:0016020)	6685	65	170.47	-	0.38	4.00E-26	9.65E-24
membrane part (GO:0044425)	6315	57	161.04	-	0.35	1.07E-26	3.24E-24
nuclear part (GO:0044428)	866	6	22.08	-	0.27	8.55E-05	3.44E-03
nuclear lumen (GO:0031981)	580	4	14.79	-	0.27	1.75E-03	4.70E-02
plasma membrane part (GO:0044459)	897	5	22.87	-	0.22	1.44E-05	9.15E-04
integral component of membrane (GO:0016021)	5897	29	150.38	-	0.19	7.52E-41	9.07E-38
intrinsic component of membrane (GO:0031224)	5905	29	150.58	-	0.19	7.63E-41	4.60E-38
integral component of plasma membrane (GO:0005887)	602	2	15.35	-	0.13	5.05E-05	2.34E-03
intrinsic component of plasma membrane (GO:0031226)	607	2	15.48	-	0.13	5.15E-05	2.30E-03
neuron projection (GO:0043005)	424	1	10.81	-	0.09	4.21E-04	1.21E-02
neuron part (GO:0097458)	535	1	13.64	-	0.07	3.06E-05	1.68E-03
transferase complex (GO:1990234)	345	0	8.8	-	< 0.01	3.20E-04	9.89E-03

Table 2.3 Over or under represented Cellular Component GO terms. Fisher's exact test was used in enrichment analysis. Results were shown with $p<0.05$.

GO biological process complete	Ce - REFLIST (19921)	Ce_ESP (508)	Ce_ESP (expected)	Ce_ESP (over/ under)	(fold Enrichment)	Ce_ESP(raw P-value)	Ce_ESP (FDR)
peptidoglycan metabolic process (GO:0000270)	4	4	0.1	+	39.21	2.44E-05	1.60E-03
peptidoglycan catabolic process (GO:0009253)	4	4	0.1	+	39.21	2.44E-05	1.59E-03
glycosaminoglycan catabolic process (GO:0006027)	5	4	0.13	+	31.37	4.31E-05	2.69E-03
cell wall macromolecule catabolic process (GO:0016998)	7	4	0.18	+	22.41	1.09E-04	6.09E-03
aminoglycan catabolic process (GO:0006026)	11	6	0.28	+	21.39	2.25E-06	1.85E-04
cellular response to superoxide (GO:0071451)	6	3	0.15	+	19.61	1.15E-03	4.48E-02
cellular response to oxygen radical (GO:0071450)	6	3	0.15	+	19.61	1.15E-03	4.45E-02
gland morphogenesis (GO:0022612)	18	9	0.46	+	19.61	1.07E-08	2.86E-06
gland development (GO:0048732)	18	9	0.46	+	19.61	1.07E-08	2.74E-06
pharyngeal gland morphogenesis (GO:1905905)	18	9	0.46	+	19.61	1.07E-08	2.63E-06
removal of superoxide radicals (GO:0019430)	6	3	0.15	+	19.61	1.15E-03	4.43E-02
proteasomal ubiquitin-independent protein catabolic process (GO:0010499)	17	8	0.43	+	18.45	1.03E-07	1.87E-05
ceramide catabolic process (GO:0046514)	9	4	0.23	+	17.43	2.26E-04	1.15E-02
glycosphingolipid catabolic process (GO:0046479)	9	4	0.23	+	17.43	2.26E-04	1.14E-02
cell wall macromolecule metabolic process (GO:0044036)	9	4	0.23	+	17.43	2.26E-04	1.13E-02
glycolipid catabolic process (GO:0019377)	9	4	0.23	+	17.43	2.26E-04	1.12E-02
cell wall organization or biogenesis (GO:0071554)	10	4	0.26	+	15.69	3.10E-04	1.48E-02
chitin metabolic process (GO:0006030)	21	8	0.54	+	14.94	3.75E-07	4.92E-05
glucose metabolic process (GO:0006006)	21	8	0.54	+	14.94	3.75E-07	4.82E-05
ATP generation from ADP (GO:0006757)	24	9	0.61	+	14.71	7.68E-08	1.48E-05
glycolytic process (GO:0006096)	24	9	0.61	+	14.71	7.68E-08	1.44E-05
pyruvate metabolic process (GO:0006090)	30	11	0.77	+	14.38	3.25E-09	1.05E-06
ADP metabolic process (GO:0046031)	26	9	0.66	+	13.57	1.34E-07	2.30E-05
nucleoside diphosphate phosphorylation (GO:0006165)	26	9	0.66	+	13.57	1.34E-07	2.24E-05
glucosamine-containing compound metabolic process (GO:1901071)	24	8	0.61	+	13.07	8.60E-07	9.00E-05
nucleotide phosphorylation (GO:0046939)	28	9	0.71	+	12.6	2.27E-07	3.25E-05
eggshell formation (GO:0030703)	17	5	0.43	+	11.53	1.73E-04	9.04E-03
sphingolipid catabolic process (GO:0030149)	17	5	0.43	+	11.53	1.73E-04	8.96E-03
membrane lipid catabolic process (GO:0046466)	17	5	0.43	+	11.53	1.73E-04	8.89E-03
glycosphingolipid metabolic process (GO:0006687)	14	4	0.36	+	11.2	8.76E-04	3.75E-02
cellular response to heat (GO:0034605)	14	4	0.36	+	11.2	8.76E-04	3.73E-02
purine ribonucleoside diphosphate metabolic process (GO:0009179)	32	9	0.82	+	11.03	5.84E-07	6.44E-05
purine nucleoside diphosphate metabolic process (GO:0009135)	32	9	0.82	+	11.03	5.84E-07	6.33E-05
amino sugar metabolic process (GO:0006040)	29	8	0.74	+	10.82	2.83E-06	2.30E-04
cellular carbohydrate catabolic process (GO:0044275)	15	4	0.38	+	10.46	1.09E-03	4.36E-02
defense response to Gram-positive bacterium (GO:0050830)	30	8	0.77	+	10.46	3.51E-06	2.81E-04
response to oxygen radical (GO:0000305)	15	4	0.38	+	10.46	1.09E-03	4.33E-02
response to superoxide (GO:0000303)	15	4	0.38	+	10.46	1.09E-03	4.30E-02
carbohydrate catabolic process (GO:0016052)	49	13	1.25	+	10.4	3.26E-09	1.01E-06
ribonucleoside diphosphate metabolic process (GO:0009185)	34	9	0.87	+	10.38	9.00E-07	9.25E-05
ATP hydrolysis coupled proton transport (GO:0015991)	24	6	0.61	+	9.8	8.20E-05	4.82E-03
ATP hydrolysis coupled transmembrane transport (GO:0090662)	24	6	0.61	+	9.8	8.20E-05	4.77E-03
ATP hydrolysis coupled ion transmembrane transport (GO:0099131)	24	6	0.61	+	9.8	8.20E-05	4.73E-03
nucleoside diphosphate metabolic process (GO:0009132)	37	9	0.94	+	9.54	1.65E-06	1.43E-04
hexose metabolic process (GO:0019318)	33	8	0.84	+	9.51	6.42E-06	4.89E-04
protein refolding (GO:0042026)	21	5	0.54	+	9.34	3.98E-04	1.87E-02
energy coupled proton transmembrane transport, against electrochemical gradient (GO:0015988)	27	6	0.69	+	8.71	1.44E-04	7.64E-03
monosaccharide metabolic process (GO:0005996)	37	8	0.94	+	8.48	1.33E-05	9.41E-04
innate immune response (GO:0045087)	220	45	5.61	+	8.02	1.10E-24	1.13E-21
immune response (GO:0006955)	222	45	5.66	+	7.95	1.53E-24	1.34E-21
immune system process (GO:0002376)	227	46	5.79	+	7.95	4.61E-25	5.68E-22
aminoglycan metabolic process (GO:0006022)	63	12	1.61	+	7.47	3.09E-07	4.15E-05
defense response (GO:0006952)	277	52	7.06	+	7.36	7.56E-27	2.33E-23
ceramide metabolic process (GO:0006672)	27	5	0.69	+	7.26	1.08E-03	4.34E-02
carbohydrate derivative catabolic process (GO:1901136)	65	12	1.66	+	7.24	4.16E-07	5.03E-05
defense response to Gram-negative bacterium (GO:0050829)	39	7	0.99	+	7.04	1.30E-04	7.11E-03
defense response to bacterium (GO:0042742)	69	12	1.76	+	6.82	7.32E-07	7.78E-05
cellular oxidant detoxification (GO:0098869)	58	10	1.48	+	6.76	6.60E-06	4.97E-04
ATP hydrolysis coupled cation transmembrane transport (GO:0099132)	47	8	1.2	+	6.67	6.03E-05	3.58E-03
response to bacterium (GO:0009617)	71	12	1.81	+	6.63	9.58E-07	9.69E-05
cellular detoxification (GO:1990748)	61	10	1.56	+	6.43	9.81E-06	7.04E-04
response to external biotic stimulus (GO:0043207)	80	13	2.04	+	6.37	5.05E-07	5.99E-05
response to other organism (GO:0051707)	80	13	2.04	+	6.37	5.05E-07	5.88E-05
response to biotic stimulus (GO:0009607)	80	13	2.04	+	6.37	5.05E-07	5.77E-05
negative regulation of peptidase activity (GO:0010466)	69	11	1.76	+	6.25	4.49E-06	3.51E-04
defense response to other organism (GO:0098542)	76	12	1.94	+	6.19	1.81E-06	1.51E-04
cellular response to toxic substance (GO:0097237)	64	10	1.63	+	6.13	1.43E-05	9.90E-04
detoxification (GO:0098754)	64	10	1.63	+	6.13	1.43E-05	9.79E-04
negative regulation of endopeptidase activity (GO:0010951)	58	9	1.48	+	6.09	4.02E-05	2.53E-03
negative regulation of proteolysis (GO:0045861)	71	11	1.81	+	6.08	5.74E-06	4.43E-04
regulation of endopeptidase activity (GO:0052548)	67	10	1.71	+	5.85	2.04E-05	1.35E-03
regulation of peptidase activity (GO:0052547)	82	12	2.09	+	5.74	3.69E-06	2.92E-04
organonitrogen compound catabolic process (GO:1901565)	436	62	11.12	+	5.58	5.34E-26	8.23E-23
protein catabolic process (GO:0030163)	306	41	7.8	+	5.25	1.14E-16	6.39E-14
ATP metabolic process (GO:0046034)	105	14	2.68	+	5.23	1.57E-06	1.38E-04
negative regulation of hydrolase activity (GO:0051346)	85	11	2.17	+	5.07	2.65E-05	1.70E-03
macromolecule catabolic process (GO:0009057)	413	51	10.53	+	4.84	4.14E-19	2.84E-16
organic substance catabolic process (GO:1901575)	693	81	17.67	+	4.58	1.32E-28	8.13E-25
proteolysis involved in cellular protein catabolic process (GO:0051603)	261	30	6.66	+	4.51	5.16E-11	2.27E-08
carbohydrate metabolic process (GO:0005975)	210	24	5.36	+	4.48	4.98E-09	1.46E-06
cellular protein catabolic process (GO:0044257)	265	30	6.76	+	4.44	7.24E-11	2.98E-08
generation of precursor metabolites and energy (GO:0006091)	168	19	4.28	+	4.43	2.31E-07	3.24E-05
animal organ morphogenesis (GO:0009887)	115	13	2.93	+	4.43	1.87E-05	1.27E-03
proton transmembrane transport (GO:1902600)	73	8	1.86	+	4.3	9.08E-04	3.83E-02
cellular carbohydrate metabolic process (GO:0044262)	73	8	1.86	+	4.3	9.08E-04	3.81E-02
negative regulation of catalytic activity (GO:0043086)	111	12	2.83	+	4.24	5.79E-05	3.50E-03
catabolic process (GO:0009056)	802	84	20.45	+	4.11	1.15E-26	2.37E-23
anatomical structure formation involved in morphogenesis (GO:0048646)	100	10	2.55	+	3.92	4.21E-04	1.97E-02
drug metabolic process (GO:0017144)	190	19	4.85	+	3.92	1.28E-06	1.18E-04
cellular macromolecule catabolic process (GO:0044265)	361	36	9.21	+	3.91	2.45E-11	1.16E-08
proteolysis (GO:0006508)	599	59	15.27	+	3.86	1.13E-17	6.96E-15
response to toxic substance (GO:0009636)	112	11	2.86	+	3.85	2.54E-04	1.23E-02
negative regulation of molecular function (GO:0044092)	123	12	3.14	+	3.83	1.42E-04	7.64E-03
purine-containing compound metabolic process (GO:0072521)	161	15	4.11	+	3.65	3.56E-05	2.27E-03
regulation of proteolysis (GO:0030162)	145	13	3.7	+	3.52	1.66E-04	8.77E-03
lipid catabolic process (GO:0016042)	113	10	2.88	+	3.47	1.01E-03	4.16E-02
response to stress (GO:0006950)	831	73	21.19	+	3.44	3.91E-19	3.02E-16
proteasome-mediated ubiquitin-dependent protein catabolic process (GO:0043161)	114	10	2.91	+	3.44	1.08E-03	4.37E-02
monocarboxylic acid metabolic process (GO:0032787)	151	13	3.85	+	3.38	2.41E-04	1.18E-02
cellular catabolic process (GO:0044248)	674	58	17.19	+	3.37	5.72E-15	2.94E-12
purine ribonucleotide metabolic process (GO:0009150)	143	12	3.65	+	3.29	5.12E-04	2.37E-02
purine nucleotide metabolic process (GO:0006163)	147	12	3.75	+	3.2	6.42E-04	2.89E-02
carboxylic acid metabolic process (GO:0019752)	345	28	8.8	+	3.18	2.38E-07	3.27E-05
ribonucleotide metabolic process (GO:0009259)	156	12	3.98	+	3.02	1.04E-03	4.26E-02
negative regulation of cellular protein metabolic process (GO:0032269)	170	13	4.34	+	3	6.89E-04	3.06E-02
negative regulation of protein metabolic process (GO:0051248)	171	13	4.36	+	2.98	7.25E-04	3.17E-02
ribose phosphate metabolic process (GO:0019693)	160	12	4.08	+	2.94	1.28E-03	4.89E-02
cellular homeostasis (GO:0019725)	187	14	4.77	+	2.94	5.29E-04	2.42E-02
oxoacid metabolic process (GO:0043436)	380	28	9.69	+	2.89	1.44E-06	1.30E-04
organic acid metabolic process (GO:0006082)	420	28	10.71	+	2.61	8.55E-06	6.28E-04
nucleobase-containing small molecule metabolic process (GO:0055086)	258	17	6.58	+	2.58	5.61E-04	2.54E-02
carbohydrate derivative metabolic process (GO:1901135)	476	31	12.14	+	2.55	6.83E-06	5.08E-04
multi-organism process (GO:0051704)	371	24	9.46	+	2.54	5.84E-05	3.50E-03
organonitrogen compound metabolic process (GO:1901564)	2499	115	63.73	+	1.8	5.82E-10	2.11E-07
response to stimulus (GO:0050896)	2693	100	68.67	+	1.46	1.40E-04	7.58E-03
metabolic process (GO:0008152)	4751	170	121.15	+	1.4	1.56E-06	1.40E-04
organic substance metabolic process (GO:0071704)	4276	147	109.04	+	1.35	1.02E-04	5.78E-03
biological_process (GO:0008150)	9106	287	232.21	+	1.24	1.65E-06	1.41E-04
Unclassified (UNCLASSIFIED)	10815	221	275.79	-	0.8	1.65E-06	1.39E-04
regulation of biological process (GO:0050789)	4016	73	102.41	-	0.71	1.10E-03	4.31E-02
regulation of cellular process (GO:0050794)	3639	60	92.8	-	0.65	1.12E-04	6.23E-03
cell communication (GO:0007154)	1850	26	47.18	-	0.55	7.77E-04	3.38E-02
organic cyclic compound metabolic process (GO:1901360)	1737	24	44.29	-	0.54	9.69E-04	4.01E-02
cellular aromatic compound metabolic process (GO:0006725)	1692	23	43.15	-	0.53	8.43E-04	3.64E-02
biosynthetic process (GO:0009058)	1708	23	43.56	-	0.53	6.63E-04	2.96E-02
signal transduction (GO:0007165)	1726	23	44.01	-	0.52	5.21E-04	2.40E-02
signaling (GO:0023052)	1819	24	46.39	-	0.52	2.89E-04	1.39E-02
cellular biosynthetic process (GO:0044249)	1637	21	41.74	-	0.5	3.85E-04	1.83E-02
organic substance biosynthetic process (GO:1901576)	1672	21	42.64	-	0.49	2.32E-04	1.15E-02
cellular nitrogen compound biosynthetic process (GO:0044271)	1181	13	30.12	-	0.43	7.23E-04	3.19E-02
heterocycle biosynthetic process (GO:0018130)	909	7	23.18	-	0.3	1.28E-04	7.05E-03
organic cyclic compound biosynthetic process (GO:1901362)	939	7	23.95	-	0.29	9.59E-05	5.48E-03
aromatic compound biosynthetic process (GO:0019438)	903	6	23.03	-	0.26	4.46E-05	2.75E-03
cellular macromolecule biosynthetic process (GO:0034645)	1224	8	31.21	-	0.26	1.07E-06	1.02E-04
gene expression (GO:0010467)	1382	9	35.24	-	0.26	1.61E-07	2.49E-05
macromolecule biosynthetic process (GO:0009059)	1232	8	31.42	-	0.25	1.10E-06	1.03E-04
protein modification process (GO:0036211)	1271	8	32.41	-	0.25	3.90E-07	4.91E-05
cellular protein modification process (GO:0006464)	1271	8	32.41	-	0.25	3.90E-07	4.81E-05
regulation of gene expression (GO:0010468)	1335	8	34.04	-	0.23	1.47E-07	2.39E-05
macromolecule modification (GO:0043412)	1379	8	35.17	-	0.23	5.04E-08	1.07E-05
nucleobase-containing compound biosynthetic process (GO:0034654)	873	5	22.26	-	0.22	1.97E-05	1.32E-03
regulation of biosynthetic process (GO:0009889)	1123	6	28.64	-	0.21	5.61E-07	6.29E-05
regulation of cellular macromolecule biosynthetic process (GO:2000112)	1092	5	27.85	-	0.18	2.18E-07	3.20E-05
regulation of macromolecule biosynthetic process (GO:0010556)	1107	5	28.23	-	0.18	1.51E-07	2.39E-05
regulation of cellular biosynthetic process (GO:0031326)	1118	5	28.51	-	0.18	1.04E-07	1.83E-05
system process (GO:0003008)	690	3	17.6	-	0.17	5.12E-05	3.13E-03
nucleic acid metabolic process (GO:0090304)	1385	6	35.32	-	0.17	1.83E-09	6.26E-07
regulation of transcription, DNA-templated (GO:0006355)	977	3	24.91	-	0.12	7.32E-08	1.46E-05
regulation of RNA biosynthetic process (GO:2001141)	991	3	25.27	-	0.12	4.98E-08	1.14E-05
regulation of nucleic acid-templated transcription (GO:1903506)	991	3	25.27	-	0.12	4.98E-08	1.10E-05
regulation of RNA metabolic process (GO:0051252)	1060	3	27.03	-	0.11	1.14E-08	2.72E-06
regulation of nucleobase-containing compound metabolic process (GO:0019219)	1088	3	27.74	-	0.11	5.20E-09	1.46E-06
RNA metabolic process (GO:0016070)	1178	3	30.04	-	0.1	5.09E-10	1.96E-07
RNA processing (GO:0006396)	401	1	10.23	-	0.1	9.19E-04	3.83E-02
transcription, DNA-templated (GO:0006351)	674	1	17.19	-	0.06	9.81E-07	9.76E-05
nucleic acid-templated transcription (GO:0097659)	678	1	17.29	-	0.06	9.91E-07	9.71E-05
RNA biosynthetic process (GO:0032774)	684	1	17.44	-	0.06	1.03E-06	9.97E-05
G protein-coupled receptor signaling pathway (GO:0007186)	787	1	20.07	-	0.05	6.55E-08	1.35E-05
nervous system process (GO:0050877)	635	0	16.19	-	< 0.01	1.68E-07	2.53E-05
sensory perception of chemical stimulus (GO:0007606)	465	0	11.86	-	< 0.01	1.33E-05	9.34E-04
sensory perception (GO:0007600)	487	0	12.42	-	< 0.01	9.50E-06	6.89E-04

Table 2.4 Over or under represented Biological Process GO terms. Fisher's exact test was used in enrichment analysis. Results are shown with $p<0.05$.

GO molecular function complete	Ce - REFLIST (19921)	Ce_ESP (508)	Ce_ESP(expected)	Ce_ESP (over/under)	(fold Enrichment)	Ce_ESP (raw P-value)	Ce_ESP (FDR)
glyceraldehyde-3-phosphate dehydrogenase (NAD+) (phosphorylating) activity (GO:0004365)	4	4	0.1	+	39.21	2.44E-05	1.57E-03
glyceraldehyde-3-phosphate dehydrogenase (NAD(P)+) (phosphorylating) activity (GO:0043891)	4	4	0.1	+	39.21	2.44E-05	1.53E-03
galactosidase activity (GO:0015925)	3	3	0.08	+	39.21	2.89E-04	1.32E-02
serine-type carboxypeptidase activity (GO:0004185)	9	7	0.23	+	30.5	5.31E-08	8.51E-06
serine-type exopeptidase activity (GO:0070008)	9	7	0.23	+	30.5	5.31E-08	8.01E-06
lysozyme activity (GO:0003796)	10	7	0.26	+	27.45	8.84E-08	1.26E-05
peptidoglycan muralytic activity (GO:0061783)	10	7	0.26	+	27.45	8.84E-08	1.19E-05
threonine-type endopeptidase activity (GO:0004298)	15	8	0.38	+	20.91	4.88E-08	8.94E-06
threonine-type peptidase activity (GO:0070003)	15	8	0.38	+	20.91	4.88E-08	8.35E-06
nutrient reservoir activity (GO:0045735)	6	3	0.15	+	19.61	1.15E-03	4.53E-02
aspartic-type endopeptidase activity (GO:0004190)	25	11	0.64	+	17.25	6.91E-10	1.61E-07
aspartic-type peptidase activity (GO:0070001)	25	11	0.64	+	17.25	6.91E-10	1.48E-07
dipeptidyl-peptidase activity (GO:0008239)	12	5	0.31	+	16.34	4.50E-05	2.62E-03
proton-transporting ATPase activity, rotational mechanism (GO:0046961)	18	6	0.46	+	13.07	2.11E-05	1.42E-03
hydrolase activity, hydrolyzing O-glycosyl compounds (GO:0004553)	64	19	1.63	+	11.64	1.38E-13	5.90E-11
proton-exporting ATPase activity (GO:0036442)	21	6	0.54	+	11.2	4.36E-05	2.60E-03
carboxypeptidase activity (GO:0004180)	28	8	0.71	+	11.2	2.27E-06	2.15E-04
acid phosphatase activity (GO:0003993)	14	4	0.36	+	11.2	8.76E-04	3.56E-02
hydrolase activity, acting on glycosyl bonds (GO:0016798)	75	20	1.91	+	10.46	1.72E-13	6.29E-11
ATPase activity, coupled to transmembrane movement of ions, rotational mechanism (GO:0044769)	31	8	0.79	+	10.12	4.32E-06	3.35E-04
oxidoreductase activity, acting on the aldehyde or oxo group of donors, NAD or NADP as acceptor (GO:0016620)	20	5	0.51	+	9.8	3.28E-04	1.48E-02
exopeptidase activity (GO:0008238)	75	17	1.91	+	8.89	1.03E-10	2.93E-08
NADP binding (GO:0050661)	24	5	0.61	+	8.17	6.76E-04	2.89E-02
aminopeptidase activity (GO:0004177)	39	8	0.99	+	8.04	1.85E-05	1.28E-03
cysteine-type endopeptidase activity (GO:0004197)	60	12	1.53	+	7.84	1.95E-07	2.49E-05
serine hydrolase activity (GO:0017171)	62	12	1.58	+	7.59	2.66E-07	3.10E-05
serine-type peptidase activity (GO:0008236)	62	12	1.58	+	7.59	2.66E-07	2.96E-05
endopeptidase inhibitor activity (GO:0004866)	59	11	1.5	+	7.31	1.16E-06	1.24E-04
oxidoreductase activity, acting on the aldehyde or oxo group of donors (GO:0016903)	27	5	0.69	+	7.26	1.08E-03	4.32E-02
endopeptidase regulator activity (GO:0061135)	62	11	1.58	+	6.96	1.78E-06	1.83E-04
active ion transmembrane transporter activity (GO:0022853)	46	8	1.17	+	6.82	5.27E-05	3.00E-03
cation-transporting ATPase activity (GO:0019829)	46	8	1.17	+	6.82	5.27E-05	2.93E-03
ATPase coupled ion transmembrane transporter activity (GO:0042625)	46	8	1.17	+	6.82	5.27E-05	2.87E-03
serine-type endopeptidase inhibitor activity (GO:0004867)	47	8	1.2	+	6.67	6.03E-05	3.22E-03
peptidase inhibitor activity (GO:0030414)	66	11	1.68	+	6.54	3.06E-06	2.45E-04
chitin binding (GO:0008061)	43	7	1.1	+	6.38	2.23E-04	1.08E-02
endopeptidase activity (GO:0004175)	243	38	6.2	+	6.13	1.66E-17	8.50E-15
peptidase regulator activity (GO:0061134)	74	11	1.89	+	5.83	8.18E-06	6.17E-04
peptidase activity, acting on L-amino acid peptides (GO:0070011)	375	55	9.56	+	5.75	1.04E-23	8.93E-21
carbohydrate binding (GO:0030246)	309	45	7.88	+	5.71	1.89E-19	1.21E-16
peptidase activity (GO:0008233)	397	57	10.12	+	5.63	3.91E-24	5.01E-21
antioxidant activity (GO:0016209)	49	7	1.25	+	5.6	4.56E-04	1.98E-02
cysteine-type peptidase activity (GO:0008234)	100	12	2.55	+	4.71	2.28E-05	1.50E-03
proton transmembrane transporter activity (GO:0015078)	70	8	1.79	+	4.48	7.06E-04	2.92E-02
enzyme inhibitor activity (GO:0004857)	97	11	2.47	+	4.45	7.92E-05	3.98E-03
identical protein binding (GO:0042802)	124	12	3.16	+	3.79	1.53E-04	7.53E-03
coenzyme binding (GO:0050662)	203	15	5.18	+	2.9	3.86E-04	1.70E-02
hydrolase activity (GO:0016787)	1688	124	43.05	+	2.88	3.84E-26	9.84E-23
catalytic activity, acting on a protein (GO:0140096)	1389	63	35.42	+	1.78	1.51E-05	1.11E-03
catalytic activity (GO:0003824)	4106	171	104.71	+	1.63	1.52E-11	4.88E-09
molecular_function (GO:0003674)	9293	291	236.98	+	1.23	2.64E-06	2.41E-04
Unclassified (UNCLASSIFIED)	10628	217	271.02	-	0.8	2.64E-06	2.33E-04
heterocyclic compound binding (GO:1901363)	3004	48	76.6	-	0.63	2.48E-04	1.18E-02
organic cyclic compound binding (GO:0097159)	3015	48	76.88	-	0.62	2.48E-04	1.16E-02
transferase activity (GO:0016740)	1629	17	41.54	-	0.41	1.55E-05	1.10E-03
signaling receptor activity (GO:0038023)	896	6	22.85	-	0.26	6.42E-05	3.36E-03
molecular transducer activity (GO:0060089)	917	6	23.38	-	0.26	3.21E-05	1.96E-03
transmembrane signaling receptor activity (GO:0004888)	807	5	20.58	-	0.24	7.55E-05	3.87E-03
nucleic acid binding (GO:0003676)	1687	9	43.02	-	0.21	2.93E-10	7.50E-08
DNA binding (GO:0003677)	945	2	24.1	-	0.08	2.08E-08	4.09E-06
G protein-coupled receptor activity (GO:0004930)	538	0	13.72	-	< 0.01	1.81E-06	1.78E-04
sequence-specific DNA binding (GO:0043565)	527	0	13.44	-	< 0.01	2.74E-06	2.26E-04
DNA-binding transcription factor activity (GO:0003700)	521	0	13.29	-	< 0.01	2.69E-06	2.30E-04
ion channel activity (GO:0005216)	311	0	7.93	-	< 0.01	6.89E-04	2.89E-02
transcription regulator activity (GO:0140110)	613	0	15.63	-	< 0.01	2.46E-07	3.01E-05

Table 2.5 Over or under represented Molecular Function GO terms. Fisher's exact test was used in enrichment analysis. Results are shown with $p<0.05$.

Gene_ID	HMMER	Hotpep	DIAMOND	Signalp	#ofTools
sp|O16580|GLCM1_CAEEL	GH30_1(86-517)	N	GH30_1	Y(1-24)	2
sp|Q11174|CHIT_CAEEL	GH18(57-403)	GH18	CBM14	N	3
sp|Q20967|LYS5_CAEEL	GH25(22-202)	N	GH25	Y(1-18)	2
sp|Q22492|HEXA_CAEEL	GH20(166-504)	GH20	GH20	N	3
sp|Q9UB00|GLCM4_CAEEL	GH30_1(84-518)	GH30	GH30_1	N	3
tr|O76632|O76632_CAEEL	GH35(35-353)	GH35	GH35	N	3
tr|P91982|P91982_CAEEL	GH13_15(57-355)	GH13	GH13_15	N	3
tr|Q17816|Q17816_CAEEL	GH19(60-361)	GH19	GH19	Y(1-24)	3
tr|Q19004|Q19004_CAEEL	GH31(281-783)	GH31	GH31	Y(1-25)	3
tr|Q19874|Q19874_CAEEL	GH152(26-233)	N	GH152	Y(1-20)	2
tr|Q19876|Q19876_CAEEL	GH152(26-233)	N	GH152	Y(1-20)	2
tr|Q20964|Q20964_CAEEL	GH25(32-200)	N	N	Y(1-17)	1
tr|Q20968|Q20968_CAEEL	GH25(32-199)	N	GH25	N	2
tr|Q21750|Q21750_CAEEL	GH31(268-779)	GH31	GH31	Y(1-19)	3
tr|Q21801|Q21801_CAEEL	GH27(122-310)	GH27	GH27	N	3
tr|Q27526|Q27526_CAEEL	GH35(51-369)	GH35	GH35	N	3
tr|Q95Q32|Q95Q32_CAEEL	GH2(25-609)	GH2	GH2	N	3
tr|O76358|O76358_CAEEL	N	GH22	N	Y(1-19)	1
tr|A0A0K3AYJ1|A0A0K3AYJ1_CAEEL	N	GH22	N	N	1
tr|O76357|O76357_CAEEL	N	GH22	N	Y(1-19)	1
tr|O45599|O45599_CAEEL	N	CBM14	CBM14	Y(1-28)	2
sp|P41996|CPG2_CAEEL	N	CBM14	CBM14	Y(1-19)	2
sp|Q17802|CPG1_CAEEL	N	CBM14	CBM14	Y(1-18)	2
tr|Q18529|Q18529_CAEEL	N	CBM14	CBM14	N	2
tr|Q09975|Q09975_CAEEL	N	N	GH25	Y(1-24)	1
sp|O16202|LYS7_CAEEL	N	N	GH25	Y(1-20)	1
tr|O01780|O01780_CAEEL	N	N	CBM14	Y(1-19)	1
sp|O62415|LYS1_CAEEL	N	N	GH25	N	1
sp|O62416|LYS2_CAEEL	N	N	GH25	Y(1-20)	1
sp|Q9XXK1|ATPA_CAEEL	N	N	GT4	N	1
tr|Q21650|Q21650_CAEEL	N	N	CBM14	Y(1-22)	1

Table 2.6 Predicted Carbohydrate-active enzymes (CAZyme) in *C. elegans* ESPs. GH, Glycoside Hydrolase. CBM, Carbohydrate-Binding Module. GT, GlycosylTransferase.

RNAi targeting genes	Defective rate(%)	P-value
cpz-1	0.1	0.2747
Y52B11A.8	0	0.7553
clec-15	0.5	0.0001
F37H8.5	0	0.7553
unc-52	0.1	0.2747
far-2	0	0.7553
Y37D8A.2	0	0.7553
ilys-3	0.2	0.0567
cpi-1	0.3	0.0084
ttr-16	0.6	0.0000
ttr-17	0.4	0.001
cpl-1	0.1	0.2747
mlt-11	0.2	0.0567
cpi-2	0.3	0.0084
asp-6	0.2	0.0567
mig-6	0.6	0.0000
nas-31	0	0.8663
nas-20	0.1	0.2747
cpr-1	0.1	0.2747
cpr-6	0	0.7553
vit-1	0.2	0.0567
tbh-1	0.1	0.2747
Y52B11A.8	0.1	0.2747
far-2	0	0.7553
F37H8.5	0	0.7163
F41C3.5	0	0.7553
hsp-3	0.2	0.0567
asp-6	0.3	0.0084
ttr-17	0.2	0.0567
cpi-1	0.2	0.0567
ttr-16	0	0.7973
ilys-3	0.1	0.2747
cpr-1	0.25	0.1371
cpr-6	0	0.7553
ost-1	0.1	0.2747
unc-52	0.2	0.0567
Y37D8A.2	0	0.7553
nas-31	0.1	0.2747
mig-6	0.8	0.0000
vit-1	0.1	0.2747

Table 2.7 List of screened proteases and proteases inhibitors using RNAi. The deficiency rate of utse development is recorded.

***B. subtilis*/OP50**	**Fold change**	***B. mycoides*/OP50**	**Fold change**
ilys-2	288	*ilys-2*	1155
ilys-3	74	*ilys-3*	183
ttr-15	7.8	*Pud-2.1*	22.11
clec-50	6.7	*Pud-1.2*	5.9
Pud-2.1 (protein upregulated in daf-2)	6.6	*ttr-15*	4.6
cpr-5	6	*cpr-1*	4.1
K12H4.7 (putative serine protease)	5.2		

Table 2.8 Up-regulated genes in *C. elegans* ESPs when culturing nematodes on *Bacillus subtilis* and *Bacillus mycoides* diet.

2.8 References

[1] S. Bennuru, R. Semnani, Z. Meng, J. M. C. Ribeiro, T. D. Veenstra, and T. B. Nutman, "Brugia malayi Excreted/Secreted Proteins at the Host/Parasite Interface: Stage- and Gender-Specific Proteomic Profiling," *PLoS Negl Trop Dis*, vol. 3, no. 4, Apr. 2009.

[2] J. Geary *et al.*, "First analysis of the secretome of the canine heartworm, Dirofilaria immitis," *Parasites & Vectors*, vol. 5, no. 1, p. 140, Jul. 2012.

[3] D. Ditgen, E. M. Anandarajah, K. A. Meissner, N. Brattig, C. Wrenger, and E. Liebau, "Harnessing the Helminth Secretome for Therapeutic Immunomodulators," *BioMed Research International*, 2014.

[4] R. Cassini *et al.*, "A new approach to outbreak management for bovine Cystic Echinococcosis cases in hypo-endemic areas," *Vet Parasitol Reg Stud Reports*, vol. 16, p. 100269, Apr. 2019.

[5] D. J. Jenkins, T. Williams, S. Raidal, C. Gauci, and M. W. Lightowlers, "The first report of hydatid disease (Echinococcus granulosus) in an Australian water buffalo (Bubalus bubalis)," *Int J Parasitol Parasites Wildl*, vol. 8, pp. 256–259, Apr. 2019.

[6] D. O. Famakinde, "Mosquitoes and the Lymphatic Filarial Parasites: Research Trends and Budding Roadmaps to Future Disease Eradication," *Trop Med Infect Dis*, vol. 3, no. 1, Jan. 2018.

[7] R. Shinya, H. Morisaka, T. Kikuchi, Y. Takeuchi, M. Ueda, and K. Futai, "Secretome Analysis of the Pine Wood Nematode Bursaphelenchus xylophilus Reveals the Tangled Roots of Parasitism and Its Potential for Molecular Mimicry," *PLoS One*, vol. 8, no. 6, Jun. 2013.

[8] B.-W. Li, A. C. Rush, D.-J. Jiang, M. Mitreva, S. Abubucker, and G. J. Weil, "Gender-associated genes in filarial nematodes are important for reproduction and potential intervention targets," *PLoS Negl Trop Dis*, vol. 5, no. 1, p. e947, Jan. 2011.

[9] M. Breloer and D. Abraham, "Strongyloides infection in rodents: immune response and immune regulation," *Parasitology*, vol. 144, no. 3, pp. 295–315, Mar. 2017.

[10] G. Dong *et al.*, "Novel non-peptidic small molecule inhibitors of secreted aspartic protease 2 (SAP2) for the treatment of resistant fungal infections," *Chem. Commun. (Camb.)*, vol. 54, no. 96, pp. 13535–13538, Dec. 2018.

[11] S. Bellafiore, Z. Shen, M.-N. Rosso, P. Abad, P. Shih, and S. P. Briggs, "Direct identification of the Meloidogyne incognita secretome reveals proteins with host cell reprogramming potential," *PLoS Pathog.*, vol. 4, no. 10, p. e1000192, Oct. 2008.

[12] J. Murray, B. Manoury, A. Balic, C. Watts, and R. M. Maizels, "Bm-CPI-2, a cystatin from Brugia malayi nematode parasites, differs from Caenorhabditis elegans cystatins in a specific site mediating inhibition of the antigen-processing enzyme AEP," *Mol. Biochem. Parasitol.*, vol. 139, no. 2, pp. 197–203, Feb. 2005.

[13] H. Craig, J. M. Wastling, and D. P. Knox, "A preliminary proteomic survey of the in vitro excretory/secretory products of fourth-stage larval and adult Teladorsagia circumcincta," *Parasitology*, vol. 132, no. Pt 4, pp. 535–543, Apr. 2006.

[14] J. P. Hewitson *et al.*, "The secretome of the filarial parasite, Brugia malayi: proteomic profile of adult excretory-secretory products," *Mol. Biochem. Parasitol.*, vol. 160, no. 1, pp. 8–21, Jul. 2008.

[15] X. Zang *et al.*, "Homologues of human macrophage migration inhibitory factor from a parasitic nematode. Gene cloning, protein activity, and crystal structure," *J. Biol. Chem.*, vol. 277, no. 46, pp. 44261–44267, Nov. 2002.

[16] M. W. Robinson, R. Greig, K. A. Beattie, D. J. Lamont, and B. Connolly, "Comparative analysis of the excretory-secretory proteome of the muscle larva of Trichinella pseudospiralis and Trichinella spiralis," *Int. J. Parasitol.*, vol. 37, no. 2, pp. 139–148, Feb. 2007.

[17] D. Ditgen *et al.*, "Comparative characterization of two galectins excreted-secreted from intestine-dwelling parasitic versus free-living females of the soil-transmitted nematode Strongyloides," *Molecular and Biochemical Parasitology*, vol. 225, pp. 73–83, Oct. 2018.

[18] S. Gomez *et al.*, "Genome analysis of Excretory/Secretory proteins in *Taenia solium* reveals their Abundance of Antigenic Regions (AAR)," *Scientific Reports*, vol. 5, p. 9683, May 2015.

[19] N. Santarém *et al.*, "Immune response regulation by leishmania secreted and nonsecreted antigens," *J. Biomed. Biotechnol.*, vol. 2007, no. 6, p. 85154, 2007.

[20] C. Dieterich and R. J. Sommer, "How to become a parasite - lessons from the genomes of nematodes," *Trends Genet.*, vol. 25, no. 5, pp. 203–209, May 2009.

[21] "High-Throughput Proteomics | Annual Review of Analytical Chemistry."

[22] A. R. Jex *et al.*, "*Ascaris suum* draft genome," *Nature*, vol. 479, no. 7374, pp. 529–533, Nov. 2011.

[23] G. Garg and S. Ranganathan, “In silico secretome analysis approach for next generation sequencing transcriptomic data,” *BMC Genomics*, vol. 12, no. Suppl 3, p. S14, Nov. 2011.

[24] J. Borloo *et al.*, “In-Depth Proteomic and Glycomic Analysis of the Adult-Stage *Cooperia oncophora* Excretome/Secretome,” *J. Proteome Res.*, vol. 12, no. 9, pp. 3900–3911, Sep. 2013.

[25] S. Bennuru, R. Semnani, Z. Meng, J. M. C. Ribeiro, T. D. Veenstra, and T. B. Nutman, “Brugia malayi excreted/secreted proteins at the host/parasite interface: stage- and gender-specific proteomic profiling,” *PLoS Negl Trop Dis*, vol. 3, no. 4, p. e410, 2009.

[26] H. Soblik *et al.*, “Life cycle stage-resolved proteomic analysis of the excretome/secretome from Strongyloides ratti--identification of stage-specific proteases,” *Mol. Cell Proteomics*, vol. 10, no. 12, p. M111.010157, Dec. 2011.

[27] J. Mulvenna, B. Hamilton, S. H. Nagaraj, D. Smyth, A. Loukas, and J. J. Gorman, “Proteomics analysis of the excretory/secretory component of the blood-feeding stage of the hookworm, Ancylostoma caninum,” *Mol. Cell Proteomics*, vol. 8, no. 1, pp. 109–121, Jan. 2009.

[28] “WormBase : Nematode Information Resource.”

[29] J. J. Almagro Armenteros *et al.*, “SignalP 5.0 improves signal peptide predictions using deep neural networks,” *Nat. Biotechnol.*, vol. 37, no. 4, pp. 420–423, 2019.

[30] H. Saverwyns *et al.*, “Identification and characterization of a novel specific secreted protein family for selected members of the subfamily Ostertagiinae (Nematoda),” *Parasitology*, vol. 135, no. Pt 1, pp. 63–70, Jan. 2008.

[31] Y. M. Harcus *et al.*, "Signal sequence analysis of expressed sequence tags from the nematode Nippostrongylus brasiliensis and the evolution of secreted proteins in parasites," *Genome Biol*, vol. 5, no. 6, p. R39, 2004.

[32] J. A. Gadahi *et al.*, "Proteomic Analysis of the Excretory and Secretory Proteins of Haemonchus contortus (HcESP) Binding to Goat PBMCs In Vivo Revealed Stage-Specific Binding Profiles," *PLoS One*, vol. 11, no. 7, Jul. 2016.

[33] G. Coakley *et al.*, "Extracellular Vesicles from a Helminth Parasite Suppress Macrophage Activation and Constitute an Effective Vaccine for Protective Immunity," *Cell Rep*, vol. 19, no. 8, pp. 1545–1557, May 2017.

[34] A. H. Buck *et al.*, "Exosomes secreted by nematode parasites transfer small RNAs to mammalian cells and modulate innate immunity," *Nature Communications*, vol. 5, p. 5488, Nov. 2014.

[35] F. Riaz and G. Cheng, "Exosome-like vesicles of helminths: implication of pathogenesis and vaccine development," *Ann Transl Med*, vol. 5, no. 7, Apr. 2017.

[36] M. Siles-Lucas, R. Morchon, F. Simon, and R. Manzano-Roman, "Exosome-transported microRNAs of helminth origin: new tools for allergic and autoimmune diseases therapy?," *Parasite Immunol.*, vol. 37, no. 4, pp. 208–214, Apr. 2015.

[37] A. Marcilla *et al.*, "Extracellular Vesicles from Parasitic Helminths Contain Specific Excretory/Secretory Proteins and Are Internalized in Intestinal Host Cells," *PLOS ONE*, vol. 7, no. 9, p. e45974, Sep. 2012.

[38] A. J. Knight, N. M. Johnson, and C. A. Behm, "VHA-19 Is Essential in Caenorhabditis elegans Oocytes for Embryogenesis and Is Involved in Trafficking in Oocytes," *PLoS One*, vol. 7, no. 7, Jul. 2012.

[39] V. Spiegler, A. Hensel, J. Seggewiß, M. Lubisch, and E. Liebau, "Transcriptome analysis reveals molecular anthelmintic effects of procyanidins in C. elegans," *PLOS ONE*, vol. 12, no. 9, p. e0184656, Sep. 2017.

[40] F. K. Nelson, P. S. Albert, and D. L. Riddle, "Fine structure of the Caenorhabditis elegans secretory-excretory system," *J. Ultrastruct. Res.*, vol. 82, no. 2, pp. 156–171, Feb. 1983.

[41] "http://www.wormatlas.org/ver1/handbook/anatomyintro/anatomyintr," p. 15.

[42] M. Ashburner *et al.*, "Gene ontology: tool for the unification of biology. The Gene Ontology Consortium," *Nat. Genet.*, vol. 25, no. 1, pp. 25–29, May 2000.

[43] H. Mi *et al.*, "PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements," *Nucleic Acids Res.*, vol. 45, no. D1, pp. D183–D189, 04 2017.

[44] The Gene Ontology Consortium, "The Gene Ontology Resource: 20 years and still GOing strong," *Nucleic Acids Res.*, vol. 47, no. D1, pp. D330–D338, Jan. 2019.

[45] S. Carbon *et al.*, "AmiGO: online access to ontology and annotation data," *Bioinformatics*, vol. 25, no. 2, pp. 288–289, Jan. 2009.

[46] J. Day-Richter, M. A. Harris, M. Haendel, and S. Lewis, "OBO-Edit—an ontology editor for biologists," *Bioinformatics*, vol. 23, no. 16, pp. 2198–2200, Aug. 2007.

[47] D. W. Huang, B. T. Sherman, and R. A. Lempicki, "Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists," *Nucleic Acids Res.*, vol. 37, no. 1, pp. 1–13, Jan. 2009.

[48] D. W. Huang, B. T. Sherman, and R. A. Lempicki, "Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources," *Nat Protoc*, vol. 4, no. 1, pp. 44–57, 2009.

[49] M. Kanehisa and S. Goto, "KEGG: Kyoto Encyclopedia of Genes and Genomes," *Nucleic Acids Res*, vol. 28, no. 1, pp. 27–30, Jan. 2000.

[50] M. Kanehisa, M. Furumichi, M. Tanabe, Y. Sato, and K. Morishima, "KEGG: new perspectives on genomes, pathways, diseases and drugs," *Nucleic Acids Res.*, vol. 45, no. D1, pp. D353–D361, 04 2017.

[51] M. Kanehisa, Y. Sato, M. Kawashima, M. Furumichi, and M. Tanabe, "KEGG as a reference resource for gene and protein annotation," *Nucleic Acids Res*, vol. 44, no. D1, pp. D457–D462, Jan. 2016.

[52] A. Fabregat *et al.*, "The Reactome Pathway Knowledgebase," *Nucleic Acids Res.*, vol. 46, no. D1, pp. D649–D655, Jan. 2018.

[53] S. Hunter *et al.*, "InterPro: the integrative protein signature database," *Nucleic Acids Res*, vol. 37, no. Database issue, pp. D211–D215, Jan. 2009.

[54] S. El-Gebali *et al.*, "The Pfam protein families database in 2019," *Nucleic Acids Res*, vol. 47, no. D1, pp. D427–D432, Jan. 2019.

[55] D. Szklarczyk *et al.*, "The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible," *Nucleic Acids Res*, vol. 45, no. Database issue, pp. D362–D368, Jan. 2017.

[56] N. D. Rawlings, A. J. Barrett, P. D. Thomas, X. Huang, A. Bateman, and R. D. Finn, "The MEROPS database of proteolytic enzymes, their substrates and inhibitors in

2017 and a comparison with peptidases in the PANTHER database," *Nucleic Acids Res*, vol. 46, no. D1, pp. D624–D632, Jan. 2018.

[57] Y. Yin, X. Mao, J. Yang, X. Chen, F. Mao, and Y. Xu, "dbCAN: a web resource for automated carbohydrate-active enzyme annotation," *Nucleic Acids Res*, vol. 40, no. Web Server issue, pp. W445–W451, Jul. 2012.

[58] S. Ghosh and P. W. Sternberg, "Spatial and molecular cues for cell outgrowth during C. elegans uterine development," *Dev. Biol.*, vol. 396, no. 1, pp. 121–135, Dec. 2014.

[59] E. Gómez-Orte *et al.*, "Effect of the diet type and temperature on the C. elegans transcriptome," *Oncotarget*, vol. 9, no. 11, pp. 9556–9571, Dec. 2017.

[60] K. K. Brooks, B. Liang, and J. L. Watts, "The Influence of Bacterial Diet on Fat Storage in C. elegans," *PLOS ONE*, vol. 4, no. 10, p. e7545, Oct. 2009.

[61] B. B. Shtonda and L. Avery, "Dietary choice behavior in Caenorhabditis elegans," *J Exp Biol*, vol. 209, no. Pt 1, pp. 89–102, Jan. 2006.

[62] E. A. Abada, H. Sung, M. Dwivedi, B.-J. Park, S.-K. Lee, and J. Ahnn, "C. elegans behavior of preference choice on bacterial food," *Mol. Cells*, vol. 28, no. 3, pp. 209–213, Sep. 2009.

[63] J. Suh and H. Hutter, "A survey of putative secreted and transmembrane proteins encoded in the C. elegans genome," *BMC Genomics*, vol. 13, p. 333, Jul. 2012.

[64] K. L. Hoang, N. M. Gerardo, and L. T. Morran, "The effects of Bacillus subtilis on Caenorhabditis elegans fitness after heat stress," *Ecology and Evolution*, vol. 9, no. 6, pp. 3491–3499, 2019.

[65] J. Hodgkin, P. E. Kuwabara, and B. Corneliussen, "A novel bacterial pathogen, Microbacterium nematophilum, induces morphological change in the nematode C. elegans," *Current Biology*, vol. 10, no. 24, pp. 1615–1618, Dec. 2000.

[66] L. M. Parsons *et al.*, "Caenorhabditis elegans Bacterial Pathogen Resistant bus-4 Mutants Produce Altered Mucins," *PLoS ONE*, vol. 9, no. 10, p. e107250, Oct. 2014.

[67] H. R. Nicholas and J. Hodgkin, "The ERK MAP Kinase Cascade Mediates Tail Swelling and a Protective Response to Rectal Infection in C. elegans," *Current Biology*, vol. 14, no. 14, pp. 1256–1261, Jul. 2004.

[68] M. J. Gravato-Nobre, F. Vaz, S. Filipe, R. Chalmers, and J. Hodgkin, "The Invertebrate Lysozyme Effector ILYS-3 Is Systemically Activated in Response to Danger Signals and Confers Antimicrobial Protection in C. elegans," *PLOS Pathogens*, vol. 12, no. 8, p. e1005826, Aug. 2016.

[69] C. R. Caffrey, L. Goupil, K. M. Rebello, J. P. Dalton, and D. Smith, "Cysteine proteases as digestive enzymes in parasitic helminths," *PLOS Neglected Tropical Diseases*, vol. 12, no. 8, p. e0005840, Aug. 2018.

[70] S. Hashmi, J. Zhang, Y. Oksov, and S. Lustigman, "The Caenorhabditis elegans Cathepsin Z-like Cysteine Protease, Ce-CPZ-1, Has a Multifunctional Role during the Worms' Development," *J. Biol. Chem.*, vol. 279, no. 7, pp. 6035–6045, Feb. 2004.

[71] L.-P. Leroux *et al.*, "Analysis of the Trichuris suis excretory/secretory proteins as a function of life cycle stage and their immunomodulatory properties," *Scientific Reports*, vol. 8, no. 1, p. 15921, Oct. 2018.

[72] J. Höflich *et al.*, "Loss of srf-3-encoded nucleotide sugar transporter activity in Caenorhabditis elegans alters surface antigenicity and prevents bacterial adherence," *J. Biol. Chem.*, vol. 279, no. 29, pp. 30440–30448, Jul. 2004.

[73] J. F. Cipollo, A. M. Awad, C. E. Costello, and C. B. Hirschberg, "srf-3, a mutant of Caenorhabditis elegans, resistant to bacterial infection and to biofilm binding, is deficient in glycoconjugates," *J. Biol. Chem.*, vol. 279, no. 51, pp. 52893–52903, Dec. 2004.

[74] Y. Moreno, J. F. Nabhan, J. Solomon, C. D. Mackenzie, and T. G. Geary, "Ivermectin disrupts the function of the excretory-secretory apparatus in microfilariae of Brugia malayi," *Proceedings of the National Academy of Sciences*, vol. 107, no. 46, pp. 20120–20125, Nov. 2010.

[75] S. Brenner, "The genetics of Caenorhabditis elegans," *Genetics*, vol. 77, no. 1, pp. 71–94, May 1974.

[76] J. Cox and M. Mann, "MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification," *Nature Biotechnology*, vol. 26, no. 12, pp. 1367–1372, Dec. 2008.

[77] J. Cox, N. Neuhauser, A. Michalski, R. A. Scheltema, J. V. Olsen, and M. Mann, "Andromeda: a peptide search engine integrated into the MaxQuant environment," *J. Proteome Res.*, vol. 10, no. 4, pp. 1794–1805, Apr. 2011.

[78] P. J. Boersema, R. Raijmakers, S. Lemeer, S. Mohammed, and A. J. R. Heck, "Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics," *Nature Protocols*, vol. 4, no. 4, pp. 484–494, Apr. 2009.

[79] "NCBI BLAST: a better web interface."

[80] D. Angeles-Albores, R. Y. N. Lee, J. Chan, and P. W. Sternberg, "Tissue enrichment analysis for C. elegans genomics," *BMC Bioinformatics*, vol. 17, no. 1, p. 366, Sep. 2016.

[81] D. Angeles-Albores, R. Y. N. Lee, J. Chan, and P. W. Sternberg, "Two new functions in the WormBase Enrichment Suite," *Open Access*, p. 3.

[82] A. Conesa and S. Götz, "Blast2GO: A Comprehensive Suite for Functional Analysis in Plant Genomics," *Int J Plant Genomics*, vol. 2008, 2008.

[83] A. Conesa, S. Götz, J. M. García-Gómez, J. Terol, M. Talón, and M. Robles, "Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research," *Bioinformatics*, vol. 21, no. 18, pp. 3674–3676, Sep. 2005.

[84] G. Jékely, "Global view of the evolution and diversity of metazoan neuropeptide signaling," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 110, no. 21, pp. 8702–8707, May 2013.

[85] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, "Basic local alignment search tool," *J. Mol. Biol.*, vol. 215, no. 3, pp. 403–410, Oct. 1990.

[86] G. Csárdi and T. Nepusz, "The igraph software package for complex network research," 2006.

[87] UniProt Consortium, "UniProt: a worldwide hub of protein knowledge," *Nucleic Acids Res.*, vol. 47, no. D1, pp. D506–D515, Jan. 2019.

[88] Y. Zhang, D. Chen, M. A. Smith, B. Zhang, and X. Pan, "Selection of Reliable Reference Genes in Caenorhabditis elegans for Analysis of Nanotoxicity," *PLOS ONE*, vol. 7, no. 3, p. e31849, Mar. 2012.

Chapter 3

IDENTIFICATION OF STAGE SPECIFIC AND PATHOGEN RESPONSE SPECIFIC EXCRETED/SECRETED PROTEINS OF CAENORHABDITIS ELEGANS BY COMPARATIVE ANALYSIS OF DIFFERENTIAL GENE EXPRESSION

3.1 Abstract

The compositions of *C. elegans* excretome/secretome is dynamic through developmental life stages and in response to environmental changes. We took advantage of the deeply sequenced transcriptome of *C. elegans* under many conditions including stages and upon pathogen infections to identify stage-specific and pathogen-associated ESP encoding genes. By this simple comparative method, we successfully characterized L1, L3, adult, L2 dauer and post dauer related ESPs. We demonstrated that expressions of proteases inhibitors are up-regulated during dauer formation while proteases expressions are increased during dauer exit. We performed nano LC-MS/MS to directly characterize excreted/secreted proteins with dauer cultures. The comparison between proteomics and RNAseq dauer expression profiles revealed 91 ESP encoding genes that are highly expressed in dauers. We performed dauer formation assay to these dauer-associated gene mutants. The excellent prediction rate confirmed that our comparative method is the simplest way to identify candidates for functional assays. Similarly, we employed this comparative method to pathogen-induced transcriptomes. We reported a group of genes that are associated with *Serratia marcescens* infection and a group of bacterial/pathogen responding genes. We confirmed the roles of *C. elegans* ESPs in immunoregulation by infection assays with various pathogens. Lysosomes and cysteine protease inhibitor are among the most important genes in the innate immune response pathway of *C. elegans* defending pathogen infection.

3.2 Introduction

We reported the standard protocol to characterize protein components of *C. elegans* excretome/secretome using mixed staged worms cultured at 20 °C with *E. coli* HB101 as diet. However, the composition of ESPs is dynamic and is under accurate transcriptional and translational regulation [1]. ESPs function directly at the interfaces between free-living nematodes/parasites and environment/host [2]–[5]. They need to quickly respond to the environmental stimuli to maintain *C. elegans* homeostasis [6].

The dynamics in ESPs composition can be studied by comparing ESPs of different developmental stages and under different conditions. The model organism Encyclopedia of DNA Elements (modENCODE) project on *C. elegans* allowed in depth systematical annotation of functional genomic elements [7]. Up to now, 669 *C. elegans* datasets have been deposited. Transcriptome profiles of *C. elegans* from a variety of developmental stages, conditions and tissues would provide a solid foundation for comparative functional studies [8], [9]–[14].

C. elegans, as a free-living nematode, undergoes four larval stages after hatch and reach reproductive stage in about 3 days under favorable condition [15]. Under unfavorable conditions, *C. elegans* would enter an alternative development cycle, forming non-aging dauer larvae that can survive for months [16]. In parasites, each species has its distinct life cycle. Parasites life cycles involve the invasion and exploitation of more than one host [17]. Each life cycle stage of parasites has both common and unique features [4], [18]. Despite the complexity in a parasites' life

cycles, nematodes typically undergo four larval molts, suggesting *C. elegans* can serve as the reference model to study environment regulated developmental changes [2], [17], [19].

Dauer is of particular interest because the well-accepted hypothesis that parasite infective larvae has evolved from free-living nematode dauer [20]–[22]. Dauer share a lot of common features with parasite infective larvae, including slim bodies, constricted esophagus, closed mouth [16], [20], [22], [23]. Dafachronic acid (DA) is required in both dauer and parasites for dauer/infective larvae formation [24]–[27]. This shared mechanism strongly supports that parasitism may be acquired from dauer to infective larvae evolution.

Thus, we took advantage of the deeply sequenced *C. elegans* transcriptomes though all developmental stages and extracted the expression profiles of *C. elegans* ESPs [7]. We performed clustering to identify co-expressed genes associated with different stages [28], [29]. These stage-specific ESPs would provide insights towards the mechanisms of nematodes life stage transition. We were able to characterize several groups of genes related with L3, adult, L2d dauer, and post dauer. Interestingly, L2d dauer has up-regulated expression of many proteases inhibitors and post dauer has up-regulated expression of many proteases. This indicates the formation and exit of dauer developmental stage is under accurate regulation of enzyme activates.

C. elegans encounters a wide variety of bacteria and fungi in nature [30]. Among these, several bacteria and fungi have been proved to be pathogenic to *C. elegans*, causing shortened life span or non-lethal disease [31]–[34]. These

infection models have been studied intensively to study the host-pathogen interaction and to identify host pathways exploited by pathogens. *C. elegans* ESPs are the major players in immunoregulation. RNAi experiments targeting several ESPs including *lys-1, lys-8, cpi-1, cpi-2, spp-5,* and *asp-3* showed increased pathogen susceptibility or altered life span, supporting the hypothesis that one major role of *C. elegans* ESPs is to defend against pathogen infection [35]–[38].

Therefore, we examined the expressional regulation of *C. elegans* ESPs with transcriptome data upon bacterial or fungal infections [8]. The *C. elegans* ESPs fall into five common expression groups. We were able to identify a group of genes that is more up-regulated in *Serratia marcescens* infection and a group of genes that is more "bacterial pathogen responding" rather than "fungal pathogen responding". Among these genes, lysozymes are also among the most abundant ESPs. We further confirmed their functions in innate immune response by infecting *C. elegans* with *Serratia marcescens, Pseudomonas aeruginosa* PA14 and *Microbacterium nematophilum.*

3.3 Results

ESP-encoding genes can be assigned into seven groups by clustering based on expression similarity through developmental stages

We utilized the RNA-seq expression profiles associated with different *C. elegans* developmental stages from Gerstein *et al* [7]. Using the excretome/secretome profile of mix-staged *C. elegans*, we pooled the expression profiles of these ESP-encoding genes from the 244 expression experiments reported in the paper. We combined and collapsed the experiments results from the same developmental stage. Thus, data were assigned to 10 developmental stage categories (embryo, L1-stage larvae, L2-stage larvae, L3-stage larvae, L4-stage larvae, young adult, adult, L2d dauer, dauer, and post-dauer). The median value of TPM (Transcipts Per Kilobase Million) was taken in each stage categories. Soft clustering was performed to cluster genes with common expression profiles **(Figure 3.1)** [28], [29].

The genes encoding *C. elegans* excreted/secreted proteins can be sorted into 7 clusters **(Appendix table 3.1)**. We examined the enriched Gene Ontology (GO) terms and enriched biochemical pathways with KEGG biochemical pathway data [39]–[45].

55 Genes were assigned into cluster 1 and they have higher expressions in L1-stage animals. They are enriched in “Glycolysis/Gluconeogenesis” and “Phagosome” KEGG pathways.

In cluster 2, 86 genes have a relatively higher expression in L3 larva stage than the others. The enriched KEGG pathway include “Lysosome” and “Autophagy-animal”, as may be expected for their immune regulation roles. This cluster of genes contains enriched group of aspartic peptidase, cysteine-type endopeptidase, lysozyme, and hydrolase, indicating that this cluster of genes is enzymatically active. Enriched biological process GO terms include “innate immune response”, “defense response to bacterium”, “response to stress”, “peptidoglycan metabolic process”, “cell wall metabolic process”, and “response to external stimulus”. This is consistent with our finding that one major role of excreted/secreted proteins is to regulate innate immune response to bacteria/pathogen/environmental changes. Enriched cellular component GO terms include “lysosome”, “vacuole”, and “extracellular region”, as may be expected due to the secreted nature of the proteins.

Cluster 3 contains 74 genes. These genes are expressed more in the late developmental stage like adult stage. The enriched GO terms include “chitin binding”, “eggshell formation”, “oogenesis”. In fact, several proteoglycan proteins including *cpg-1* and *cpg-3* were included in this cluster. *cpg-1* is involved in eggshell formation and is expressed in eggshell and germ line [46]. This is consistent with the active reproduction process in adult hermaphrodite. GO terms of “nutrient reservoir activity”, and “defense response to Gram-positive bacterium” are also enriched. This is consistent with the hypothesis that excreted/secreted proteins are important players in helping nutrition uptake and food processing [5], [6].

Cluster 4 genes are more like post-dauer related (74 genes). Thus, this group of gene may function as the key players at the dauer to developmental stage transition. “Lysosome” is the enriched KEGG pathway in this cluster. This cluster contains genes from several protein families, including cysteine peptidase, threonine peptidase, and zinc metalloproteinase.

In cluster 5, 72 genes are highly expressed in L2d dauer. Studying this cluster of genes may provide clues for decision making in early dauer. The biological process GO terms enriched in this group are “negative regulation of endopeptidase activity”, “negative regulation of proteolysis”, and “negative regulation of hydrolase activity”. This cluster contains a lot of proteases inhibitors, in contrast to the enriched proteases during dauer exit in cluster 4. This contrast is of particular interest, suggesting the enzyme active levels are down-regulated during dauer and would be turned up again during dauer exit.

Cluster 6 has fewer genes (only 18 genes) than other clusters and the genes in this group are slightly expressed more in embryo but in general the expressions are relatively stable across development stages.

Cluster 7 contains 125 genes with clearly higher expression in L3 larva stage. The enriched KEGG pathways are “Lysosome” and “Biosynthesis of amino acids”. The enriched biological process GO terms include “innate immune response”, “proteolysis”, “response to stress”, and “catabolic process”. Cluster 7 shares a lot in common with cluster 2, which also has higher expression level in L3. But they differ in the type of active enzymes. Cluster 7 contains more serine-type exopeptidases and serine-type hydrolase.

In short, we identified stage-specific ESPs based on their expression profiles. Studying stage-specific ESPs would provide information to characterize ESPs function related with unique biological features in each developmental stage. This could also be used to compare with stage-specific ESPs from parasites, which comprise more distinct and complicated life styles [2]–[4].

Direct profiling of dauer excreted/secreted proteins with proteomics

To validate the clustering method, we wanted to directly characterize the composition of stage-specific excretome/secretome. We found dauer is of particular interest since dauer is believed to be equivalent to infective larvae in parasites in development stages. The similarities between dauer and infective larvae lead to the hypothesis that dauer is a pre-adaptation for parasitism [20], [22]. Parasitism is closely related to series of discrete immune responses [47]. Thus, identification of dauer-specific ESPs would help to find key functional regulators in the innate immune response pathways.

To investigate the composition of dauer excretome/secretome, we employed similar proteomic method as previously reported. *C. elegans* dauers were obtained in liquid culture by adding dauer-inducing pheromone [48]. The dauer excreted/secreted proteins were collected after incubation and were subjected to nanoLC-MS/MS. A total of 430 proteins were identified through Mass-Spec in dauer excretome/secretome.

Validation of dauer ESPs with dauer RNA-seq expression result

Previous works in this lab had established a sophisticated design to pinpoint key time points during dauer formation and dauer exit. Briefly, *daf-9 (dh6)* null mutant with deficiency to produce growth hormone dafachronic acid (DA) was used. This strain is dauer-constitutive if not introducing synthetic DA to the culture. The RNA samples of *daf-9 (dh6)* mutant were taken at the following timepoints: with no DA, "L2d" at 24 hours post hatch (hph), "L2d" at 26 hph, "dauer-committed L2d" at 34 hph, "fully developed dauers" at 60 hph; with DA added at 24 hph, "L3-committing larvae" at 26 hph, "L4" at 34 hph [49].

We extracted the expression profiles of dauer ESPs from the transcription data **(Figure 3.3) [49]**. In summary, 91 ESP-encoding genes were highly expressed in "dauer-committed L2d" and "fully developed dauers" experiments. We considered this subset of ESPs to be dauer-specific. This group of genes is enriched in "Longevity regulating pathway" and "Phagosome" pathway. This is consistent with the extended life span of dauers. The enriched molecular function GO terms are "endopeptidase inhibitor activity", "endopeptidase regulator activity", "peptidase inhibitor activity", and "peptidase regulator activity", suggesting the proteases inhibitors are up-regulated in expression during dauer stage. This is consistent with the observation in Cluster 5 in last section.

Screen of candidate genes for deficiency in dauer entry

To investigate the function of dauer-specific ESPs, we performed a screen for dauer entry deficiency on available mutants. We first narrowed down the list by picking out genes with some level of annotation related to secretion. Then we

chose genes with available knockout strains in CAENORHABDITIS GENETICS CENTER (CGC). A total of 13 genes were screened for dauer entry phenotype. Among these, 10 strains showed a decreased dauer formation **(Figure 3.4)**. Our result suggested we successfully predict dauer-specific ESPs. Among these, *cpi-1* and *cpi-2*, two cysteine protease inhibitors, promote dauer formation.

Clustering revealed pathogen-associated ESP encoding genes

C. elegans responds to distinct pathogens by regulating the expression of specific subset of genes [12], [13], [33], [50], [51]. Transcriptional profiles of *C. elegans* in response to infection by three bacterial (*Serratia marcescens, Enterococcus faecalis* and *otorhabdus luminescens*) and two fungal pathogens (*Drechmeria coniospora* and *Harposporium sp.*) were taken for comparative studies [8].

We extracted the expression of *C. elegans* ESPs from the 5 transcriptional profiles mentioned above. Similar clustering method was conducted to identify co-expressing genes **(Figure 3.5)**.

The genes encoding *C. elegans* excreted/secreted proteins were sorted into 5 clusters **(Appendix table 3.2)**. Similarly, GO term and KEGG biochemical pathway enrichment analysis were checked within each cluster [42], [44].

Cluster 1 contains 132 genes. This group of genes is specifically responding to *Serratia marcescens* infection. The enriched KEGG pathways in this group are “Lysosome” and “Autophagy-animal”, as expected for the up-regulated

bacteria/pathogen defending reactions. The group contains aspartic peptidase, hydrolase, serine peptidase, and galactosidase.

Cluster 2 (65 genes) has the highest level of expression in infection of *Enterococcus faecalis* and relatively high expression with *otorhabdus luminescens, Drechmeria coniospora and Harposporium sp.* infections. The enriched biological process GO terms in this cluster are "gland morphogenesis", "gland development", and "pharyngeal gland morphogenesis". In fact, a peroxidase, HPX-2, was expressed in pharyngeal to protect *C. elegans* from *Enterococcus faecalis* infection [52]. The enriched proteins feature in "superoxide dismutase activity", consistent with literature report.

Cluster 3 (88 genes) defines a group of "bacteria pathogen responding" genes with high expression level in bacterial infection but low expression level in fungal infection. They are enriched in "biocynthesis of amino acids", "lysosome", "phagosome" and "Carbon metabolism" pathways.

Cluster 4 (101 genes) and Cluster 5 (104 genes) have opposite directions in expression. Cluster 4 genes have a higher expression when challenging with *Enterococcus faecalis* and *otorhabdus luminescens* while cluster 5 genes are down regulated when facing *Enterococcus faecalis* and *otorhabdus luminescens.* However, there was no enriched GO terms or KEGG pathways in these two clusters.

Infection assays by pathogens to validate ESPs function

To validate the pathogen-specific ESPs and their function, we challenged *C. elegans* with *Serratia marcescens, Pseudomonas aeruginosa* PA14 and *Microbacterium nematophilum.* All three are well-studied bacterial pathogens to *C. elegans* [31], [33, p. 14], [50], [53]–[59].

The *Serratia marcescens* infection assay revealed that *grd-10, nas-20, cpi-2* promote pathogen defense in *C. elegans* **(Figure 3.6)**. Noticeably, *grd-10* and *nas-20* are Cluster 1 members, which was defined as *Serratia marcescens* response genes. *cpi-2* was found in Cluster 5, which is also highly expressed upon *Serratia marcescens* exposure.

We screened the same set of candidate genes to check the pathogen specific effect with *Pseudomonas aeruginosa* PA14 **(Figure 3.7)**. Indeed, *grd-10* and *nas-20* mutants are not significantly more vulnerable to PA14 infection. Instead, *cpi-2, asp-6, clec-15, ilys-3,* and *spp-8* may be involved in defense mechanism upon PA14 infection.

Microbacterium nematophilum is a non-lethal pathogen causing tail swelling and constipation [31], [53]. Thus, the innate immune system of *C. elegans* should be active and up-regulated all the time during the whole infection process [31]. Indeed, we observed a very large group of genes that are associated with *Microbacterium nematophilum* infection **(Figure 3.8)**, including *ilys-3, cpi-2, dur-1, cpr-2, grd-10, cdr-1, C25E10.8, clec-65, dgk-3,* and *dpy-5*.

3.4 Discussion

Stage-specific proteomic profilings of excretome/secretome have been investigated in *Strongyloides ratti* and *Brugia malayi [4], [18], [60]*. In *Strongyloides ratti*, proteomic characterizations were performed to infective larvae, parasitic females and free-living stages. 140 proteins were shared in all stages and 196 infective larvae specific ESPs were identified [18]. In *Brugia malayi*, ESPs of L3, L3 to L4 molting, adult male, adult female, and microfilarial stage were profiled, leading to an identification of 852 proteins in total. The functional and component distribution remains similar across the developmental stages. Several immunologically important proteins were identified in the L3 to L4 transition excretome/secretome, including abundant larval transcripts (ALT), a nematode polyprotein allergen (NPA) LL20, and DJ-1 family protein [4]. Stress-response related proteins are also abundant in L3 to L4 infectome, including thioredoxin peroxidases and glutathione peroxidases [4].

We compared the stage-specific ESPs identified in our clustering method with the reported stage-specific ESPs from parasites. In general, the function annotation distributions in terms of protein families are very similar between species. It is hard to compare particular protein due to the missing annotation and function characterization. One successful example is that the intermediate filament protein is abundant in L3/L4 molting excretome/secretome of *B. malayi* [4]. Its *C. elegans* homologue, IFA-1, is characterized as a L3 stage specific ESP in cluster 7. Galectins are also abundant in L3/L4 molting ESPs in *B. malayi* while several

lectins members are L3 stage-specific ESPs of *C. elegans* [4]. The consistence in both protein function and expression stage suggest that some of the excretome/secretome functions may be shared between species.

We also demonstrated a trend of up-regulated proteases inhibitor expression in dauer formation and up-regulated proteases expression in dauer exit. Our results provided insights for dynamic dauer regulation. During dauer formation, cystatins, *cpi-1,* and *cpi-2*, are up regulated. We predicted the protein interactions with *cpi-1* and *cpi-2* and found the predicted proteins that closely interact with *cpi-1* and *cpi-2* are also abundant ESPs in proteomic results, including *far-2, ttr-51, cpz-1,* and *cpl-1* [61]. This strongly supports the notion that the interaction networks around *cpi-1* and *cpi-2* play important role in ESPs. How this group of ESPs functions in dauer formation needs further investigation. K11D12.7 and C02F12.5 are another two serine-type endopeptidase inhibitors that are up-regulated in dauer formation. C02F12.5 is expressed in reproductive system suggesting that it may suppress the reproductive system activity and promote dauer formation [62].

C. elegans maintains homeostasis through innate immune system upon the infection of pathogens [33], [50]. A number of distinct protein families are involved in protective response to infections. Indeed, we observed all types of reported antimicrobial proteins are abundant in *C. elegans* ESPs, including nematode specific peptide group C families (nspc genes), saposin-like amoebapores (spp genes) as Antimicrobial peptides (AMPs), defensins (abf-2), and lysozymes (lys genes). AMPs and other secreted signals trigger the downstream signaling pathways in immune response [33], [63], [64].

Therefore, we proved a broad presence of all types of antimicrobial proteins in *C. elegans* excretome/secretome, suggesting ESPs are a complicated mixture and function to protect pathogen infection through innate immune response. Although there are shared pathogen defense signaling pathways, *C. elegans* also co-evolved with different pathogens and can exhibit pathogen-specific responses. Evidence has shown the *C. elegans* strains that are resistant to some natural pathogens may at the same time more sensitive to other pathogen infection, suggesting *C. elegans* may respond to specific pathogen with different signaling pathways [33], [63], [64].

We examined the expression profiles of ESPs upon five bacterial/fungal pathogen infections and clustered ESPs into five co-expression groups [8]. One co-expression groups seems to be more sensitive in reaction to *Serratia marcences*. The other co-expression group seems to utilize a shared mechanism in defending bacterial pathogen other than fungal pathogens.

We confirmed the antimicrobial roles of several important protein family members in *C. elegans* ESPs by infection assays. Among these, lysozymes and cysteine protease inhibitors are two of the major groups of antimicrobial proteins. We have shown *cpi-2* mutants are more sensitive to infection of all three bacterial pathogens while *ilys-3* mutants are more vulnerable to PA14 and *Microbacterium nematophilum* infection but not to *Serratia marcences. ilys-3* was identified in cluster 5, which has relatively low expression levels upon bacterial pathogen infections. However, we can observe a lot of lysozyme members, including *lys-1, lys-2, lys-4, lys-7,* and *lys-8*, are highly expressed during *Serratia marcences*

infection. The distribution of other key protein families also shows that each cluster contains different protein family members, suggesting that although the anti-infection mechanisms may be shared, *C. elegans* could response to specific pathogens with corresponding specialized family members.

In summary, our comparative method successfully characterized stage- and pathogen- specific excreted/secreted proteins. The stage-specific ESPs may be used to allow cross-species comparison with other parasites, helping to elucidate parasitism evolution and life stage transition mechanism. The pathogen-specific ESPs will promote the study of host-pathogen relationship and identify specific signaling pathways in innate immune system upon infection. Our study also demonstrated the possibility to extend this comparative method to expression profiles under other conditions, for example, under stress or chemical stimuli. Another pivotal role of excreted/secreted proteins is to maintain homeostasis under stress, which is supported by the abundant presentence of peroxidases and superoxide dismutases to neutralize reactive oxygen species in *C. elegans* ESPs. Our comparative method allows quick selection of candidate genes to allow functional characterization. Together, studying dynamics of *C. elegans* ESPs composition is a key to understand how nematodes respond and adapt to fast changing environment and to elucidate the strategies to survive and reproduce in natural niche.

3.5 Materials and Methods

Dauer Excreted/Secreted Proteins sample preparation

Dauers were cultured in liquid culture by adding dauer-inducing pheromone as previously described. Dauers were then harvested by centrifuge and washed intensively for six times in M9 buffer. Dauers were left in 1 ml M9 buffer for incubation at 20°C for another 4 hours. Supernatant was collected and filtered by a 0.22μm syringe filter unit.

nanoLC-MS/MS and data process

All excreted/secreted protein samples were digested by LysC and Trypsin enzymes after the reduction and alkylation of Cysteines. In-solution digested C. elegans ESP samples were chemically labeled using dimethyl labeling strategy [ ]. Briefly digested tryptic peptides are tagged on primary amines (N-terminus and amino group of lysine) using a mixture of cyanoborohydride and formaldehyde in their unlabeled and stable isotope-labeled forms. In this study samples were labeled with regular formaldehyde and cyanoborohydride for the light label (+28) and deuterated formaldehyde and cyanoborohydride to generates a mass increase of +32 for the intermediate label. Then the samples were mixed desalted and subjected to LC-MS/MS analysis on a nanoflow LC system, EASY-nLC II, (Thermo Fisher Scientific) coupled to a LTQ Orbitrap mass spectrometer (Thermo Fisher Scientific, Bremen, Germany) equipped with a Nanospray Flex ion source (Thermo Fisher Scientific).

For the EASY-nLC II system, solvent A consisted of 97.8% H2O, 2% ACN, and 0.2% formic acid and solvent B consisted of 19.8% H2O, 80% ACN, and 0.2% formic acid. Samples were directly loaded onto a 16-cm analytical HPLC column (75 mm ID) packed in-house with ReproSil-Pur C18AQ 3 um resin (120A° pore size, Dr. Maisch, Ammerbuch, Germany). The column was heated to 45° C. The peptides were separated with a 120 min gradient at a flow rate of 350 nL/min. The gradient was as follows: 2–30% Solvent B (110 min), 30–100% B (1 min), and 100% B (9 min). Eluted peptides were then ionized using a standard coated silica tip (New Objective, Woburn, MA) as an electrospray emitter and introduced into the mass spectrometer. The LTQ Orbitrap was operated in a data-dependent mode, automatically alternating between a full-scan (m/z 300-1700) in the Orbitrap and subsequent MS/MS scans of the 15 most abundant peaks in the linear ion trap (Top15 method). Data acquisition was controlled by Xcalibur 2.0.7 and Tune 2.4 software (Thermo Fisher Scientific).

Raw data was analyzed using MaxQuant (v. 1.5.3.30) [65], [66] Spectra were searched against *C. elegans* entries in UniProt as well as a contaminant database containing common proteins like trypsin and keratins. Precursor mass tolerance was 4.5 ppm after recalibration and fragment tolerance was 0.5 Da. Carbamidomethylation of cysteine was specified as a fixed modification and protein N-terminal acetylation and oxidation of methionine were specified as variable modifications. Trypsin was the specified digestion enzyme and up to two missed cleavages were allowed. Score were thresholded so as to achieve a 1%

FDR at the PSM, peptide, and protein levels as estimated by a decoy database. Match-between-runs, iBAQ, and LFQ quantitation were enabled.

Soft clustering

Soft clustering was performed with the mFuzz package in R [1], [2]. Gene expression data were standardized before clustering, and cluster numbers were chosen based on cluster stability, minimum cluster centroid distance, and visual clarity of the clusters. Over-represented gene classes in clusters were determined by Fisher's exact test by $p<0.05$.

Heat maps

RNA-seq data of *C. elegans* dauer developmental stages were used as published. Expression values were centered and scaled for each gene. Heatmaps were drawn using the gplots and RColorBrewer packages in R. Heatmap dendrograms were drawn using correlation distances and average-linkage hierarchical clustering.

Functional enrichment analysis

Gene ontology and KEGG analysis was performed with g:Profiler (e94_eg41_p11_9f195a1) [41]–[45], [67]. Significance was tested with g_SCS multiple testing correction with a threshold of 0.05.

Bacteria strains

E. coli HB101, E. coli OP50, *Bacillus subtilis, Bacillus mycoides* were cultured in LB at 37°C. *Bacillus simplex, Bacillus megaterium, Microbacterium nematophilum CBX102* were obtained from Caenorhabditis Genetics Center (CGC) and cultured at 37°C. *Pseudomonas aeruginosa* PA14 was a gift from Dianne Newman lab and was cultured in LB at 37°C. *Serratia marcescens* db11was a gift from Elizabeth Glater lab.

Animal strains

C. elegans wild type strain, N2 (Bristol), was maintained in lab followed the standard protocols on NGM plates with the *E. coli* strain OP50 as a food source (91). Strains obtained from the *Caenorhabditis* Genetics Center (CGC) were: RB667 *far-2(ok435) III*, RB1207 *cpi-2(ok1256) V*, RB2365 *vit-2 (ok3211) X*, VC2496 *ilys-3 (ok3222) IV*, VC687 *dur-1(ok1010) IV*, RB2129 *cpr-2(ok2833) V*, RB2223 *grd-10(ok3008) IV*, RB2564 *nas-20(ok3572)V*, RB2496 *cdr-1(ok3456)V*, VC1277 *C25E10.8(ok1753)V*, CB5635 *bus-13(e2710)V*, JN214 *iff-2(tm393)/mIn1 [dpy-10(e128) mIs14]II*, RB859 *Y57A10C.6(ok693) II*, RB1803 *clec-65(ok2337)II*, VC218 *dgk-3(gk110)III*, BC14615 *dpy-5(e907) I;sEx14615 [rCes C02F12.5::GFP + pCeh361]*. FX02213 *asp-6 (tm2213)* and FX03833 *clec-15 (tm3833)* were obtained from National Bioresource Project in Japan.

Dauer entry assay on pheromone plates

Dauer entry assay was performed as previously described. Crude pheromone was extracted from exhausted liquid culture medium, re-suspended with distilled water and stored at -20°C. Fresh pheromone plates (NGM-agar with added crude pheromone and no peptone) were made one day before experiment and left to dry overnight at room temperature. Overnight culture of E. coli OP50 was resuspended to 8 g/100 mL in S Basal buffer and heat at 100°C for 5 minutes. On the day of assay setup, 10 young adults were picked onto each plate, and allowed to lay eggs for 1 hour (approximately 50-60 eggs). After removing young adults, exact number of laid eggs was counted for each plate. 20 µl of heat-killed OP50 was added as food source. The numbers of dauers and non-dauers were counted after 48 hours of incubation at 25.5°C. Each strain had 3 replicates.

***Pseudomonas aeruginosa* PA14 slow killing assay**

PA14 slow killing assay was performed as previously described. 10 µL overnight culture of PA14 was spread to each Slow-Killing agar plate. Plates were left to dry for 20 min at room temperature and followed by incubation at 37°C for 24h. Plates were transferred to incubate at 25°C for 24h. 40 µL of 100 X 5-fluoro-2'-deoxyuridine (FUDR) was added to the edge of Slow-Killing plate and wait for diffusion for about 30 mins. About 50 synchronized L4 larval stage worms were transferred to each assay plate and incubate at 25°C. Number of living worms were counted every 24 hours for about a week. Each condition had at least 3 replicates. Graphpad Prism 6 was used to plot survival curves and perform two-way ANOVA statistical tests.

Serratia marcescens db11 infection assay

Serratia marcescens db11 infection assay was done in a similar way to PA14 infection assay. 10 μL of overnight cultere of db11 strain was spread to NGM plate and incubated at 37°C for 24h and 25°C for another 24h. 40 μL of 100 X 5-fluoro-2'-deoxyuridine (FUDR) was added to the edge of NGM plate and wait for diffusion for about 30 mins. About 50 synchronized L4 larval stage worms were transferred to each assay plate and incubate at 25°C. Number of living worms were counted every 24 hours for about a week. Each condition had at least 3 replicates. Graphpad Prism 6 was used to plot survival curves and perform two-way ANOVA statistical tests.

Microbacterium nematophilum CBX102 infection assay

Microbacterium nematophilum CBX102 infection assay was performed as previously described with modifications. 6 mL overnight culture of CBX102 was centrifuged at 3000g for 3 mins and resuspended in 1 mL M9 buffer. One pre-packaged tube of Cy3 monoreactive dye (GE healthcare PA23001) was mixed with CBX102 and incubated for 1h at 4°C. Pellet was washed for 3 times with M9 and resuspended to a final concentration of 200 mg/mL. 16 mg/mL *Microbacterium nematophilum* CBX102 was mixed with 16 mg/mL E. coli OP50 at a 1:9 ratio. 20 μL of 10% M. nematophilum mixture was seeded to NGM plate and left to dry for a day at room temperature. About 30 L2/L3 larval stage worms were transferred to each mixed bacteria lawn plate. After 24 hours, worms were

scored for a tail-swelling phenotype. Bump formation was checked if necessary under Zeiss LSM 710 Inverted confocal microscope. Each strain had 3 replicates.

3.6 Figures

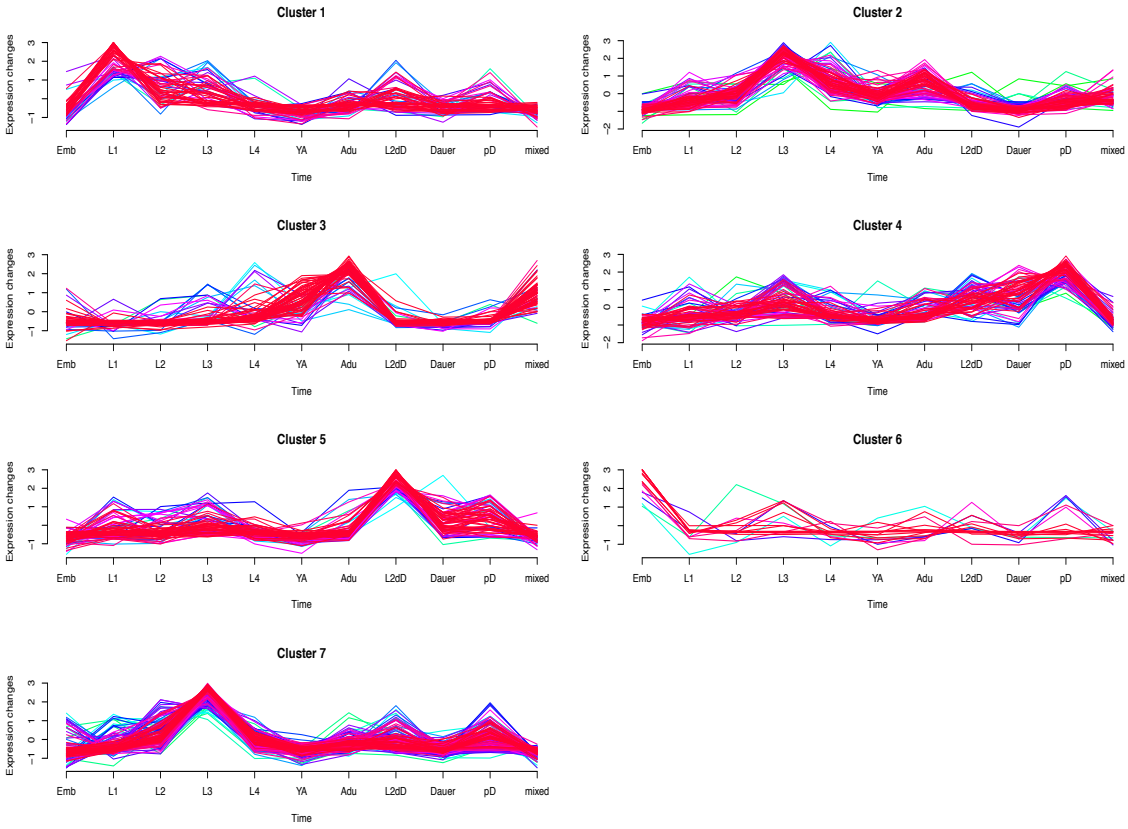


Figure 3.1. *C. elegans* ESPs can be classified into seven common expression profiles through developmental stages. Red lines are expression profiles for genes with high membership score.

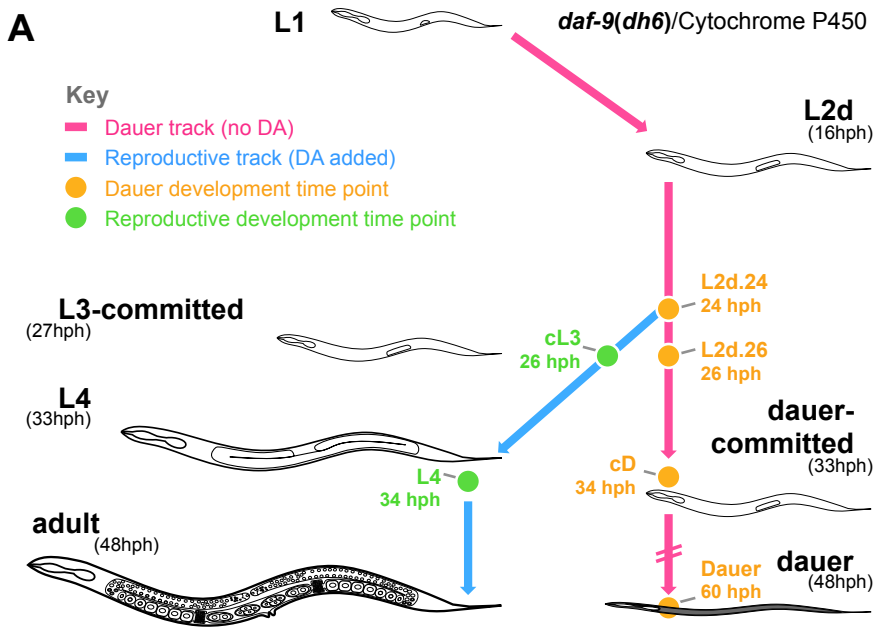


Figure 3.2 Experimental design for dauer and developemental expression profiles with RNA-seq using *daf-9(dh6)* mutant. Samples were taken at: with no DA, "L2d" at 24 hours post hatch (hph), "L2d" at 26 hph, "dauer-committed L2d" at 34 hph, "fully developed dauers" at 60 hph; with DA added at 24 hph, "L3-committing larvae" at 26 hph, "L4" at 34 hph. This is an unmodified figure from [49].

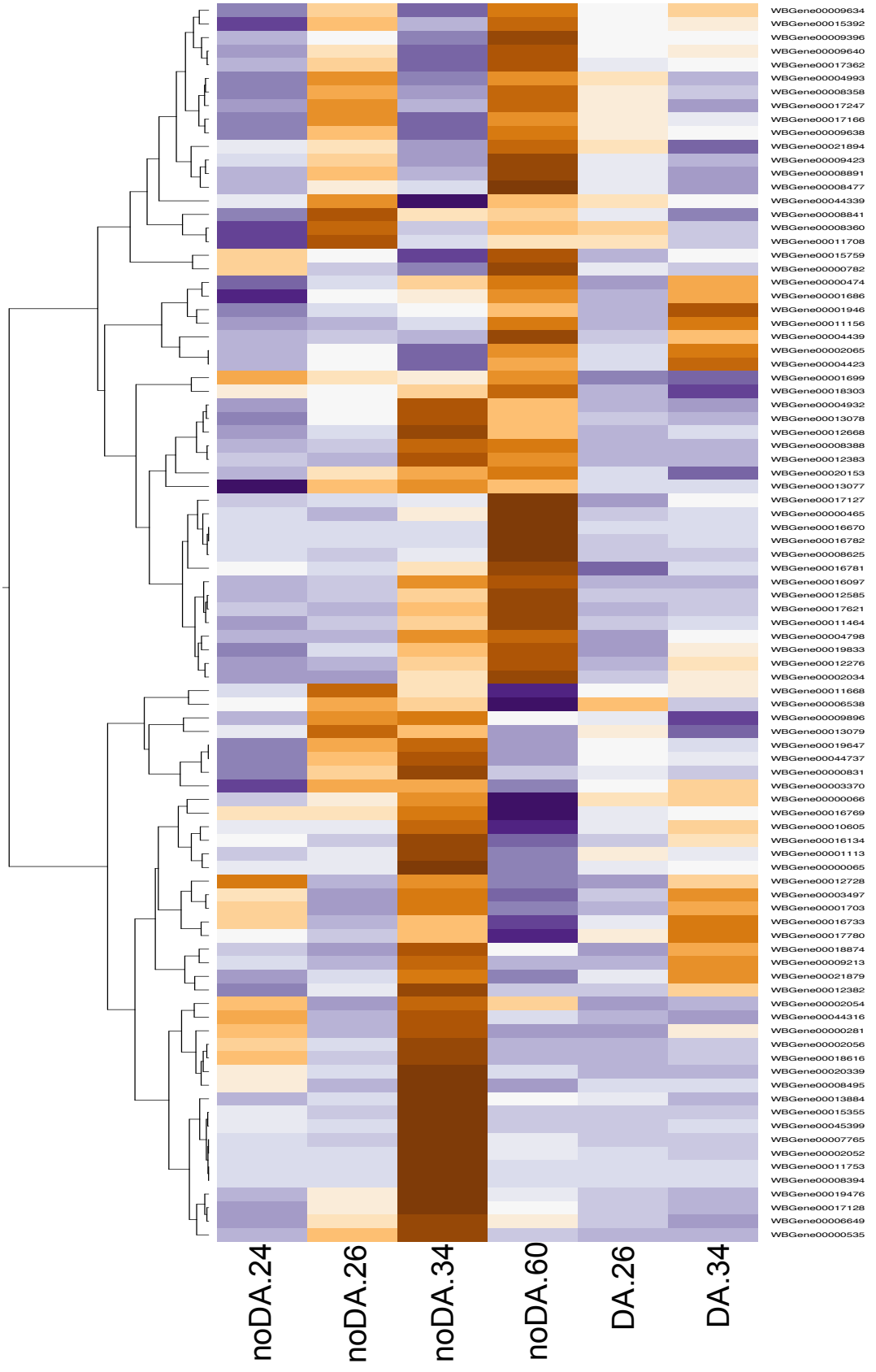

WBGene00009634
WBGene00015392
WBGene00009396
WBGene00009640
WBGene00017362
WBGene00004993
WBGene00008358
WBGene00017247
WBGene00017166
WBGene00009638
WBGene00021894
WBGene00009423
WBGene00008891
WBGene00008477
WBGene00044339
WBGene00008841
WBGene00008360
WBGene00011708
WBGene00015759
WBGene00000782
WBGene00000474
WBGene00001686
WBGene00001946
WBGene00011156
WBGene00004439
WBGene00002065
WBGene00004423
WBGene00001699
WBGene00018303
WBGene00004932
WBGene00013078
WBGene00012668
WBGene00008388
WBGene00012383
WBGene00020153
WBGene00013077
WBGene00017127
WBGene00000465
WBGene00016670
WBGene00016782
WBGene00008625
WBGene00016781
WBGene00016097
WBGene00012585
WBGene00017621
WBGene00011464
WBGene00004798
WBGene00019833
WBGene00012276
WBGene00002034
WBGene00011668
WBGene00006538
WBGene00009896
WBGene00013079
WBGene00019647
WBGene00044737
WBGene00000831
WBGene00003370
WBGene00000066
WBGene00016769
WBGene00010605
WBGene00016134
WBGene00001113
WBGene00000065
WBGene00012728
WBGene00003497
WBGene00001703
WBGene00016733
WBGene00017780
WBGene00018874
WBGene00009213
WBGene00021879
WBGene00012382
WBGene00002054
WBGene00044316
WBGene00000281
WBGene00002056
WBGene00018616
WBGene00020339
WBGene00008495
WBGene00013884
WBGene00015355
WBGene00045399
WBGene00007765
WBGene00002052
WBGene00011753
WBGene00008394
WBGene00019476
WBGene00017128
WBGene00006649
WBGene00000535
noDA.24
noDA.26
noDA.34
noDA.60
DA.26
DA.34

Figure 3.3 Hierarchical heat map for Dauer ESPs. Genes that were identified in dauer excretome/secretome and also were highly expressed in dauer expression profiles by RNA-seq (noDA.34 and noDA.60) were selected. They were further clustered by expression profile in heat map.

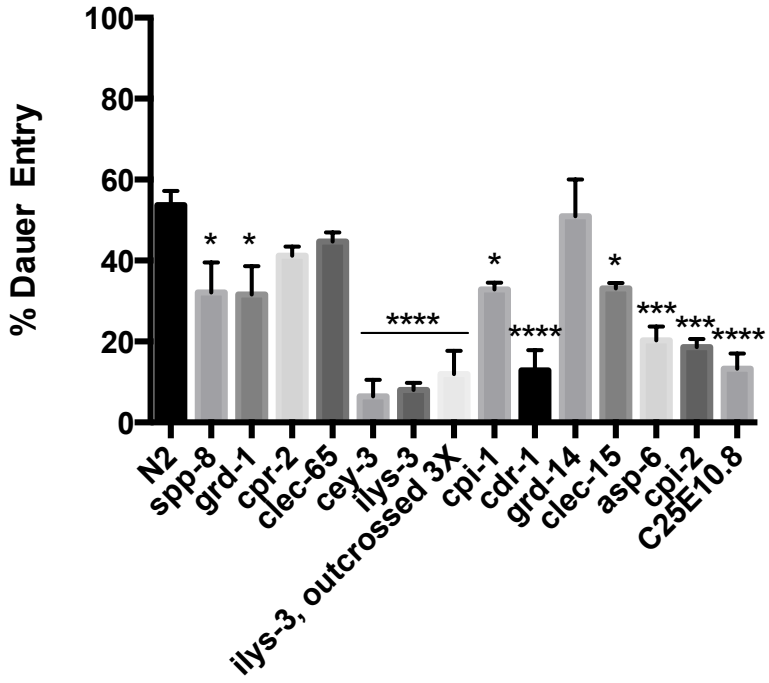


Figure 3.4 Dauer entry scores for predicted dauer-specific genes. Statistics: One-way ANOVA with Dunnett's multiple comparisons. Ns, not significant, *, P <= 0.05, **, P <= 0.01, ***, P<= 0.001 ****, P <= 0.0001.

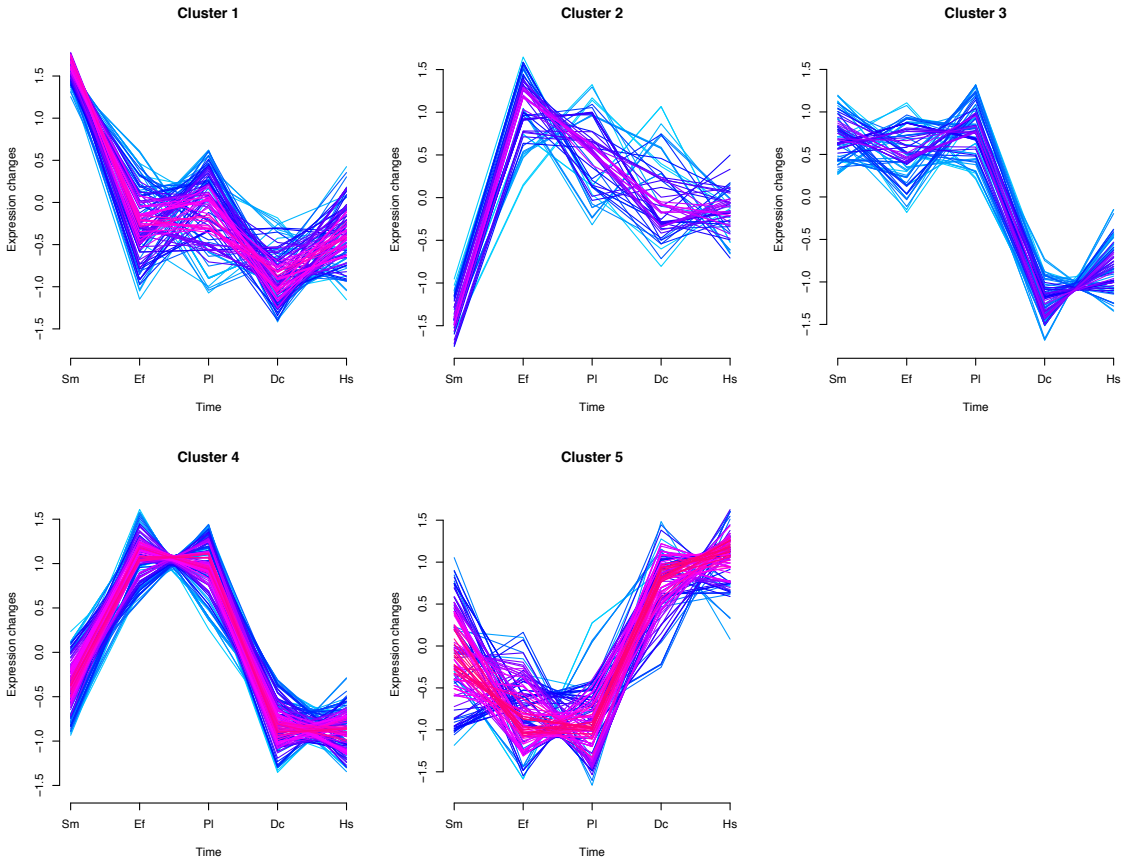


Figure 3.5 *C. elegans* ESPs can be classified into five common expression profiles in response to bacterial and fungal pathogens. Red lines are expression profiles for genes with high membership score.

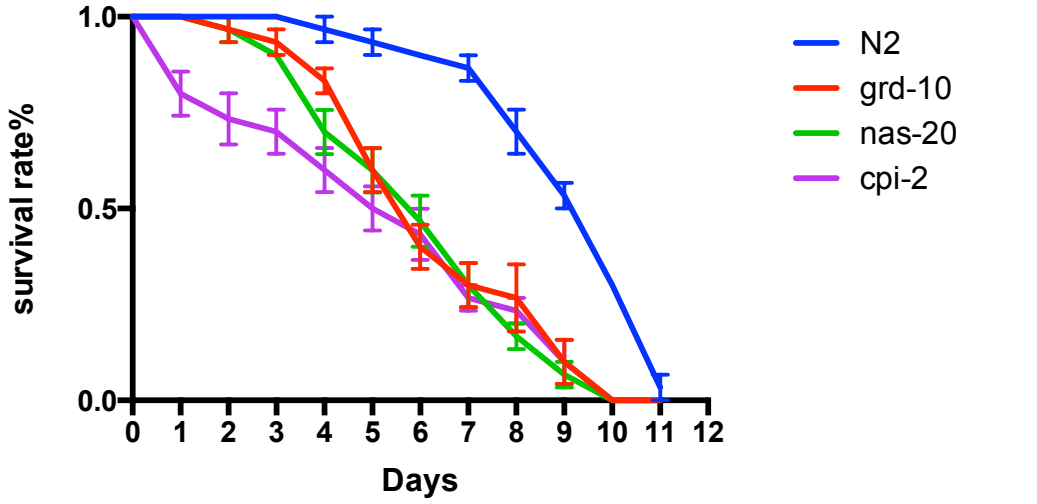


Figure 3.6 *Serratia marcences* killing assay revealed several genes in innate immune response pathway. Statistics: two-way ANOVA with Sidak's multiple comparisons test. Ns, not significant, *, P <= 0.05, **, P <= 0.01, ***, P<= 0.001 ****, P <= 0.0001.

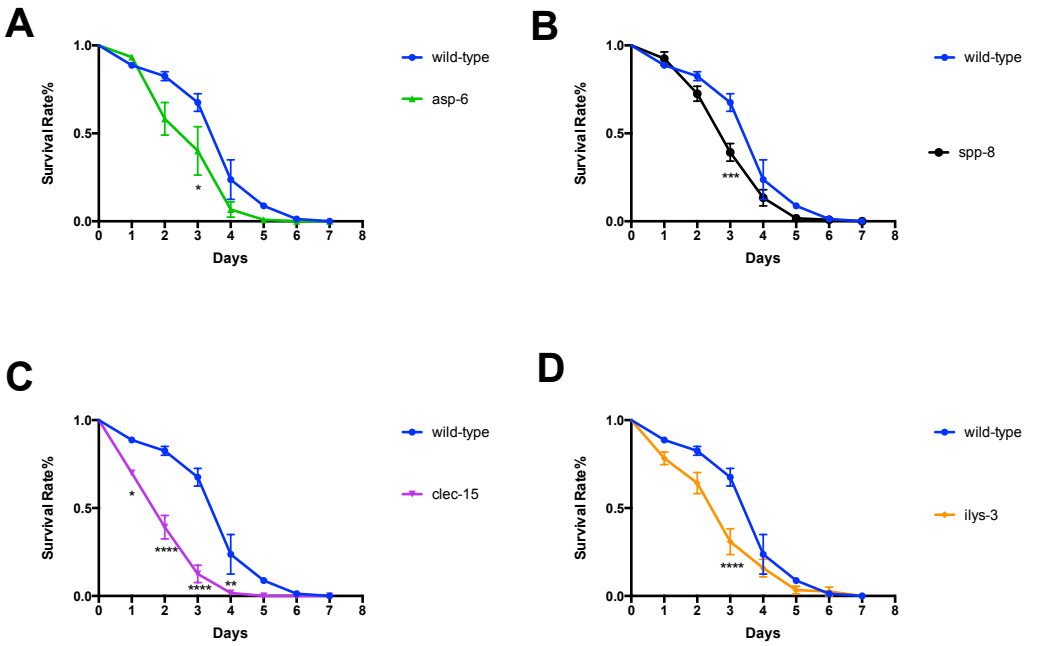


Figure 3.7 *Pseudomonas aeruginosa* PA14 slow killing assay revealed several genes in defending PA14 infection. (A) infection of *asp-6* knock out mutant and wild-type N2. (B) infection of *spp-8* knock out mutant and wild-type N2. (C) infection of *clec-15* knock out mutant and wild-type N2. (D) infection of *ilys-3* knock out mutant and wild-type N2. Statistics: two-way ANOVA with Sidak's multiple comparisons test. Ns, not significant, *, P <= 0.05, **, P <= 0.01, ***, P<= 0.001 ****, P <= 0.0001.

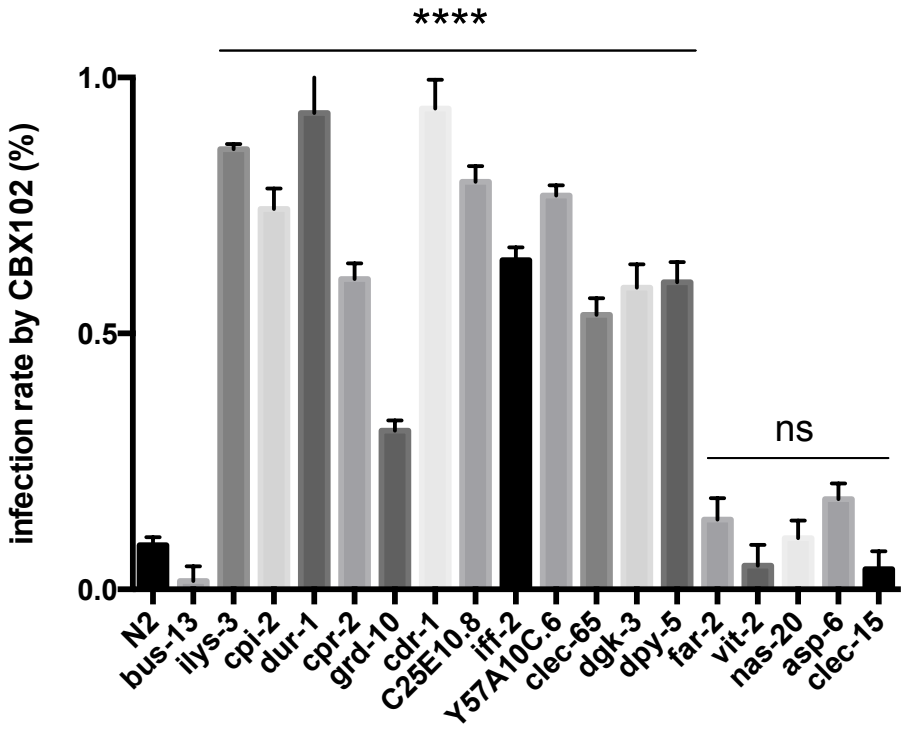


Figure 3.8 ***Microbacterium nematophilum*** **infections of ESP mutants.** Statistics: One-way ANOVA with Dunnett's multiple comparisons. Ns, not significant, *, P <= 0.05, **, P <= 0.01, ***, P<= 0.001 ****, P <= 0.0001.

3.7 References:

[1] E. Gómez-Orte *et al.*, "Effect of the diet type and temperature on the C. elegans transcriptome," *Oncotarget*, vol. 9, no. 11, pp. 9556–9571, Dec. 2017.

[2] R. Shinya, H. Morisaka, T. Kikuchi, Y. Takeuchi, M. Ueda, and K. Futai, "Secretome Analysis of the Pine Wood Nematode Bursaphelenchus xylophilus Reveals the Tangled Roots of Parasitism and Its Potential for Molecular Mimicry," *PLoS One*, vol. 8, no. 6, Jun. 2013.

[3] S. Bellafiore, Z. Shen, M.-N. Rosso, P. Abad, P. Shih, and S. P. Briggs, "Direct identification of the Meloidogyne incognita secretome reveals proteins with host cell reprogramming potential," *PLoS Pathog.*, vol. 4, no. 10, p. e1000192, Oct. 2008.

[4] S. Bennuru, R. Semnani, Z. Meng, J. M. C. Ribeiro, T. D. Veenstra, and T. B. Nutman, "Brugia malayi Excreted/Secreted Proteins at the Host/Parasite Interface: Stage- and Gender-Specific Proteomic Profiling," *PLoS Negl Trop Dis*, vol. 3, no. 4, Apr. 2009.

[5] D. Ditgen, E. M. Anandarajah, K. A. Meissner, N. Brattig, C. Wrenger, and E. Liebau, "Harnessing the Helminth Secretome for Therapeutic Immunomodulators," *BioMed Research International*, 2014.

[6] G. Coakley, A. H. Buck, and R. M. Maizels, "Host parasite communications—Messages from helminths for the immune system," *Mol Biochem Parasitol*, vol. 208, no. 1, pp. 33–40, Jul. 2016.

[7] M. B. Gerstein *et al.*, "Integrative analysis of the Caenorhabditis elegans genome by the modENCODE project," *Science*, vol. 330, no. 6012, pp. 1775–1787, Dec. 2010.

[8] I. Engelmann *et al.*, "A comprehensive analysis of gene expression changes provoked by bacterial and fungal infection in C. elegans," *PLoS ONE*, vol. 6, no. 5, p. e19055, 2011.

[9] K. Jovic *et al.*, "Temporal dynamics of gene expression in heat-stressed Caenorhabditis elegans," *PLOS ONE*, vol. 12, no. 12, p. e0189445, Dec. 2017.

[10] M. E. Boeck *et al.*, "The time-resolved transcriptome of C. elegans," *Genome Res.*, vol. 26, no. 10, pp. 1441–1450, Oct. 2016.

[11] R. Di, H. Zhang, and M. A. Lawton, "Transcriptome Analysis of C. elegans Reveals Novel Targets for DON Cytotoxicity," *Toxins (Basel)*, vol. 10, no. 7, Jun. 2018.

[12] J. W. Lightfoot, V. M. Chauhan, J. W. Aylott, and C. Rödelsperger, "Comparative transcriptomics of the nematode gut identifies global shifts in feeding mode and pathogen susceptibility," *BMC Res Notes*, vol. 9, Mar. 2016.

[13] R. P. Shivers, M. J. Youngman, and D. H. Kim, "Transcriptional responses to pathogens in Caenorhabditis elegans," *Curr Opin Microbiol*, vol. 11, no. 3, pp. 251–256, Jun. 2008.

[14] N. Wittenburg and R. Baumeister, "Thermal avoidance in Caenorhabditis elegans: An approach to the study of nociception," *PNAS*, vol. 96, no. 18, pp. 10477–10482, Aug. 1999.

[15] "C elegans anatomy and lifecycle.pdf." .

[16] N. Fielenbach and A. Antebi, "C. elegans dauer formation and the molecular basis of plasticity," *Genes Dev*, vol. 22, no. 16, pp. 2149–2165, Aug. 2008.

[17] B. F. Kochin, J. J. Bull, and R. Antia, "Parasite Evolution and Life History Theory," *PLOS Biology*, vol. 8, no. 10, p. e1000524, Oct. 2010.

[18]H. Soblik *et al.*, "Life Cycle Stage-resolved Proteomic Analysis of the Excretome/Secretome from Strongyloides ratti—Identification of Stage-specific Proteases," *Mol Cell Proteomics*, vol. 10, no. 12, Dec. 2011.

[19]D. Duneau and D. Ebert, "The role of moulting in parasite defence," *Proc Biol Sci*, vol. 279, no. 1740, pp. 3049–3054, Aug. 2012.

[20]M. Crook, "The dauer hypothesis and the evolution of parasitism: 20 years on and still going strong," *Int. J. Parasitol.*, vol. 44, no. 1, pp. 1–8, Jan. 2014.

[21]W. Wong, "How to Evolve into a Parasite," *Sci. Signal.*, vol. 2, no. 54, pp. ec26–ec26, Jan. 2009.

[22]C. Dieterich and R. J. Sommer, "How to become a parasite - lessons from the genomes of nematodes," *Trends Genet.*, vol. 25, no. 5, pp. 203–209, May 2009.

[23]A. Ogawa, A. Streit, A. Antebi, and R. J. Sommer, "A Conserved Endocrine Mechanism Controls the Formation of Dauer and Infective Larvae in Nematodes," *Current Biology*, vol. 19, no. 1, pp. 67–71, Jan. 2009.

[24]S. S. Lee and F. C. Schroeder, "Steroids as Central Regulators of Organismal Development and Lifespan," *PLOS Biology*, vol. 10, no. 4, p. e1001307, Apr. 2012.

[25]H. Aguilaniu, P. Fabrizio, and M. Witting, "The Role of Dafachronic Acid Signaling in Development and Longevity in Caenorhabditis elegans: Digging Deeper Using Cutting-Edge Analytical Chemistry," *Front Endocrinol (Lausanne)*, vol. 7, Feb. 2016.

[26]T.-M. Li *et al.*, "No Significant Increase in the Δ4- and Δ7-Dafachronic Acid Concentration in the Long-Lived glp-1 Mutant, nor in the Mutants Defective in Dauer Formation," *G3: Genes, Genomes, Genetics*, vol. 5, no. 7, pp. 1473–1479, Jul. 2015.

[27]K. J. Dumas *et al.*, "Functional divergence of dafachronic acid pathways in the control of C. elegans development and lifespan," *Dev Biol*, vol. 340, no. 2, pp. 605–612, Apr. 2010.

[28]L. Kumar and M. E. Futschik, "Mfuzz: A software package for soft clustering of microarray data," *Bioinformation*, vol. 2, no. 1, pp. 5–7, May 2007.

[29]M. E. Futschik and B. Carlisle, "Noise-robust soft clustering of gene expression time-course data," *J. Bioinform. Comput. Biol.*, vol. 03, no. 04, pp. 965–988, Aug. 2005.

[30]E. A. Abada, H. Sung, M. Dwivedi, B.-J. Park, S.-K. Lee, and J. Ahnn, "C. elegans behavior of preference choice on bacterial food," *Mol. Cells*, vol. 28, no. 3, pp. 209–213, Sep. 2009.

[31]J. Hodgkin, P. E. Kuwabara, and B. Corneliussen, "A novel bacterial pathogen, Microbacterium nematophilum, induces morphological change in the nematode C. elegans," *Current Biology*, vol. 10, no. 24, pp. 1615–1618, Dec. 2000.

[32]J. Höflich *et al.*, "Loss of srf-3-encoded nucleotide sugar transporter activity in Caenorhabditis elegans alters surface antigenicity and prevents bacterial adherence," *J. Biol. Chem.*, vol. 279, no. 29, pp. 30440–30448, Jul. 2004.

[33]C. Darby, "Interactions with microbial pathogens," *WormBook*, 2005.

[34]H. R. Nicholas and J. Hodgkin, "The ERK MAP Kinase Cascade Mediates Tail Swelling and a Protective Response to Rectal Infection in C. elegans," *Current Biology*, vol. 14, no. 14, pp. 1256–1261, Jul. 2004.

[35]S. Miyata, J. Begun, E. R. Troemel, and F. M. Ausubel, "DAF-16-Dependent Suppression of Immunity During Reproduction in Caenorhabditis elegans," *Genetics*, vol. 178, no. 2, pp. 903–918, Feb. 2008.

[36]M. E. Yanos, C. F. Bennett, and M. Kaeberlein, "Genome-Wide RNAi Longevity Screens in Caenorhabditis elegans," *Curr Genomics*, vol. 13, no. 7, pp. 508–518, Nov. 2012.

[37]A. V. Samuelson, R. R. Klimczak, D. B. Thompson, C. E. Carr, and G. Ruvkun, "Identification of Caenorhabditis elegans genes regulating longevity using enhanced RNAi-sensitive strains," *Cold Spring Harb. Symp. Quant. Biol.*, vol. 72, pp. 489–497, 2007.

[38]C. L. Kurz and M.-W. Tan, "Regulation of aging and innate immunity in C. elegans," *Aging Cell*, vol. 3, no. 4, pp. 185–193, 2004.

[39]M. Ashburner *et al.*, "Gene ontology: tool for the unification of biology. The Gene Ontology Consortium," *Nat. Genet.*, vol. 25, no. 1, pp. 25–29, May 2000.

[40]H. Mi *et al.*, "PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements," *Nucleic Acids Res.*, vol. 45, no. D1, pp. D183–D189, 04 2017.

[41]"g:Profiler—a web server for functional interpretation of gene lists (2016 update) | Nucleic Acids Research | Oxford Academic."

[42]J. Reimand, T. Arak, and J. Vilo, "g:Profiler—a web server for functional interpretation of gene lists (2011 update)," *Nucleic Acids Res*, vol. 39, no. suppl_2, pp. W307–W315, Jul. 2011.

[43]M. Kanehisa and S. Goto, "KEGG: Kyoto Encyclopedia of Genes and Genomes," *Nucleic Acids Res*, vol. 28, no. 1, pp. 27–30, Jan. 2000.

[44]M. Kanehisa, M. Furumichi, M. Tanabe, Y. Sato, and K. Morishima, "KEGG: new perspectives on genomes, pathways, diseases and drugs," *Nucleic Acids Res.*, vol. 45, no. D1, pp. D353–D361, 04 2017.

[45]M. Kanehisa, Y. Sato, M. Kawashima, M. Furumichi, and M. Tanabe, "KEGG as a reference resource for gene and protein annotation," *Nucleic Acids Res*, vol. 44, no. D1, pp. D457–D462, Jan. 2016.

[46]S. K. Olson, G. Greenan, A. Desai, T. Müller-Reichert, and K. Oegema, "Hierarchical assembly of the eggshell and permeability barrier in C. elegans," *J Cell Biol*, vol. 198, no. 4, pp. 731–748, Aug. 2012.

[47]M. Crook, "The dauer hypothesis and the evolution of parasitism: 20 years on and still going strong," *Int J Parasitol*, vol. 44, no. 1, pp. 1–8, Jan. 2014.

[48]X. Karp, "Working with dauer larvae," *WormBook*, pp. 1–19, Aug. 2018.

[49]J. S. Lee, P.-Y. Shih, O. N. Schaedel, P. Quintero-Cadena, A. K. Rogers, and P. W. Sternberg, "FMRFamide-like peptides expand the behavioral repertoire of a densely connected nervous system," *PNAS*, vol. 114, no. 50, pp. E10726–E10735, Dec. 2017.

[50]L. B. Cohen and E. R. Troemel, "Microbial pathogenesis and host defense in the nematode C. elegans," *Curr Opin Microbiol*, vol. 0, pp. 94–101, Feb. 2015.

[51]N. Pujol *et al.*, "Anti-Fungal Innate Immunity in C. elegans Is Enhanced by Evolutionary Diversification of Antimicrobial Peptides," *PLoS Pathog*, vol. 4, no. 7, Jul. 2008.

[52]Y. Liu, K. G. Kaval, A. van Hoof, and D. A. Garsin, "Heme peroxidase HPX-2 protects Caenorhabditis elegans from pathogens," *PLOS Genetics*, vol. 15, no. 1, p. e1007944, Jan. 2019.

[53]M. J. Gravato-Nobre, F. Vaz, S. Filipe, R. Chalmers, and J. Hodgkin, "The Invertebrate Lysozyme Effector ILYS-3 Is Systemically Activated in Response to Danger Signals and Confers Antimicrobial Protection in C. elegans," *PLOS Pathogens*, vol. 12, no. 8, p. e1005826, Aug. 2016.

[54]H. Schulenburg and J. J. Ewbank, "Diversity and specificity in the interaction between Caenorhabditis elegans and the pathogen Serratia marcescens," *BMC Evolutionary Biology*, vol. 4, no. 1, p. 49, Nov. 2004.

[55]E. Pradel, Y. Zhang, N. Pujol, T. Matsuyama, C. I. Bargmann, and J. J. Ewbank, "Detection and avoidance of a natural product from the pathogenic bacterium Serratia marcescens by Caenorhabditis elegans," *PNAS*, vol. 104, no. 7, pp. 2295–2300, Feb. 2007.

[56]R. L. Feinbaum *et al.*, "Genome-Wide Identification of Pseudomonas aeruginosa Virulence-Related Genes Using a Caenorhabditis elegans Infection Model," *PLOS Pathogens*, vol. 8, no. 7, p. e1002813, Jul. 2012.

[57]N. V. Kirienko, B. O. Cezairliyan, F. M. Ausubel, and J. R. Powell, "Pseudomonas aeruginosa PA14 pathogenesis in Caenorhabditis elegans," *Methods Mol. Biol.*, vol. 1149, pp. 653–669, 2014.

[58]M. W. Robinson, R. Greig, K. A. Beattie, D. J. Lamont, and B. Connolly, "Comparative analysis of the excretory-secretory proteome of the muscle larva of Trichinella pseudospiralis and Trichinella spiralis," *Int. J. Parasitol.*, vol. 37, no. 2, pp. 139–148, Feb. 2007.

[59]M.-W. Tan, S. Mahajan-Miklos, and F. M. Ausubel, “Killing of Caenorhabditis elegans by Pseudomonas aeruginosa used to model mammalian bacterial pathogenesis,” *Proc Natl Acad Sci U S A*, vol. 96, no. 2, pp. 715–720, Jan. 1999.

[60]J. P. Hewitson *et al.*, “The secretome of the filarial parasite, Brugia malayi: proteomic profile of adult excretory-secretory products,” *Mol. Biochem. Parasitol.*, vol. 160, no. 1, pp. 8–21, Jul. 2008.

[61]D. Szklarczyk *et al.*, “The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible,” *Nucleic Acids Res*, vol. 45, no. Database issue, pp. D362–D368, Jan. 2017.

[62]J. W. Ziel, D. Q. Matus, and D. R. Sherwood, “An expression screen for RhoGEF genes involved in C. elegans gonadogenesis,” *Gene Expr Patterns*, vol. 9, no. 6, pp. 397–403, Sep. 2009.

[63]J. Ewbank, “Signaling in the immune response,” *WormBook*, 2006.

[64]D. H. Kim, “Signaling in the innate immune response,” *WormBook*, pp. 1–35, Aug. 2018.

[65]J. Cox and M. Mann, “MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification,” *Nature Biotechnology*, vol. 26, no. 12, pp. 1367–1372, Dec. 2008.

[66]J. Cox, N. Neuhauser, A. Michalski, R. A. Scheltema, J. V. Olsen, and M. Mann, “Andromeda: a peptide search engine integrated into the MaxQuant environment,” *J. Proteome Res.*, vol. 10, no. 4, pp. 1794–1805, Apr. 2011.

[67]J. Reimand, M. Kull, H. Peterson, J. Hansen, and J. Vilo, “g:Profiler—a web-based toolset for functional profiling of gene lists from large-scale experiments,” *Nucleic Acids Res*, vol. 35, no. suppl_2, pp. W193–W200, Jul. 2007.

C h a p t e r 4

CHARACTERIZING EXCRETED/SECRETED PROTEINS OF A *CAENORHABDITIS ELEGANS* SIBLING SPECIES, *CAENORHABDITIS INOPINATA*

4.1 Abstract

The recent discovery of a *Caenorhabditis elegans* sibling species, *Caenorhabditis inopinata,* allows deeply comparative study for evolutional interpretation. The excretome/secretome of *C. inopinata* has not been characterized. We took advantage of the sensitive and high-throughput technique of nanoscale liquid chromatography coupled to tandem mass spectrometry (nano LC-MS/MS) to directly characterize the protein components of *C. inopinata* excretome/secretome. Functional annotations reveal several protein families including C-type lectins, Cathepsin Z, Cathepsin B family, transthyretin, and saposin-like families, suggesting ESPs play critical roles in regulating innate immune response. We compared *C. inopinata* excretome/secretome with *C. elegans*. The compositions of protein families are highly conserved across species, suggesting that the sibling species shares common mechanism with *C. elegans* to respond to environmental stimuli.

4.2 Introduction

Caenorhabditis inopinata n. sp. was found very recently in the fresh fig syconia in Japan [1], [2]. The phylogenetic analyses found that *C. inopinata* is the closest species to *C. elegans* in evolution, which diverged about 10.5 million years ago [1]. *C. inopinata* feeds on the bacteria in the fig syconia and proliferates. Dauer of *C. inopinata* can translocate to new habitat by attaching to mutualistic pollinating wasp. *C. inopinata* shares several morphological features with other fig-associated nematodes, indicating that *C. inopinata* coevolved with Fig *Ficus syconia* and is adapted to fig syconia environment. The genome model of *C. inopinata* found more than 95% of genes have homologs in *C. elegans* and the essential pathways, including dauer formation and insulin/insulin like growth factor signaling, are highly conserved [1], [2].

Since *C. inopinata* is the sibling species to *C. elegans*, we wanted to characterize *C. inopinata* ESPs using the same HPLC-MS/MS method to allow direct comparison of ESPs between species and to elucidate the evolutions of how species adapt to their habit environments [1]. The study of *C. inopinata* excretome/secretome itself also provides rich information on how this nematode regulates the metabolism and morphology and how they benefit from the mutualism.

Thus, we performed a high throughput proteomic characterization of *C. inopinata* excreted/secreted proteins. A total number of 585 proteins were identified using a mix stage population of *C. inopinata*. The size of the excretome/secretome is very close to *C. elegans*. We performed functional annotations of Gene Ontology and KEGG pathways [3]–[5], [6, p. 2], [7]–[9]. The structure of *C. inopinata* excretome/secretome in terms of GO

molecular functions, biological processes and protein families is highly conserved. This is also a great example demonstrating how annotation of *C. elegans* ESPs can benefit annotation to proteins in the new gene model.

The main pathways associated with *C. inopinata* are "innate immune response" and "metabolism", in consistent with the hypothesis that ESPs function continuously to fight against environmental changes. Major protein families are also conserved, including peptidases, lectins, and lysozymes. By orthologous comparison, majority of ESPs have homologues in both species. We only observed a slight expansion in cuticle collagen proteins and a slight shrink in pharyngeal gland toxin related proteins in *C. inopinata*. In general, the highly conserved *C. inopinata* and *C. elegans* excretomes/secretomes suggest they share common mechanisms to maintain homeostasis. However, particular protein functions need to be investigated more carefully in specific signaling pathways since modification of essential functions is the key to adapt to specialized niche [10]–[12] .

We further expanded the comparisons to other parasitic nematodes in order to explain the evolution of parasitism [11]–[16]. We demonstrated that the functional structures of ESPs are similar for parasitic nematodes with distinct lifestyles and for free-living nematodes. Majority of the protein families are shared among all the species, although the numbers of protein family members inside each species could vary. Among these, cysteine-type peptidase, transthyretin-like superfamily, pharyngeal gland toxin related proteins, and lysosomes vary the most across species. *C. elegans* C-type lectin family is expanded in ESPs, suggesting the life style of *C. elegans* with complicated soil bacteria populations may require a broader spectrum of immune-regulation molecules [17]–[20].

4.3 Results

Profiling the components of *C. inopinata* Eecreted/Secreted Proteins (ESPs) with nanoLC-MS/MS

Nano-liquid chromatography coupled with tandem mass spectrometry (nanoLC-MS/MS) was performed to directly characterize the protein components of *C. inopinata*.

The wild type of *C. inopinata* strain was grown on 50 large NGM plates with pre-seeded *E. coli* OP50 in each Mass-Spec sample preparation batch. Worms were growing at 25 °C to the point of food depletion. Contaminated plates were discarded. Mixed staged *C. inopinata* worms were collected and washed intensively six times with M9 buffer. Then nematodes were incubated in 1 mL M9 at 25 °C for 4 hours. Supernatant was collected by centrifuge and filtered by a 0.22μm syringe filter. We treated this as the unconcentrated *C. inopinata* excreted/secreted proteins sample. The concentration of unconcentrated *C. inopinata* was measured and followed by lysyl endopeptidase/trypsin digestion and was subjected to nanoLC-MS/MS for peptide characterization.

In summary, a total of 585 *C. inopinata* excreted/secreted proteins were characterized with a majority of proteins identified with two or more specific peptides. The detailed composition of *C. inopinata* ESPs was provided in a concise form in Appendix table 4.1 with all details from Mass-Spec (**Appendix table 4.1**). The size of detected *C. inopinata* excretome/secretome was similar to what we found about *C. elegans* excretome/secretome. Also, similar proteomic methods performed to other parasites reported a similar size of parasites excretome/secretome. An excretome/secretome proteomic study with

Strongyloides ratti identified 586 proteins in total and 852 proteins were characterized in *Brugia malayi* ESPs [15], [16].

The characterized *C. inopinata* ESPs covered about 2.7% (585 out of 21608) of the *C. inopinata* genes while *C. elegans* ESPs covered 2.5% of the *C. elegans* genome. In general, *C. inopinata* genome is very similar to its sibling species, *C. elegans*. More than 95 percent of the *C. inopinata* genes have orthologues in *C. elegans*. The main difference between these two genomes was the sex determination genes since *C. inopinata* has a XO male/female sexual system while *C. elegans* is hermaphrodite/male. Considering the high conservation between genomes and similar free-living life styles, we would expect the sizes of excretome/secretome to be close to each other.

Gene ontology analysis

We conducted Gene Ontology (GO) term analysis to the list of 585 *C. inopinata* excreted/secreted proteins [4], [21]. Each of the excreted/secreted proteins was BLASTed against NCBI with BLASTp. The annotations to the top blast hits were added and InterProScan was performed by Blast2GO **(Figure 4.1)** [5, p. 2], [6], [22].

Majority of the excreted/secreted proteins are around a length of 300 amino acids. By BLAST search, about 78% of the excreted/secreted proteins were successful annotated. 9% of the excreted/secreted proteins were not returned with a BLAST best hit **(Figure 4.2)**. For the sequences with returned BLAST hits, the coverage is around 97% on average and most of the sequences returned multiple BLAST results, suggesting the identified excreted/secreted proteins are evolutionally conserved between species **(Figure 4.3)**. The annotation to *C. inopinata* ESPs benefits significantly from the homologous models to *C.*

elegans ESPs. The annotation levels in some parasites are limited by the lack of whole genome sequence and gene models [23]–[27]. This demonstrated how characterization of *C. elegans* ESPs can be used to help annotate the parasites ESPs.

The BLAST results were mainly from the closest nematodes relatives including *C. elegans, C. briggsae, C. remanei,* and *C. brenneri*. The BLAST top-Hits were mainly from *C. elegans*, confirming that *C. inopinata* indeed is the sibling species of *C. elegans* **(Figure 4.4)**.

GO terms were successfully assigned to *C. inopinata* sequences. About half of the sequences in excretome/secretomes were assigned with more than one GO term. A total of 1468 GO terms were annotated to *C. inopinata* ESPs. About 75% of the GO terms were annotated to level-3 GO terms or above, suggesting the annotation level in *C. inopinata* is good **(Figure 4.5)**.

GO terms of biological processes include several interesting major categories. The first group contains GO terms of "immune system process", "response to stress". This presence of this group of GO terms is consistent with the finding of similar GO terms enriched in *C. elegans* ESPs. The equivalent *C. elegans* GO term group contains GO terms like "defense response to Gram-positive bacterium", "innate immune response" and "defense response to Gram-negative bacterium". This group of excreted/secreted proteins is expected to regulate innate immune response and help nutrition uptake or removing pathogens/cell debris probably through lysosome and phagolysosome. The second interesting group contains GO terms of "cell adhesion", and "vesicle-mediated transport", indicating the proteins in this group may be involved in extracellular functions. This is consistent with our expectation that ESPs are mainly active directly at the nematode-

environment surface. Another major group of GO terms are associated with metabolic pathways, including “small molecule metabolic process”, “carbohydrate metabolic process”, and “lipid metabolic process”. This subgroup of proteins may function as critical enzymes allowing the modification of metabolism to adapt to fast changing environment **(Figure 4.6)**.

GO terms of molecular function include “peptidase activity”, “lyase activity”, and “hydrolase activity” **(Figure 4.7)**. In *C. elegans* excretome/secretome, GO terms enriched in molecular function are mainly lysozymes and proteases and proteases inhibitors, including galactosidase activity, serine-type exopeptidase activity, lysozyme activity, peptidoglycan muralytic activity, aspartic-type endopeptidase activity, and cysteine-type endopeptidase activity. The molecular function GO terms in *C. inopinata* suggest that proteases and proteases inhibitors are important components in *C. inopinata* excretome/secretome.

KEGG pathway analysis was performed to identify the related pathways to *C. inopinata* ESPs. Majority of the result are related to metabolism and protein degradation **(Figure 4.8 and Appendix table 4.2)**. In *C. elegans*, innate immune system pathway and protein degradation pathway are the most over-represented KEGG pathway. This similarity suggests one major role of ESPs is to continuously fight with the changing environment, especially defending infection of bacteria pathogens.

Brief overview of *C. inopinata* Excretome/Secretome structure shows several protein families are main players

The most abundant proteins identified in *C. inopinata* ESPs under the standard culture condition were listed (**Table 4.1 and Figure 4.9**). *C. inopinata* excretome/secretome shared many common features with *C. elegans* ESPs. Both excretome/secretome contains proteins from lysozyme family, transthyretin-like family, lectins, proteases, and proteases inhibitors. The similarity in *C. inopinata* ESPs and *C. elegans* ESPs indicated that these two species are highly conserved in evolution.

Signal peptide detection in *C. inopinata*

The presence of a signal peptide and presence of signal peptide cleavage sites can be used to predict whether a protein is secreted or not [28], [29]. Thus, we used SignalP5.0 to predict signal peptide presence for each excreted/secreted protein in *C. inopinata [28], [29]*. 163 of 585 genes were predicted to contain a signal peptide sequence, representing 27.8% of the total identified *C. inopinata* ESP. In *C. elegans*, however, 72.3% of identified *C. elegans* ESPs were predicted to contain signal peptide. In other parasitic nematodes, 55% of *B. malayi* ESPs are predicted to have signal peptide and 41% of *B. xylophilus* ESPs are predicted to have signal peptide [13], [15], [30].

Surprisingly, the signal peptide presence is much lower in *C. inopinata* than other species. The main difference in excretome/secretome preparation between *C. inopinata* and *C. elegans* is that *C. inopinata* was cultured on plates instead of liquid culture. However, in both cases, we checked the worm viabilities before collection and followed by intensive wash to minimize the content of soluble worm lysate. Thus, the possibility of all these proteins without signal peptide are from worm lysate contamination is low.

The top GO terms showed that actins, RNA binding proteins and ribosomes are abundant in *C. inopinata* ESPs. This may be surprising and can be easily interpreted to cellular components. However, studies in either *C. elegans* ESPs and parasitic nematodes ESPs showed that the presence of actins, histones, and RNA binding proteins are common. By comparing with whole worm lysate proteomic results, these are proved to be truly enriched in excreted/secreted proteins and may be associated with exosome secretion pathway in various organisms [13], [15], [31], [32].

Literatures reported many parasitic excreted/secreted proteins do not have a classical signal peptide. These proteins may contain unknown secretory signals or they could be secreted through non-canonical secretory pathways. For example, macrophage migration inhibitory factor homolog in *B. malayi* was proven to be exported despite lacking a signal peptide [15], [30]. Helminth parasites were shown to produce exosome, carrying immunoglobulins and metabolic enzymes. Exosome proteome of *E. caproni* even explained 54% of the characterized secretome [32]–[35]. Thus, non-canonical secretion mechanism might be a common feature in nematodes. Whether the secretion of *C. inopinata* rely more the non-classical secretion needs to be further investigated.

Pfam domain search and InterproScan sequence search

In order to compare the similarities and differences between *C. inopinata* and *C. elegans*, we searched against Pfam and InterPro database to identify domain structure and protein family information for each protein sequence **(Figure 4.10 and Figure 4.11)** [22], [36], [37]. Transthyretin-like, Aspartic peptidase A1 family, C-type lectins, and glycoside hydrolase are the major protein familes in *C. inopinata*. The plot from enzyme annotation

gave back the same result **(Figure 4.12)**. Network of protein-protein interactions was also plotted. And in consistent with *C. elegans* protein-protein network, Cathepsin Z family and Cathepsin B family members are in the center position to interact with other ESPs.

In order to visualize the protein family distribution in *C. inopinata* and *C. elegans*, we took protein sequences from two species and ran all versus all BLAST. Based on sequence similarities, a BLAST map clearly showed the protein families in two species **(Figure 4.13)**. The *C. inopinata* proteins were labeled in yellow. From the map, the *C. inopinata* proteins and the *C. elegans* proteins were evenly distributed, suggesting the high similarity in the composition of excretomes/secretomes. There are two groups of proteins that showed slight differences between these two species. *C. elegans* excretome/secretome has more C-type lectins while *C. inopinata* has more cuticle collagen proteins. The ten most abuandant protein families are labeled in Figure 4.13, showing the highly conserved structures between excretomes/secretomes.

Protease and proteases inhibitors in *C. inopinata*

Peptidases contribute to host specificity, host range and virulence. To detect putative proteases (also termed peptidase or proteinases) and proteases inhibitors, a total of 585 excreted/secreted protein sequences were subjected to the MEROPS BLAST search and classified into detailed MEROPS proteases or proteases inhibitor families (E-value cutoff of 1e-4). This approach predicted 110 putative proteases and 37 proteases inhibitors, contributing to 25.1% of annotated genes in list **(Appendix table 4.3)**. The percentages of peptidases in *C. elegans, B. xylophilus, M. incognita, B. malayi* are 13.3%, 10.6%, 6.4%, and 5.2%, respectively [6,8].

Detection of Carbohydrate-active enzymes

Carbohydrate-active enzymes (CAZymes) are key regulators to metabolism of glycoconjugates, oligosaccharides and polysaccharides [38]. The size and diversity of CAZymes can provide clues to nematode nutritional strategy and host specificity [1], [13]. dbCAN (automated CAZyme annotation) database was searched to identify CAZymes in *C. inopinata* ESPs. With HMMER3 search, 17 glycoside hydrolases (GH) were identified in *C. elegans* ESPs, belonging to Glycoside Hydrolase, GlycosylTransferase and Carbohydrate Esterase families **(Table 4.2)**. The total number of identified CAZymes is very close to what is found in *C. elegans* (17 glycoside hydrolases). However, the CAZymes are more diverse in *C. inopinata*. Whether the CAZymes in *C. inopinata* have associated functions with degradation of bacterium in fig or binding to wasps needs to be further investigated.

Orthogroups between *C. inopinata* and *C. elegans* for cross-species comparison

Orthologous groups permit comparative evolutionary and functional analyses [39]. A combined 1098 genes from *C. inopinata* and *C. elegans* excretomes/secretomes were used to find orthogroups. 473 genes, which represent 43.1% of the total number of genes, were successfully assigned into 182 orthogroups with a mean orthogroup size of 2.6. There was no species-specific orthogroups, suggesting *C. inopinata* ESPs composition is very similar to *C. elegans* ESPs composition **(Table 4.3)**.

Detailed orthogroups between *C. inopinata* and *C. elegans* were provided in Appendix table 4.3 **(Appendix Table 4.3)**. Almost all orthogroups have similar numbers of proteins

across species. We can only observe a slight expansion in cuticle collagen proteins (OG0000001) and a slight shrink in pharyngeal gland toxin related proteins (OG0000006) in *C. inopinata*. In general, the orthogroups provided strong evidence that *C. inopinata* excretome/secretome share common features with *C. elegans*.

Comparisons with other parasitic nematodes ESPs

Similar comparisons of secretome profiles of *Bruga malayi, Bursaphelenchus xylophilus, Strongyloide ratti* were performed to *C. elegans* ESPs in order to investigate species differences [13], [15], [16]. These three species were picked out since their excretome/secretome profiles were of the best qualities, characterizing at least around 500 proteins. The phylogenetic relationships of nematodes were shown **(Figure 4.14, right side)**.

Orthogroups between *B. malayi* and *C. elegans*

1365 genes from *B. malayi* and *C. elegans* excretomes/secretomes were used to find orthogroups. 221 genes, which represent only 16.2% of the total number of genes, were assigned into 78 orthogroups with a mean orthogroup size of 2.8.

Among the orthogroups, we can observe *B. malayi* has expanded protein families in cuticlin and endochitinase. *C. elegans* ESPs contains expansion in lectins, trypsin inhibitors, and SCP-like extracellular proteins.

Orthogroups between *B. xylophilus* and *C. elegans*

2028 genes from *B. malayi* and *C. elegans* excretomes/secretomes were used to find orthogroups. 508 genes, which represent 25% of the total number of genes, were assigned into 179 orthogroups with a mean orthogroup size of 2.8.

5 species-specific orthogroups were identified. Two *C. elegans* specific orthogroups are lysozymes (*lys-1, lys-2, lys-7, lys-8*) and SCP-Like extracellular proteins expressed in pharyngeal gland cell (*scl-14, scl-2, scl-3, scl-5*). Three *B. xylophilus* specific orthogroups are lysozyme-like proteins, aspartic proteases and SCP-Like extracellular protein. Shared orthogroup OG0000005 contains 1 *B. xylophilus* gene and 8 *C. elegans* aspartic protease members. These observations suggest that, in general, major protein families are conserved between *B.xylophilus* and *C. elegans*. But the protein structures diverged between species and evolved to function in different life styles.

Among the shared orthogroups, we can observe that *B. xylophilus* has several expanded protein families. OG0000000 contained 24 *B. xylophilus* cysteine-type peptidases genes and 1 *C. elegans* gene. OG0000001 contained 11 *B.xylophilus* aminopeptidase genes and 1 *C. elegans* gene. OG0000004 contained 7 *B. xylophilus* and 1 *C. elegans* gene. OG0000006 contained 6 *B. xylophilus* glutathione S- transferase genes and 1 *C. elegans* gene. OG0000007 contained 5 *B. xylophilus* Neprillysin metallopeptidase family genes and 1 *C. elegans* gene. OG0000009 contained 5 *B. xylophilus* cuticlin-like genes and 1 *C. elegans* gene.

C. elegans excretome/secretome has an expansion in pharyngeal gland toxin-related gene. OG0000002 includes 9 *C. elegans* pharyngeal gland toxin-related gene and 1 *B. xylophilus* gene.

Orthogroups between *S. ratti* and *C. elegans*

946 genes from *S. ratti* and *C. elegans* excretomes/secretomes were used to find orthogroups. 297 genes, which represent 31.4% of the total number of genes, were assigned to 115 orthogroups with a mean orthogroup size of 2.6. There was no species-specific orthogroup identified.

Among the shared orthogroups, we can observe *S. ratti* has expanded protein families in transthyretin-like superfamily. *C. elegans* ESPs contains expansion in a group of hydrolases that are expressed in intestine and SCP-like extracellular proteins.

Orthogroups across 5 nematodes species

3898 genes from *Caenorhabditis elegans, Caenorhabditis inopinata, Bruga malayi, Bursaphelenchus xylophilus, Strongyloide ratti* excretomes/secretomes were combined together and used to find orthogroups [13], [16], [30]. 2059 genes, which represent 52.8% of the total number of genes, were assigned into 523 orthogroups with a mean orthogroup size of 3.9. There was no species-specific orthogroup identified.

Among the shared orthogroups, several gene families obviously expanded in one species than the others **(Appendix Table 4.4)**. Gene families expanded in *B. xylophilus* are cysteine-type peptidase, transthyretin-like superfamily and aminopeptidase. A group of immunoglobulin I-set domain containing proteins expanded in *B. malayi. C. elegans* has more C-type lectins genes than other species.

Based on the excretome/secretome orthogroups, we could also rebuild a phylogenetic tree of nematodes. The rebuilt tree **(Figure 4.14, left side)** is slightly different from species phylogenetic relationships based on small subunit (SSU) homologies **(Figure 4.14, right**

side) [13]. *B. malayi* and *S. ratti* are both vertebrate parasites and in the rebuilt ESPs tree, these two vertebrate parasites share more similarities than others. This is an interesting observation showing the composition of ESPs may be associated with life styles. This hypothesis can only be tested further if there are more well-characterized excretomes/secretomes of parasites.

4.4 Discussion

We used the same proteomic method to characterize ESPs of *C. inopinata*, the sibling species of *C. elegans*. Comparison of ESPs between two species confirmed that ESPs functional structures are similar across species and are highly conserved in terms of protein family categories and related pathways. We took a further step to expand the comparison to many characterized parasites excretomes/secretomes. Again, half of the ESPs have at least one homolog in at least one species.

How does parasitism evolve? We first take a look at the genes that failed to be assigned into orthologous groups. However, majority of the unassigned genes are uncharacterized and very little information is known about their immune regulation roles. One direction to study parasitism is to utilize this subset of unassigned genes in parasites and characterize their functions, as they may conduct novel mechanisms in evading host immune system. Secondly, we carefully examined the conserved gene groups since the general mechanism may be shared across species but parasites may evolve and make critical modifications to gain novel feature to allow infection and invasion of host. One excellent example is demonstrated by *B. malayi cpi-2*. Bm-CPI-2 possess two conserved functional domains, one to inhibit papain-like proteases and the other one to inhibit asparaginyl endopeptidase (AEP) and therefore inhibits antigen processing by human B cells. In contrast, *C. elegans* homologs of cystatins, *cpi-1* and *cpi-2*, are not functional to block AEP. Thus, *B. malayi* CPI-2 may convergently evolved to be able to function in mammalian environment [40]. Thirdly, gene duplication events for major protein families in excretome/secretome provide another possibility to allow species-specific antigen

recognition and process. For example, lysozymes and proteases are broadly present in many free-living and parasitic nematodes. Gene duplication events happened several times and may provide clues for how parasite genes diverged to gain parasitism.

In summary, the proteomic characterizations of *C. elegans* and *C. inopinata* serve as a great platform for comparative studies in parasites. First of all, many parasite genomes were poorly annotated, limiting the power to identify excreted/secreted proteins. Based on *C. elegans* ESPs, we could help to annotate the earlier hypothetical proteins in parasites and also confirm their presence in excretome/secretome. Second, studying the ESPs evolution between free-living nematodes and parasites would provide clues for how parasitism is acquired and evolved.

4.5 Materials and Methods

Excreted/Secreted Proteins Preparation

The wild type of *C. inopinata* strain, NK74SC, was a gift from Taisei Kikuchi. *C. inopinata* strain was grown using a similar protocol to standard lab maintain method of *C. elegans*. For preparing the large number of animals required for Mass-spec, 10 gravid females from a 1-week-old culture were picked to d=10cm NGM plates with pre-seeded *E. coli* OP50. 50 large NGM plates were used in each Mass-spec sample preparation batch. Plates with *C. inopinata* were kept at 25 degree and left to grow to the point of almost depleting of bacteria food. Contaminated plates were discarded and M9 buffer was used to wash *C. inopinata* worms off the rest of the plates. The collected worms were washed intensively for six times with M9 and were left in 1ml M9 in the last step and incubated at 25 degree for 4 hours. The supernatant was collected and filtered by a 0.22μm syringe filter.

NanoLC-MS/MS and data process

In-solution digested *C. inopinata* ESP samples were subjected to LC-MS/MS analysis on a nanoflow LC system, EASY-nLC 1000, (Thermo Fisher Scientific) coupled to a LTQ Orbitrap Elite mass spectrometer (Thermo Fisher Scientific, Bremen, Germany).

For the EASY-nLC 1000 system, solvent A consisted of 97.8% H2O, 2% ACN, and 0.2% formic acid and solvent B consisted of 19.8% H2O, 80% ACN, and 0.2% formic acid. Samples were directly loaded onto a 25-cm analytical HPLC column (50 μm ID) packed in-house with ReproSil-Pur C18AQ 3 μm resin (120A° pore size, Dr. Maisch, Ammerbuch,

Germany). The column was heated to 55° C. The peptides were separated with a 125min gradient at a flow rate of 220 nL/min. The gradient was as follows: 2–30% Solvent B (115 min), 30–100% B (1 min), and 100% B (9 min). Eluted peptides were then ionized using a Nanospray Flex ion source (Thermo Fisher Scientific) and introduced into the mass spectrometer. The LTQ Orbitrap Elite was operated in a data-dependent mode, automatically alternating between a full-scan (m/z 400-1600, 120K resolution) in the Orbitrap and subsequent MS/MS scans of the 20 most abundant peaks in the linear ion trap (Top20 method). Data acquisition was controlled by Xcalibur 2.2 and LTQ Tune Plus 2.7 software (Thermo Fisher Scientific).

Raw data was analyzed using MaxQuant (v. 1.5.3.30) [41], [42]. Spectra were searched against *C.inopinata* genome assembly [1] as well as a contaminant database containing common proteins like trypsin and keratins. Precursor mass tolerance was 4.5 ppm after recalibration and fragment tolerance was 0.5 Da. Carbamidomethylation of cysteine was specified as a fixed modification and protein N-terminal acetylation and oxidation of methionine were specified as variable modifications. Trypsin was the specified digestion enzyme and up to two missed cleavages were allowed. Score were thresholded so as to achieve a 1% FDR at the PSM, peptide, and protein levels as estimated by a decoy database. Match-between-runs, iBAQ, and LFQ quantitation were enabled.

BLAST

Each protein sequence in the ESP was BLASTed against NCBI using BLASTp search with the following parameters: BLAST expectation value (e-value) 1.0E-3, number of BLAST hits 20, HSP length cutoff 33.

Bioinformatics

The protein sequences were further searched against Pfam search (version 32.0) with HMMER3 [36], SignalP (version 5.0) [28], [29], InterProScan [22], dbCAN2 for automated Carbonhydrate-active enzyme annotation [38]. Gene Ontology (GO) terms of molecular function, cellular component and biological process were assigned by transferring the GO terms from the top BLAST hit, mainly *C. elegans* orthologues. Blast2GO was used to add more GO and InterProScan results [5], [6].

BLAST map

The *C. elegans* and *C. inopinata* ESPs BLAST map was created similarly to [43]. All-to-all BLAST [44] was performed with *C. elegans* ESPs and *C. inopinata* identified in this work, with a threshold of E-value < 0.1. The map was created using igraph package in R [45]. Edges between nodes (proteins) were plotted as percent identity scores, scaled between 0 and 1, in a force-directed graph. Nodes without edges were removed from the graphs. Annotations for protein classes were pulled from WormBase [46] and Uniprot [47].

Orthologues and phylogenetic tree construction

In order to compare orthologues between two species, Orthofinder version 2.3.1 was used to assign proteins from two species into different orthologous groups [39]. Maximum-likehood phylogenetic trees were constructed to interesting gene families by first doing multiple sequence alignment using EMBL-EBI MUSCLE [48]. The aligned sequences

were further trimmed by Trimal (v1.4.rev15) [49]; FastTree (2.1.7) [50], [51] and FigTree 1.4 [52] were followed to construct and view the phylogenetic trees.

4.6 Figures

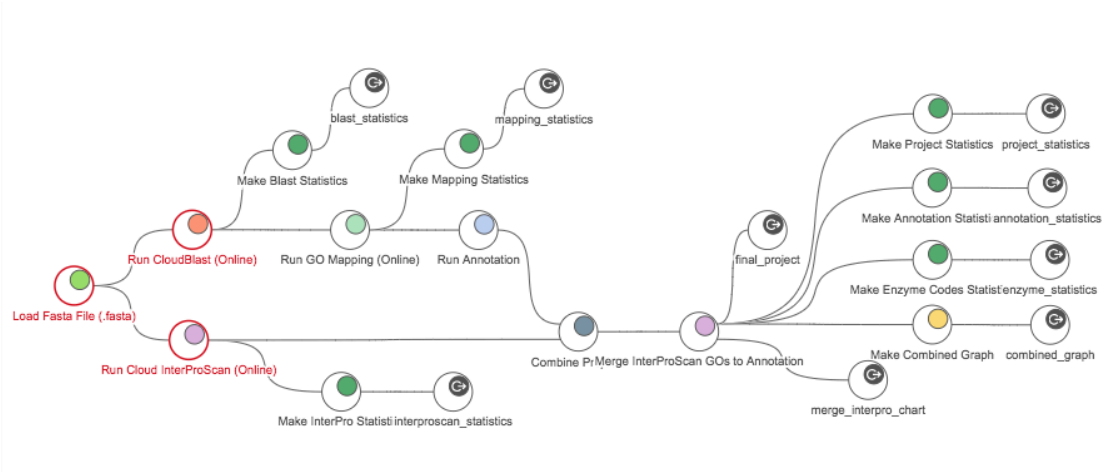


Figure 4.1 Workflow of functional annotation of *C. inopinata* ESPs. Each entry was first BLASTed against NCBI then followed by searching InterProScan and Gene Ontology. The results from different databases were mapped and merged.

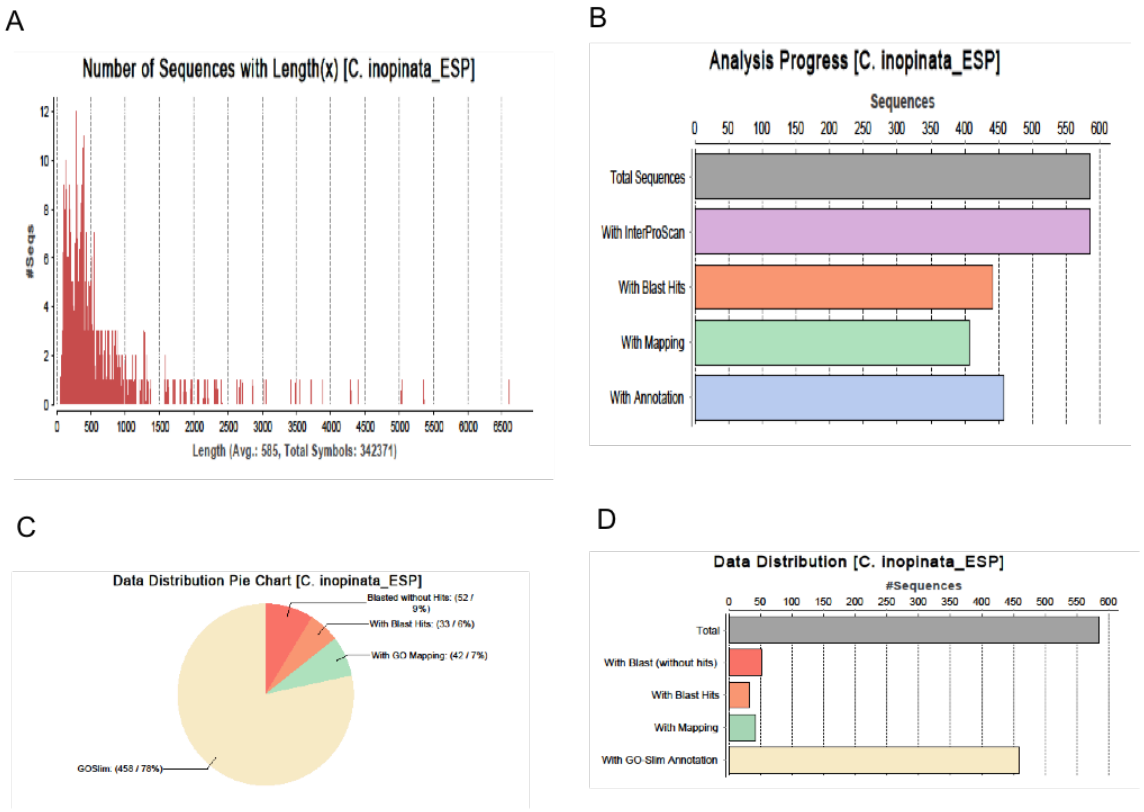
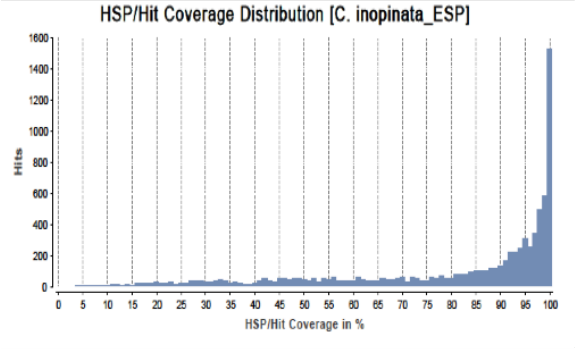
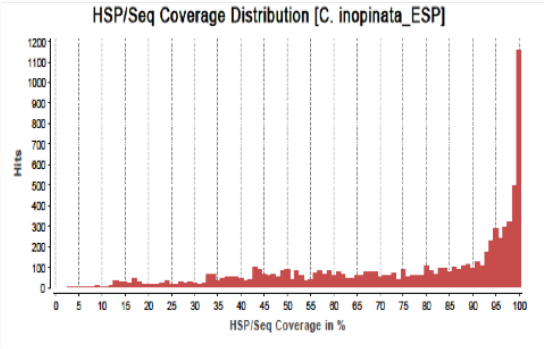


Figure 4.2 Statistics for annotation levels of *C. inopinata* ESPs. (A) Length distribution of Mass-Spec identified excreted/secreted proteins. (B) Number of sequences in each level of annotations. (C) Pie chart of sequences annotated to different levels. (D) Bar chart of sequences annotated to different levels.

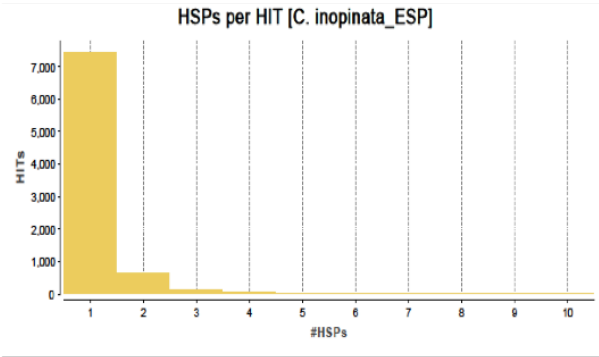
A


HSP/Hit Coverage Distribution [C. inopinata_ESP]
Hits
HSP/Hit Coverage in %

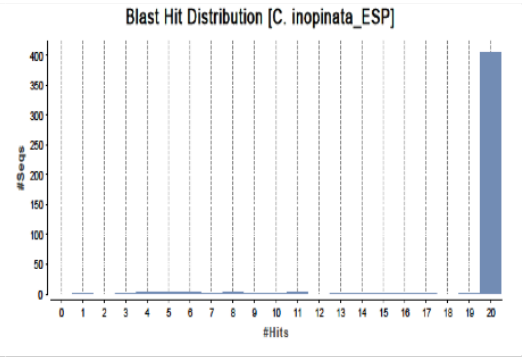
B


HSP/Seq Coverage Distribution [C. inopinata_ESP]
Hits
HSP/Seq Coverage in %

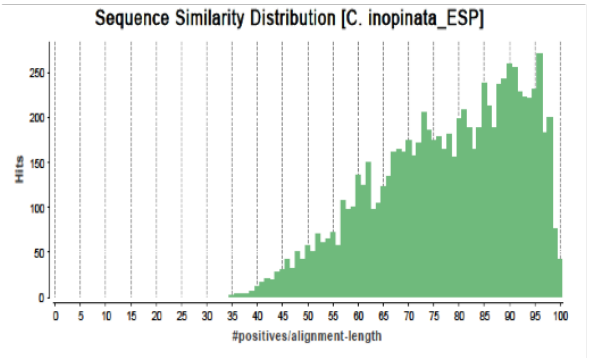
C


HSPs per HIT [C. inopinata_ESP]
HITs
#HSPs

D


Blast Hit Distribution [C. inopinata_ESP]
#Seqs
#Hits

E


Sequence Similarity Distribution [C. inopinata_ESP]
Hits
#positives/alignment-length

F

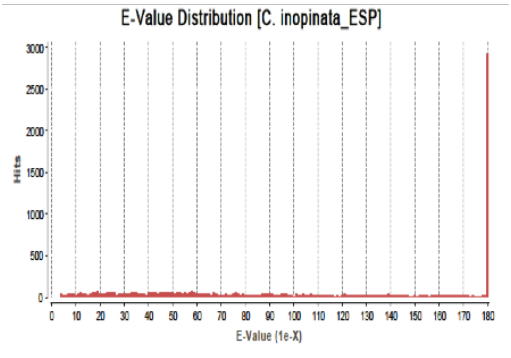

E-Value Distribution [C. inopinata_ESP]
Hits
E-Value (1e-X)

Figure 4.3 Statistics for BLAST performed to each *C. inopinata* ESP. (A) High-scoring Segment Pair (HSP)/Hit coverage distribution. (B) High-scoring Segment Pair (HSP)/Seq Coverage distribution. (C) Statistic for number of HSP in each Hit. (D) Statistic for number of Hits returned in each sequence. (E) Statistic for positive/alignment-length in each Hit. (F) Statistic for e-value for each Hit.

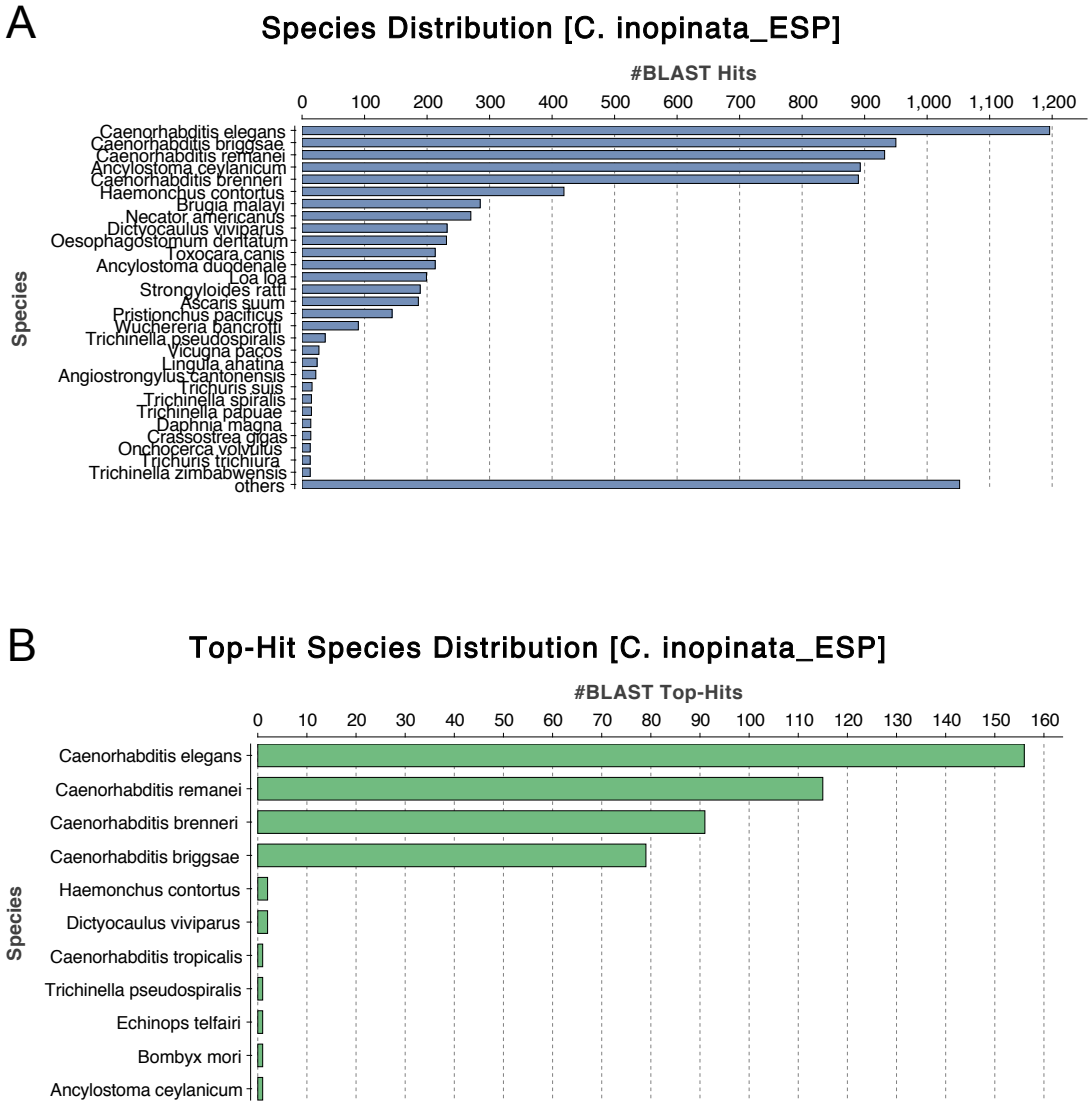
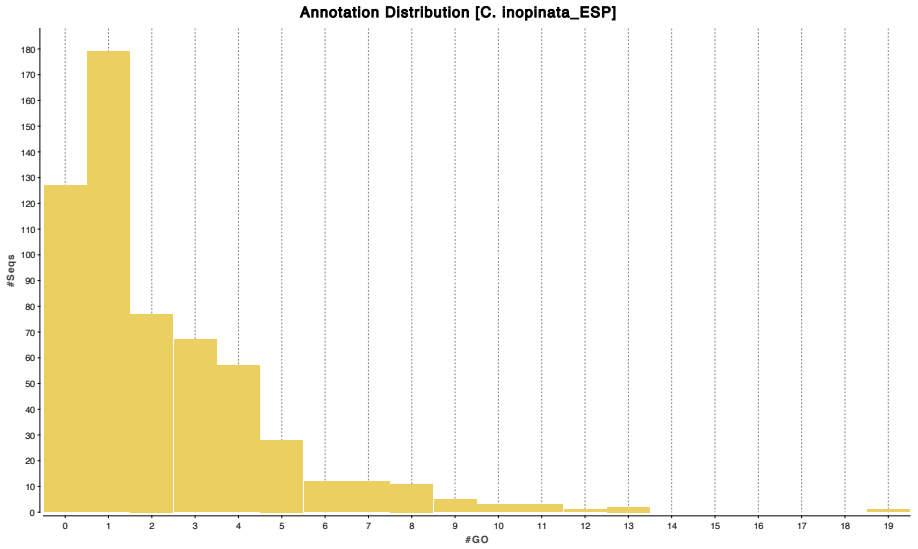


Figure 4.4 Species distribution of BLAST results of *C. inopinata* ESPs. (A) Species distribution for all BLAST Hits. (B) Species distribution for only top BLAST Hits.

A



B

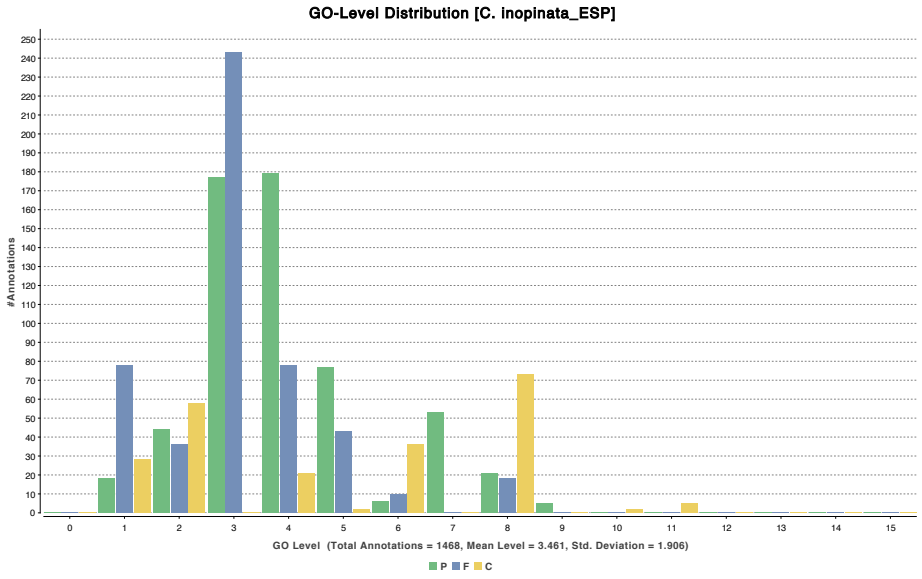


Figure 4.5 GO terms distribution of *C. inopinata* ESPs (A) Statistics for number of GO terms annotated to each sequence (B) GO-level distribution. P = biological process, F = molecular function, C = cellular component.

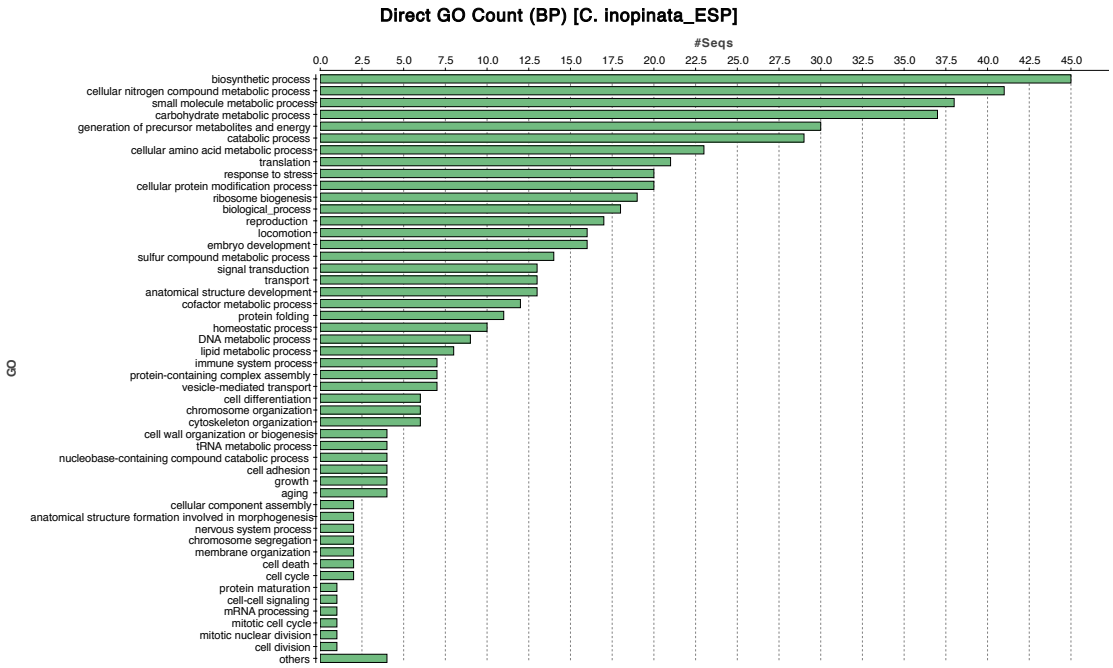


Figure 4.6 Count of biological process GO terms of *C. inopinata* ESPs

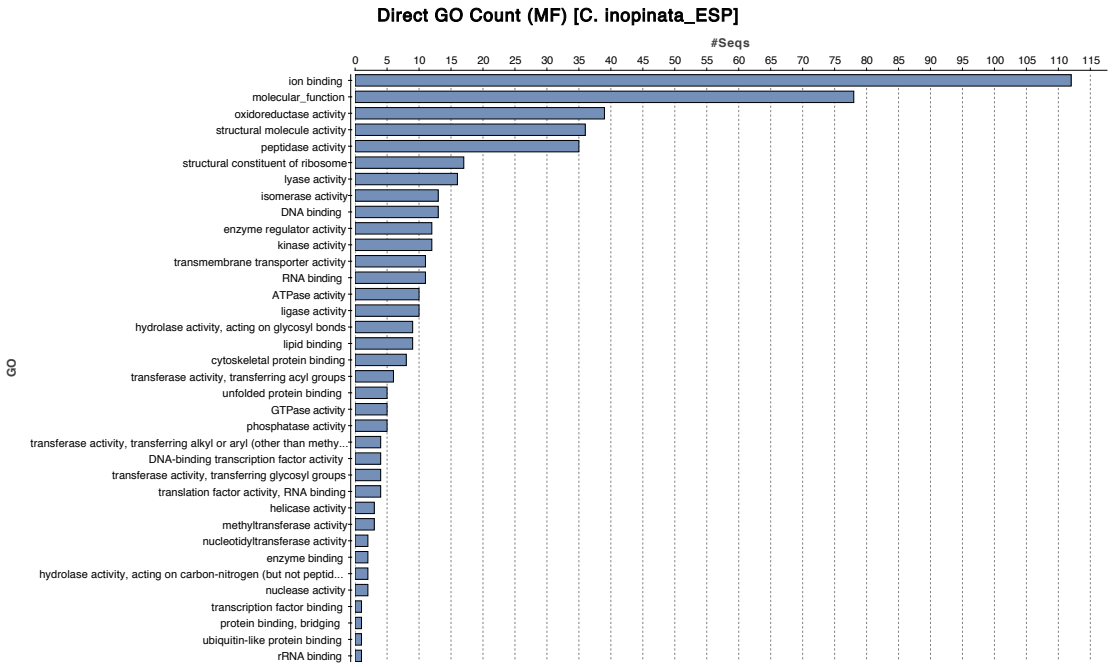


Figure 4.7 Count of molecular function GO terms in *C. inopinata* ESPs

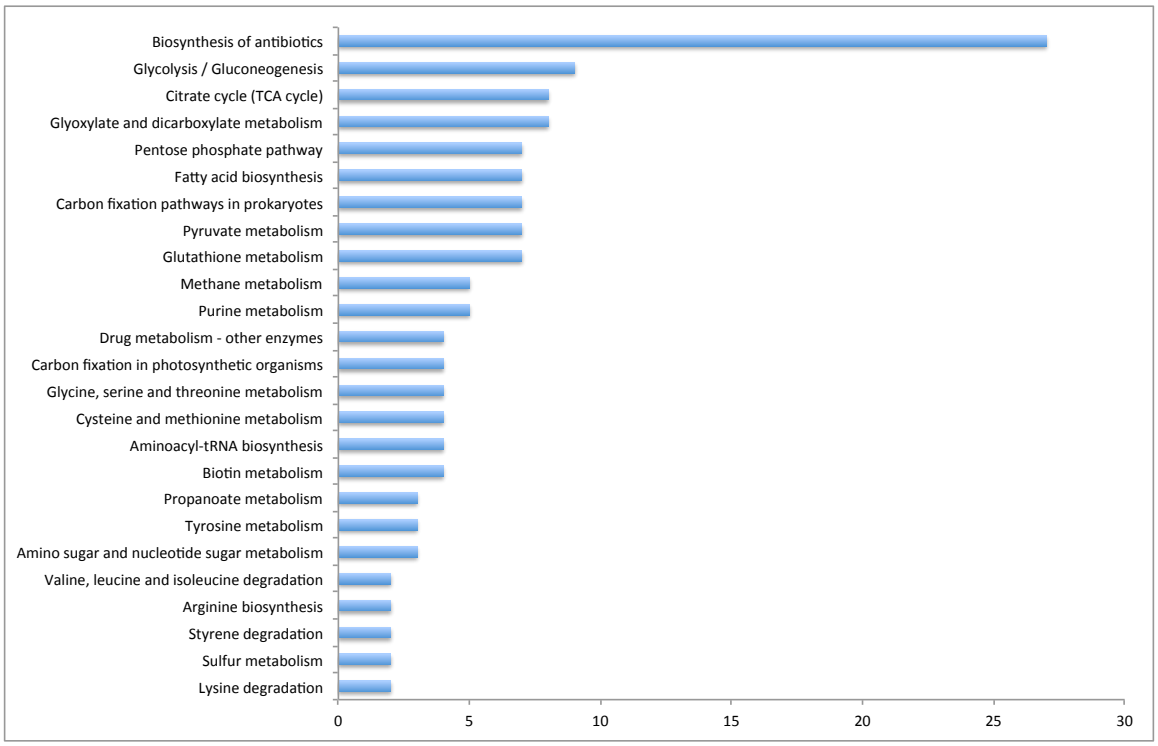


Figure 4.8 KEGG pathways annotated to *C. inopinata* ESPs

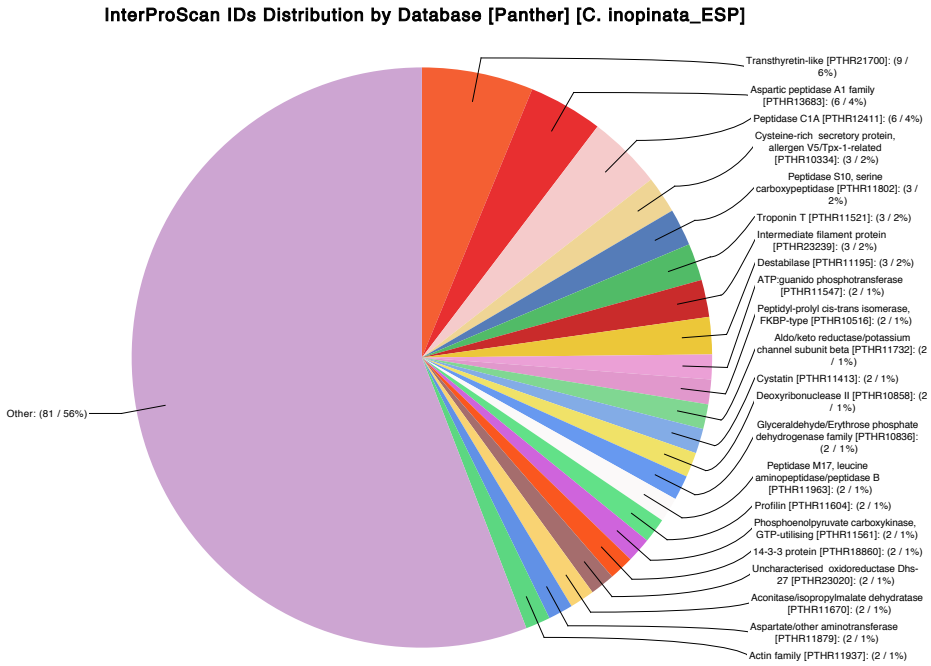


Figure 4.9 Structure of *C. inopinata* Excretome/Secretome. Pie chart described the distribution of protein families

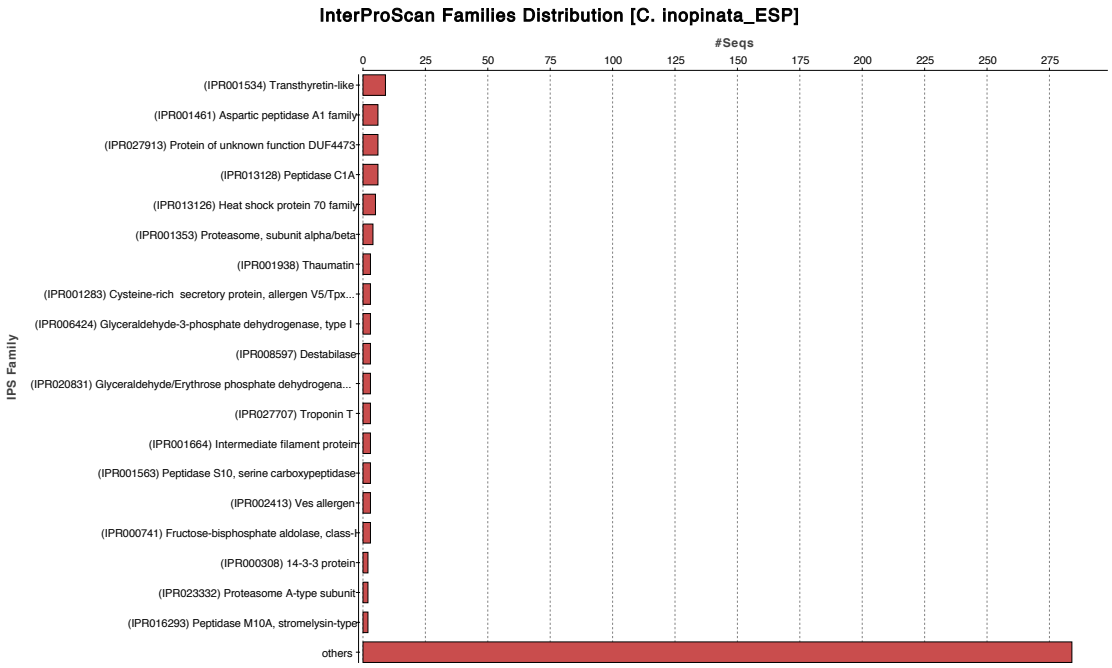


Figure 4.10 InterPro family distribution of annotated *C. inopinata* ESPs.

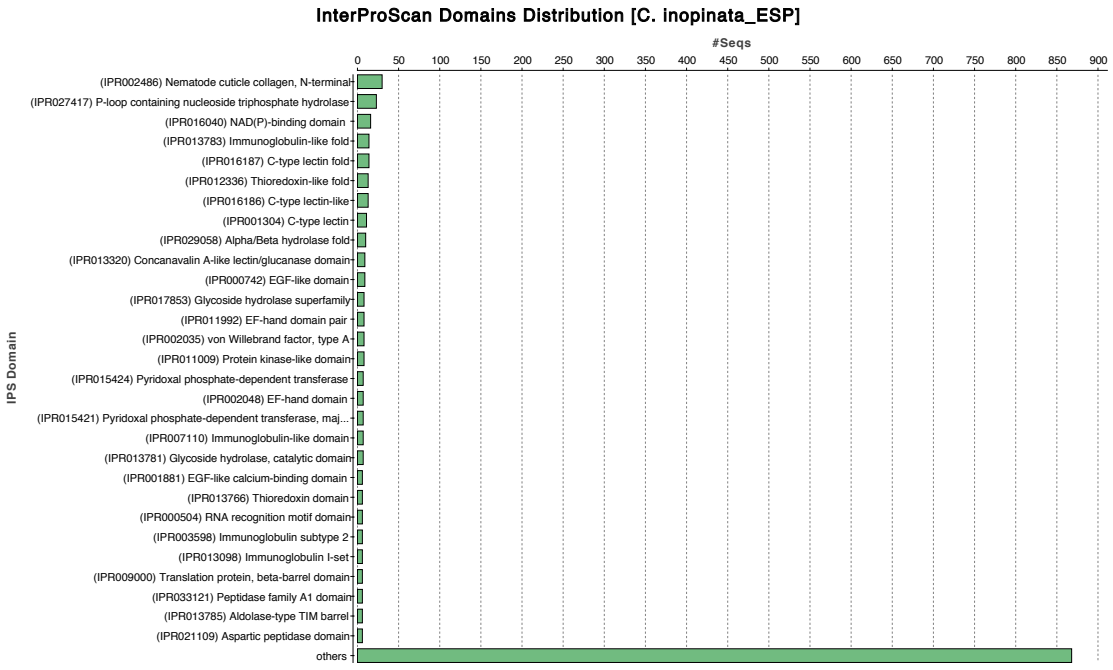


Figure 4.11 InterPro domain distribution of annotated *C. inopinata* ESPs.

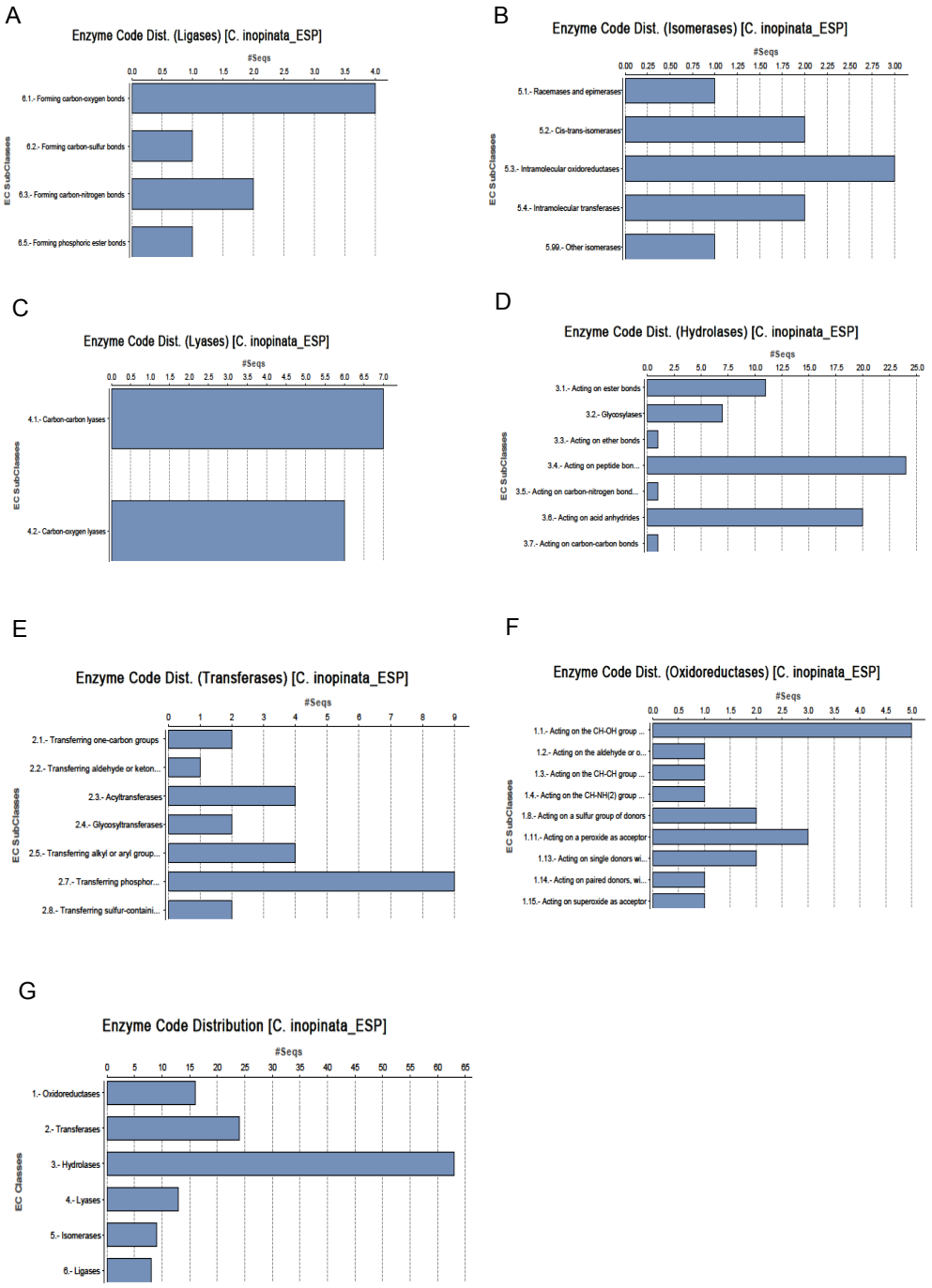


Figure 4.12 Annotated enzymes in *C. inopinata* excretome/secretome. (A-F) detailed distribution for each identified enzyme class. A, ligases. B, isomerases. C, lyases. D, Hydrolases. E, Transferases. F, Oxidoreductases. (G) Distribution of six identified enzyme classes in C. inopinata excretome/secretome

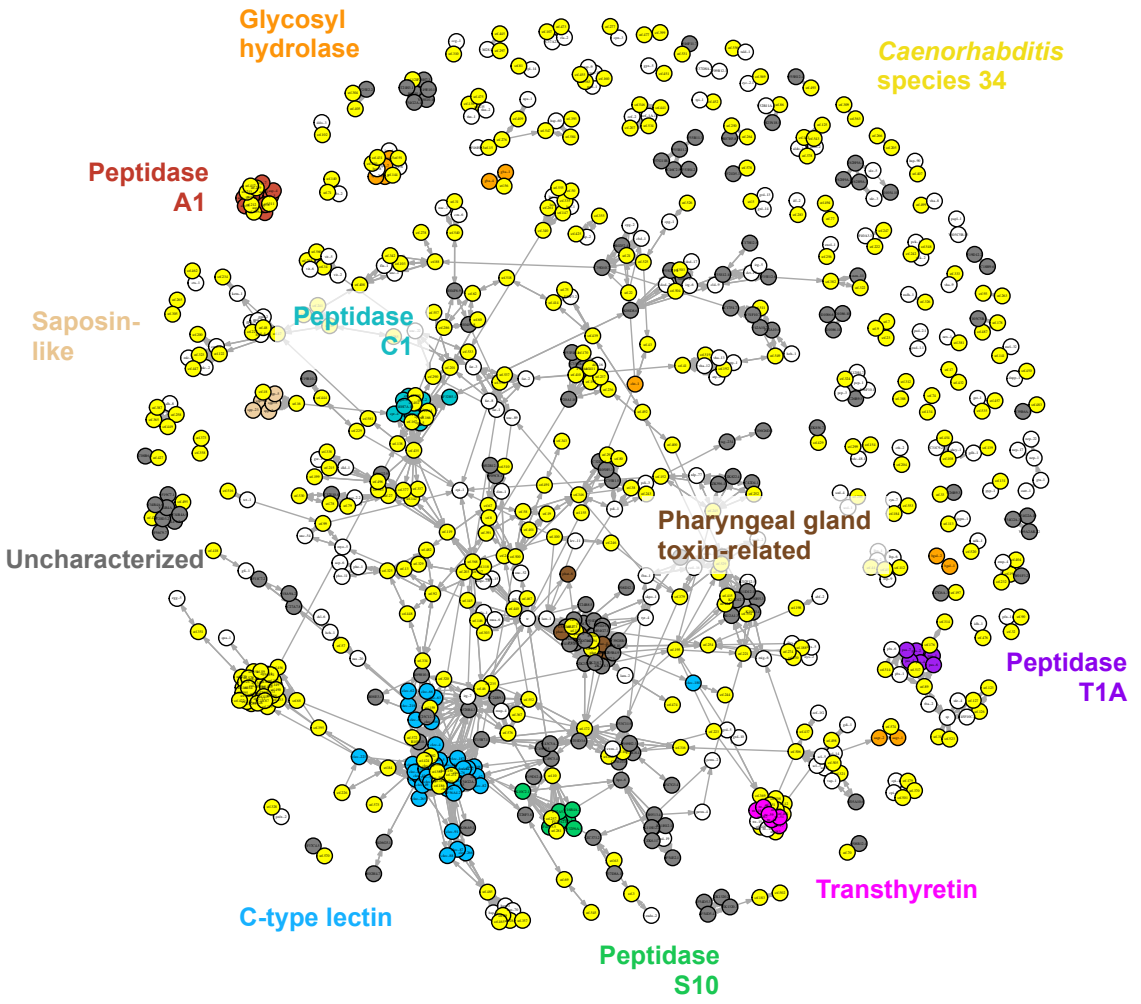


Figure 4.13 BLAST map of combined *C. inopinata* ESPs and *C. elegans* ESPs. All ESP sequences from two species were BLASTed in all-against-all manner. Edges indicate similarity by BLAST, with an E-value < 0.1. Proteins of *C. inopinata* origin were yellow dots and Proteins of *C. elegans* origin were circles. Color-coded are the ten protein classes with the highest numbers among the ESPs.

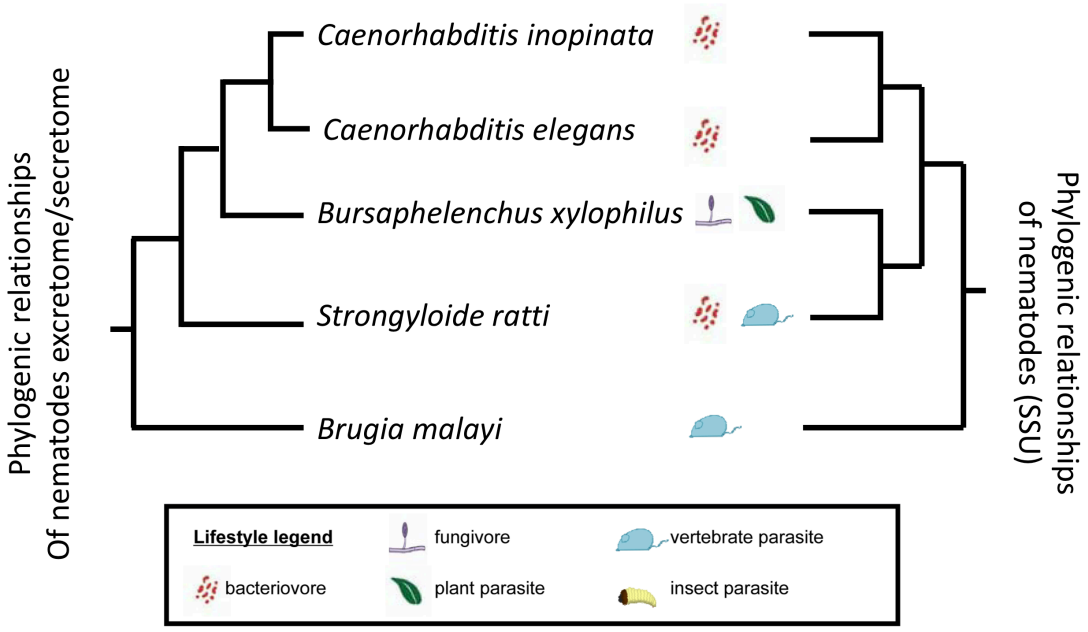


Figure 4.14 Phylogenetic relationships of nematodes based on ESPs similarities and SSU. Left side is the reconstructed phylogenetic tree of five nematodes based on excretome/secretome similarities. Right side is the phylogenetic tree constructed by comparing small subunit (SSU). This figure is modified from [13].

4.7 Tables

SeqName	Description	GONames
Sp34_scaffold1.g1606.t1	---NA---	F:kinase activity
Sp34_scaffold1.g286.t1	nematode cuticle collagen domain	F:structural constituent of cuticle
Sp34_scaffold1.g521.t1	Hypothetical protein CBG17344	C:nucleus
Sp34_scaffold10.g799.t1	lipocalin cytosolic fatty-acid binding	F:transporter activity; P:transport; F:lipid binding
Sp34_scaffold11.g164.t1	hypothetical protein CRE_27387	C:integral component of membrane
Sp34_scaffold11.g641.t1	14-3-3 zeta	F:protein domain specific binding
Sp34_scaffold12.g432.t1	Ani s 9 allergen precursor	P:response to gamma radiation
Sp34_scaffold12.g8.t1	Fructose-bisphosphate aldolase 2	F:fructose-bisphosphate aldolase activity
Sp34_scaffold13.g705.t1	cystatin domain	F:cysteine-type endopeptidase inhibitor activity
Sp34_scaffold14.g122.t1	peptidyl-prolyl cis-trans cyclophilin-type	P:protein peptidyl-prolyl isomerization; P:protein folding
Sp34_scaffold16.g122.t1	---NA---	F:aspartic-type endopeptidase activity; P:proteolysis
Sp34_scaffold17.g183.t1	Invertebrate LYSozyme	F:lysozyme activity; P:carbohydrate metabolic process
Sp34_scaffold17.g64.t1	DAF-16 FOXO germline Tumor affecting	
Sp34_scaffold18.g27.t1	Phosphoenolypyruvate inase	F:phosphoenolpyruvate carboxykinase (GTP) activity
Sp34_scaffold2.g513.t1	translation initiation factor eIF-5A	F:translation elongation factor activity; C:ribosome
Sp34_scaffold3.g1197.t1	nematode cuticle collagen domain	F:structural constituent of cuticle
Sp34_scaffold3.g521.t1	---NA---	F:structural constituent of cuticle
Sp34_scaffold4.g1259.t1	major allergen	C:extracellular region
Sp34_scaffold4.g194.t1	---NA---	
Sp34_scaffold4.g202.t1	---NA---	
Sp34_scaffold4.g400.t1	COLlagen	F:structural constituent of cuticle
Sp34_scaffold4.g742.t1	collagen	F:structural constituent of cuticle
Sp34_scaffold4.g945.t1	muscle actin	F:isopentenyl-diphosphate delta-isomerase activity;
Sp34_scaffold5.g421.t1	---NA---	C:Golgi apparatus
Sp34_scaffold6.g436.t1	heat shock 70	
Sp34_scaffold7.g282.t1	MD-2-related lipid-recognition domain-containing	
Sp34_scaffold7.g656.t1	Major sperm 19 31 40 45 50 51 53 59 61 65 81 113 142	
Sp34_scaffold7.g776.t1	nematode cuticle collagen domain	F:structural constituent of cuticle
Sp34_scaffold8.g126.t1	---NA---	F:translation elongation factor activity;
Sp34_scaffold8.g185.t1	---NA---	F:lipid binding

Table 4.1 The top 30 abundant excreted/secreted proteins identified in Mass-Spec BLASTp search against NCBI was done to each protein. GO terms was assigned by Blast2GO. F, molecular function; P, biological process; C, cellular component.

Gene_ID	HMMER	Hotpep	DIAMOND	Signalp	#ofTools
Sp34_scaffold1.g2282.t1	GH18(56-404)	GH18	CBM14	Y(1-25)	3
Sp34_scaffold1.g2610.t1	GH30_1(89-487)	GH30	GH30_1	N	3
Sp34_scaffold1.g2795.t1	GT31(80-264)	GT31	GT31	N	3
Sp34_scaffold10.g308.t1	GH25(37-207)	N	N	Y(1-27)	1
Sp34_scaffold11.g356.t1	CE10(495-750)	N	N	N	1
Sp34_scaffold11.g530.t1	GH152(26-233)	N	GH152	N	2
Sp34_scaffold11.g531.t1	GH152(26-233)	N	GH152	Y(1-25)	2
Sp34_scaffold17.g112.t1	GT35(145-839)	GT35	GT35	N	3
Sp34_scaffold3.g321.t1	GH13_8(558-851)	GH13+CBM48	CBM48	N	3
Sp34_scaffold4.g1177.t1	GT1(375-559)+GT1(985-1156)+GT1(1464-1635)	N	GT1	N	2
Sp34_scaffold4.g26.t1	CE10(71-231)	N	N	N	1
Sp34_scaffold4.g646.t1	GH19(59-361)	GH19	GH19	Y(1-24)	3
Sp34_scaffold6.g140.t1	GH152(59-231)	N	N	N	1
Sp34_scaffold9.g638.t1	GH31(69-572)	GH31	GH31	N	3
Sp34_scaffold17.g183.t1	N	GH22	N	Y(1-19)	1
Sp34_scaffold2.g18.t1	N	GH22	N	Y(1-19)	1
Sp34_scaffold9.g503.t1	N	GT1	N	N	1
Sp34_scaffold10.g536.t1	N	N	GT13	N	1
Sp34_scaffold12.g107.t1	N	N	GH25	N	1
Sp34_scaffold19.g285.t1	N	N	CBM13	N	1
Sp34_scaffold1.g1713.t1	N	N	CBM14	N	1
Sp34_scaffold4.g685.t1	N	N	GH25	N	1
Sp34_scaffold7.g104.t1	N	N	GT22	N	1
Sp34_scaffold9.g363.t1	N	N	GT30	N	1

Table 4.2 The list of carbohydrate active enzymes in the excretome/secretome of *C. inopinata*

Average number of genes per-species in orthogroup	Number of orthogroups	Percentage of orthogroups	Number of genes	Percentage of genes
<1	0	0	0	0
'1	157	86.3	335	70.8
'2	17	9.3	74	15.6
'3	5	2.7	33	7
'4	1	0.5	9	1.9
'5	1	0.5	10	2.1
'6	1	0.5	12	2.5
'7	0	0	0	0
'8	0	0	0	0
'9	0	0	0	0
'10	0	0	0	0
10-15	0	0	0	0
16-20	0	0	0	0
21-50	0	0	0	0
51-100	0	0	0	0
101-150	0	0	0	0
151-200	0	0	0	0
201-500	0	0	0	0
501-1000	0	0	0	0
'1001+	0	0	0	0

Number of species in orthogroup	Number of orthogroups
1	0
2	182

Table 4.3 Statistics of orthogroup assignments between *C. inopinata* and *C. elegans*

	Bursaphelenchus xylophilus_ESP	Bursaphelenchus xylophilus_ESP	Celegans_ESP	Celegans_ESP	Total
OG0000000	BUX.c07587.1, BUX.s00083.30, BUX.s00083.31, BUX.s00083.32, BUX.s00116.607, BUX.s00713.1009, BUX.s00713.1010, BUX.s00713.1011, BUX.s00713.1016, BUX.s00713.1034, BUX.s00713.1076, BUX.s00813.52, BUX.s00813.53, BUX.s00813.54, BUX.s01063.86, BUX.s01109.631, BUX.s01147.156, BUX.s01147.175, BUX.s01147.176, BUX.s01147.177, BUX.s01147.64, BUX.s01259.43, BUX.s01259.45, BUX.s01288.15	24	CELE_Y40H7A.10	1	25
OG0000001	BUX.s00508.67, BUX.s00508.71, BUX.s00974.24, BUX.s01198.139, BUX.s01198.141, BUX.s01198.143, BUX.s01281.328, BUX.s01281.329, BUX.s01337.113, BUX.s01337.114, BUX.s01653.360	11	CELE_T16G12.1	1	12
OG0000002	BUX.s00110.75	1	phat-1, phat-2, C14C6.2, C14C6.5, C54D10.3, CELE_F15A4.6, CELE_F41G3.10, CELE_F48G7.5, ZK673.1,	9	10
OG0000003	BUX.s00560.3, BUX.s00609.105, BUX.s01063.177, BUX.s01063.193, BUX.s01063.6, BUX.s01063.7	6	ttr-18, ttr-59	2	8
OG0000004	BUX.s00116.457, BUX.s00647.119, BUX.s00647.122, BUX.s00961.40, BUX.s00961.41, BUX.s00961.42, BUX.s01254.333	7	gst-1	1	8
OG0000005	BUX.s01143.209	1	asp-1, asp-10, asp-12, asp-13, asp-2, asp-5, asp-6	7	8
OG0000006	BUX.s00116.908, BUX.s00139.52, BUX.s01092.201, BUX.s01281.110, BUX.s01281.237, BUX.s01281.52	6	msp-49	1	7
OG0000007	BUX.s00139.135, BUX.s01063.66, BUX.s01063.80, BUX.s01661.62, BUX.s01661.67	5	nep-17	1	6
OG0000008	BUX.s01143.109, BUX.s01143.142, BUX.s01281.44	3	ttr-44, ttr-45, ttr-51	3	6
OG0000009	BUX.s00333.141, BUX.s00422.384, BUX.s01078.13, BUX.s01109.44, BUX.s01109.45	5	cutl-16	1	6
OG0000010	BUX.s00252.35, BUX.s00333.33	2	C10G8.4, C25E10.8, CELE_F53C11.9, swm-1	4	6
OG0000011	BUX.s01268.52, BUX.s01662.95	2	vit-2, vit-5, vit-6	3	5
OG0000012	BUX.s01063.203, BUX.s01063.204, BUX.s01063.205, BUX.s01063.206	4	F37H8.5	1	5
OG0000013	BUX.c00054.1, BUX.s01066.1, BUX.s01066.3, BUX.s01066.8, BUX.s01505.4	5		0	5
OG0000014	BUX.s01281.46	1	gpd-1, gpd-2, gpd-3, gpd-4	4	5
OG0000015	BUX.s01038.28, BUX.s01653.296, BUX.s01653.306, BUX.s01653.308	4	col-90	1	5
OG0000016	BUX.s01109.74	1	C39D10.7, CELE_R02F2.4, cpg-1, cpg-2	4	5
OG0000017	BUX.s00351.323, BUX.s00351.324, BUX.s00351.325	3	pdi-1	1	4
OG0000018	BUX.s00422.202	1	far-1, far-2, far-3	3	4
OG0000019	BUX.s00579.461, BUX.s00713.520, BUX.s01656.92	3	anc-1	1	4

Table 4.4 Top 20 orthogroups between *B. xylophilus* and *C. elegans*. Numbers represent the count of genes in each category.

4.8 References:

[1] N. Kanzaki *et al.*, "Biology and genome of a newly discovered sibling species of Caenorhabditis elegans," *Nat Commun*, vol. 9, no. 1, p. 3216, 10 2018.

[2] G. C. Woodruff and P. C. Phillips, "Field studies reveal a close relative of C. elegans thrives in the fresh figs of Ficus septica and disperses on its Ceratosolen pollinating wasps," *BMC Ecology*, vol. 18, no. 1, p. 26, Aug. 2018.

[3] M. Ashburner *et al.*, "Gene ontology: tool for the unification of biology. The Gene Ontology Consortium," *Nat. Genet.*, vol. 25, no. 1, pp. 25–29, May 2000.

[4] S. Carbon *et al.*, "AmiGO: online access to ontology and annotation data," *Bioinformatics*, vol. 25, no. 2, pp. 288–289, Jan. 2009.

[5] A. Conesa, S. Götz, J. M. García-Gómez, J. Terol, M. Talón, and M. Robles, "Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research," *Bioinformatics*, vol. 21, no. 18, pp. 3674–3676, Sep. 2005.

[6] A. Conesa and S. Götz, "Blast2GO: A Comprehensive Suite for Functional Analysis in Plant Genomics," *Int J Plant Genomics*, vol. 2008, 2008.

[7] M. Kanehisa and S. Goto, "KEGG: Kyoto Encyclopedia of Genes and Genomes," *Nucleic Acids Res*, vol. 28, no. 1, pp. 27–30, Jan. 2000.

[8] M. Kanehisa, M. Furumichi, M. Tanabe, Y. Sato, and K. Morishima, "KEGG: new perspectives on genomes, pathways, diseases and drugs," *Nucleic Acids Res.*, vol. 45, no. D1, pp. D353–D361, 04 2017.

[9] M. Kanehisa, Y. Sato, M. Kawashima, M. Furumichi, and M. Tanabe, "KEGG as a reference resource for gene and protein annotation," *Nucleic Acids Res*, vol. 44, no. D1, pp. D457–D462, Jan. 2016.

[10] V. L. Hunt *et al.*, "The genomic basis of parasitism in the *Strongyloides* clade of nematodes," *Nature Genetics*, vol. 48, no. 3, pp. 299–307, Mar. 2016.

[11] M. Crook, "The dauer hypothesis and the evolution of parasitism: 20 years on and still going strong," *Int. J. Parasitol.*, vol. 44, no. 1, pp. 1–8, Jan. 2014.

[12] C. Dieterich and R. J. Sommer, "How to become a parasite - lessons from the genomes of nematodes," *Trends Genet.*, vol. 25, no. 5, pp. 203–209, May 2009.

[13] R. Shinya, H. Morisaka, T. Kikuchi, Y. Takeuchi, M. Ueda, and K. Futai, "Secretome Analysis of the Pine Wood Nematode Bursaphelenchus xylophilus Reveals the Tangled Roots of Parasitism and Its Potential for Molecular Mimicry," *PLoS One*, vol. 8, no. 6, Jun. 2013.

[14] S. Bellafiore, Z. Shen, M.-N. Rosso, P. Abad, P. Shih, and S. P. Briggs, "Direct identification of the Meloidogyne incognita secretome reveals proteins with host cell reprogramming potential," *PLoS Pathog.*, vol. 4, no. 10, p. e1000192, Oct. 2008.

[15] S. Bennuru, R. Semnani, Z. Meng, J. M. C. Ribeiro, T. D. Veenstra, and T. B. Nutman, "Brugia malayi Excreted/Secreted Proteins at the Host/Parasite Interface: Stage- and Gender-Specific Proteomic Profiling," *PLoS Negl Trop Dis*, vol. 3, no. 4, Apr. 2009.

[16] H. Soblik *et al.*, "Life Cycle Stage-resolved Proteomic Analysis of the Excretome/Secretome from Strongyloides ratti—Identification of Stage-specific Proteases," *Mol Cell Proteomics*, vol. 10, no. 12, Dec. 2011.

[17] L. C. Clark and J. Hodgkin, "Commensals, probiotics and pathogens in the Caenorhabditis elegans model," *Cellular Microbiology*, vol. 16, no. 1, pp. 27–38, 2014.

[18] J. E. E. U. Hellberg, M. A. Matilla, and G. P. C. Salmond, “The broad-spectrum antibiotic, zeamine, kills the nematode worm Caenorhabditis elegans,” *Front. Microbiol.*, vol. 6, 2015.

[19] H. Schulenburg and M.-A. Félix, “The Natural Biotic Environment of Caenorhabditis elegans,” *Genetics*, vol. 206, no. 1, pp. 55–86, May 2017.

[20] B. S. Samuel, H. Rowedder, C. Braendle, M.-A. Félix, and G. Ruvkun, “Caenorhabditis elegans responses to bacteria from its natural habitats,” *PNAS*, vol. 113, no. 27, pp. E3941–E3949, Jul. 2016.

[21] H. Mi *et al.*, “PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements,” *Nucleic Acids Res.*, vol. 45, no. D1, pp. D183–D189, 04 2017.

[22] S. Hunter *et al.*, “InterPro: the integrative protein signature database,” *Nucleic Acids Res*, vol. 37, no. Database issue, pp. D211–D215, Jan. 2009.

[23] R. Eberle *et al.*, “Isolation, identification and functional profile of excretory–secretory peptides from Onchocerca ochengi,” *Acta Tropica*, vol. 142, pp. 156–166, Feb. 2015.

[24] “Needles in the EST Haystack: Large-Scale Identification and Analysis of Excretory-Secretory (ES) Proteins in Parasitic Nematodes Using Expressed Sequence Tags (ESTs).”.

[25] J. Geary *et al.*, “First analysis of the secretome of the canine heartworm, Dirofilaria immitis,” *Parasites & Vectors*, vol. 5, no. 1, p. 140, Jul. 2012.

[26] M. W. Robinson, R. Greig, K. A. Beattie, D. J. Lamont, and B. Connolly, “Comparative analysis of the excretory-secretory proteome of the muscle larva of

Trichinella pseudospiralis and Trichinella spiralis," *Int. J. Parasitol.*, vol. 37, no. 2, pp. 139–148, Feb. 2007.

[27] M. W. Robinson and B. Connolly, "Proteomic analysis of the excretory-secretory proteins of the Trichinella spiralis L1 larva, a nematode parasite of skeletal muscle," *Proteomics*, vol. 5, no. 17, pp. 4525–4532, Nov. 2005.

[28] Y. M. Harcus *et al.*, "Signal sequence analysis of expressed sequence tags from the nematode Nippostrongylus brasiliensis and the evolution of secreted proteins in parasites," *Genome Biol*, vol. 5, no. 6, p. R39, 2004.

[29] J. J. Almagro Armenteros *et al.*, "SignalP 5.0 improves signal peptide predictions using deep neural networks," *Nat. Biotechnol.*, vol. 37, no. 4, pp. 420–423, 2019.

[30] J. P. Hewitson *et al.*, "The secretome of the filarial parasite, Brugia malayi: proteomic profile of adult excretory-secretory products," *Mol. Biochem. Parasitol.*, vol. 160, no. 1, pp. 8–21, Jul. 2008.

[31] C. M. Atyame Nten *et al.*, "Excreted/Secreted Proteins from Trypanosome Procyclic Strains," *J Biomed Biotechnol*, vol. 2010, 2010.

[32] C. Olver and M. Vidal, "Proteomic analysis of secreted exosomes," *Subcell. Biochem.*, vol. 43, pp. 99–131, 2007.

[33] A. H. Buck *et al.*, "Exosomes secreted by nematode parasites transfer small RNAs to mammalian cells and modulate innate immunity," *Nature Communications*, vol. 5, p. 5488, Nov. 2014.

[34] F. Riaz and G. Cheng, "Exosome-like vesicles of helminths: implication of pathogenesis and vaccine development," *Ann Transl Med*, vol. 5, no. 7, Apr. 2017.

[35] M. Siles-Lucas, R. Morchon, F. Simon, and R. Manzano-Roman, "Exosome-transported microRNAs of helminth origin: new tools for allergic and autoimmune diseases therapy?," *Parasite Immunol.*, vol. 37, no. 4, pp. 208–214, Apr. 2015.

[36] S. El-Gebali *et al.*, "The Pfam protein families database in 2019," *Nucleic Acids Res*, vol. 47, no. D1, pp. D427–D432, Jan. 2019.

[37] S. El-Gebali *et al.*, "The Pfam protein families database in 2019," *Nucleic Acids Res*, vol. 47, no. D1, pp. D427–D432, Jan. 2019.

[38] Y. Yin, X. Mao, J. Yang, X. Chen, F. Mao, and Y. Xu, "dbCAN: a web resource for automated carbohydrate-active enzyme annotation," *Nucleic Acids Res*, vol. 40, no. Web Server issue, pp. W445–W451, Jul. 2012.

[39] D. M. Emms and S. Kelly, "OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy," *Genome Biology*, vol. 16, no. 1, p. 157, Aug. 2015.

[40] J. Murray, B. Manoury, A. Balic, C. Watts, and R. M. Maizels, "Bm-CPI-2, a cystatin from Brugia malayi nematode parasites, differs from Caenorhabditis elegans cystatins in a specific site mediating inhibition of the antigen-processing enzyme AEP," *Mol. Biochem. Parasitol.*, vol. 139, no. 2, pp. 197–203, Feb. 2005.

[41] J. Cox and M. Mann, "MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification," *Nature Biotechnology*, vol. 26, no. 12, pp. 1367–1372, Dec. 2008.

[42] J. Cox, N. Neuhauser, A. Michalski, R. A. Scheltema, J. V. Olsen, and M. Mann, "Andromeda: a peptide search engine integrated into the MaxQuant environment," *J. Proteome Res.*, vol. 10, no. 4, pp. 1794–1805, Apr. 2011.

[43] G. Jékely, “Global view of the evolution and diversity of metazoan neuropeptide signaling,” *Proc. Natl. Acad. Sci. U.S.A.*, vol. 110, no. 21, pp. 8702–8707, May 2013.

[44] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, “Basic local alignment search tool,” *J. Mol. Biol.*, vol. 215, no. 3, pp. 403–410, Oct. 1990.

[45] G. Csárdi and T. Nepusz, “The igraph software package for complex network research,” 2006.

[46] “WormBase : Nematode Information Resource.”

[47] UniProt Consortium, “UniProt: a worldwide hub of protein knowledge,” *Nucleic Acids Res.*, vol. 47, no. D1, pp. D506–D515, Jan. 2019.

[48] R. C. Edgar, “MUSCLE: multiple sequence alignment with high accuracy and high throughput,” *Nucleic Acids Res*, vol. 32, no. 5, pp. 1792–1797, 2004.

[49] S. Capella-Gutiérrez, J. M. Silla-Martínez, and T. Gabaldón, “trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses,” *Bioinformatics*, vol. 25, no. 15, pp. 1972–1973, Aug. 2009.

[50] M. N. Price, P. S. Dehal, and A. P. Arkin, “FastTree: Computing Large Minimum Evolution Trees with Profiles instead of a Distance Matrix,” *Mol Biol Evol*, vol. 26, no. 7, pp. 1641–1650, Jul. 2009.

[51] M. N. Price, P. S. Dehal, and A. P. Arkin, “FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments,” *PLoS One*, vol. 5, no. 3, Mar. 2010.

[52] “FigTree.” [Online]. Available: http://tree.bio.ed.ac.uk/software/figtree/. [Accessed: 07-May-2019].