Characterization of Chitin Synthases and Chitin Deacetylases that Function During the Development of the Nematode *Caenorhabditis elegans*

A dissertation

submitted by

Ronald Jason Heustis

In partial fulfillment of the requirements
for the degree of

Doctor of Philosophy

in

Biology

TUFTS UNIVERSITY

August 2012

ADVISER: Juliet A. Fuhrman

ABSTRACT

Chitin (β-1,4-linked-N-acetylglucosamine) is a major structural macromolecule in nematodes. Its presence in the eggshell and the lining of the pharynx confers protective benefits and also facilitates important steps of development. Chitin metabolism is considered to be highly conserved within roundworms – but not in the many plants and vertebrates that are parasitized by some of them – so research on chitin metabolism will provide knowledge on basic aspects of invertebrate development and may also provide a foundation from which to develop new drugs or vaccines targeting parasitic nematodes. This dissertation encompasses various projects aiming to deepen our understanding of the chitin synthases and chitin deacetylases that function during nematode development.

This body of work is the first showing that polysaccharide deacetylases (which can convert chitin to chitosan) are encoded in a diverse array of nematode genomes. Two copies of these genes are found in *C. elegans* where they are somatically expressed and affect development since their loss results in a retardation of growth. Our knowledge of transcript numbers and sequences, as well as on the dynamics of their expression, is incomplete. Proteins with *bona fide* chitin deacetylase activity are produced in *C. elegans*, and while low levels of carbohydrate limit our ability to demonstrate the presence of chitosan using standard biochemical techniques, we show that chitosan is localized to the *C. elegans* pharynx because it stains with the dye Eosin Y.

Chitin synthases convert UDP-N-acetylglucosamine into the high molecular weight homopolymer chitin and previously published work has shown that the two genes act non-redundantly to deposit chitin in the nematode eggshell and pharynx. The work in this dissertation further characterizes the germline expression of the *chs-1* gene and investigates novel roles and expression patterns for both *chs-1* and *chs-2*. The enzymatic properties of the germline and somatic chitin synthase proteins are also studied and presented along with attempts to confirm the presence of novel chitinous structures.

This dissertation closes with a perspective on future directions of study for nematode chitin synthases and chitin deacetylases and outlines the significance of these enzymes in the context of biochemistry, developmental biology and parasitology.

ACKNOWLEDGEMENTS

I would like to thank my adviser Dr. Juliet Fuhrman who has been steadfast in her attention guiding my experiments in the lab, my engagement with the scientific literature, and my progress with writing. My understanding of good experimental work, good scientific writing and high quality communication have all benefited from her immeasurable attention and guidance. I sincerely appreciate her hands-on approach and accessibility throughout my time as a student in her lab.

My development as a scientist has also benefited from the continuous contributions of three great internal Committee Members – Dr. Susan Ernst, Dr. Mitch McVey and Dr. Harry Bernheim. Over many years of Committee Meetings, their feedback has deepened my understanding of my own work and helped me to properly streamline my many research questions as I sought to arrive at my current point: a completed body of work in the form of this dissertation. Though I have met him only briefly, I also need to thank Dr. John Samuelson, for his service as the external reviewer of my dissertation.

My progress with research also benefited from many sources of outside advice and I continuously benefited from the kind help of many members of the *C. elegans* research community. In addition, Dr. Charles Specht was of invaluable help in establishing some of my initial research questions, in assistance with experiments, and in preparation of a published manuscript.

Prior to graduate school, I benefited from the mentorship of two great advisers, Dr. Douglas Coulter and Dr. Pamela Hoppe, and I owe them gratitude for cultivating my early interest in science and for their continued willingness to provide advice as I pondered steps before, during and in departing graduate school.

Of course, I am also thankful to the many fellow students (graduate and undergraduate) with whom I have been fortunate to share my time in the lab. Thank you all the help, all the thought-provoking questions, and all the stimulating discussion that has strengthened the quality of my dissertation and the published portions of this work. To that end, I am also similarly indebted to the entire Biology Department at Tufts University – faculty, staff and students. Together, they have created an incredibly collegial environment for work helping me as I developed my research, teaching and general communication skills in preparation for the next step of my career.

As I have pursued this intellectual journey, I have remained grounded through tremendous family support and by amazing friends. For my friends from Belize, St. Louis, Kalamazoo, and the Boston-Cambridge-Medford-Somerville area, I am thankful for all the conversations, meals, and outings to keep me attached to the world – and for all the calls to check on me when I disappeared.

For my family all over the U.S., I am thankful for your willingness to provide refuge when I needed to get away from school. For my family in the Caribbean, I am grateful for your diligence in keeping me abreast of everything I have missed at home as I continued on with work. And, of course, for my parents

Audrey Engleton and Ronald P. Heustis (both of whom are still in Belize), I am thankful for the values that they instilled in me – most of all commitment – that allowed me to see this work to completion.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

BLAST	Basic Local Alignment Search Tool
CBD/CBM	chitin-binding domain/chitin-binding motif
CDA	chitin deacetylase
cDNA	complementary DNA
CHS	chitin synthase
CHT	chitinase
gDNA	genomic DNA
GlcNAc	N-acetylglucosamine
LDL	low-density lipoprotein
mf	microfilariae
ORF	open reading frame
NEXTDB	The Nematode Expression Pattern DataBase
UDP	uridine diphosphate
PCR	polymerase chain reaction
PDA	polysaccharide deacetylase
PSI-BLAST	Position-Specific Iterative Basic Local Alignment Search Tool
RNAi	RNA interference
RT	reverse transcriptase
SL1/SL2	Spliced Leader 1/ Spliced Leader 2
SP	signal peptide
ts	temperature-sensitive

Characterization of Chitin Synthases and Chitin Deacetylases that Function During the Development of the Nematode *Caenorhabditis elegans*

Chapter 1

The Chitin Metabolism Pathway in Nematodes

INTRODUCTION

A general overview of chitin metabolism: Important structures and enzymes

Chitin (β-1,4-linked-N-acetylglucosamine) is the second most abundant carbohydrate in nature (after cellulose) owing to its presence as a major structural component in diverse life forms such as protists, fungi, and invertebrates. Since chitin is essential during the development of many invertebrates, the chitin metabolism pathway is studied as an avenue for the development of insecticides against many pests or for the development of drug or vaccines against infectious roundworms. Chitin metabolism is considered a putative target for the development of selective interventions that combat nematode infections such as dog heartworm, river blindness and lymphatic filariasis – all of which are caused by filarial ("threadlike") roundworms.

Key mediators of chitin metabolism include chitin synthases, chitinases and chitin deacetylases and these enzymes interact to facilitate chitin synthesis, degradation or modification as depicted in **Figure 1.1**. Uridine diphosphate-N-acetylglucosamine (UDP-GlcNAc) is used as a substrate for chitin synthase (CHS) enzymes which polymerize the macromolecule. Once synthesized, chitin may be broken down by the activity of chitinase (CHT) enzymes that act in highly regulated processes to catalyze the hydrolysis of the β-1,4-linkage between two GlcNAc residues and so degrade chitin. Alternatively, some chitin is converted to chitosan when chitin deacetylase (CDA) enzymes remove a variable number of acetyl groups from their chitin substrate.

Chitin is an inert, uncharged, insoluble homopolymer of GlcNAc with a high molecular weight. It undergoes a limited set of interactions with other molecules and is capable of maintaining a rigid structure ideal for a protective covering. Chitin is resistant to physical interactions with solvents such as water.

Chitosan, a derivative of chitin produced by partial or complete deacetylation of its GlcNAc residues, is also found in nature. Chitosan has notably different physical and chemical properties owing to the presence of many reactive amine ($-NH_2$) groups generated through deacetylation. Chitosan is reactive, positively charged and soluble. Its amine groups can form both covalent and non-covalent linkages: nitrogen in the amine group is a strong nucleophile and attracts other atoms to share electrons and form bonds. One example of this is seen when the amine group is the recipient of an electrophile such as H^+, which is readily accepted by chitosan in water and results in the presence of an NH_3^+ group. This transition also makes chitosan a positively-charged molecule and the accumulation of positive charge has important implications for spatial rearrangements of the macromolecule that are distinct from the natural conformation of chitin.

Chitin Metabolism in Nematodes

Various molecular and biochemical properties of the enzymes in the chitin metabolism pathway were initially studied in parasitic roundworms. Chitin was first characterized as a component of the nematode eggshell in the nematode *Aspiculuris tetraptera* (Wharton, 1979); and subsequent identification of chitinase

activity from *Onchocerca gibsoni* (Gooday *et al.*, 1988) laid the foundation for the study of nematode chitin metabolism. Chitin is present not only in the nematode eggshell but also in the pharyngeal lining, a localization first described using the parasitic nematode *Oesophagostomum dentatum* (Neuhaus *et al.*, 1996). The parasitic nematodes were used as study systems to elucidate the molecular properties of the genes, the biochemical properties of the enzymes, and – by way of microscopy – to understand the structures built from chitin. A cohesive view of the role of the pathway during development, though, came from studies using *Caenorhabditis elegans* (Fanelli *et al.*, 2005; Veronico *et al.*, 2001; Zhang *et al.*, 2005).

C. elegans is the best-characterized nematode system. It was the first nematode to have its genome fully sequenced and annotated (Consortium, 1998) and provides a wealth of resources in the form of various mutant strains. As a free-living nematode, *C. elegans* is easy to maintain. It also has a short generation time, with gravid adults observed 2 – 3 days after hatching when development proceeds at 25°C. Together, these characteristics make *C. elegans* an ideal model for research on the various stages of nematode development.

In *C. elegans*, *gna-1* and *gna-2* (the glucosamine-6-phosphate N-acetyltransferase genes) encode the enzymes which catalyze one of the final steps in the synthesis of UDP-GlcNAc. GNA-2 functions in the *C. elegans* germline and produces this substrate that is incorporated into eggshell chitin by the enzyme CHS-1 (Johnston *et al.*, 2006). *gna-1* is expressed in the *C. elegans* pharynx and GNA-1 is homologous to GNA-2; therefore, GNA-1 is presumed to produce the

monomer that is incorporated into the chitinous pharyngeal lining by CHS-2 (Wormbase WS227). This is depicted in **Figure 1.1**.

As is true in other nematodes, multiple chitinases (CHTs) are found in *C. elegans*, but specific functions for each isoform have not been comprehensively assigned and described. Chitin deacetylases (CDAs) have been shown to convert chitin to chitosan in many species, including protists (Das *et al.*, 2006), fungi (Baker *et al.*, 2007; Banks *et al.*, 2005) and insects (Arakane *et al.*, 2009; Dixit *et al.*, 2008; Luschnig *et al.*, 2006; Wang *et al.*, 2006); however, no chitin deacetylases have been previously identified in any nematode, including *C. elegans*.

C. elegans continues to provide an unparalleled platform for dissecting fundamental aspects of the chitin metabolism pathway in nematodes. Much of the accepted facts about chitin metabolism in parasitic roundworms extrapolate from findings in the free-living species. However, our ability to directly address questions in the parasites – and to pursue questions on the commonalities and differences in the use of the pathway among nematodes – has increased as resources for study of the parasitic roundworms have expanded. Of these parasitic roundworms, *Brugia malayi* is the best characterized of the filarial ("threadlike") nematodes: it is the only filarial worm with a fully sequenced genome (Ghedin *et al.*, 2007) and for which a stage-specific proteome has been generated (Bennuru *et al.*, 2011). *Trichinella spiralis* is a non-filarial parasite with a fully sequenced genome (Mitreva *et al.*, 2011). Information on parasites

like *B. malayi* and *T. spiralis* create increasing power to address questions from the perspective of comparative biology.

The Life Cycles of *C. elegans* and *B. malayi*

At its core, the work discussed in this dissertation discusses a biochemical pathway; however, research questions were informed and are discussed in the context of their relevance to nematode development. The empirical data presented in this dissertation is all based in the free-living nematode *C. elegans*; however, research questions often originate by observations from parasites like *B. malayi*. This approach assumes analogies between the development of the two organisms and hence a review of the life cycles of the two organisms, including references to analogous structures and functions, will be helpful. An overview of the life cycles of *C. elegans* and *B. malayi* is presented in **Figure 1.2**.

C. elegans is a sexually dioecious species, having hermaphrodites and males as adults. Self-fertilizing hermaphrodites produce both oocytes and sperm, while the males which can cross-fertilize hermaphrodites, produce sperm only. Once fertilized, the embryo develops under the protection of a chitinous eggshell before it emerges to progress through four larval stages en route to adulthood. The four larval stages (L1, L2, L3 and L4) are observed in worms growing under ideal conditions, but under extremes of temperature and overpopulation or under conditions of limited food, *C. elegans* L1 larvae molt to successive L2d and L3d stages. The L3d dauer larvae persist for extended periods using an altered

metabolic profile, molting back to the normal L4 worms when conditions again favor proper survival and reproduction.

B. malayi is transmitted between vertebrate hosts by mosquitoes and key stages of development depend on the alternation between vector and host. *B. malayi* proceeds through four larval stages (microfilaria/L1, L2, L3, L4) before maturing into male or female worms. There is no clear delineation between the microfilaria and L1, and the terms are sometimes used interchangeably. In this dissertation, the term "microfilariae" will be used to describe the early L1 that are enclosed in a protective chitinous sheath (derived from the remodeled eggshell); after exsheathment, this chitinous covering surrounding the cuticle is lost, and these worms will be called the "L1" of *B. malayi* (**Figure 1.2**). *B. malayi* is also a sexually dioecious species, although reproduction only occurs through insemination since adults are either male or female. Once copulation takes place between adult males and females within the human lymphatic system, successful fertilization and embryogenesis leads to the production of microfilariae (mf) that leave the lymphatics and enter the bloodstream. Mf are taken up by mosquitoes during a bloodmeal and progress from mf to the third larval stage (L3) within this vector. The L3 larvae are deposited onto the skin and enter through the opening created from a mosquito bite. The L3 enter the narrow branches of the lymphatic system and migrate to larger vessels as they mature from the L3 to adults. Like the dauer larvae of *C. elegans*, infective L3 larvae (that are held in mosquitoes) can remain at this life stage for extended periods of time. The longevity of the infective L3 is limited by the lifespan of the vector and, as such, these worms can

persist at this stage for 1 – 2 weeks (depending on temperature) in mosquitoes. The analogies between life stages in *C. elegans* and *B. malayi* are emphasized in **Figure 1.2**.

FIGURES

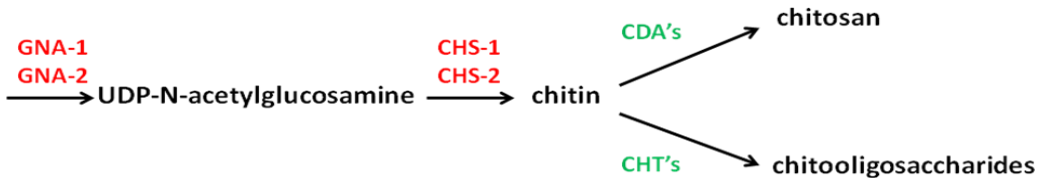


Figure 1.1. The chitin metabolism pathway. Chitin, the homopolymer of β-1,4-linked-N-acetylglucosamine, is produced under the action of chitin synthases (CHS proteins) using UDP-activated GlcNAc as a substrate. In nematodes, the two chitin synthases CHS-1 and CHS-2 act non-redundantly to produce chitin in the eggshell and pharynx, respectively. In *C. elegans*, the enzyme GNA-2 catalyzes one of the final steps to produce UDP-GlcNAc in the germline while GNA-1 is presumed to produce the monomer necessary for the pharyngeal chitin synthesis.

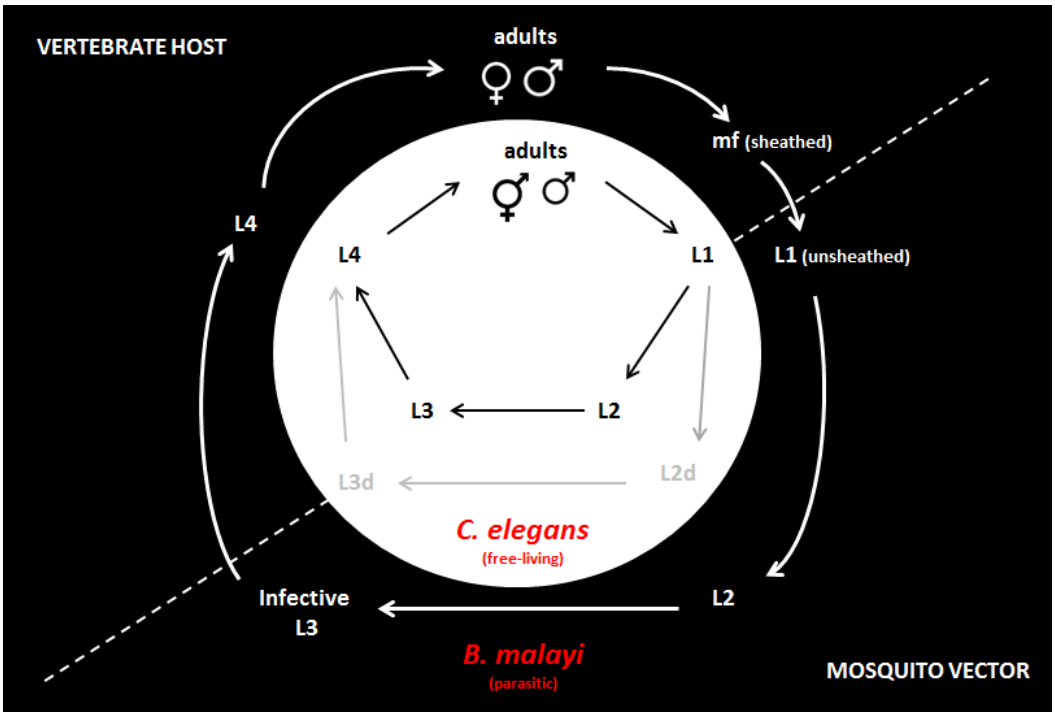


Figure 1.2. Overview of the development of *C. elegans* and *B. malayi*. Nematodes progress through four larval stages in maturation from hatched progeny to adults. In *C. elegans*, the L1 larvae emerge from eggs and progress through four molts before becoming adult hermaphrodites or males. Under non-ideal conditions of temperature, over-crowding or food scarcity, *C. elegans* assumes an alternative profile of development when L1 larvae molt to two successive stages of L2d and L3d worms. The L3d worms are distinguishable from normal L3 worms and employ different metabolic profiles for survival. *B. malayi* adults are male or female and reside in the lymphatics of the human lymphatic system. Successful mating between *B. malayi* adults produces circulating mf that can be found in the bloodstream where they are picked up during a mosquito bloodmeal. Mf develop to infective L3 larvae in this insect vector and this stage can be passed onto humans during a subsequent bloodmeal. *B. malayi* require the human host for their progress from the L3 to adulthood.

Chapter 2

Pharyngeal Polysaccharide Deacetylases Affect Development in the Nematode *C. elegans* and Deacetylate Chitin *in vitro*

TITLE

Pharyngeal Polysaccharide Deacetylases Affect Development in the Nematode *C. elegans* and Deacetylate Chitin *In Vitro*

AUTHORS

Ronald J. Heustis[1,*], Hong K. Ng[1,a], Kenneth J. Brand[1,b], Meredith C. Rogers[1,c], Linda T. Le[1], Charles A. Specht[2], Juliet A. Fuhrman[1]

[1]*Department of Biology, Tufts University, Medford, MA, 02155, USA*

[2]*Department of Medicine, University of Massachusetts, Worcester, MA, 01605, USA*

[a] Current Address: *School of Medicine, St. George's University, Grenada, West Indies*

[b] Current Address: *College of Veterinary Medicine, Cornell University, Ithaca, NY, 14850, USA*

[c] Current Address: *Department of Pathology, Microbiology and Immunology, Vanderbilt University, Nashville, TN, 37232, USA*

*Corresponding author:	*Telephone Phone Number*:	617-627-4038
	E-mail address:	ronald.heustis@tufts.edu
*Other E-mail addresses:	Hong K. Ng:	hng1@sgu.edu
	Kenneth J. Brand:	kjb252@cornell.edu
	Meredith C. Rogers:	meredith.rogers@vanderbilt.edu
	Linda T. Le:	linda.le@tufts.edu
	Charles A. Specht	charles.specht@umass.med.edu
	Juliet A. Fuhrman	juliet.fuhrman@tufts.edu

ABSTRACT

Chitin (β-1,4-linked-N-acetylglucosamine) provides structural integrity to the nematode eggshell and pharyngeal lining. Chitin is synthesized in nematodes, but not in plants and vertebrates which are often hosts to parasitic roundworms; hence, the chitin metabolism pathway is considered a potential target for selective interventions. Polysaccharide deacetylases (PDAs), including those that convert chitin to chitosan, have been previously demonstrated in protists, fungi and insects. We show that genes encoding PDAs are distributed throughout the phylum *Nematoda*, with the two paralogs F48E3.8 and C54G7.3 found in *C. elegans*. We confirm that the genes are somatically expressed and show that RNAi knockdown of these genes retards *C. elegans* development. Additionally, we show that proteins from the nematode deacetylate chitin *in vitro*, we quantify the substrate available *in vivo* as targets of these enzymes, and we show that Eosin Y (which specifically stains chitosan in fungal cells walls) stains the *C. elegans* pharynx. Our results suggest that one function of PDAs in nematodes may be deacetylation of the chitinous pharyngeal lining.

KEY WORDS

chitin; polysaccharide deacetylase; chitosan; nematode; pharynx

INTRODUCTION

Chitin, the homopolymer of β-1,4-linked-N-acetylglucosamine (GlcNAc), contributes to the mechanical strength and chemical impermeability of both the embryonic eggshell and the pharyngeal lining in nematodes. As such, this carbohydrate is of major structural importance during nematode development. Chitin is an insoluble, neutrally charged, chemically inert carbohydrate well-suited for its protective functions. It is synthesized in nematodes, but not in plants and vertebrates, many of which serve as hosts for parasitic roundworms. Understanding chitin metabolism will inform the development of interventions that selectively target nematodes important to agriculture and to the health of domesticated animals and humans.

Enzymes that function during major steps of chitin metabolism have been identified in a wide variety of species. Chitin synthases (which polymerize and deposit chitin) and chitinases (which hydrolyze chitin to its subunits) have previously been characterized in protists, fungi, and invertebrates including nematodes and insects. Recently, chitin deacetylases (which convert chitin to chitosan) have been identified in a range of organisms. Deacetylation has been demonstrated to be a versatile mechanism that influences cellular and organismal growth, as it converts chitin to a charged polymer that has increased solubility in aqueous environments and increased pliability. Chitin deacetylases alter the composition of the cyst wall in the protist *Entamoeba invadens* (Das *et al.*, 2006) and the cell wall of the fungus *Cryptococcus neoformans* (Baker *et al.*, 2007; Banks *et al.*, 2005). Chitin deacetylases also play essential roles during insect

development (Arakane *et al.*, 2009; Dixit *et al.*, 2008; Luschnig *et al.*, 2006; Wang *et al.*, 2006). In *Drosophila melanogaster*, *serpentine* (*serp*) and *vermiform* (*verm*) genes encode chitin deacetylases that function in the chitin-lined embryonic tracheal tubes and that are essential for survival. This finding first established a role for chitin deacetylases in animals. Luschnig *et al.* (2006) noted a homolog of the *serp* and *verm* genes found in *C. elegans*: *Ce-lgx-1*/C54G7.3 bears both a predicted polysaccharide deacetylase (PDA) and predicted chitin-binding (CBD) domain; however, no functional work has established a role for chitin deacetylases in nematodes.

Chitin deacetylases, and all other PDAs, belong to the CAZY (http://www.cazy.org) carbohydrate esterase 4 family of enzymes. This family of enzymes acts on an array of different carbohydrates catalyzing the cleavage of ester linkages in xylan or amide linkages of GlcNAc in peptidoglycan and chitin. These enzymes function through a catalytic site termed the PDA domain, which is also referred to as the NodB homology domain. In *Rhizobium*, the NodB protein processes lipooligosaccharide signaling molecules. It serves the role of a polysaccharide deacetylase when it deacetylates chitooligosaccharides (short stretches of repeating GlcNAc monomers) during the production of these signaling molecules (John *et al.*, 1993). We used the protein sequences of this domain from previously characterized peptidoglycan and chitin deacetylases as query sequences in a comprehensive bioinformatic search to identify polysaccharide deacetylases among the *Nematoda* and by sequence similarity comparisons suggest that chitin is the substrate.

In this paper, we show that in addition to the previously noted homolog of *serp* and *verm* (C54G7.3), a second PDA-encoding gene (F48E3.8) is found in *C. elegans*. We also demonstrate the presence of homologs in many other species of *Nematoda*, including important parasites of plants and vertebrates. We took advantage of *C. elegans* to investigate a role for this class of enzymes, building on existing publicly available data (NextDB, http://nematode.lab.nig.ac.jp) and published information (Gaudet and Mango, 2002), showing that the genes are expressed in the pharynx. We confirm that the genes are somatically expressed and present evidence that disruption of these genes results in a delay in developmental timing. Our results also include a demonstration that proteins derived from the nematode deacetylate chitin *in vitro*.

MATERIALS AND METHODS

Strains

The wild-type *C. elegans* Bristol N2 strain, the SS104 strain (carrying the temperature-sensitive *glp-4* allele, bn2), and the NL2099 strain (carrying the *rrf*-3 allele, pk1426) were used for these experiments.

Worm Growth and Sampling

Worms for developmental time course experiments were grown on solid media at 25°C with sampling performed as previously described (Johnstone and Barry, 1996; Veronico et al., 2001), after the following minimum numbers of arrested L1's were plated: 25,000 worm/plate for plates to be sampled for L1's

and L2's; 12,500 worms/plate for plates to be sampled for L3's and L4's; 6250 worms/plate for plates sampled to gather adults. (A larger number of worms were used for time courses with the SS104 strain to account for the absence of a germline.) Under our conditions, 75% of all hermaphrodites had identifiable vulvas by 72 h and 100% had identifiable vulvas by 96 h. At specific time points, worms were harvested by washing plates using ice-cold 0.1M NaCl, separated from bacteria by flotation on 30% sucrose, suspended in 1 mL Trizol reagent, and immediately frozen at -80°C.

Worms for enzyme activity assays were grown in liquid culture using complete S-Basal (Wood, 1988) and collected as above. 5 mL aliquots of the wet worm pellets were stored with 1 µM leupeptin, 0.2 µM pepstatin, 1 µg/mL aprotinin and 500 µM PMSF at -20°C. Thawed worm samples were extracted by sonification and cleared by centrifugation for 15 min at 12,500xg.

Worms used for chitin/chitosan quantification were similarly grown and purified from liquid culture. 1.25 mL-1.50 mL wet worm pellets were stored in 2.0-mL screw cap tubes (with no protease inhibitors) at -20°C. Samples were lyophilized for ~48 hours before dry mass was determined.

Identification and phylogenetic analysis of nematode CDAs

We used the predicted or confirmed catalytic domain peptide sequence of *Ce*-C54G7.3 and *Sp*PDA (a peptidoglycan deacetylase characterized from the bacterium *Streptococcus pneumonia*) to search for additional genes with predicted polysaccharide deacetylase domains in *C. elegans*, the parasitic nematode *Brugia*

malayi, and the Brugian endosymbiotic bacterium, *Wolbachia sp*. C54G7.3. Three newly identified PDA-encoding sequences from *C. elegans* (F48E3.8) and *B. malayi* (Bml_33340 and WolBm0147) were then used to search for additional nematode homologs. For all searches we used a combination of BLAST and PSI-BLAST to mine the NCBI non-redundant sequences for the phylum *Nematoda* as a whole, for select nematodes, and (where relevant) for the endosymbiont *Wolbachia*. Additional searches were conducted against Sanger Institute genome sequencing data for *Onchocerca volvulus* (http://www.sanger.ac.uk/cgi-bin/blast/submitblast/o_volvulus) and for other species accessible through http://www.nematode.net. We also searched for homologs in humans. (The accession numbers for all protein sequences as well as the specific region of each used for alignments are provided in **Table 2.1** and **Supplementary Table 2S.1**.)

The predicted or confirmed catalytic domain peptide sequences for the nematode homologs were used to create an alignment using ClustalW, which included sequences from bacteria (*S. pneumonia* peptidoglycan deacetylase), fungi *(C. neoformans* chitin deacetylases), and insects (*D. melanogaster* and *T. castaneum* chitin deacetylases). The alignment was created with the following parameters: pairwise alignment was generated using a gap opening penalty of 10 and gap extension penalty of 0.1, while multiple alignment was generated using a gap opening penalty of 3.0 and gap extension penalty of 1.8. This alignment was used as the basis for creating a minimum evolution tree in Mega4.0 (Kumar *et al.*, 2004).

RT-PCR

RNA was isolated from worms by homogenization in Trizol reagent (Invitrogen), treated with Turbo DNAse (Ambion) and quantified using a NanoDrop 2000 Spectrophotometer (Thermo Scientific). We diluted samples to 0.2 µg/µL and performed cDNA synthesis for each time point using the Protoscript M-MuLV First Strand cDNA Synthesis Kit (New England BioLabs) using 2 µg of RNA (+/- RT) in a 40 µL reaction.

PCR amplifications used 5 µL of each cDNA template in a 50 µL reaction along with 0.8 mM dNTP mix (USB), 1 µM of each primer (Operon), and Taq with ThermoPol Buffer (NEB). All primers and PCR conditions are listed in **SupplementaryTable 2S.2**.

RNAi

In order to generate dsRNA targeting *lgx-1*/C54G7.3, a large fragment of the gene from the yk1621b03 clone (kindly provided by Dr. Yuji Kohara) was subcloned to the v28i Litmus Vector (NEB) using EcoRI and XbaI sites. In order to generate dsRNA targeting the F48E3.8 gene, a 0.9-kb fragment of the yk1130a03 clone (kindly provided by Dr. Yuji Kohara) was subcloned to the v28i Litmus Vector (NEB) using EcoRI and PstI sites. RNAi experiments utilized the feeding protocol previously described (Kamath and Ahringer, 2003). *E. coli* (HT115[DE3]) expressing dsRNA targeting one or both PDA genes were fed to individual L3 worms. The parent was removed from each plate at 48 h. The F1

generation of each was then screened at 48 h and again at 72 h using blind scoring. Control experiments used the v28i vector with no insert.

Enzyme Activity Assays

CDA activity was visualized using the substrate gel method as described by Trudel and Asselin (1990). SDS-PAGE was used to separate soluble worm extracts (14 μg per lane) under reducing or non-reducing conditions in gels containing 0.01% glycol chitin. In order to visualize CDA activity following electrophoresis, the gels were incubated for 24 h at room temperature in a renaturation solution containing 1% Triton X-100 in 50 mM Hepes-KOH (pH 7.0), on a shaker. The gels were then stained with 0.01% Calcofluor in 0.5 M Tris-HCl (pH 9) solution for approximately 2 hours, followed by constant washes of picopure water. Deacetylated chitin generates hyperfluorescent bands over the background staining of chitin under UV illumination (Trudel and Asselin, 1990). Replicate gels developed by staining without renaturation were used as negative controls to demonstrate dependence of the signal on enzymatic digestion.

Chitin and Chitosan Detection

The chitin and chitosan content from lyophilized worm pellets was quantified using *in vitro* reacetylation, chitinase degradation and a Morgan-Elson assay modified for a 96-well plate format as previously described (Bulik *et al.*, 2003; Das *et al.*, 2006).

For eosin Y staining, mixed stage N2 worms were washed and resuspended in 500 µL citrate-phosphate buffer, pH 6.0 (0.2 M NaH_2PO_4 and 0.1 M KCitrate) before 15 µL of eosin Y stock (5mg/ml in 70% ethanol) was added. Tubes were incubated at RT in the dark for 10 min, then spun for 5 min at 600 x g to wash. The supernatant was replaced with fresh citrate-phosphate buffer, and this wash step was repeated for a total of 5 washes. The samples were observed and imaged using an Olympus BX40 epifluorescent microscope with a Chroma 31001 filter. Non-specific staining through ingestion of fluorescent dye was tested using 4-methylumbelliferone at 0.15 mg/mL as a control.

RESULTS

The first metazoan chitin deacetylase was identified as a novel peritrophic membrane protein from the cabbage looper *Trichoplusia ni* (Guo *et al.*, 2005) and subsequent work has shown that insects have a large family of PDA-domain encoding genes, with nine homologs found in the red flour beetle *Tribolium castaneum* (Dixit *et al.*, 2008). In *D. melanogaster*, *serp* and *verm* encode two chitin deacetylases which function during embryonic development (Luschnig *et al.*, 2006; Wang *et al.*, 2006). Luschnig *et al.* (2006) noted that these genes shared homology with the *C. elegans* gene *lgx-1*/C54G7.3 (hereafter referred to as C54G7.3), which bears a predicted chitin-binding domain (CBD) and a predicted polysaccharide deacetylase domain (PDA). Using the C54G7.3 predicted PDA protein sequence as a basis for bioinformatic analysis, we identified homologs in other nematodes, including important parasites of plants and animals and four

free-living species from the genus *Caenorhabditis* **(Table 2.1)**. In some nematode species, we identified a single predicted PDA sequence, but noted the presence of two homologs in all four *Caenorhabdita*. In *C. elegans*, F48E3.8 also encodes a putative polysaccharide deacetylase. Although we observed only a single homolog in the genome of the filarial parasite *B. malayi* (Bml_33340), using *SpPDA* as a search sequence against the Brugian endosymbiont *Wolbachia* returned the gene WolBm0147, from the TRS strain of this bacterium. As such, two PDAs may be expressed in *B. malayi*, one by the nematode and one by its endosymbiotic bacterium. In *Onchocerca volvulus*, a related filarial parasite, we identified a sequence apparently related to the *B. malayi* genomic sequence (contig33574) and one related to the *Wolbachia*-derived sequence (contig48143) based on an analysis of Sanger sequence repositories. PDAs appear highly conserved across the nematode phylum since at least one homologous sequence was found in species from Clades I, III, IVb and V. We were unable to identify homologs in other important vertebrate parasites: *Acanthocheilenema viteae*, *Brugia pahangi*, *Brugia timori*, *Dirofilaria immitis* and *Wuchereria bancrofti*. This may reflect a lack of available sequence data rather than the absence of genes or transcripts in these species. No human homolog was detected in our searches, although vertebrate homologs of glycosaminoglycan deacetylases (bearing protein sequence highly divergent from the NodB domain) have been identified (Berninsone and Hirschberg, 1998).

We created a phylogenetic tree based on ClustalW alignment of the PDA catalytic domains (**Figure 2.1)**. We used a subset of our newly identified

nematode sequences and included only those predicted protein sequences generated from reliable transcript sequence data. In *C. elegans*, the exon-intron boundaries covering the predicted PDA-encoding domain of F48E3.8 were previously verified in ESTs; however, exon-intron boundaries spanning the predicted C54G7.3 PDA-encoding domain have not been verified (Wormbase WS229). We sequenced a region of C54G7.3 and found that the predicted 69-bp exon 28 was not expressed. We have therefore omitted 23 amino acids (SFKFKIKNFKKVIPNTLSLKNTI) from the protein associated with the NCBI Accession No. CCD67046. All our alignments and the phylogenetic tree reflect this correction.

The phylogenetic tree, generated using the minimum evolution (ME) method, shows that the nematode sequences are more closely related to each other than to the PDA sequences from bacteria, protists, fungi and insects (**Figure 2.1**). The sequence WolBm0147 derived from the endosymbiont of *B. malayi*, however, is most closely related to the only other bacterial sequence, *S. pneumoniae* PgDA (peptidoglycan deacetylase) included in our analysis. Our results replicate the branching pattern previously described for insect PDA proteins where the nine *T. castaneum* polysaccharide deacetylases were categorized into five groups. This organization was based on homology within the catalytic domain and domain architecture within the full primary sequence (Dixit *et al.*, 2008). (In insects, all members of this family of proteins have been named chitin deacetylases although functional work has not specifically elucidated chitinous targets for some of these enzymes.) In *T. castaneum*, Group 1 and

Group 2 CDA proteins have a chitin-binding domain (CBD), a low-density lipoprotein receptor domain (LDLa) and a chitin deacetylase domain (CDA). Group 3 and Group 4 CDA proteins have a CBD and a CDA domain (but no LDLa) and Group 5 CDA proteins have a CDA domain (but neither of the other 2 domains). Our phylogeny suggests that nematode PDAs are more closely related to the Groups 1-4 insect CDAs (*Tc*CDA1-5), than they are to the Group 5 CDAs (*Tc*CDA6-9). PFam analysis predicts the presence of a weak-scoring CBD in *C. elegans* F48E3.8 and C54G7.3, but not LDLa domains in these nematode sequences.

The relationship established here among *C. elegans* F48E3.8, its orthologs *C. brenneri* CAEBREN_29772, *C. briggsae* CBG_16715, and *C. remanei* CRE_00058, and the single genes detected from *A. suum*, *B. malayi*, *L. loa* and *T. spiralis*, reflects the same evolutionary profile described for these species when phylogenetic analysis is based on the ribosomal DNA sequences (Blaxter *et al.*, 1998). The filarial sequences are most closely related, share a most recent common ancestor with *A. suum*, an earlier common ancestor with the *Caenorhabdita*, and an even earlier common ancestor with *T. spiralis* (**Figure 2.1**). This is similar to the relationship among these species in the Blaxter *et al.* (1998) analysis. The distance between F48E3.8 orthologs and C54G7.3 orthologs suggests that significant variation occurred within the sequence of one or both genes following a likely gene duplication event that created paralogs in a common ancestor of the *Caenorhabdita*.

The PDA domain has been designated as the NodB homology domain in the PFam database (NCBI Accession No. PF01522), and defines a catalytic site including five functional motifs that were first characterized in bacteria (Blair *et al.*, 2005; Blair and van Aalten, 2004) and later in fungi (Blair *et al.*, 2006). These motifs are also conserved in the insect CDA proteins (Dixit *et al.*, 2008). Based on the alignment of the NodB domains from the newly identified nematode PDA proteins included in our tree, these five motifs are also conserved in predicted nematode PDA proteins (**Figure 2.2**). In bacterial and fungal sequences, specific residues in Motifs 1 and 2 have been assigned roles in binding to metal co-factors while specific residues in Motifs 1, 3, 4 and 5 have been shown to function as part of the catalytic acetyltransferase activity. The presence of the five motifs in all nematode-derived sequences suggests that they retain the ability to act on carbohydrates to deacetylate them.

Motif 1 has been identified as a TFDD sequence in other organisms and is highly conserved in nematodes as a (T/S)FDD sequence, where a threonine-to-serine substitution characterizes *Ce*C54G7.3 and its orthologs. In nematodes, motif 2 is an NSI(T/S)X. This is a notable exception to other organisms where the first and last residues of motif 2 have been reported as two histidine residues as they are in *S. pneumoniae* peptidoglycan deacetylase and the *Wolbachia* sequence WolBm0147. In the *Caenorhabdita*, NSISH is the fully conserved Motif 2 sequence. The nematode motif 3 sequence is an R(S/A)PX: the fourth residue is a variable amino acid which replaces the tyrosine residue found in that position for bacterial and insect sequences. Motif 4 in nematodes is an FXXDN

and Motif 5 is a more variable, though not exceptionally divergent sequence, with a fully conserved tryptophan residue replacing the histidine found in bacterial and insect sequences at position 7. Among *Ce*C54G7.3 and its orthologs in the *Caenorhabdita*, the residues in Motifs 3 – 5 are fully conserved just as they are in Motifs 1 and 2. This alignment also underscores the similarity between our identified WolBm0147 sequence and that of the other representative bacterial sequence.

Functions for these newly classified nematode PDAs have not been previously established. *Ce*F48E3.8 and *Ce*C54G7.3 have been shown to be upregulated in the embryonic pharynx and both genes have multiple TRTTKRY binding sites for the PHA-4 transcription factor within their upstream regulatory regions (Gaudet and Mango, 2002). The Nematode Expression Database (NEXTDB, http://nematode.lab.nig.ac.jp/) provides *in situ* hybridization data confirming that both genes are expressed in the developing embryonic pharynx. F48E3.8 (cosmid F48E3, clone 272e1 in the database) is detected in embryos, but not in other stages. C54G7.3 (cosmid C54G7, clones 325b6 and 208c10) is abundantly expressed in the pharynx of embryos, larvae and adults, with a spatiotemporal expression pattern similar to that of the pharyngeal chitin synthase *chs-2* (cosmid F48A11, clone 316g4), although *chs-2* is also detected in non-pharyngeal tissue (NEXTDB, http://nematode.lab.nig.ac.jp/). The expression of F48E3.8 and C54G7.3 in the *C. elegans* pharynx suggested that one possible target of these polysaccharide deacetylases could be the chitinous lining of the pharyngeal tract. In *D. melanogaster*, the two chitin deacetylases Serp and Verm

target the chitin lining the embryonic trachea and act to regulate longitudinal (and possibly radial) growth of these epithelial tubes. This regulation is crucial for permitting efficient gas diffusion for the developing embryo (Luschnig *et al.*, 2006; Wang *et al.*, 2006). This mechanism of regulating epithelial tube growth may be applicable to other organisms and we reasoned that nematode homologs may affect development through their action on pharyngeal chitin.

Using semi-quantitative RT-PCR, we show that both genes are somatically expressed as they were detected in the SS104 strain grown under conditions that restrict the production of a germline (**Figure 2.3**). In our time course, both genes were first detected at 24 h following hatching, when grown at 25°C. Both PDA transcripts continued to be expressed at 48 h and 72 h. Adults first appeared at the 72 h time point, marked by the onset of expression of the adult specific collagen *col-19* (Thein *et al.*, 2003) and consistent with our cytological observations (as approximately 75% of worms on the plates had an identifiable vulva at 72 h while none had this at 48 h). F48E3.8 and C54G7.3 genes showed clear expression despite the significant attenuation of the germline, evidenced by the trace expression of the eggshell chitin synthase gene *chs-1*. *chs-1* is abundantly expressed in the germline and can be easily detected by RT-PCR in similar time courses from wild-type worms (Veronico *et al.*, 2001). In contrast to *chs-1*, the pharyngeal chitin synthase *chs-2* was previously shown to be expressed at early stages of development, to appear in the pharynx prior to the L1/L2 molt when assayed using GFP reporters, and to decline following the molt to adulthood (72 h) (Fanelli *et al.*, 2005; Veronico *et al.*, 2001; Zhang *et al.*, 2005). The

expression patterns for the two PDAs were similar to *chs-2* from 24 h to 72 h in our experiment (**Figure 2.3**). However, C54G7.3 and *chs-2* expression declined following onset of adulthood (72 h), while F48E3.8 expression persisted through to 120 h as did *col-19* expression. Both PDAs, then, were expressed in the absence of a germline, but F48E3.8 appeared to persist into adulthood while C54G7.3 expression declined once all the worms reached adulthood. All three predicted transcripts of F48E3.8, but only one of the two predicted transcripts of C54G7.3, include sequence coding for the PDA domain. Importantly, our RT-PCR experiments were designed to detect all transcripts predicted to encode PDA domains; thus, our results show that transcripts capable of producing active deacetylase enzymes are expressed in the *C. elegans* soma. The demonstration of F48E3.8 expression in larval and adult stages contrasts with *in situ* data that showed expression was restricted to embryos (NEXTDB, http://nematode.lab.nig.ac.jp/). This may reflect increased sensitivity afforded by RT-PCR or differences in expression of alternative transcripts of F48E3.8.

We used RNAi to determine whether these PDAs were essential for *C. elegans* development. For each gene, we targeted regions of the gene that would permit disruption of all predicted transcripts. Co-feeding dsRNA targeting both F48E3.8 and C54G7.3 resulted in a modest delay in the development of F1 worms from N2 wild-type parents. This was evident from the presence of L3 progeny as the most advanced stage on experimental plates at 48 h in contrast to L4 progeny on matched control plates (**Figure 2.4A**). This effect was corroborated at 72 h when a smaller proportion of worms from experimental plates had attained

adulthood in contrast to F1 adults producing eggs on control plates (**Figure 2.4B**). We tested whether both genes contributed to this effect, by conducting similar experiments for F48E3.8 and C54G7.3 separately using N2 wild-type worms; however, we could not detect a consistent change in development (data not shown).

To enhance the likelihood of detecting a developmental defect caused by each gene, we repeated experiments using the NL2099 strain, which carries an *rrf-3* allele rendering the strain hypersensitive to exogenous RNAi (Simmer *et al.*, 2002). Again, we could detect no consistent effect when testing F48E3.8 alone in the *rrf-3* background (data not shown). Knockdown of C54G7.3 alone in *rrf-3* worms, though, caused a developmental delay. At 48 h, L3 progeny were the latest stage evident on control plates while L2 were the latest stage observed on experimental plates (**Figure 2.4C**) and at 72 h a greater proportion of F1 worms had attained adulthood in control than in experimental treatments (**Figure 2.4D**).

This delay in development is similar to what was observed for RNAi (by feeding) against the pharyngeal chitin synthase *chs-2* (Veronico *et al.*, 2001), where the knockdown impaired the capacity of *chs-2* to produce the new lining of the pharynx prior to each molt. Although PDA knockdown gives a less dramatic effect, the similarity with the *chs-2* data strengthens the argument that these novel PDAs may act in the *C. elegans* pharynx to modify the chitinous lining produced by *chs-2*.

None of our RNAi experiments led to a disruption in the number of intact eggs produced; eggs with normal morphology were visible on all plates.

Additionally, brood sizes were normal and equivalent for all treatments (**Table 2.2**). Disruption of *chs-1* expression by RNAi disrupts the formation of eggshells and affects brood size (Zhang *et al.*, 2005). Our results do not provide any evidence that these PDAs are critical to eggshell development in *C. elegans*. The integrity of these eggs and embryos is further validated by the similarity in hatching rates (measured by the presence of viable progeny at 72 hours) for matched treatments (**Table 2.2**).

The overlap in expression patterns for these genes and the RNAi results support a role for these polysaccharide deacetylases in modifying the chitinous pharyngeal lining. We next asked whether these genes were capable of utilizing chitin as a substrate to produce chitosan. Soluble extracts of mixed stages of *C. elegans* (strain N2) deacetylated chitin *in vitro*, as demonstrated by several bands of activity in substrate gels (**Figure 2.5**). The hyperfluorescent bands in calcofluor-stained gels are characteristic of deacetylated regions in the glycol chitin matrix. This activity was detected in samples after separation under non-reducing or reducing conditions. Under non-reducing conditions, bands of activity appeared at 178, 165, and 132 kDa and under reducing conditions appeared at 160, 130 and 104 kDa. Comparable bands were visualized in soluble extracts from germline-ablated SS104 adult worms, although the 178 kDa band was not visualized under reducing conditions, and an additional band at 118 kDa was detected (data not shown). Hyperfluorescent bands were absent in control gels stained immediately after separation without a renaturation step, and nitrous acid-mediated depolymerization of chitosan in the gel abolished the hyperfluorescent

signal as well (data not shown). This confirms the dependence of the hyperfluorescent signal on renaturation of the active enzymes and on the generation of chitosan.

In these *in vitro* experiments, the substrate is present in excess and it is noteworthy that polysaccharide deacetylases often act non-specifically on multiple substrates presented *in vitro* (Caufrier *et al.*, 2003). We attempted to quantify the extent of chitin deacetylation *in vivo* by measuring chitin and chitosan in the nematode using a differential reacetylation assay (Baker *et al.*, 2007; Banks *et al.*, 2005; Das *et al.*, 2006). Following alkali treatment, chitinase digestion of the insoluble fraction of N2 worms yielded 0.499-1.511 nmol GlcNAc per mg lyophilized worm tissue, and chitinase digestion of the equivalent fraction from germline-ablated SS104 adult worms yielded 0.050-0.170 nmol GlcNAc per mg lyophilized worm tissue (**Table 2.3**). Together, these results show that < 10% of the chitin found in *C. elegans* is localized to somatic tissue. When mixed-stage SS104 worms grown under conditions with normal germline activity were subjected to chitinase treatment, similar yields of GlcNAc were obtained with or without prior reacetylation *in vitro* (**Table 2.3**). Thus any chitosan present, which would be converted to chitin by the *in vitro* acetylation, was below the level of detection. SS104 germline-ablated worms yielded product below the level of detection for the reacetylation assay (data not shown).

Although we could not quantify chitosan in the nematodes using this assay, our results show that small proportion of the total endogenous chitin in *C. elegans* is found in the pharynx and the vast majority is used in the chitinous

eggshells. If these PDAs act as chitin deacetylases *in vivo*, they target a small proportion of the total endogenous chitin in fertile hermaphrodites.

Eosin Y has previously been reported to stain chitosan specifically in the fungal cell wall of unfixed *C. neoformans* (Baker *et al.*, 2007). We stained living, unfixed worms with this dye and noted a highly reproducible staining apparent in the lumen of the buccal cavity, procorpus, metacorpus and isthmus, and the flaps of the terminal bulb (**Figure 2.6**). This staining occurs in various developmental stages and overlaps with the pattern seen in fixed worms stained for chitin (Zhang *et al.*, 2005). Solutions of a non-specific fluorescent dye, 4-methlyumbelliferone (4-MU), were used in the same staining protocol to test whether ingestion or trapping of dye, rather than specific staining, would generate an equivalent fluorescent pattern in the unfixed worms. No such luminal staining was detected in worms treated with 4-MU under matching conditions.

DISCUSSION

We have conducted the first functional analysis of polysaccharide deacetylases in nematodes, starting with the genes F48E3.8 and C54G7.3 in *C. elegans*. The somatic expression of PDA domain-encoding transcripts of F48E3.8 and C54G7.3 in *C. elegans*, demonstrated by our RT-PCR, corroborates the pharyngeal expression of these genes described by Gaudet and Mango (2002) and observed in publicly available *in situ* hybridization data sets for the genes (NEXTDB, http://nematode.lab.nig.ac.jp/). Our results demonstrate that somatic

tissues contain gene products capable of deacetylating chitin when presented with the substrate *in vitro*.

Genes encoding PDAs are widely distributed in the *Nematoda* as we have identified homologs in representatives from all clades except Clade IVa, where sequence availability may be a limiting factor (**Table 2.1**). In *C. elegans*, F48E3.8 and C54G7.3 are the two representatives of the polysaccharide deacetylase family. These are large genes of ~9.7 kb and 13.3 kb, respectively. In the closely related *C. briggsae*, we have identified CBG14800 as the ortholog of *Ce*-C54G7.3 gene. CBG14800 (which encodes a predicted PDA domain) and CBG14801 (which encodes a predicted CBD domain) are annotated as separate but adjacent genes in *C. briggsae*; however, they likely represent one chitin deacetylase-encoding gene that is similar to *Ce*-C54G7.3. In C54G7.3, predicted exons 11 and 12 are separated by an unusually large ~4kb intron. CBG14801 aligns with *Ce*-C54G7.3 predicted exons 1-11; CBG14800 aligns with *Ce*-C54G7.3 predicted exons 12-33. Clones generated from *C. elegans* cDNA include transcripts spanning this large intron (Wormbase 229) confirming that these units represent a single gene that includes a large intron in *C. elegans*. Using RT-PCR, we have confirmed the existence of transcripts spanning this intron (data not shown). The small size of the other C54G7.3 orthologs, *C. brenneri* CBN-LGX-1 and *C. remanei* CRE_00232 (**Table 2.1**), suggests that the annotation of this locus as two separate genes has also been repeated during transcript prediction in these other species, and should be revisited.

Two PDA-encoding homologs have been identified in all the *Caenorhabdita* (**Table 2.1** and **Figure 2.1**) suggesting that a gene duplication event occurred following the separation of the genus from other nematodes and prior to the speciation leading to four separate species. The relative proximity of the two genes on the X chromosome in *C. elegans* supports the argument that paralogs in each of the *Caenorhabdita* may have resulted from a relatively recent duplication event within this genus. The high level of sequence similarity among *Ce*-C54G7.3 and its orthologs (**Figure 2.2**) suggests the possibility of a higher level of functional constraint in residues among these proteins when compared to *Ce*-F48E3.8 and its orthologs.

Luschnig *et al.* (2006) and Wang *et al.* (2006) first elucidated a role for chitin deacetylases in invertebrates. They demonstrated that the Serp and Verm proteins of *Drosophila* act on the chitinous lining of the embryonic dorsal trunk to regulate longitudinal extension of this tracheal tube, following the deposition of luminal chitin by the chitin synthase Kkv/CS-1. Although more complex in form than the *Drosophila* dorsal trunk, the nematode pharynx also contains a tubular structure lined by epithelial cells from which chitin synthase secretes a luminal chitin lining. The wide distribution of PDAs in nematodes suggests that they may be involved in fundamental, highly conserved processes that are key to roundworm development. We propose a model wherein nematode PDAs act on the pharyngeal lining, consequently affecting development in roundworms. Deacetylation converts chitin, an inert, uncharged homopolymer, to chitosan, a partially charged molecule, with different fibrillar structure. In the *C. elegans*

pharynx, this transition could alter the proximity of the pharyngeal lining to the underlying epithelium as changes in charge or hydration cause conformational change. The creation of free amine groups by deacetylation may also allow new covalent linkages to be formed. This raises the prospect of signaling between extracellular chitosan and underlying cells through covalent linkages between chitosan and proximal proteins (integral or peripheral membrane proteins or other proteins of the ECM). Such signaling could control growth of the pharynx, either during organogenesis or in the intermolt periods. The chitinous lining of the pharynx is re-synthesized by CHS-2 prior to each molt (Fanelli *et al.*, 2005; Veronico *et al.*, 2001; Zhang *et al.*, 2005); but unlike insects, nematodes undergo considerable growth between molts, and the pharynx lengthens as well. Deacetylation of the pharyngeal lining may directly regulate radial or longitudinal growth, through physical and chemical changes that alter the interaction between this extracellular lining and the underlying myoepithelium. Deacetylation may be an important mechanism for coordinating growth between pharyngeal cells and their protective barrier.

Regulating pharyngeal growth is a particularly complex problem in organisms like the filarial parasites, where adult worms may be many centimeters long and the pharynx consequently undergoes significant longitudinal extension following the final molt. A stage specific proteome for *B. malayi* has recently been published (Bennuru *et al.*, 2011) and the results demonstrate that the homolog Bml_33340, which we have identified (**Table 2.1**), has a protein product present in stages ranging from microfilaria (equivalent to the *C. elegans* L1)

through to adult male and female worms. Bml_33340 is a somatic PDA and it remains possible that this protein may act to regulate pharyngeal development in *B. malayi*.

Our results show that PDAs affect nematode development and we propose that this results from the action of the PDAs on the chitin lining the pharynx. However, deacetylation of pharyngeal chitin may have other implications for nematodes, and indeed, PDAs may have alternative endogenous or exogenous targets in various free-living or parasitic nematodes (**Table 2.4**).

For example, pharyngeal PDAs may serve to protect the chitinous lining from degradation by chitinases. In *S. pneumoniae*, the peptidoglycan deacetylase *Sp*PgDA is a virulence factor that deacetylates peptidoglycan in the cell wall, protecting the bacterium from degradation by lysozymes (Blair *et al.*, 2005). Pharyngeal PDAs could serve an analogous role in parasitic nematodes by protecting the pharyngeal lining from degradation by chitinases produced as part of the innate immune response of plants and vertebrates. Similarly, free-living nematodes may utilize this deacetylation of pharyngeal chitin to confer protection from bacterial and fungal chitinases produced by environmental microbes.

Other acetylated polysaccharides may be targets for the nematode PDAs we have identified. Caufrier *et al.* (2003) showed that members of this enzyme family may act on various N- and O-acetylated substrates. Heparan sulfate, a glycosaminoglycan (GAG) is required for *C. elegans* pharyngeal morphogenesis (Franks *et al.*, 2006) and members of the PDA family of enzymes have been shown to deacetylate the GAGs heparan sulfate and heparin (Berninsone and

Hirschberg, 1998). Just as endogenous GAGs could be targets of these deacetylases, so too might GAGs in vertebrate host tissues be targets for parasite enzymes, facilitating infection.

Plant parasitic nematodes invade tissues fortified by various carbohydrates including xylan (a complex variety of polysaccharides including the pentose sugar xylose). Members of the PDA family may act to weaken targeted plant tissue, facilitating invasion by plant parasites. Intriguingly, the homolog we have identified from the plant parasitic nematode *Meloidogyne incognita* (Msp9) is likely produced and secreted from esophageal gland cells, since it bears a signal peptide and is expressed in these cells (Huang *et al.*, 2003). The function of this enzyme has not been elucidated, but it is presumed to function as a plant virulence factor like other proteins secreted from the gland cells.

We have demonstrated the expression of both *C. elegans* genes, F48E3.8 and C54G7.3, in the soma, but we cannot exclude the possibility that the genes may also function in the germline, as eggshell chitin represents roughly 90% of all chitin synthesized in the nematode. We were unable to confirm significant levels of chitosan in fertile hermaphrodites, but cannot rule out a role for PDAs in eggshell development. Among the filarial parasites, certain species (including *Brugia spp.*) exhibit a remodeled eggshell, or sheath, surrounding their first stage larvae. Two PDAs were identified in *B. malayi*, including one encoded in the genome of the *Wolbachia* endosymbiont known to be essential to the nematode's fecundity. WolBm0147 is a *bona fide* product found in the *Wolbachia* proteome

(Bennuru *et al.*, 2011) and it is an intriguing prospect that the WolBm0147 protein (which has a signal peptide) may participate in remodeling the eggshell.

Using a differential reacetylation assay, we were unable to biochemically identify chitosan in *C. elegans* (with or without the abundant chitin present from the germline). Two possibilities may explain our results: first, chitosan may be produced in the worms then coupled to some other molecule rendering the amino group inaccessible for *in vitro* reacetylation and chitinase degradation in our experiments; alternatively, the level of chitosan present may be below the level of detection afforded by a colorimetric assay. In the protist *E. invadens*, Das et al. (2006) have detected approximately 60 nmol GlcNAc released per mg cyst (wet weight), attributing roughly 15 nmol to chitin and 45 nmol to chitosan. These quantities are 40-fold greater than the levels of the carbohydrate we observe in wild-type *C. elegans* and 400-fold greater than the levels when compared to germline-deficient worms; given that our values are normalized to lyophilized weight, it suggests that the difference in levels of the carbohydrate per unit mass are far greater than this direct comparison. Our inability to detect chitosan by this method, then, may simply reflect the low levels of the carbohydrate present in the pharynx.

Nematodes may utilize PDAs to target endogenous chitin and GAGS or exogenous substrates, thereby influencing morphogenesis, immunity and virulence. Our results suggest that disrupting the function of polysaccharide deacetylases disrupts nematode development irrespective of their target. Importantly, we have found no close homologs in human sequence, suggesting

that polysaccharide deacetylases (including chitin deacetylases) may prove to be a valid target for the development of interventions that selectively affect human parasites.

ACKNOWLEDGEMENTS

We would like to thank the Caenorhabditis Genetics Center for providing all worm strains used in these experiments. Dr. Yuji Kohara (Center for Genetic Resource Information, National Institute of Genetics, Japan) provided the yk1621b03 and yk1130a03 which were of invaluable use for our RNAi experiments. We thank Lawrence Chen and Ha Cho for preparation of Litmus constructs.

FIGURES

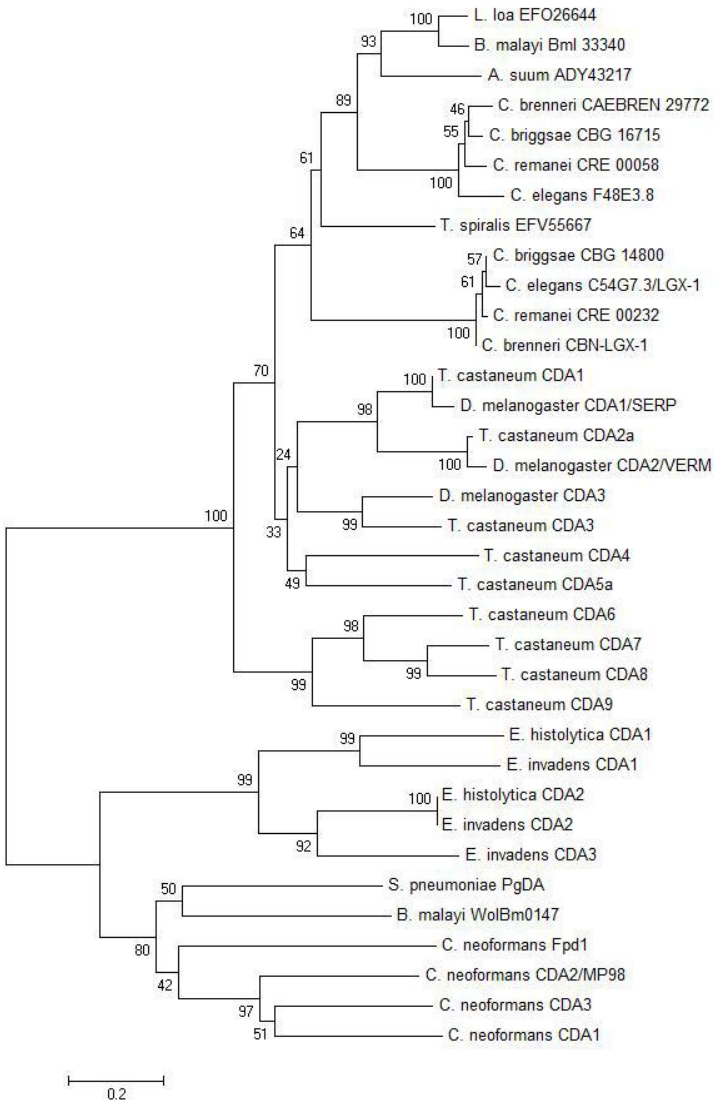


Figure 2.1. Phylogenetic tree showing the relationship of the putative *C. elegans* PDAs to previously characterized peptidoglycan or chitin deacetylases from bacteria, fungi and insects. A ClustalW alignment of residues predicted to encode the polysaccharide deacetylase domain of each protein was used to generate a phylogenetic tree by applying the minimum evolution method using the tree-drawing software Mega 4.0 (Kumar *et al.*, 2004). A bootstrap consensus tree based on 5000 replicates is depicted and bootstrapping values are shown. C54G7.3 and F48E3.8 are two predicted homologs in *C. elegans* and are related to distinct homologs in three other species from the genus *Caenorhabditis.* Other sequences are from the filarial parasitic nematodes *Brugia malayi* and *Loa loa* and from the non-filarial parasitic nematodes *Ascaris suum* and *Trichinella spiralis*. A homologous sequence derived from the *B. malayi* bacterial endosymbiont *Wolbachia sp.* TRS (WolBm0147) is also included along with sequences from the bacterium *Streptococcus pneumoniae*, the protists *Entamoeba histolytica* and *Entamoeba invadens*, the fungus *Cryptococcus neoformans*, and the insects *Drosophila melanogaster* and *Tribolium castaneum*

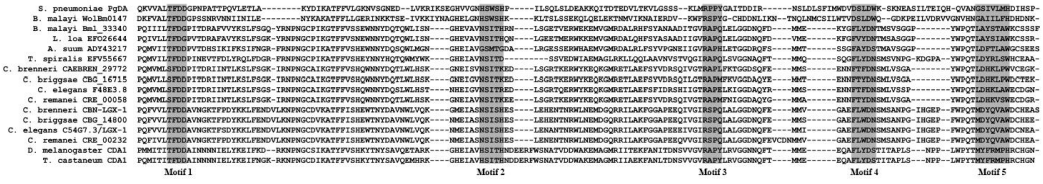


Figure 2.2. Alignment of protein sequences spanning the polysaccharide deacetylase domain (Nod B homology domain) for some newly identified nematode polysaccharide deacetylases and representative sequences from bacteria and insects. The catalytic domain of this family of enzymes has been previously characterized by the inclusion of five motifs essential for activity. Motifs 1 and 2 contribute residues necessary for co-factor binding; Motifs 1, 3, 4 and 5 contribute residues essential for acetyltransferase activity (Blair et al., 2006; Blair et al., 2005; Blair and van Aalten, 2004). All five motifs are conserved in nematode PDAs. Some residues are highly conserved among nematodes but differ from the canonical residues observed in bacterial and/or insect sequences. In general, nematode sequences are more similar to the other metazoan homologs from insects – with large stretches of identity observed among nematode and insect sequences even outside the motifs. Our identified sequence WolBm0147, derived from *Wolbachia sp*. TRS, is more similar to the representative bacterial sequence from *Streptococcus pneumoniae* supporting a bacterial origin of this sequence.

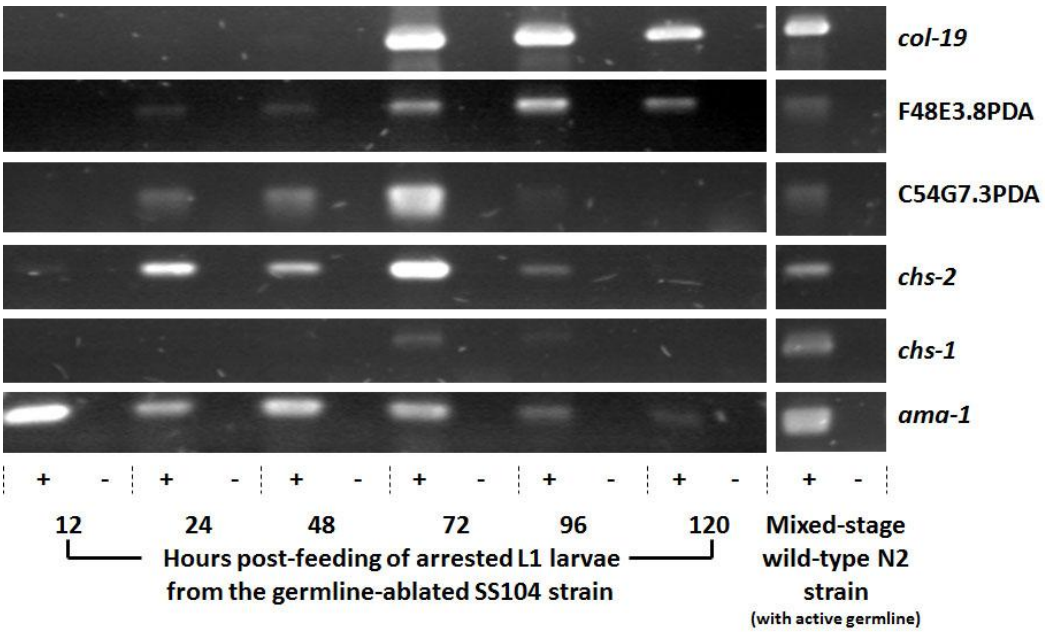


Figure 2.3. Developmental time course of F48E3.8 and C54G7.3 gene expression in germline-ablated *C. elegans* using RT-PCR. Synchronous non-permissive (germline-ablated) worms of this strain were grown on solid media and used to collect RNA from various time-points following L1 arrest at hatching and continuing into adulthood. First strand cDNA (+) was generated from each RNA extract (including No RT controls, -) and used as a template for PCR to analyze expression of the genes F48E3.8, C54G7.3, the eggshell chitin synthase *chs-1* and the pharyngeal chitin synthase *chs-2*. The gene *ama-1*, encoding the large subunit of RNA polII, was used as a positive control (as done by Johnstone and Barry 1996 and Veronico *et. al.* 2001) although we note that expression of the gene is not consistent as previously described. Ablation of the germline was confirmed by the attenuation in expression of *chs-1*. The transition to adult stages was confirmed by the expression of *col-19*, an adult specific collagen gene. All primers and PCR conditions are listed in Supplementary Table S2.

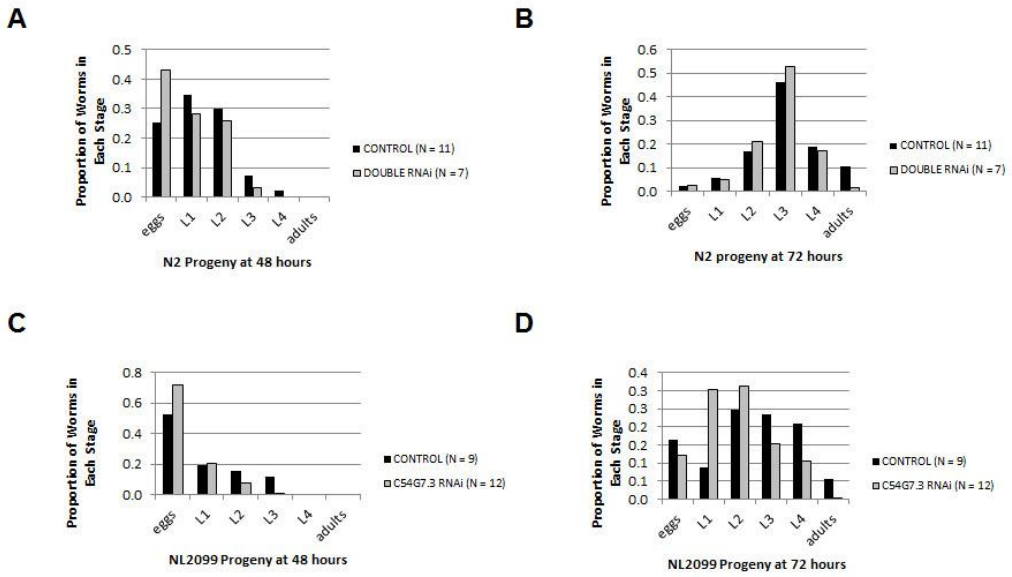


Figure 2.4. RNAi targeting chitin deacetylases affect developmental timing. dsRNA targeting both putative polysaccharide deacetylases F48E3.8 and C54G7.3 was introduced to the wild-type N2 strain (A,B) by feeding. dsRNA targeting the C54G7.3 gene alone was introduced to the RNAi-hypersensitive strain NL2099 (C,D) by feeding. Single L3 larvae were allowed to proceed through development while feeding before progeny were scored at 48 hours (A,C) and 72 hours (B,D). RNAi targeting both genes results in a delay in wild-type worm development that can also be observed when C54G7.3 is targeted through RNAi in a strain hypersensitive to exogenous RNAi.

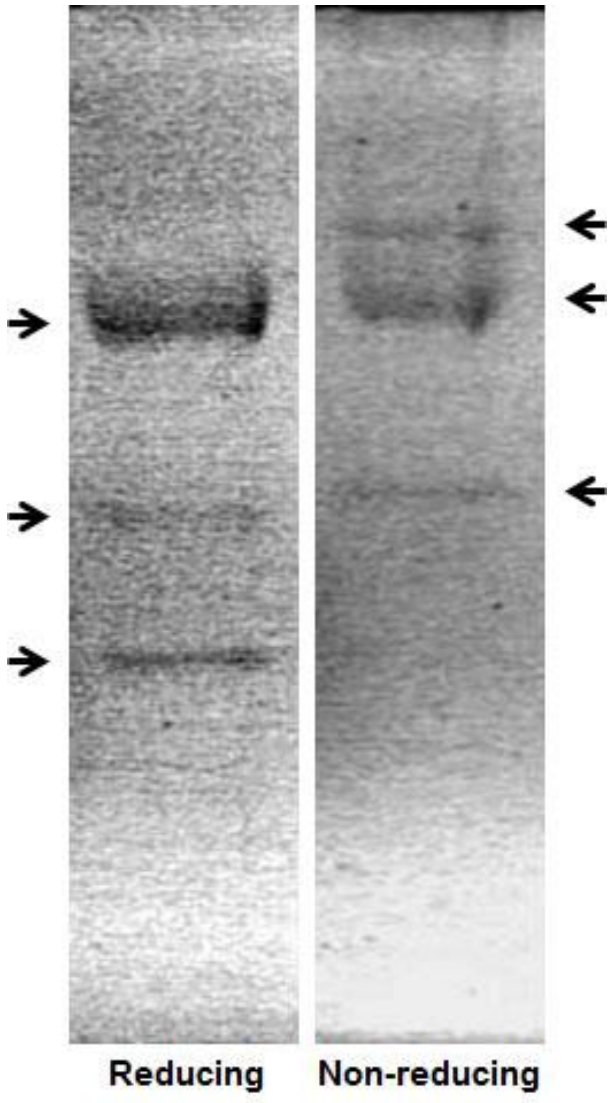


Figure 2.5. Soluble extracts of mixed-stage N2 worms deacetylate chitin *in vitro*. *C. elegans* soluble extract (14 µg total protein per lane) was denatured at 80°C and separated by SDS-PAGE under reducing and non-reducing conditions. Proteins were renatured for 24 hours at RT, enabling enzymes to act on glycol chitin embedded within the gel. Gels were stained with calcoflour. Hyperfluorescent bands mark zones where chitosan was produced. (Image is black/white reversed)

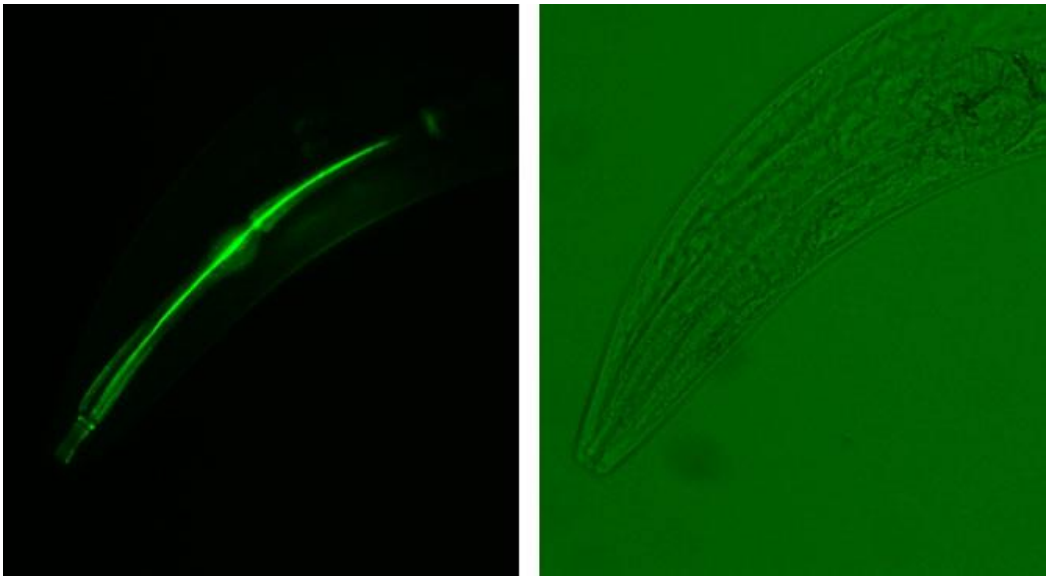


Figure 2.6. Eosin Y, a dye specific for chitosan in the fungal cell wall, stains the *C. elegans* pharynx. N2 worms were stained without fixation using Eosin Y and visualized with an Olympus BX40 epifluorescent microscope with a Chroma 31001 filter (left) or with bright field (right).

TABLES

Table 2.1. Nematode PDAs Identified Through Bioinformatics using resources from NCBI, Sanger, and www.nematode.net

Species	Clade	NCBI Accession #	Previously Assigned Names or Retrieval Information Relevant for Other Databases	Total # of Residues in Archived Sequence	Residues Included in Alignments and Tree Relative to Archived Sequence
Ancylostoma caninum	V	N/A	Acan_isotig17215 (Nematode.net)	-	-
Ascaris suum	III	ADY42356[a]	ADY42356	-	-
		ADY43217[b]	ADY43217	735	578-733
Brugia malayi	III	AAW70738	WolBm0147, YP_197908	264	55-214
		EDP33092	Bml_33340; 14961.m05161 (TIGR)	1802	1469-1624
Caenorhabditis brenneri	V	EGT33229	CAEBREN_29772	2693	2373-2526
		EGT30475	CBN-LGX-1	666	321-481
Caenorhabditis briggsae	V	CAP34613	CBG_16715	2523	2202-2355
		CAP33224	CBG_14800	699	354-513
Caenorhabditis elegans	V	CCD67046	C54G7.3	1876	1500-1533, 1557-1650[c]
		ABX00803	F48E3.8	2444	2124-2277
Caenorhabditis remanei	V	EFO83068	CRE_00058	2545	2225-2378
		EFO82567	CRE_00232	682	335-498
Loa loa	III	EFO26644	LOAG_01839	416	83-237
Meloidogyne incognita[d]	IVb	AF531169	Msp9, MI00182 (Nematode.net), MI04101 (Nematode.net)	-	-
Oesophagostomum dentatum	V	N/A	Oden_isotig26807 (Nematode.net)	-	-
Onchocerca flexuosa	III	N/A	OF01019 (Nematode.net)	-	-
Onchocerca volvulus	III	N/A	contig33574 (Sanger)	-	-
		N/A	contig48143 (Sanger)	-	-
Ostertagia ostertagi	V	N/A	OS02548 (Nematode.net)	-	-
Teladorsagia circumtincta	V	N/A	Tcir_isotig26540 (Nematode.net)	-	-
Trichinella spiralis	I	EFV55667	Tsp_03980	1524	1190-1343

a, b The sequences ADY42356 and ADY43217 share 99% sequence identity within the predicted NodB homology domain, but only 32% sequence identity outside this region. ADY42356 has a truncated portion of the catalytic domain. We are unable to discern these as 1 or 2 homologs, but given the catalytic domain identity we use only ADY43217 for alignments.

c Based on our sequencing of this region of the gene we have shown that predicted residues 1534-1556 (relative to predicted sequence associated with Accession No. CCD67046) would not be present in the protein sequence.

d *M. incognita* Msp9 is a potential homolog which has sequence verification from two clones available from www.nematode.net. Msp30 (AY142120) shows some homology as well, but there is less supporting data.

Table 2.2. Brood Size and Progeny Viability for Reported RNAi Experiments

TREATMENT		AVERAGE # OF PROGENY WITHIN 72 HRS OF PLATING P1 WORMS	
		Average Progeny	Progeny Viability
F48E3.8 & C54G7.3 (N2)	Control (N = 11)	40	97.7%
	Experimental (N = 7)	37	97.7%
C54G7.3 (*rrf-3*)	Control (N = 9)	42	83.5%
	Experimental (N = 12)	32	88.0%

Table 2.3. Chitin and chitosan measured in *C. elegans* tissues

SAMPLE & TREATMENT		GLCNAC RECOVERED FROM CHITINASE DIGESTION (NMOL/MG WORM DRY WEIGHT)
Mixed-stage worms, including adults with active germline: *measures germline and somatic chitin in various stages* (4 independent trials)		0.499, 0.508, 1.099, 1.511
Adult worms, germline-ablated: *measures somatic chitin in adults* (3 independent trials)		0.050, 0.055, 0.170
Mixed-stage worms, including adults with active germline	Control (mock reacetylated): *measures germline and somatic chitin in various stages*	0.538 +/- 0.061
	Reacetylated: *measures germline and somatic chitin and chitosan in various stages*	0.551 +/- 0.091

Table 2.4. Potential roles for nematode PDAs

TARGET		EFFECT	OUTCOME
Endogenous	**Pharyngeal chitin**	Facilitates signaling to underlying epithelium; control of length/diameter of chitinous lining	Pharyngeal morphogenesis
		Affects recognition of the pharyngeal lining by other enzymes	Protection of gut from exogenous chitinases
	Eggshell chitin	Alters physical properties of the macromolecule	Hatching or eggshell remodeling
	GAGS or other acetylated hexosamines	Affects various context-specific developmental functions	Morphogenesis
Exogenous	**Xylan (Plant hosts)**	Permits penetration of host tissues	Virulence
	GAGs (Vertebrate hosts)	Permits penetration of host tissues	Virulence

Chapter 2 SUPPLEMENT

Pharyngeal Polysaccharide Deacetylases Affect Development in the Nematode *C. elegans* and Deacetylate Chitin *In Vitro*

STATEMENT OF AUTHOR CONTRIBUTION

RJH completed all bioinformatics searches (**Table 2.1**), phylogenetic analyses (**Figure 2.1**), sequence alignments (**Figure 2.2**), RT-PCR experiments (**Figure 2.3**) and biochemical attempts to identify chitosan *in vitro* (**Table 2.3**). KJB executed all the RNAi experiments presented here (**Figure 2.4** and **Table 2.2**). HKN and LTL completed substrate gel analysis (**Figure 2.5**).
MCR completed Eosin Y staining (**Figure 2.6**). RJH completed all data analysis and prepared all figures for final presentation with the help of JAF and CAS.

SUPPLEMENTARY MATERIALS AND METHODS

Identification and phylogenetic analysis of nematode CDAs

In addition to the *Sp*PDA, *B. malayi* and *C. elegans* peptide sequences corresponding to the PDA domain, bioinformatics searches were also done using the PDA-domain sequences for *Dm*CDA1-3, *Tc*CDA1-9, *Cn*CDA1-3 and *Cn*FPD1 (accession numbers listed in **Table 2S.1**) to ensure that sequence divergence could not account for otherwise unidentified nematode homologs.

The alignments and trees are based exclusively on the sequences predicted to encode the PDA domain in each protein. This catalytic domain from *Tribolium castaneum*, *Brugia malayi* and the *Wolbachia* endosymbiont of *B. malayi* proteins are depicted along with other relevant functional domains in relevant isoforms (**Figure 2S.1**).

When calculating bootstrapping values under the minimum evolution method employed for the creation of the phylogenetic tree, I used the Poisson Correction for the Substitution model (a user-variable parameter).

RT-PCR

For all cDNA preparations, the major source of non-*C. elegans* RNA/cDNA material would be derived from the OP50 strain of *E. coli* bacteria used to culture worms. In order to ensure that amplification would not reflect bacterial genes or transcripts, all primers were tested for the likelihood of annealing to bacterial sequence (in addition to alternative *C. elegans* sites) using a BLASTn search to ensure that the primers did would not align over 16 contiguous bases including one of three terminal 3' bases of the primer. All primers used in this study were established to have negligible likelihood of annealing to bacterial target sequences.

Prior to application in time-course experiments, all primer combinations were tested for the ability to efficiently amplify from gDNA and cDNA templates. Additional information on all primer sequences used in these studies is available in **Appendix A**.

Chitin and Chitosan Detection

The method used in attempts to quantify chitin and chitosan in nematodes is summarized in **Figure 2S.14**. This method is a differential reacetylation assay coupled with a modified Morgan Elson assay for detection. The original protocol

was that previously used for the quantification of these carbohydrates in fungi (Charles Specht, personal communication). Original steps are presented in black text while modifications added during my optimization of the protocol for *C. elegans* are highlighted in red (**Figure 2S.14**). For drying of samples, I used lyophilization rather than desiccation at 37°C (the latter used for analysis of fungal material). Following treatment with acetic anhydride for reacetylation, I introduced a wash step using 0.1M glycine to remove residual acetic anhydride and acetic acid.

SUPPLEMENTARY RESULTS AND DISCUSSION

Analysis of the NEXTDB in situ hybridization data for F48E3.8 and C54G7.3 expression

The NextDB database is an online resource with *in situ* hybridization results for all *C. elegans* genes from all life stages spanning embryos to adults. Data in the NextDB database was cited by Gaudet and Mango (2002) as supporting evidence for the pharyngeal expression of the genes F48E3.8 and C54G7.3. The spatial expression pattern for C54G7.3 is easily distinguishable as localized to the *C. elegans* pharynx in various life stages; however, F48E3.8 expression (probe based on clone 272e1) appears only as a weak expression in a small region of the embryos (before extensive cellularization or organogenesis is apparent). This suggests a gene that is "on" early during pharyngeal morphogenesis. In order to confirm that this is, in fact, the staining pattern that would be seen for early pharyngeal expression, these images were compared to

the *in situ* data available for other genes noted as being expressed during early pharyngeal morphogenesis. In attempts to dissect the coordinated activity of *pha-4* and other transcription factors in directing the tightly regulated temporal patterns of pharyngeal gene expression, 34 genes described as "early" pharyngeal genes were identified and studied at the level of their gene regulatory networks (Gaudet *et al.*, 2004). F48E3.8, but not C54G7.3, was studied as one of these genes with early expression. I evaluated the expression patterns for the other 33 genes using the *in situ* data available from the NextDB database and found that while there is noticeable variation among the expression patterns, there are some genes that have embryonic expression sufficiently similar to F48E3.8 to support and interpretation that what is seen in the embryo is the expression in the pharyngeal primordium.

All primer pairs used in the RT-PCR time course experiment were found to efficiently amplify target sequence from either gDNA or mixed-stage *wild-type* cDNA templates prior to amplification in this study. Amplimers of the predicted sizes were obtained for *ama-1* (**Figure 2S.2**) and RNA/cDNA quality was not affected by DNAse treatment prior to the use of reverse transcriptase to generate cDNA (**Figure 2S.3**). Similarly, amplimers of the predicted sizes were obtained for *chs-1* when using newly designed primers or primers used by Veronico *et al.* (2001) (**Figures 2S.4 and 2S.5**) and for *chs-2* when using newly designed primers or primers used by Veronico *et al.* (2001) (**Figures 2S.6 and 2S.7**). [Primers used by Veronico *et al.* (2001) were not published and were obtained by personal communication.] Primers targeting the predicted PDA-encoding domain of

C54G7.3, the predicted PDA-encoding domain of F48E3.8, *egg-3* and *col-19* could all efficiently amplify target sequences (**Figures 2S.8 – 2S.11**).

Use of the SS104 strain for time course experiments

During our temporal analysis of gene expression in synchronized cultures of the germline-ablated SS104 worms (**Figure 2.3**), we noted that the vulva in some hermaphrodite worms appeared as early as the 72-hr time-point and in all hermaphrodites no later than the 96-hr time-point when reared at 25°C to ablate germline activity. This timing of the vulva appearance is significantly delayed when compared to wild-type N2 worms grown at 25°C where it can be observed in most worms as early as a 30-hr time point. The *glp-4* (bn2) allele carried in this strain has been noted to exhibit incomplete gonadal morphogenesis (Wormbase WS227), and the pleiotropy of this allele may explain the observed delay of vulva appearance in our experiments.

Additionally, we noted detection of trace levels of the *chs-1* gene in the time course presented here (72-hr) and, excluding the possibility that this gene may be expressed in somatic tissue, these results suggest that germline ablation is not complete. The temperature-sensitive *glp-4* (bn2) allele results in the production of *essentially* no germline and effectively causes sterility in both males and hermaphrodites since the gene is essential to permit *efficient progress* through mitotic prophase and continued proliferation (Pam Hoppe and Tim Schedl, personal communication; Wormbase WS227). This presents a scenario in which proliferation is reduced but in which a minimal number of gametes are still

produced. The punctuated *chs-1* expression we observe in our time course may reflect some limited germline activity. The *glp-4* gene has recently described as a valyl amino-acyl tRNA synthetase (Rastogi *et al.*, 2011).

In the results presented here, synchronous cultures of worms were cultured on solid NGM media. In initial attempts at this experiment, synchronous cultures were grown in liquid culture. When worms were grown in liquid culture, *chs-1* expression was more abundant in time points surrounding the onset of adulthood and was evident whether RNA extracts were untreated or treated with Turbo DNAse prior to cDNA synthesis (**Figure 2S.12** and **Figure 2S.13**).

Chitosan Detection using Biochemistry

Prior to use in the assay, the chitinase batch was tested for its ability to liberate GlcNAc residues from glycol chitin. The enzyme activity using this substrate was measured at 0.0160 nmol/μL/hr (when compared to a matched control with no enzyme).

During my trials of the experiment, I found that using lyophilization for sample drying resulted in a 2-fold increase in the product detected over samples that were prepared by desiccation.

Since initial trials of the experiment suggested that product detection increased with the size of pellets stored in the 2.0-mL screw cap tubes, I attempted to further increase the size of pellets (and perform the experiments in 15-mL conical tubes) to see whether absolute detection increased. This modification also required adapting spins from a microcentrifuge to a centrifuge.

I found that during the reacetylation step, when there is excessive effervescence, the worms moved along the walls of the tube and were not properly subject to the reacetylation and KOH degradation steps; hence, intact worms could be observed after treatments and the sample could not properly be retrieved.

I attempted to perform the KOH degradation step before the *in vitro* reacetylation (to test whether removal of the surrounding tissue first allowed for increased chitosan accessibility for the addition of acetyl groups), but found that I obtained similar results. Therefore, I maintained the order of steps for previously published work (Baker *et al.*, 2007; Banks *et al.*, 2005).

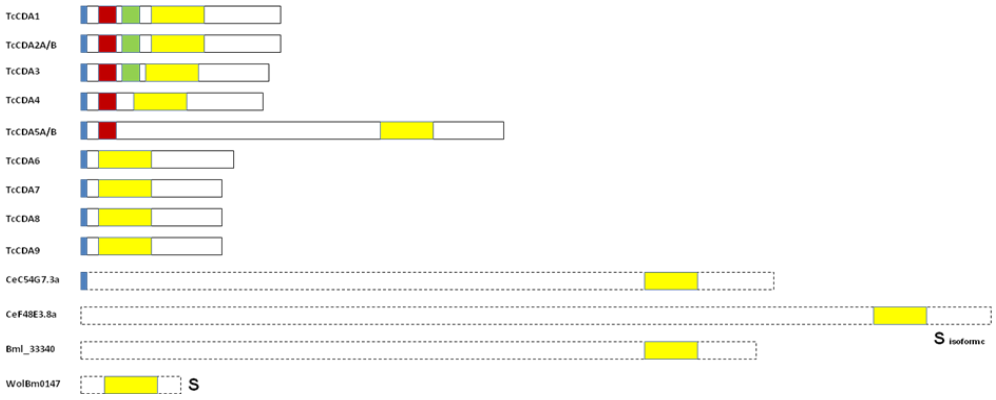
Chitosan Detection using Microscopy

We have investigated the localization of chitosan by way of staining with the dye Eosin Y, which has recently been reported to recognize chitosan (but not chitin) in fungal cell walls (Baker *et al.*, 2007; Banks *et al.*, 2005). Further attempts to investigate chitosan localization and to confirm localization in the *C. elegans* pharynx may benefit from the use Cibracron Brilliant Red [Sigma Catalog #228451 3D-A], a fluorophore that has recently been found to stain chitosan in the fungal cell wall specifically, or Solopheyl Red dye (Charles Specht, personal communication).

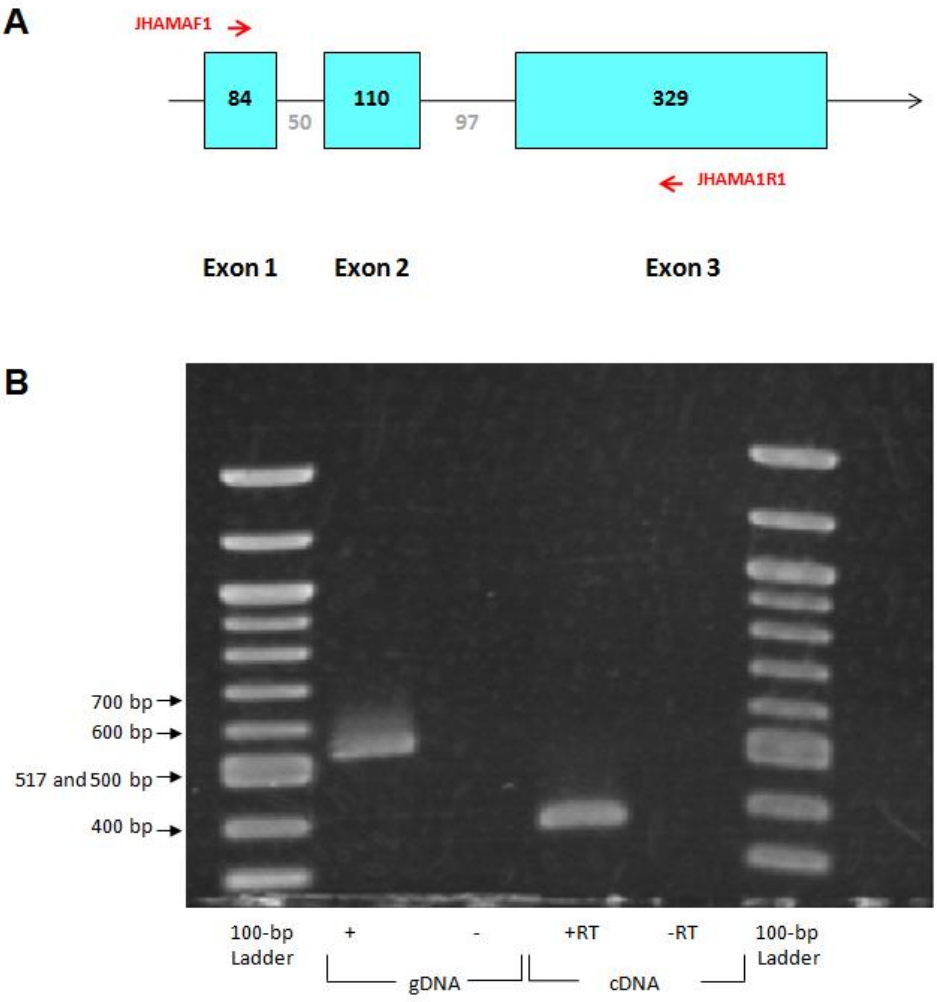
Chitosan Detection Using an *In Vivo* Assay

We have attempted to use an *in vitro* differential reacetylation assay in attempts to quantify chitin and chitosan present in the *C. elegans* germline and soma, noting that the low absolute levels of the carbohydrate limit our ability to detect and quantify chitosan using the procedure. In staining of live *C. elegans* worms, we note that Eosin Y binds to the lining of the *C. elegans* pharynx supporting the possibility that chitosan is present there. Additional methods can be employed to corroborate this finding. An *in vivo* method may be able to take advantage of the same principle of chitin (but not chitosan) susceptibility to chitinase degradation while using development and behavior as a dependent variable. If deacetylation converts pharyngeal chitin to chitosan, this modification may protect the pharyngeal lining from chitinase-mediated degradation and can be tested when asking whether worms fed bacteria – producing chitinase at natural levels or in excess – proceed through development normally when the function of polysaccharide deacetylases is attenuated. Therefore, we can ask whether PDA gene loss (through knockouts or other reverse genetic techniques) increases the survival of worms eating chitinase producing bacteria.

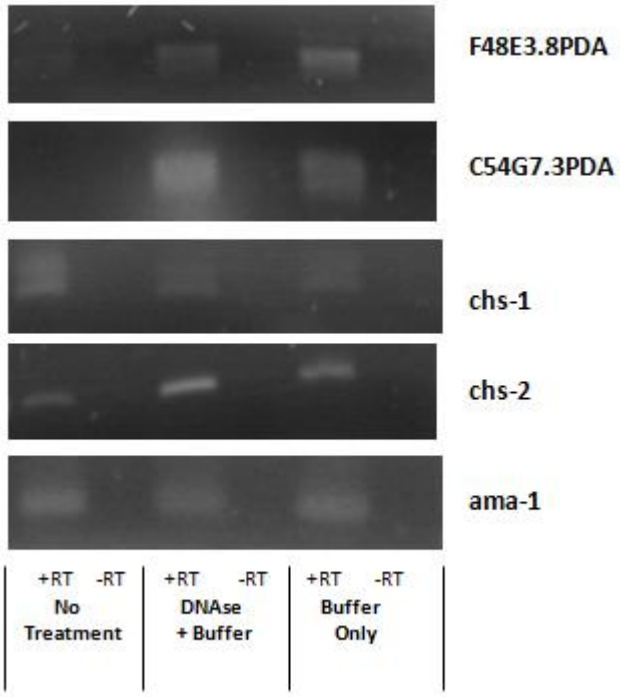
SUPPLEMENTARY FIGURES



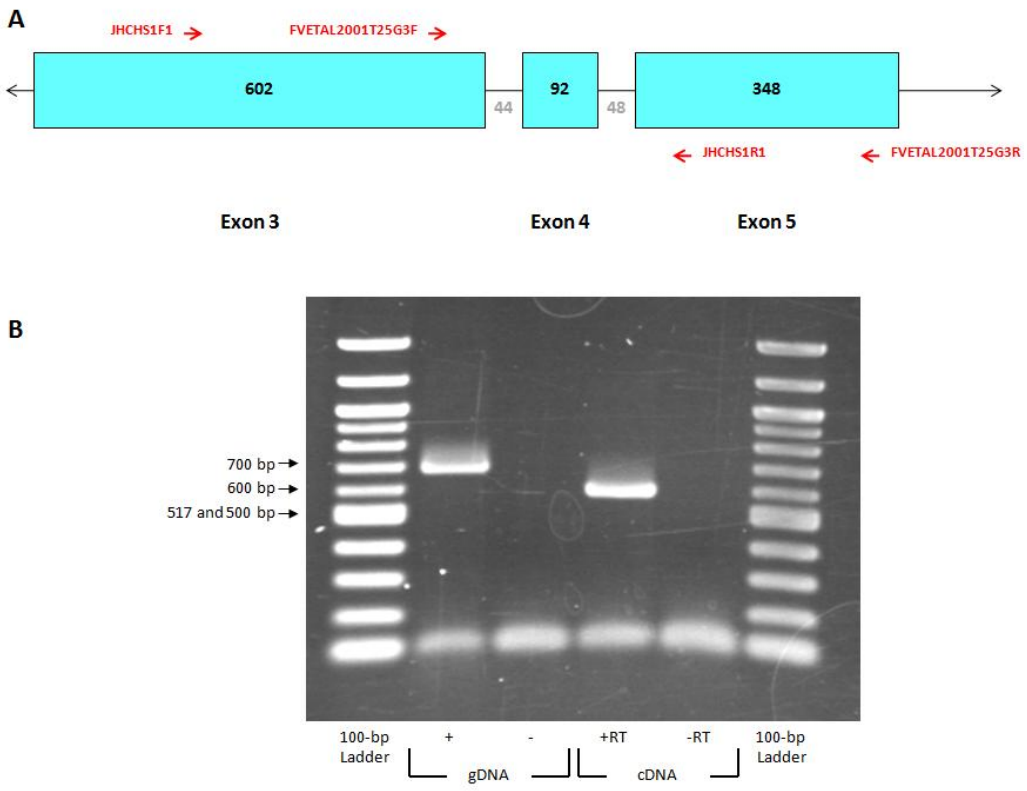
Supplementary Figure 2S.1. Comparison of protein sizes and domain architecture among the *T. castaneum* family of PDAs and newly identified nematode PDAs. Nine chitin deacetylase-encoding genes were identified in the flour beetle *Tribolium castaneum* (*Tc*) and single transcripts were identified for each gene, except for TcCDA5 where two alternate transcripts of the same length are produced (Dixit *et al.*, 2008). Representative sizes of the nine isoforms that would be produced from these 9 genes are shown here. All *Tc* proteins have a signal peptide (blue boxes) and a polysaccharide deacetylase domain (PDA, yellow boxes). *Tc*CDA1-5 all bear chitin-binding domains (burgundy boxes) while *Tc*CDA1-3 also contain LDLa receptor domains (green boxes). For comparison, the largest predicted isoforms from *C. elegans* C54G7.3 and F48E3.8 genes, the *B. malayi* gene Bml_33340, and the *B. malayi* endosymbiont *Wolbachia* gene WolBm0147 are depicted. All nematode-derived isoforms also bear a PDA domain. When all predicted isoforms of F48E3.8 and C54G7.3 are analyzed for the presence of classical or non-classical secretion signals, the C54G7.3 isoforms a and b which share the N-terminal end, are found to have a signal peptide. Isoform c of F48E3.8, the shortest isoform and only one predicted from a fully sequenced transcript, does not have a signal peptide but does have a non-classical secretion signal (S). *Bm*Bml_33340 has neither a classical nor a non-classical secretion signal. The *Wolbachia* WolBm0147 protein bears a non-classical secretion signal. *(Isoforms are drawn to comparative scale. Solid borders reflect isoforms predicted from sequenced transcripts while dashed borders reflect isoforms predicted from unsequenced or partially sequenced transcripts or the gene size for WolBm0147.)*



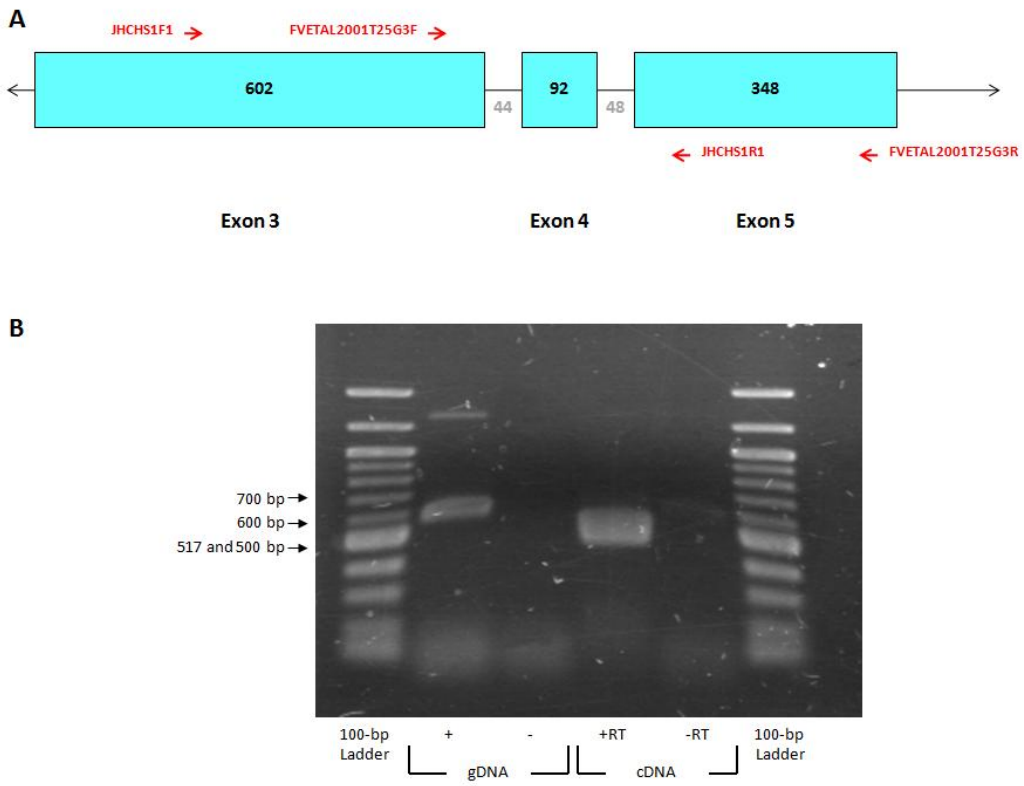
Supplementary Figure 2S.2. *ama-1* primer specificity and analysis of amplimer size. PCR was performed using the JHAMA1F1 and JHAMA1R1 primers, described by Johnstone and Barry (1996). **(A)** Schematic depicting annealing sites for these primers. **(B)** PCR was performed using these primers under the following conditions: Stage 1: 94°C, 3 minutes; Stage 2: [94°C, 1 minute; 56°C, 1 minute; 72°C, 3 minutes] x30; Stage 3:72°C, 10 minutes. Amplimers are predicted to be 501 bp (gDNA) and 354 bp (cDNA). These primer sequences and optimized PCR conditions are also listed in **Supplementary Table 2S.2**.



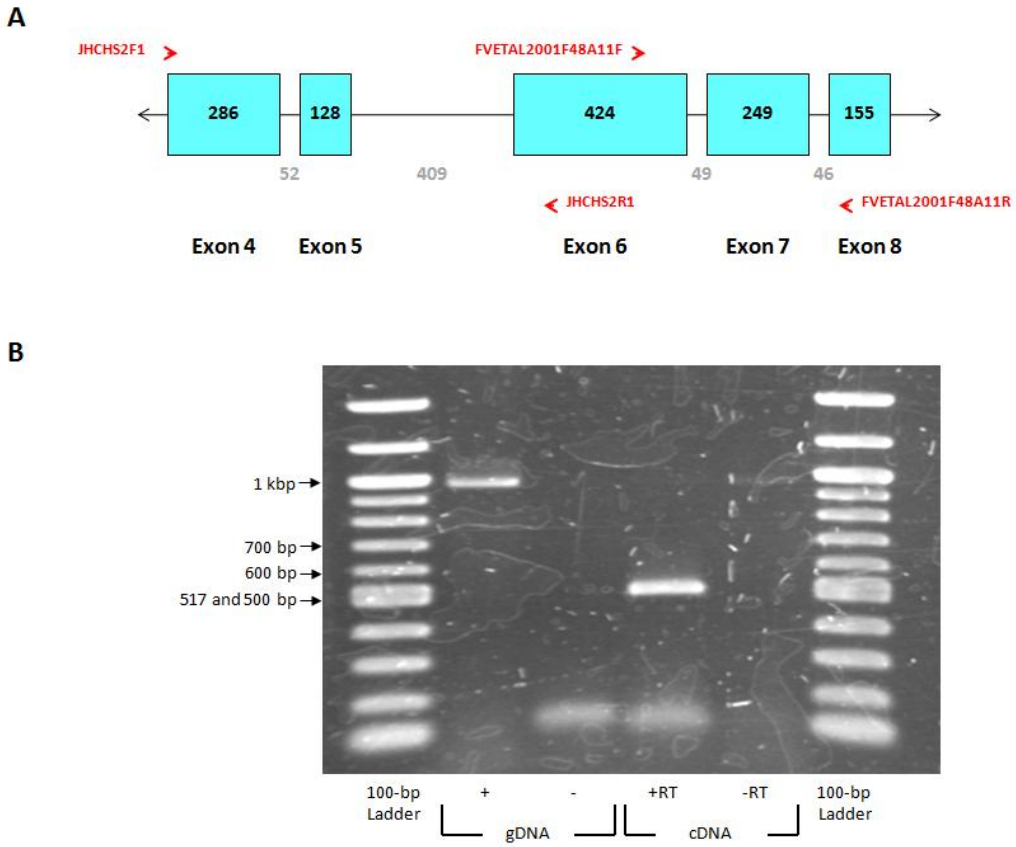
Supplementary Figure 2S.3. Verification of DNAse efficiency and post-treatment RNA integrity. ~ 2 µg of DNAse-treated RNA was used in three DNAse or Mock DNAse treatments before serving as template for cDNA synthesis (+RT) or matched controls (-RT) before amplification for *ama-1*, *chs-1*, *chs-2*, C54G7.3 and F48E3.8 were used to confirm the integrity of RNA following DNAse treatment. All amplimer sizes observed here were the same as those seen from cDNA and presented in **Figure 2.3**.



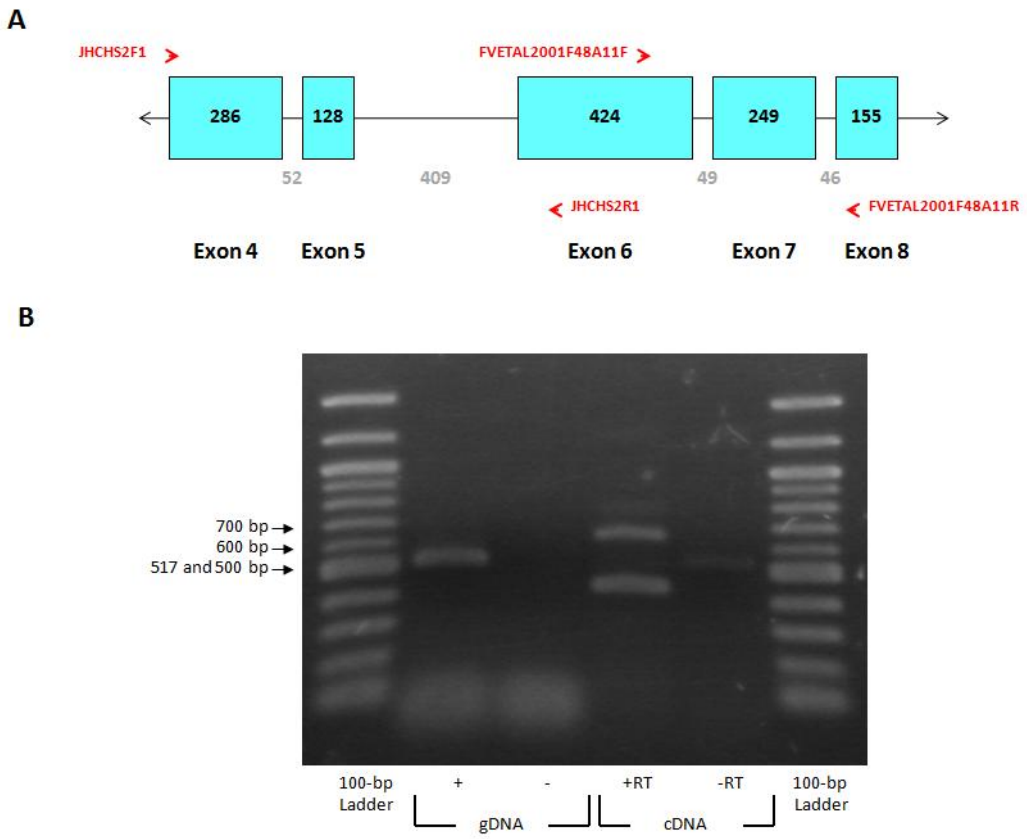
Supplementary Figure 2S.4. *chs-1* primer specificity and analysis of amplimer size. PCR was performed using the JHCHS1F1 and JHCHS1R1 primers. **(A)** Schematic depicting annealing sites for these primers. **(B)** PCR was performed using these primers under the following conditions: Stage 1: 94°C, 3 minutes; Stage 2: [94°C, 1 minute; 59.9°C, 1 minute; 72°C, 3 minutes] x2; Stage 3: [94°C, 1 minute; 63.7°C, 1 minute; 72°C, 3 minutes] x28; Stage 4:72°C, 10 minutes. Amplimers are predicted to be 673 bp (gDNA) and 581 bp (cDNA), including engineered primer ends that do not anneal to the endogenous target sequence. Optimized primer sequences and PCR conditions are also listed in **Supplementary Table 2S.2**.



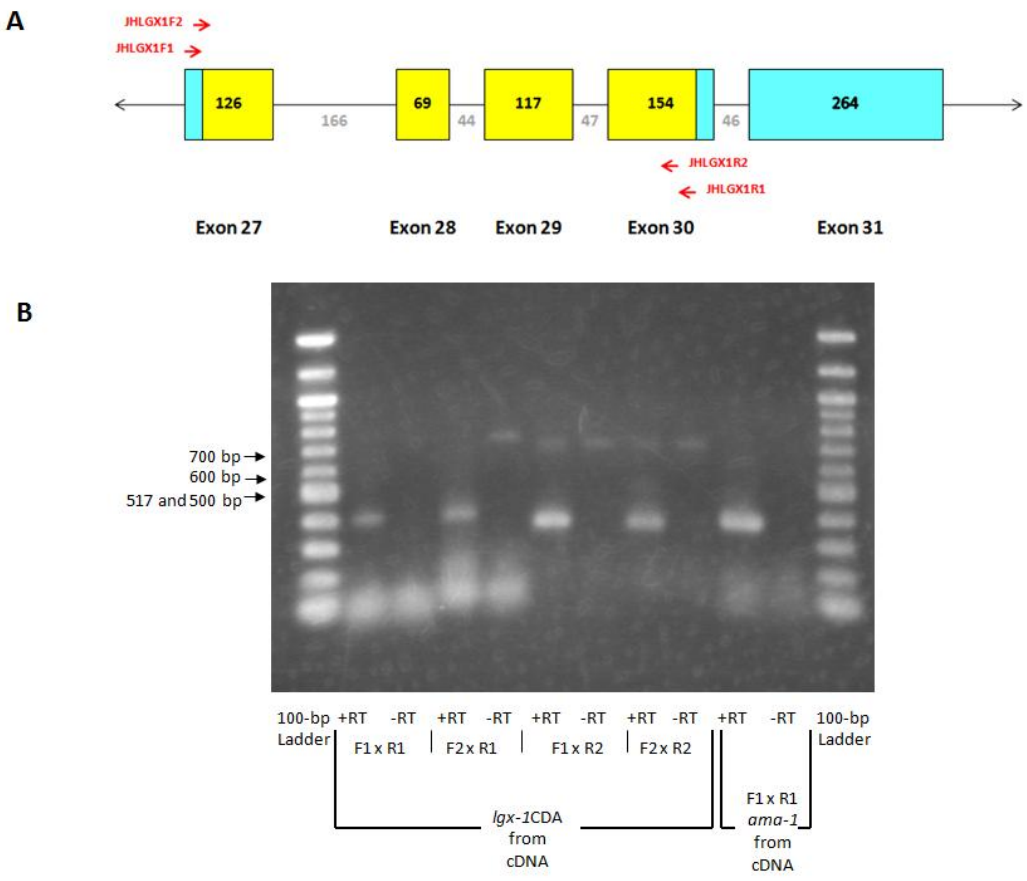
Supplementary Figure 2S.5. ***chs-1*** **primer specificity and analysis of amplimer size based on Veronico *et al.* (2001) primers.** PCRs were performed using the FVETAL2001T25G3F and FVETAL2001T25G3R primers. **(A)** Schematic depicting annealing sites for these primers. **(C)** PCR was performed using these primers under the following conditions: Stage 1: 94°C, 3 minutes; Stage 2: [94°C, 1 minute; 47.0°C, 1 minute; 72°C, 3 minutes] x30; Stage 3:72°C, 10 minutes. Amplimers are predicted to be 564 bp (gDNA) and 472 bp (cDNA). The PCR protocol used with these primers were unpublished and using this trial protocol a single, distinct band from gDNA was not obtained, so subsequent experiments used the primers studied in **Figure 2S.4**.



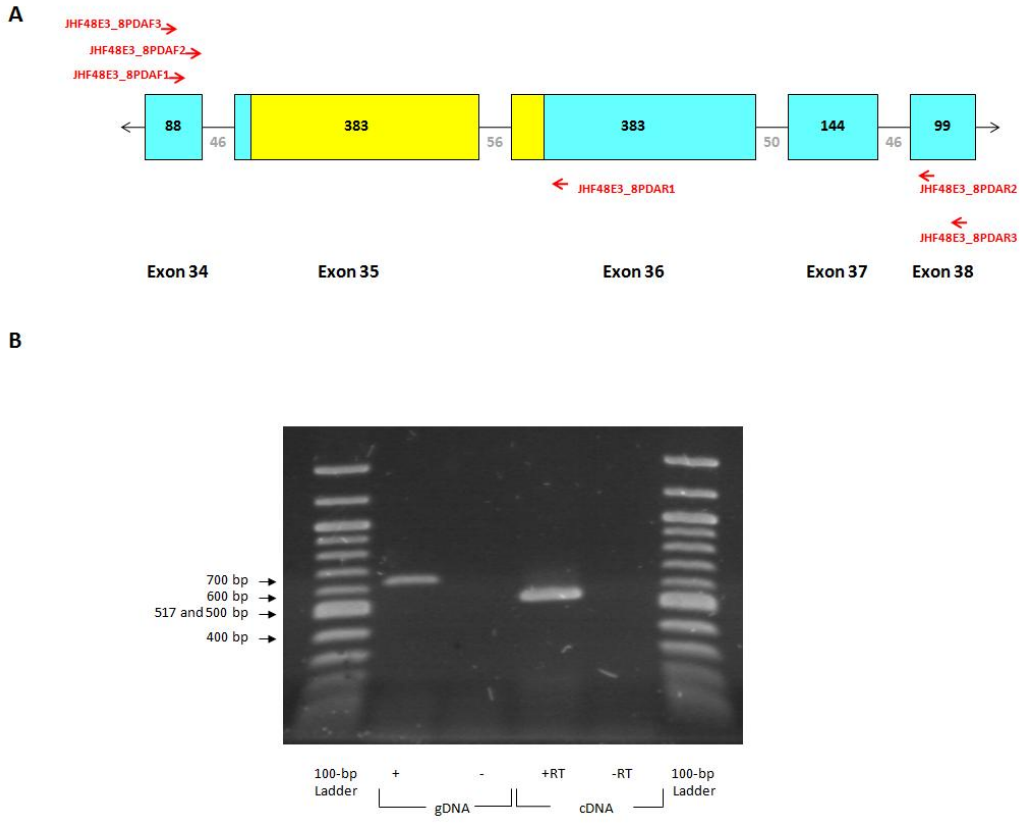
Supplementary Figure 2S.6. *chs-2* primer specificity and analysis of amplimer size. PCR was performed using the JHCHS2F1 and JHCHS2R1 primers. **(A)** Schematic depicting annealing sites for primers. **(B)** PCR was performed using these primers under the following conditions: Stage 1: 94°C, 3 minutes; Stage 2: [94°C, 1 minute; 58.5°C, 1 minute; 72°C, 3 minutes] x2; Stage 3: [94°C, 1 minute; 62°C, 1 minute; 72°C, 3 minutes] x28; Stage 4:72°C, 10 minutes. Amplimers are predicted to be 969 bp (gDNA) and 508 bp (cDNA). Optimized primer sequences and PCR conditions are also listed in **Supplementary Table 2S.2**.



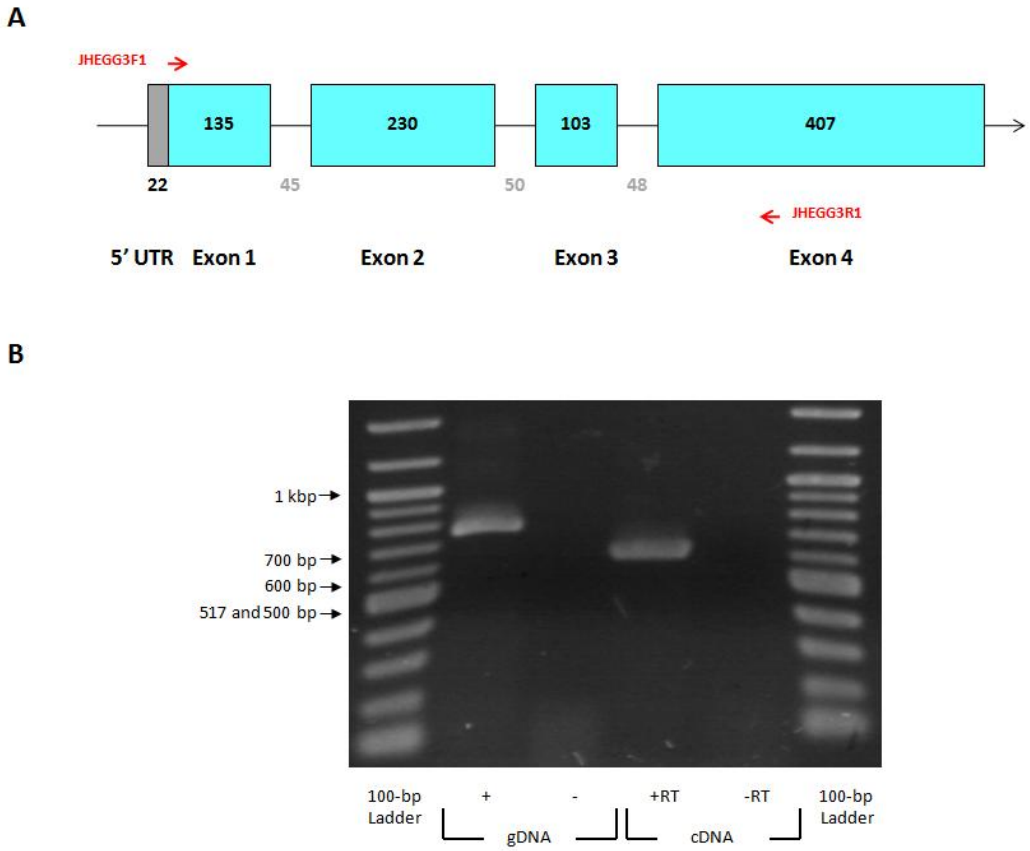
Supplementary Figure 2S.7. *chs-2* primer specificity and analysis of amplimer size based on Veronico *et al.* (2001) primers. PCRs were performed using the FVETAL2001F48A11F and FVETAL2001F48A11R primers. (**A)** Schematic depicting annealing sites for these primers. **(B)** PCR was performed using these primers under the following conditions: Stage 1: 94°C, 3 minutes; Stage 2: [94°C, 1 minute; 52.0°C, 1 minute; 72°C, 3 minutes] x30; Stage 3:72°C, 10 minutes. Amplimers are predicted to be 508 bp (gDNA) and 413 bp (cDNA). The PCR protocol used with these primers were unpublished and using this protocol no amplimer was detectable from gDNA while multiple bands were obtained from cDNA, so subsequent experiments used the primers studied in **Figure 2S.6**.



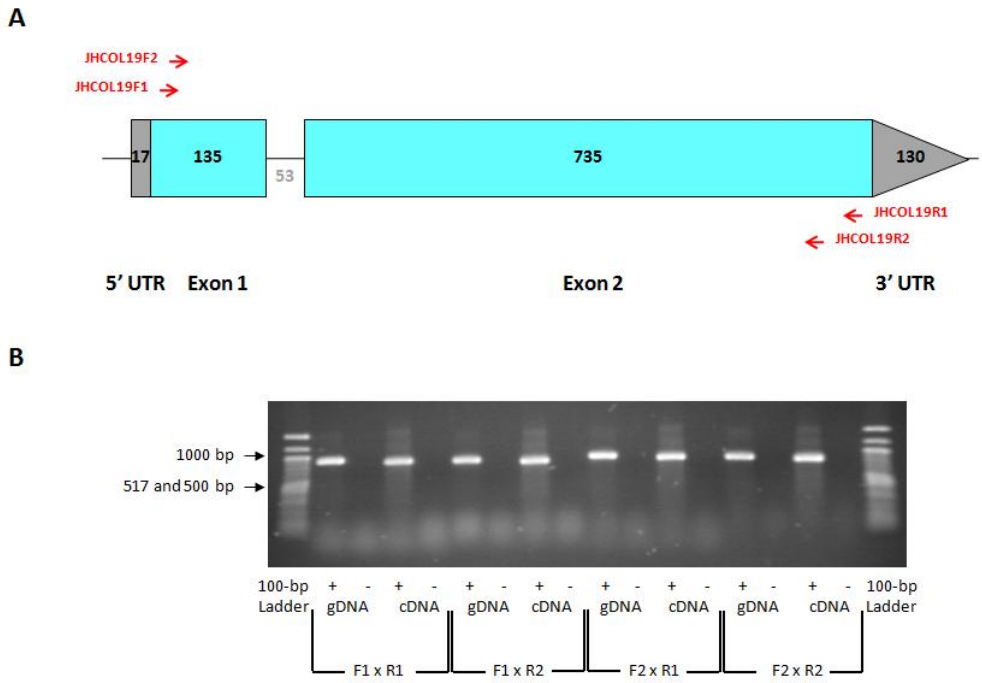
Supplementary Figure 2S.8. C54G7.3PDA primer specificity and analysis of amplimer size. PCR was performed using combinations of the forward primers JHLGX1F1 and JHLGX1F2 and the reverse primers JHLGX1R1 and JHLGX1R2. **(A)** Schematic depicting annealing sites for these primers relative to the region originally predicted to encode the PDA domain (yellow). **(B)** Amplimers of varying sizes are produced with different primer combinations. JHLGX1F1 and JHLGXR2 were selected for subsequent experiments and used with the following PCR conditions: Stage 1: 94°C, 3 minutes; Stage 2: [94°C, 1 minute; 61.4°C, 1 minute; 72°C, 3 minutes] x2; Stage 3: [94°C, 1 minute; 64.7°C, 1 minute; 72°C, 3 minutes] x28; Stage 4:72°C, 10 minutes. The optimized primer sequences and PCR conditions are also listed in **Supplementary Table 2S.2**. These primers generate amplimers predicted to be 678 bp (gDNA) and 421 bp (cDNA), including engineered primer ends that are not complementary to target sequence.



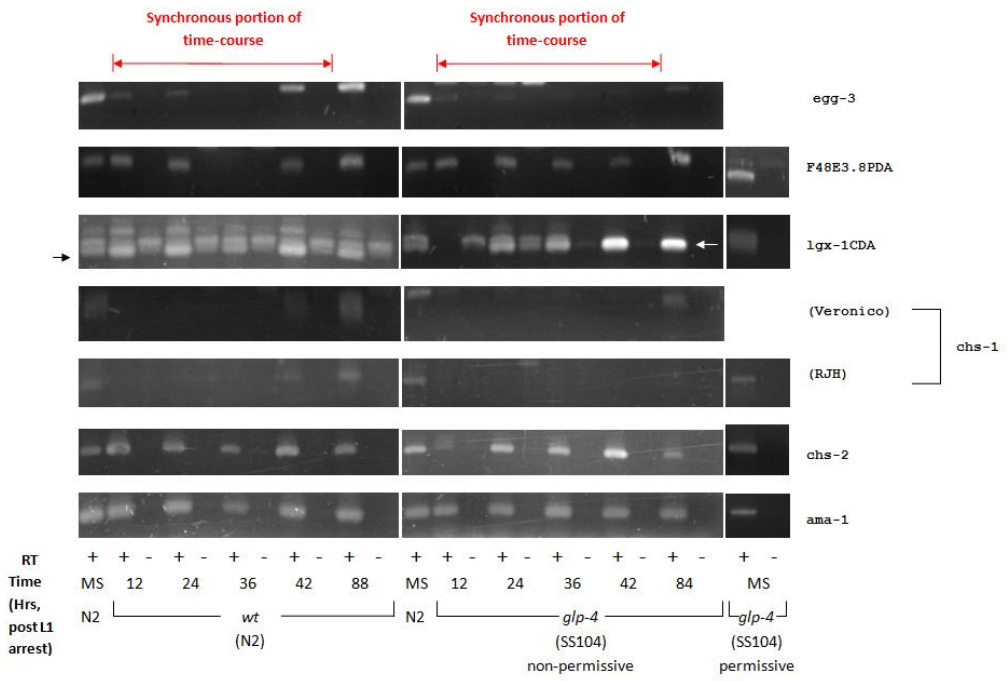
Supplementary Figure 2S.9. F48E3.8PDA primer specificity and analysis of amplimer size. PCR was originally performed using combinations of the forward primers JHF48E3_8PDAF1, JHF48E3_8PDAF2 and JHF48E3_8PDAF3, and the reverse primers JHF48E3_8PDAR1, JHF48E_8PDAR2 and JHF48E3_8PDAR3. **(A)** Schematic depicting annealing sites for these primers relative to the region originally predicted to encode the PDA domain (yellow). **(B)** Amplimers of varying sizes are produced with different primer combinations. JHF48E3_8PDAF3 and JHF48E3_8PDAR1 were selected for subsequent experiments and used with the following PCR conditions: Stage 1: 94°C, 3 minutes; Stage 2: [94°C, 1 minute; 58.0°C, 1 minute; 72°C, 3 minutes] x30; Stage 3:72°C, 10 minutes. The optimized primer sequences and PCR conditions are also listed in **Supplementary Table 2S.2**. These primers generate amplimers predicted to be 638 bp (gDNA) and 536 bp (cDNA), including engineered primer ends that are not complementary to target sequence.



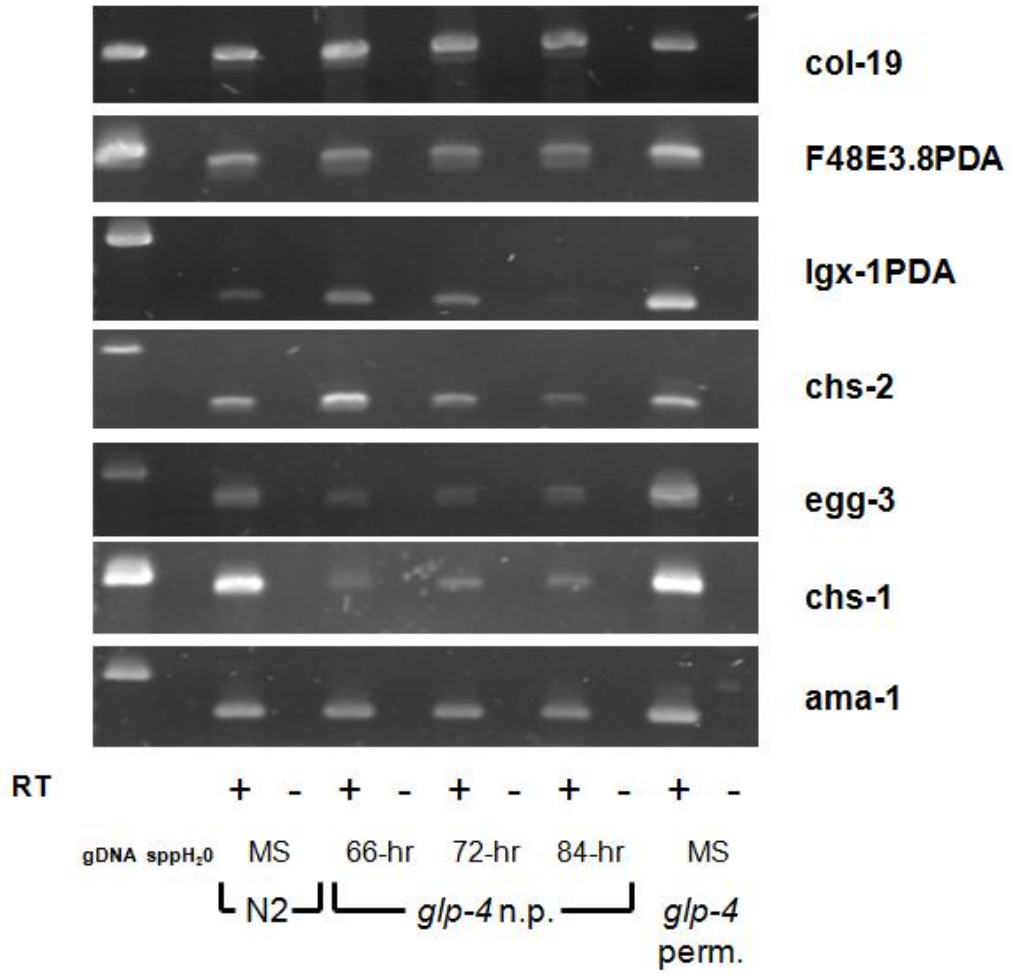
Supplementary Figure 2S.10. *egg-3* primer specificity and analysis of amplimer size. PCR was performed using the JHEGG3F1 and JHEGG3R1 primers. **(A)** Schematic depicting annealing sites for these primers. **(B)** PCR was performed using these primers under the following conditions: Stage 1: 94°C, 3 minutes; Stage 2: [94°C, 1 minute; 63.0°C, 1 minute; 72°C, 3 minutes] x30; Stage 3:72°C, 10 minutes. Amplimers are predicted to be 752 bp (gDNA) and 609 bp (cDNA). These primer sequences and optimized PCR conditions are also listed in **Supplementary Table 2S.2**.



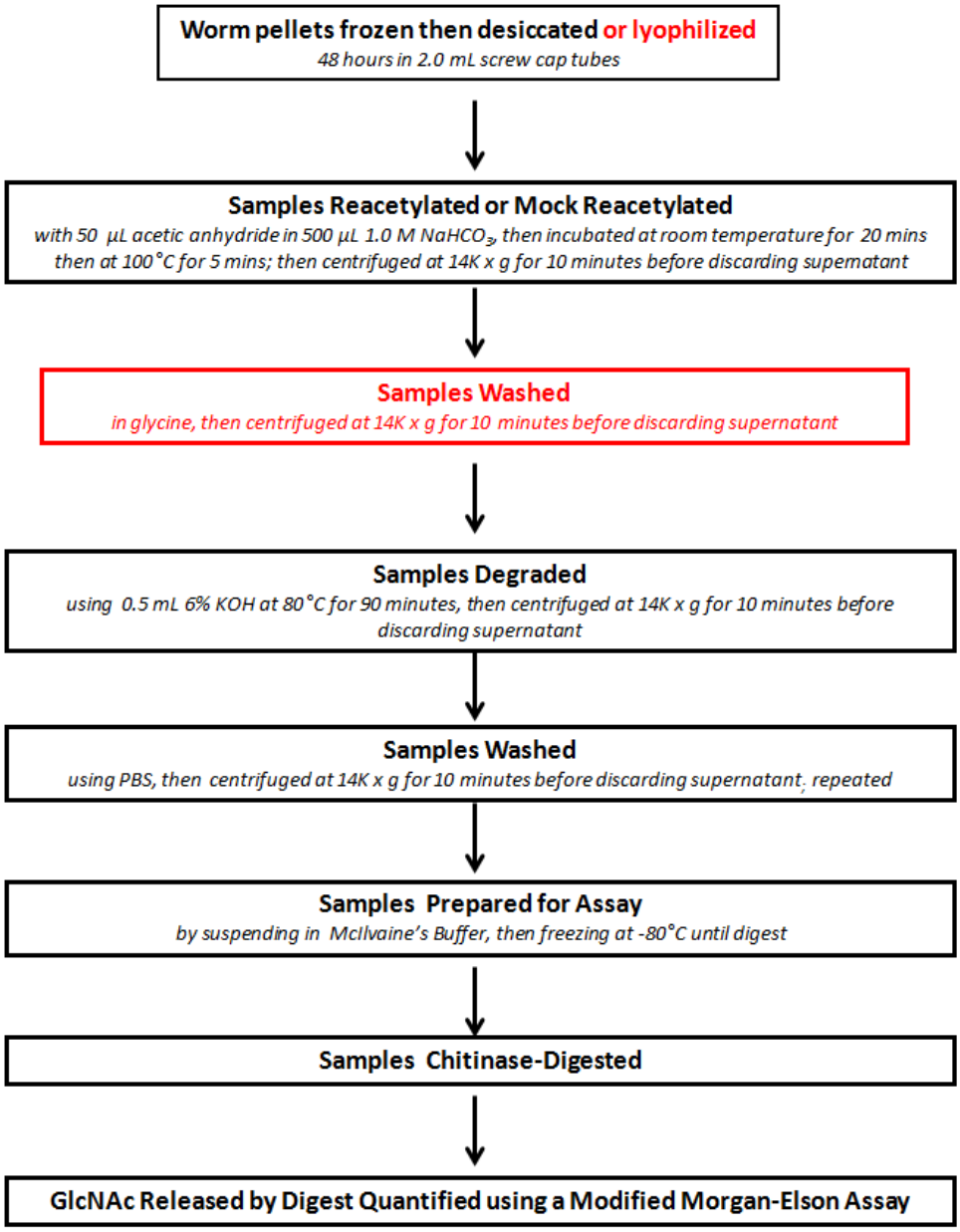
Supplementary Figure 2S.11. ***col-19*** **primer specificity and analysis of amplimer size.** PCR was performed using combinations of the forward primers JHCOL19F1 and JHCOL19F2 and the reverse primers JHCOL19R1 and JHCOL19R2. **(A)** Schematic depicting annealing sites for these primers. **(B)** PCR was performed using these primers under the following conditions: Stage 1: 94°C, 3 minutes; Stage 2: [94°C, 1 minute; 61.1°C, 1 minute; 72°C, 3 minutes] x30; Stage 3:72°C, 10 minutes. Subsequent experiments used the JHCOL19F2 and JHCOL19R2 primers. Amplimers are predicted to be 845 bp (gDNA) and 792 bp (cDNA). These primer sequences and optimized PCR conditions are also listed in **Supplementary Table 2S.2**.



Supplementary Figure 2S.12. Germline ablation of the SS104 strain is less efficient in liquid culture and evident from RNA extracts without DNAse treatment. Monoxenic worm cultures of N2 and *glp-4* were bleach-synchronized and L1 worms allowed to develop axenically for 16 hours, leading to synchronous L1 arrest. Thereafter, worms were grown in the presence of OP50. Samples of culture were taken at the time increments outlined, worms purified by sucrose flotation and stored in the presence of Trizol at -80°C. Thawed worms were homogenized and RNA extracted for use in synthesizing first-strand cDNA using RT (+) with no RT controls (-), using ~2 µg RNA (not DNAse-treated) in a 40 µL reaction diluted to a final volume of 100 µL after synthesis. PCR was conducted for segments of the genes of interest. A segment of the *ama-1* gene was amplified as a positive control. All PCR reactions were conducted using temperatures specific to annealing conditions and 30 cycles of amplication. 10 µL of 50 µL PCR reactions are loaded to each lane of a 1% agarose gel. 2 µg of permissive SS104 RNA was used as a template for cDNA synthesis where the final product was brought to 100 µL, before 5 µL was used for each PCR reaction.



Supplementary Figure 2S.13. Germline ablation of the SS104 strain is less efficient in liquid culture and evident from RNA extracts with DNAse treatment. Using the same SS104 time course presented in Figure 2S.12, additional RNA extracts were prepared for the 66-hr and 72-hr time-points. These RNA extracts were DNAse-treated in parallel with the 84-hr RNA previously utilized and mixed-stage *wild-type* N2 RNA and SS104 germline-permissive RNA. These DNAse-treated samples were quantified and RNA concentrations standardized to equivalent concentrations before the samples were used as templates for cDNA synthesis and to repeat PCR for the genes *ama-1*, *chs-1*, *egg-3*, *chs-2*, C54G7.3/*lgx-1*PDA, F48E3.8PDA and *col-19*. Amplification from gDNA was done for each primer pair to demonstrate the ability to resolve genomic and copy DNA products. These results provide clearer evidence that *chs-1* expression occurs in adult stages of the SS104 strain that was grown under germline non-permissive conditions, suggesting that liquid culture growth may not provide ideal conditions for germline-restriction.



Supplementary Figure 2S.14. Overview of differential reacetylation assay used to quantify chitin and chitosan in *C. elegans*. In each trial, four samples are processed each receiving one of four treatments: (i) reacetylation and chitinase degradation; (ii) reacetylation, no chitinase treatment; (iii) mock reacetylation with chitinase degradation; (iv) mock reacetylation with no chitinase treatment.

SUPPLEMENTARY TABLES

Supplementary Table 2S.1. Additional Non-Nematode Sequences Used in Protein Alignments and Phylogenetic Analysis

PHYLA	SPECIES	NCBI ACCESSION NUMBER	PREVIOUSLY ASSIGNED NAMES	TOTAL # OF RESIDUES IN ARCHIVED SEQUENCE	RESIDUES INCLUDED IN ALIGNMENTS AND TREE RELATIVE TO ARCHIVED SEQUENCE
Bacteria	*Streptococcus pneumoniae*	2C1G_A	SpPGDA	431	235-390
Protista	*Entamoeba histolytica*	XM_651817	EhCDA1	275	53-211
		XM_651264	EhCDA2	262	57-223
	Entamoeba invadens	DQ284499	EiCDA1	281	54-211
		DQ284500	EiCDA2	262	57-223
		DQ284501	EiCDA3	293	84-249
Fungi	*Cryptococcus neoformans*	XP_571516	CDA1	470	158-336
		XP_570561	CDA2	458	157-327
		XP_571200	CDA3	410	123-294
		XP_568540	FPD1	249	38-200
Insecta	*Drosophila melanogaster*	NP_730444	DmCDA1, SERP	541	200-359
		NP_730443	DmCDA2, VERM (Isoform B)	549	207-367
		NP_609806	DmCDA3, DmChLD3	577	243-405
	Tribolium castaneum	ABU25223	TcCDA1	534	192-352
		ABU25224	TcCDA2 (Isoform A)	535	191-351
		ABW74145	TcCDA3	505	180-336
		ABW74146	TcCDA4	490	142-297
		ABW74147	TcCDA5 (Isoform A)	1131	800-954
		ABW74149	TcCDA6	403	52-209
		ABW74150	TcCDA7	374	54-211
		ABW74151	TcCDA8	376	54-211
		ABW74152	TcCDA9	381	54-212

Supplementary Table 2S.2. Primers and PCR Conditions Used for Amplification

TARGET	PRIMERS (UNDERLINED RESIDUES ANNEAL WITH TARGET SEQUENCES)	PCR CONDITIONS
F48E3.8PDA	CCTTCAAGACTTCTGACCGAATGCCCG (F) GTTTGTGGCCAATATGGGCTGCTGACC (R)	94°C, 3 minutes; [94°C, 1 minute; 58.0°C, 1 minute; 72°C, 3 minutes]x30; 72°C, 10 minutes
C54G7.3PDA	AAGACGAGCTCCGGGTGTCTCCGC (F) CGCTAGAATTCTCCTCTTCAGGAGCTCCACCGAAC (R)	94°C, 3 minutes; [94°C, 1 minute; 61.4°C, 1 minute; 72°C, 3 minutes]x2; [94°C, 1 minute; 64.7°C, 1 minute; 72°C, 3 minutes]x28; 72°C, 10 minutes
ama-1 (F36A4.7)	TTCCAAGCGCCGCTGCGCATTGTCTC (F) CAGAATTTCCAGCACTCGAGGAGCGGA (R)	94°C, 3 minutes; [94°C, 1 minute; 60.0°C, 1 minute; 72°C, 3 minutes]x30; 72°C, 10 minutes
chs-1 (T25G3.2)	TCTTGGATCCTCTCGCCAGATGCCTTGTTCTCAG (F) ACAAGAATTCGCCAAGAAGCGTGAAATCCCTCGC (R)	94°C, 3 minutes; [94°C, 1 minute; 59.9°C, 1 minute; 72°C, 3 minutes]x2; [94°C, 1 minute; 63.7°C, 1 minute; 72°C, 3 minutes]x28; 72°C, 10 minutes
chs-2 (F48A11.1)	ACAAGGATTCACCATGTTCCGAAACGTGAAACGC (F) ACAAGAATTCAAGCAGCAGACAACGTAGATGCAG (R)	94°C, 3 minutes; [94°C, 1 minute; 58.5°C, 1 minute; 72°C, 3 minutes]x2; [94°C, 1 minute; 62.0°C, 1 minute; 72°C, 3 minutes]x28; 72°C, 10 minutes
col-19 (ZK1193.1)	TTGGATCCTGCGGAGTACTTGTGTGCG (F) CGGCATCCTCTCCCTTTTGTCCTGG (R)	94°C, 3 minutes; [94°C, 1 minute; 61.1°C, 1 minute; 72°C, 3 minutes] x30; 72°C, 10 minutes

Chapter 3

Molecular and Biochemical Characterization of *C. elegans* Chitin Deacetylases

ABSTRACT

We have previously identified F48E3.8 and C54G7.3 as two nematode polysaccharide deacetylase-encoding genes that are important for *C. elegans* development. Here I provide a comprehensive overview of our knowledge of the exon-intron boundaries of these genes, gleaned from previously existing data and new RT-PCR experiments. The genetic structure of F48E3.8 and C54G7.3 suggest they may utilize an internal promoter to drive truncated "late" transcripts and I provide sequence analysis to support this hypothesis. Efforts to further characterize transcript complexity using Northern Blots and an attempt to identify a novel transcript using a semi-nested RT-PCR approach are also reported along with results for an RT-PCR-based analysis of gene expression patterns. Attributes of the protein isoforms that would be produced from these two genes are investigated using bioinformatics.

INTRODUCTION

We have recently identified F48E3.8 and C54G7.3/*lgx-1* as two *C. elegans* genes encoding polysaccharide deacetylase enzymes (Heustis *et al.*, 2012). Previous work demonstrated that these two genes were expressed in the *C. elegans* embryonic pharynx (Gaudet and Mango, 2002) and referenced *in situ* hybridization data, available from the Nematode Expression Database (NextDB) as data supporting pharyngeal expression. Our work has shown that these two genes have one or two homologs in other nematodes, has confirmed the somatic expression of these genes, and has shown that the genes are important for *C.*

elegans development. Our work has also shown that proteins from *C. elegans* can deacetylate chitin *in vitro*, a process that leads to the formation of chitosan (Heustis *et al.*, 2012). This product has been detected in protists and fungi using a differential reacetylation assay (Baker *et al.*, 2007; Banks *et al.*, 2005; Das *et al.*, 2006); however, we have been unable to detect chitosan from *C. elegans* using this method but note that Eosin Y, a dye described as specific for chitosan in the fungal cell wall (Banks *et al.*, 2005), stains the *C. elegans* pharynx (Heustis *et al.*, 2012). In *C. elegans*, it is likely that the F48E3.8- and C54G7.3-encoded enzymes target pharyngeal chitin converting it to chitosan. This biochemical change has important consequences for development and these enzymes may serve as virulence factors in parasitic roundworms such as the human filarial parasite *Brugia malayi* and the plant parasitic nematode *Meloidogyne incognita*. If these enzymes serve essential steps in development and in maintaining virulence, the family of nematode polysaccharide deacetylases may be viable targets for the development of selective interventions.

These two genes have been the subject of limited previous study and despite the availability of some data on gene expression and protein activity – generated in large scale screens and available in published and unpublished data sets – a comprehensive analysis of the molecular properties of these genes and the biochemical properties of the enzymes which they encode has not been completed. The ability to target these enzymes will depend on increased knowledge about attributes of their regulation and activity *in vivo*.

In this study, I present a comprehensive analysis of information available for F48E3.8 and C54G7.3 in *C. elegans*, which has not previously been compiled. Additionally, I provide new information regarding transcript complexity for these genes and show that PDA-encoding transcripts of both genes are expressed throughout development. The genetic structure of C54G7.3, which is annotated to include a large ~ 4 kb intron (Wormbase WS227), suggests that it may utilize an internal promoter to drive downstream transcripts and additional evidence supporting this possibility is presented. Using bioinformatics tools, I show that isoforms from both genes are likely to be secreted, using classical and non-classical secretion mechanisms. Analysis of the peptide sequences spanning the predicted polysaccharide deacetylase (PDA) domain suggest that nematode PDAs may function in the absence of metal co-factors.

MATERIALS AND METHODS

Strains

The wild-type *C. elegans* Bristol N2 strain and the SS104 strain (carrying the temperature-sensitive *glp-4* allele, bn2) were used in these studies.

Worm Sampling

Worm growth and sampling was performed as previously described (Chapter 2).

Bioinformatics Analysis

ESTs for both the F48E3.8 and C54G7.3 genes were generated during various *C. elegans* transcriptome analysis projects and are available as alignments against the respective genes in Wormbase. These data are useful in developing a framework for previously confirmed exon-intron boundaries in the two genes. These data are available for F48E3.8 at http://www.wormbase.org/tools/genome/gbrowse/c_elegans/?name=X:7505864..7518237 and for C54G7.3 at http://www.wormbase.org/tools/genome/gbrowse/c_elegans/?name=X:5522500..5539280

[N.B. If ESTs are not visible when viewing this page, they may be retrieved by selecting the SELECT TRACKS button at the bottom of the page, locating the sub-label "Sequence Similarity Tracks", and selecting the two options "ESTs aligned by BLAT (best)" and "ESTs aligned by BLAT (other)", before returning to the viewing page.]

Analysis of sequences to determine the likelihood of including functional promoters was performed using the Promoter 2.0 Prediction Server (http://www/cbs.dtu.dk/services/Promoter). Protein sequences were analyzed for the presence of classical secretion signals using the SignalP 4.0 software (http://www.cbs.dtu.dk/services/SignalP/) or for the presence of non-classical secretion signals using the SecretomeP 2.0 software (http://www.cbs.dtu.dk/services/SecretomeP).

RT-PCR

RT-PCR was employed in amplifying segments of transcripts for sequencing, in attempts to identify novel transcripts, and in time courses to study the temporal expression patterns of genes in *C. elegans*.

RNA extracts were DNAse-treated (in some trials) using Turbo DNAse (Ambion). First strand cDNA was generated using the Protoscript First Strand cDNA Synthesis Kit (NEB) (**Figure 3.12A**) or the NEB Protoscript M-MuLV Synthesis Kit (NEB) (**Figure 3.12B**)

In attempts to identify a new SL1-led transcript produced under the control of an internal C54G7.3 promoter, a semi-nested PCR approach was employed. The procedure utilized two stages with an intermediate DNA isolation step: Stage 1 [2 cycles at 47.6°C and 28 cycles at 60.9°C for extension steps using the primers JAFSL2 and JHLGX1R1]; DNA isolation; Stage 2 [2 cycles at 47.6°C and 28 cycles at 60.9°C]. Control reactions used the JHLGX1NEWTRANSL1CONT primer in place of the SL2 primer: with the SL2 primer, the region amplified would include an SL1 leader sequence while with the JHLGX1NEWTRANSL1CONT primer the region amplified would start from exon sequence immediately adjacent to the putative SL1 sequence in a mature transcript (**Figure 3.8A**).

Additional information related to all primer sequences used these studies is available in **Appendix A**.

Northern Blots

A BamHI-linearized construct containing cDNA from the C54G7.3 gene (depicted in **Figure 3.9**) was used to generate ssRNA probe using the DIG RNA Labeling Kit (SP6/T7) (Roche), according to manufacturer's instructions. Wild-type N2 mixed-stage RNA was isolated using Trizol reagent and DNAse-treated with Turbo DNAse (Ambion). Denaturing agarose gels were prepared using the The Northern Max Kit (Ambion/Applied Biosciences) and used to separate ~ 2.5 µg or ~15 µg *C. elegans* RNA per well (in separate trials), alongside lanes with control *Neo* RNA (supplied with kit) and Millennium Markers (Ambion). Northern Blots were made onto BrightStar-Plus Positively Charged Nylon Membrane (Ambion), according to manufacturer's instructions, with cross-linking done using the UV Stratalinker 1800 (Stratagene). Pre-hybridization and hybridization was done using components of The NorthernMax Kit and probe was detected using the DIG Nucleic Acid Detection Kit (Roche) along with the DIG Wash and Block Buffer Set (Roche). Probe for the control *Neo* RNA was synthesized in parallel with C54G7.3PDA probe, using the labeling kit, as a control for probe synthesis.

RESULTS

F48E3.8 and C54G7.3 exon-intron boundary annotations are incomplete and bear inaccuracies

The F48E3.8 gene has been assigned 38 exons and the C54G7.3 gene has been assigned 33 exons based on bioinformatic analysis of the ORFs of these two genes (turquoise boxes in **Figure 3.1**). In both these genes, the exons are predicted to be used differentially in the generation of alternative transcripts. Three transcripts have been assigned for F48E3.8 (connected grey boxes in **Figure 3.1A**) and two transcripts have been assigned for C54G7.3 (connected grey boxes in **Figure 3.1B**) based on information available on Wormbase WS227.

The PFam bioinformatics tool was used to analyze the predicted peptide sequences generated from the three assigned transcripts of the F48E3.8 gene and the two assigned transcripts of the C54G7.3 gene. The current version of the PFam protein sequence analysis tool (last accessed 04/2012), predicts the presence of a PDA domain in the C54G7.3 gene but not in the F48E3.8 gene. The current version of the program also does not recognize chitin-binding domains (CBM_14) in either the F48E3.8 or C54G7.3 gene. However, using earlier versions of PFam, a PDA domain and two or three CBM_14 domains were detected in F48E3.8 (accessed 09/21/2008) and two CBM_14 domains were also detected in C54G7.3 (accessed 07/17/2006 and 01/26/2007). The PDA domains are depicted relative to the largest predicted isoform of the two genes in **Figure 3.2**.

While three transcripts (a,b,c) have been assigned for F48E3.8, only some of the exon-intron boundaries across this gene have been confirmed from ESTs (Wormbase WS227). The F48E3.8 exons with confirmed expression are depicted in **Figure 3.1A** (black boxes below the respective exons). The only fully

sequenced transcript for this gene is F48E3.8c, which comprises the terminal five exons of the gene and is predicted to encode the PDA domain. Two transcripts (a,b) have been assigned for C54G7.3 (grey boxes in **Figure 3.1B**) and again only some of the exon-intron boundaries across this gene have been confirmed via ESTs (Wormbase WS227). The C54G7.3 exons which are known to be expressed given their presence in ESTs are depicted as black boxes below the respective exons in **Figure 3.1B**. While the sequence of the PDA-encoding domain in F48E3.8 is confirmed, sequence data is lacking for the PDA-encoding domain of its homolog C54G7.3. I have sequenced the region spanning predicted exons 27 – 30 and shown that predicted exon 28 is not expressed in a transcript bearing the other three exons. (This region corresponds to that observed in **Figure 2S.8**). The amplimer used to confirm this result is denoted by connected blue boxes below the gene map in **Figure 3.1B** (Heustis *et al.*, 2012).

To gain further information about other regions of the C54G7.3 gene, I sequenced the insert present in the yk1621b03 clone (courtesy Dr. Yuji Kohara) and, in addition to 5' and 3' EST sequence for this clone which show the expression of some exons, we have found that predicted exon 16 is expressed while only a portion of the predicted exon 15 is expressed in this clone. This is presented as connected green boxes in **Figure 3.1B**. I also noted that sequence from the terminal region of the large ~4 kb intron of C54G7.3 is present in this clone and likely represents part of a mature transcript. This information was not previously captured in Wormbase WS227 annotations for the gene, although the EST including this sequence is included with the alignment there. I have depicted

the cloned sequence, showing expression of part of the annotated large intron (light green) and the connected downstream exons (dark green), in **Figure 3.1B**.

To confirm expression of the 5' end of the C54G7.3 gene, I used RT-PCR to test for the expression of a region of the gene spanning exons 2 - 4. This region includes the sequence that was first noted as encoding the highest-scoring predicted chitin-binding domain (CBM_14) identified with PFam (when accessed 07/17/2006 and 01/26/2007). Amplimers consistent with the sizes predicted for gDNA and cDNA were obtained (**Figure 3.3**).

Most introns found in *C. elegans* genes are only about 50 bp in length. The ~4kb intron present in C54G7.3 (between predicted exons 11 and 12) gives rise to a relatively unique genetic structure that has been evaluated in a number of other genes, but not C54G7.3 (Choi and Newman, 2006). In order to confirm that the two genetic regions separated by this annotated large intron were part of a single ORF, we used RT-PCR to determine whether an amplimer could be generated across the intron. We have previously noted that the PCR was successful in generating a single amplimer with cDNA sequence derived from the two sides of this large intron (Heustis *et al.*, 2012) and the results for the experiments using either Elongase or Taq Polymerase are presented in **Figure 3.4**. Elongase PCR generates discernible products from gDNA or cDNA using various concentrations of Mg^{2+} (**Figure 3.4B**), and the results suggest that only a single amplimer is produced from either template. This result is consistent with original expectations since the two transcripts (a and b) of C54G7.3 are currently annotated as identical within this region of their sequences. When Taq

polymerase is used, however, three bands are detected from the cDNA template (**Figure 3.4C**). This result may suggest an increased level of complexity with respect to C54G7.3 transcript numbers, if it suggests that at least three different transcripts are discernible within this region of the gene. However, it is noteworthy that no amplimer is detected from gDNA when the Taq polymerase is used in contrast to results using Elongase (which is better suited for amplication of longer sequences). This may suggest that the PCR conditions are not ideal for use with Taq, in which case additional bands are a product of an inefficient PCR reaction.

The large introns in F48E3.8 and C54G7.3 contain PHA-4 binding sites but not promoters

Comparison of the F48E3.8 genomic sequence to one of its transcripts (transcript b) or comparison of the C54G7.3 genomic sequence to that of the largest currently assigned transcript (transcript a), shows that both genes have a large ~4kb internal intron (Wormbase WS227). Introns of this size are extremely rare in the *C. elegans* genome.

Choi and Newman (2006) were the first to note that ~ 4 kb introns gave rise to a relatively uncommon genetic structure that had functional significance for transcript processing. They studied numerous transcripts that had been previously identified and sequenced, revealing the presence of exon sequence corresponding to the terminal end of their respective genes and, importantly, bearing an SL1 leader sequence at the 5' end. Their meta-analysis revealed that

these transcripts followed large introns that were ~ 4kb in length, in which internal promoters were predicted to regulate the production of downstream "late" transcripts.

Neither F48E3.8 nor C54G7.3 was included in the Choi and Newman (2006) analysis, since no SL1-bearing transcripts for either gene have been previously isolated and characterized (Wormbase WS227). The presence of large introns in these genes suggested to us that "late" transcripts may be produced for F48E3.8 or C54G7.3.

Based on existing Wormbase (WS227) information, a short downstream transcript (transcript c) has already been assigned for F48E3.8 but no information available from the gene annotations suggests that an SL1 leader sequence is found at the start of the F48E3.8c transcript. As such, this transcript does not fit the class of transcripts analyzed by Choi and Newman (2006) and, based on the criteria they lay out, then, may represent a splice variant driven from the canonical upstream promoter. While a short transcript has been identified for F48E3.8, no short transcript has been assigned for C54G7.3. In order to determine whether an internal promoter may be utilized to drive late "truncated" transcripts of C54G7.3, I analyzed the sequence in and around the large introns in search of key molecular markers. Various attributes suggest that a functional internal promoter may be relevant to gene regulation in C54G7.3. For C54G7.3, this hypothetical two-promoter genetic structure can be simplistically depicted as shown in **Figure 3.5A**.

In *C. elegans*, some genes are co-transcribed in operons to produce an initial polycistronic messages that is then cleaved and processed to distinct transcripts. Processing involves the addition an SL1 leader sequence to the 5' end of the first message in the transcript while all other downstream transcripts, representing additional genes in the operon, are processed to include a spliced leader 2 (SL2) sequence at their 5' ends (Blumenthal, 2005; Blumenthal *et al.*, 2002). In genes with a two promoter system, transcripts directed from the canonical promoter as well as the internal promoter are processed to include the SL1 sequence (Choi and Newman, 2006). I reasoned that the same molecular markers present to direct SL1 splicing for the initial transcript of an operon would be present in the internal promoters used to direct "late" transcripts. By comparison to SL1 processing sites in operons, the exon-intron boundary between intron 11 (the long intron) and exon 12 may be a SL1 acceptor site (**Figure 3.5B**). In transcribed pre-mRNA, the terminal long intron would end in a UUUUCAG (equivalent to the TTTTCAG in genomic sequence), which is the typical acceptor sequence for SL1 addition. Also, the docking site for the SL1 snRNP, which carries the SL1 transcript, has been deduced to sit at UC-rich sequence found at -50 bases relative to this acceptor site (Graber *et al.*, 2007). The UUUUU and UGUUC sequences present from -64 to -49 bases (equivalent to the genomic TTTTT and TGTTC) can serve as these sequences. Based on my analysis of the protein sequence resulting from a transcript that would start from predicted exon 12 and continue to the end of the gene, an open-reading frame (ORF) encoding a protein with a PDA-domain is present since an in-frame methionine is found at

the start of exon 13 (**Figure 3.5B**). [It is noteworthy that this creates a slightly longer distance between the SL1 acceptor site and translation start site than that described by Graber *et al.* (2007), however.] Using the Promoter 2.0 Prediction Server software, I have analyzed the ~ 2kb region upstream of the C54G7.3 ORF (which would include the presumptive canonical promoter of the gene) and, separately, the entire ORF of C54G7.3, including the large ~4 kb intron that could include an internal promoter. This tested for the likely presence of a promoter across the gene. For the canonical promoter, this is the ~2kb sequence immediately upstream of the presumptive translation initiation codon (ATG in the genomic sequence). While the software accurately predicts the presence of a functional promoter in the region upstream of the gene (i.e. the canonical promoter), it does not report a promoter within the ~4b region currently assigned as intron in the gene (relative to transcripts a and b). Therefore, while various molecular markers for SL1 addition suggest that the large ~4 kb intron of C54G7.3 could direct the production of an SL1-led truncated "late" transcript similar to the class analyzed by Choi and Newman (2006), bioinformatics tools suggest that no promoter is present in this intron.

The PHA-4 transcription factor drives expression of genes in the *C. elegans* embryonic pharynx. This is of particular interest because the chitinous lining of the pharynx could be a target of these polysaccharide deacetylases. Using microarray technology, a comprehensive analysis of genes upregulated during *C. elegans* embryonic pharyngeal morphogenesis has been conducted and the promoters of these genes analyzed for the presence of TRTTKRY elements,

binding sites for the PHA-4 transcription factor. F48E3.8 and C54G7.3 were both identified in the microarray analysis with 6 and 5 PHA-4 binding sites identified in their canonical promoters through bioinformatics (Gaudet and Mango, 2002). Comprehensive analysis of these binding sites has focused on canonical 5' promoters and not on the types of internal promoters I am currently positing to work in C54G7.3. I tested whether an internal promoter for C54G7.3 could be responsive to the PHA-4 transcription factor by searching for PHA-4 binding sites within the large intron. I have identified multiple putative PHA-4 binding sites within the terminal 1 kb region of the large intron of C54G7.3 (**Figure 3.6**) in addition to an extra PHA-4 binding site in the canonical promoter of the gene (**Figure 3.7**). Similarly, the terminal 1 kb stretch of the large F48E3.8 intron also includes multiple PHA-4 recognition sequences (**Figure 3.6**). About 30,000 TRTTKRY sequences are found in the *C. elegans* genome and by ChIP analysis some 2700 – 9000 PHA-4 binding sites are detected (Susan Mango, personal communication). In the putative internal promoters of F48E3.8 and C54G7.3, I have found six (6) and nine (9) binding sites respectively, suggesting that PHA-4 involvement may be important there (although in C54G7.3 many of these sites are overlapping and may not all be accessible for simultaneous binding), since it exceeds the 1.67 sites that would be found randomly if 30,000 putative binding sites were randomly distributed around the ~18,000 genes predicted from the *C. elegans* genome.

Semi-nested PCR and Northern Blot approaches to identifying novel C54G7.3 transcripts

I attempted to identify a novel SL1-spliced "late" transcript of C54G7.3 – the product of regulation by an internal promoter that would sit within the large intron of the gene – using a semi-nested PCR approach and by analyzing transcript sizes using Northern Blot. In the semi-nested PCR approach, a single SL1 sequence was used as a forward primer while two reverse primers were used in two separate 30 cycle PCR reactions – R1 then R2 – to enrich an SL1-bearing product that would contain the PDA-encoding domain of the C54G7.3 gene. The exon sequence presumed to be immediately downstream of the SL1 sequence in the processed transcript in question was targeted using a complementary primer that could serve as a control for the PCR reaction.

From the control reaction, a doublet of two amplimers, close to the size predicted for cDNA template, was obtained (**Figure 3.8**). This confirmed the validity of the PCR protocol, but the presence of two bands (rather than one) suggests that transcript complexity is indeed greater than that captured by the current annotations for the gene (**Figure 3.1**). Similarly-sized amplimers were not obtained when the SL1 sequence was used to amplify; rather a smear of product was observed. Since the SL1 sequence is found in as many as 50% of transcripts, it is possible that the cDNA representing various transcripts are being "primed" so that single complementary strands are produced. The most abundant messages can thus be observed, even in the absence of the binding of a reverse primer, with amplification proceeding in a linear rather than exponential manner.

When a single 30-cycle PCR was permitted for SL1 and the R1 primer or SL1 and the R2 primer, the same smear was observed suggesting that this not a property of either reverse primer and supporting the idea of SL1 binding to multiple targets (data not shown). If any message in the smear of product represents an SL1-led "late" transcript from C54G7.3, it suggests that this sequence varies significantly in the exons maintained in the product of the control reaction since no product from the experimental lane with cDNA as template reaches the size of the doublet obtained from the control reaction using cDNA template. The precision of this method, however, was drawn into question by the presence of product when gDNA was used as a template with either forward primer. Importantly, four distinct bands of product are observed when either the SL1 primer or the control primer are used (**Figure 3.8**). This outcome cannot be readily explained and suggests that further optimization will be necessary to ensure amplification specificity. For the SL1 primer on genomic DNA, this could reflect SL1 priming from two genetic loci, since multiple SL1 genes are found in the genome.

Only two predicted transcripts have been assigned to the C54G7.3 gene – and only transcript a is predicted to encode a PDA domain. If a "late" truncated transcript of the C54G7.3 gene is being produced, its functional relevance would be most apparent if it too encodes a PDA domain. I reasoned that a Northern Blot aimed at determining how many transcripts with a PDA-encoding domain derived from C54G7.3 were present in mixed-stage cDNA from *wild-type C. elegans* worms, could reveal the presence of a novel SL1-led PDA-encoding transcript of

the gene. In particular, the Northern Blot would reveal at least one more transcript that was significantly shorter than that produced from transcript a.

I used ssRNA probe, generated from the cloned PDA-encoding domain of C54G7.3 (**Figure 3.9**), to determine whether additional PDA-encoding transcripts of C54G7.3 could be identified. In my Northern Blot analysis, a standard control *Neo* RNA (that was supplied by manufacturer) and its probe (generated in my hands) were used to verify that all steps of the procedure worked (**Figure 3.10A and B**). Bands of the appropriate size for the *Neo* target RNA could be detected for the appropriate lanes on the blot when the blots were left in incubation buffer for as little as 2 hours (**Figure 3.10B**) while excess staining from the DIG-labeled probe was seen in when the blots were left in detection buffer for 16 hours (**Figure 3.10A**). In contrast to the abundant staining for the *Neo* target, no discernible transcripts were observed for C54G7.3PDA domain-encoding transcripts from mixed-stage *wild-type* N2 RNA (**Figure 3.10C-E**). When performed under conditions similar to the control blot shown in **Figure 3.10A**, no C54G7.3PDA transcripts were observed from ~2.5 μg of total RNA when detection was allowed to proceed for 16 hours (**Figure 3.10C**). In attempts to increase the likelihood of detecting these transcripts of interest, I tried to increase the total RNA loaded to corresponding wells of the gels by six-fold, increasing the target RNA to ~15 μg, while increasing detection time to 72 hours. This also failed to result in detection of any transcripts of C54G7.3 (**Figure 3.11D**). In an additional attempt to optimize this method, I used blots with lanes for ~15 μg RNA while using 6X probe and using a 24-hour incubation in detection buffer.

Again, no C54G7.3 transcript was detected in the appropriate lane and the increase in probe had the adverse effect of causing annealing to bands of the RNA ladder (**Figure 3.11E**). Taken together, this suggests that any PDA-encoding transcripts of the gene are present in very low copy number. This method, then, also fails to provide evidence of a novel transcript of the C54G7.3 gene and may benefit from additional attempts to optimize probe hybridization conditions.

Temporal analysis of gene expression of F48E3.8 and C54G7.3

Somatic expression of the genes F48E3.8 and C54G7.3 was confirmed using a gene expression profile based on synchronized SS104 worms, which carry the *glp-4* mutant allele (bn2) and restrict germline production at 25°C (Heustis *et al.*, 2012). In this time course, expression of F48E3.8, C54G7.3 and *chs-2* first become apparent at the 24-hr time point; however, the C54G7.3 and *chs-2* expression remain more similar as they both decline at the onset of adulthood while F48E3.8 expression persists (**Figure 2.3**). In order to further characterize similarities and differences in expression among the three genes, I repeated the experiment for an increasing number of time points.

Consistency in the presence and prevalence of *ama-1* was a limiting factor in the previously mentioned SS104 trial with worms grown on solid media; although *ama-1* is considered to have consistent levels of expression warranting its use in previous time courses (Johnstone and Barry, 1996), we note that the levels of the gene vary among samples normalized for total RNA concentration and is barely visible at some time points (**Figure 2.3**). In subsequent trials of the

experiment designed to provide greater details about changes in expression of *chs-2*, F48E3.8 and C54G7.3, we attempted to increase RNA yield – to increase likelihood of detecting *ama-1* providing a baseline for normalization – through growth and harvesting from liquid culture (**Figure 3.11** and **Figure 3.12**). In liquid culture, populations can become denser.

Using a synchronous culture of *wild-type* N2 worms, grown in liquid culture, I was able to confirm the similarity in expression between *chs-2* and C54G7.3, as the expression of both genes are high and of similar intensity in early larval stages of the worms, and gradually decline with the onset of adulthood between 36-hr and 42-hr when activity in the germline becomes evident through increasing detection of *chs-1* (**Figure 3.11**). F48E3.8 was not studied in this trial. The *chs-2* and C54G7.3 transcripts were detected after the onset of adulthood, which contrasts with the conclusions drawn from **Figure 2.3**. These represent message from the F1 generation that is produced from wild-type adults. As such, after the 42-hr time point, the culture is also no longer synchronous.

In this trial (as well as two other trials in which the *wild-type* N2 worms or germline-ablated SS104 worms were used, **Figure 2S.12** and **Figure 2S.13**), observations of the worms taken at each time point revealed less synchrony than similar time points taken from populations grown on solid media. Additionally, although some F1 worms are produced in SS104 worms grown at 25°C on solid media (evidenced by trace levels of *chs-1* in **Figure 2.3**), we observe a greater proportion of F1 worms for cultures grown in liquid media suggesting a decreased efficiency in restricting the germline for these worms. Despite the higher levels

of RNA isolated and the increased ability to detect *ama-1* at all time points when using liquid culture, subsequent trials of the experiments were again conducted on solid media in order to enhance synchrony and, where appropriate, to restrict the germline.

In order to obtain even more detailed information regarding the expression of these genes, I repeated the time course using *wild-type* N2 worms (grown on solid media at 25°C) to investigate expression at 2-hr increments over a 40-hr period. These time intervals have been previously used to detect fluctuations in expression for collagens (Johnstone and Barry, 1996) and chitin synthases (Veronico *et al.*, 2001). In two trials of the experiment under these conditions, I isolated fewer worms at each time point (relative to my previous trials) since an increased number of samples needed to be obtained over the time course and the population of the worms used to isolate samples could only be increased to a limited extent. This had the adverse effect of causing a return to the variable levels of *ama-1* (even from samples where RNA concentration was standardized) since the total RNA was reduced and RT-PCR became more sensitive to variations in transcript abundance at lower template levels. This limited the conclusions that could be drawn regarding the levels of expression of the genes being studied (**Figure 3.12**). Importantly, peaks of *ama-1* expression are always observed at or near the presumed molting times in the experiment. Two possibilities may explain this result. First, in general, it may be easier for RNA to be extracted from worms at the molting stage hence elevated levels of *ama-1* reflect ease of isolating the total pool of message. Or, alternatively, *ama-1* levels

may be higher at molts. Since *ama-1* encodes the large subunit of RNA polymerase II (Wormbase WS227), the enzyme used to synthesize other transcripts, it is possible that *ama-1* expression may be elevated at molts to facilitate increased transcription load. More worms were isolated at each time point for results shown in **Figure 3.12B** than for **Figure 3.12A**, explaining the ability to resolve *ama-1* at more time points in the **Figure 3.12B**.

Despite limits in quantitative expression analysis afforded by these experiments, some conclusions regarding the expression of the genes C54G7.3 and F48E.8 genes can be drawn. In both the trials presented here, I have confirmed the similarity in expression between *chs-2* and C54G7.3 which show some distinctions in comparison to F48E3.8 expression over time. The peaks in F48E3.8 expression are distinct from those of *chs-2* and C54G7.3 (24-hr and 34-hr/36-hr time points in **Figure 3.12B**). The expression patterns of these three genes are distinct from the patterns observed for two genes expressed in the germline, *chs-1* and *egg-3*. In these experiments, peaks in the expression of *col-12*, which has been shown to be elevated at molts, are used to confirm molting points. *col-12* and *dpy-13* have non-overlapping peaks in their cyclical patterns of expression (Johnstone and Barry, 1996) and this distinction is captured in these results (**Figure 3.12B**). *col-2* is a cuticular collagen that is highly upregulated in dauer larvae (Cox and Hirsh, 1985). I was unable to detect *col-2* here, suggesting that dauer larvae were not present and that the time course reflects growth in the absence of harsh or stressed conditions.

Although firm conclusions about variations in the gene expression levels cannot be drawn, the results provide some limited information about F48E3.8 and C54G7.3 regulation. In my analyses, F48E3.8 and C54G7.3 are first detected coincident with the L1/L2 molt in these worms. The same is observed here for *chs-2* expression and confirms previous results showing that this somatic chitin synthase is first detected at this molt (Veronico *et al.*, 2001).

However, as a whole, I was unable to re-create the punctuated expression pattern described for *chs-2* (Veronico *et al.*, 2001) and instead observe waves of expression similar to the patterns described for collagens (Johnstone and Barry, 1996). In the latter, the authors note that even for genes expressed in brief, tightly regulated windows of *C. elegans* development, a wave of expression around some peak should be expected since it reflects the normal distribution of worms around a representative (peak) stage. It is possible that changes in expression may reflect dynamic changes in the level of gene activity *in vivo*. However, it is just as likely that the patterns reflect a lack of full synchrony in the culture. Importantly, this assay provides qualitative description of gene expression but not a quantitative measure of gene expression since it does not directly assess subtle variations in template concentration. Similarly, the expression levels for two different genes cannot be compared (even at the same time-point) because the differences in expression intensity may reflect variations in PCR efficiency rather than reflect *bona fide* differences in gene activity.

Before use in these experiments, the primer specificity for each gene was confirmed through the ability to create amplimers of distinguishable size from

gDNA and cDNA, for *col-12* (**Figure 3.13**), *dpy-13* (**Figure 3.14**) and *col-2* (**Figure 3.15**). Although actin is often used a positive control for gene expression in other organisms, it does not appear possible to use actin in this way in *C. elegans*. Partial redundancy and partial homology make it difficult to discern the specific contributions of each of the five *C. elegans* actin genes in experiments of this type (**Figure 3.16**). An initial goal was to use this time course to study the relative expression patterns for the two C54G7.3 transcripts. In these results presented here, the region of C54G7.3 that is amplified spans the PDA domain and reflects expression of transcript (a), if there really are no additional transcripts of the gene. I was also interested in asking whether this time course could reveal differences in transcript expression levels between transcript a and transcript b of C54G7.3. It should be possible to take advantage of the 3' UTR of the C54G7.3 transcript b sequence to conduct transcript-specific PCR (**Figure 3.17**); however, due to limits on the total amount of template available, no attempt was made to apply these primers to the time courses presented here and the amplification specificity of the primers were not tested.

Specific isoforms of the F48E3.8 and C54G7.3 genes are likely to be secreted

I have used SignalP 4.0 to search for classical secretion signals and SecretomeP 2.0 to search for non-classical secretion signals in the deduced protein sequences of predicted F48E3.8 and C54G7.3 isoforms. In *C. elegans*, a classical signal peptide sequence is found at the N-terminal end of C54G7.3 isoforms A and B, both of which are derived from the canonical 5' end of the

gene (i.e. encoded by predicted exon 1). The hypothetical "late" isoform that could be derived from an internal promoter of the C54G7.3 gene (**Figure 3.5B**), assumed to capture all exons downstream of the large central intron as depicted in **Figure 3.2B**, would not have a classical signal peptide and additionally, the peptide sequence does not appear to have non-classical secretion signals. Non-classical secretion signals have been noted in a wide variety of species (Cleves and Kelly, 1996). Previously characterized non-classical secretion signals from bacterial and mammalian sequences were used as the basis for developing the SecretomeP2.0 bioinformatics tool (Bendtsen *et al.*, 2004; Bendtsen *et al.*, 2005). In F48E3.8, none of the isoforms currently defined by the predicted transcripts (a and b) or the one fully sequenced transcript (c) produce peptide sequences with an N-terminal classical signal peptide. It is tempting to speculate, given the similarities between the two genes and the presence of a signal peptide derived from the initial exons of C54G7.3 but not F48E3.8, that the full ORF of F48E3.8 has not been identified (i.e. that sequences upstream of the current limits of the gene might encode the real 5' end of F48E3.8). To test this hypothesis, I included the sequence for the two most proximal genes upstream of F48E3.8: the sequences F46C8.18 and F46C8.13 which are 139 bp and 70 bp in length and annotated to encode ncRNAs. Even when these sequences are included (together or separately) and a continuous ORF can be discerned moving into the F48E3.8 sequence, no signal peptide is added to the front end of the current protein sequence (although it appears that the additional sequence introduces a non-canonical secretion signal to the peptide sequence). The distance separating these

two sequences from F48E3.8 also suggests it is unlikely they serve as additional 5' exons. F48E3.8 appears to lack a signal-peptide encoding sequence at its 5' end. This is further confirmed by sequence analysis for the *C. briggsae* homologs CBG_16715 and CBG_14801 which encode protein homologs to *Ce*C54G7.3 and the N-terminal end of *Ce*F48E3.8, respectively (**Figure 2.1**). The *C. briggsae* homolog of C54G7.3 does have a signal peptide but the *C. briggsae* homolog of F48E3.8 does not. We have also reported the presence of Bml_33340 in *B. malayi* as the sole genomic homolog of the *C. elegans* PDAs. Bml_33340 is more closely related to *Ce*F48E3.8 than *Ce*C54G7.3 (**Figure 2.1**). Analysis of the *B. malayi* protein sequence using SignalP 4.0 does not reveal a signal peptide; however, a secretion signal is present in the WolBm0147 peptide sequence derived from the *Wolbachia spp.* endosymbiont of *B. malayi* (**Figure 2S.1**).

Analysis of metal-ion dependence in catalysis

We have previously reported the presence of 5 motifs, which characterize the PDA domain and share responsibility for co-factor binding or acetyltransferase activity, within the protein sequences of nematode polysaccharide deacetylase enzymes (Heustis *et al.*, 2012), noting substitutions in some residues that are highly conserved in other species. One of the changes that we have noted is a substitution of one or two histidines in Motif 2 (**Figure 2.2**). These histidine residues have been implicated in co-factor binding and, hence, this substitution may reflect an important difference in the function of nematode PDAs (Blair *et al.*, 2005). Data from other members of our lab have shown that

EDTA does not ablate the chitin deacetylase activity we observe in substrate gels similar to those from **Figure 2.5**. Further work will be necessary to develop a stringent assay to ask whether substitutions within the nematode PDA domains have resulted in major changes in enzyme function, such as co-factor dependence.

DISCUSSION

Useful information regarding the function of the F48E3.8 and C54G7.3 enzymes *in vivo* can be gleaned by a careful analysis of gene regulation and protein activity using molecular and biochemical tools to study these properties. These two genes have not previously been the subjects of any comprehensive analyses aimed at characterizing their molecular and biochemical traits; however, additional information regarding gene expression, protein processing and function has been captured in the numerous genome-wide screens for *C. elegans*. A comprehensive summary of unpublished information available on these genes will be useful to guide further inquiry (**Table 3.1**). The significance of these results is also evaluated in conjunction with information available from previously published work on genome-wide analyses and in context with the new empirical evidence obtained from my study.

We previously reported that F48E3.8 and C54G7.3 are expressed in the soma since a robust expression of the genes is detectable by RT-PCR in germline-ablated worms (**Figure 2.3**). The somatic expression of these genes is corroborated by results of *in situ* hybridization, available on NextDB (http://nematode.lab.nig.ac.jp/), for F48E3.8 (cosmid F48E3, clone 272e1) and

C54G7.3 (cosmid C54G7, clones 325b6 and 208c10). Since both genes are expressed in the soma, a larger repertoire of techniques to study gene expression and protein function will be available than if they were confined to the germline since gene and protein reporter constructs are often deactivated in the germline.

Kim and colleagues (2006) used a meta-analysis of previously reported microarray results to generate a topographical representation of the spatiotemporal co-regulation of genes during *C. elegans* development (Kim *et al.*, 2001). In their analysis, F48E3.8 is clustered in Mount 1 (genes enriched in muscle, in neurons, and in generating proteins with PDZ protein-protein interaction domains; Wormbase WS227). *chs-2*, the chitin synthase responsible for synthesizing pharyngeal chitin, is also placed in Mount 1. C54G7.3 was placed in Mount 0 (an uncharacterized mountain). These are all distinct from the placement of *chs-1*, the eggshell chitin synthase, which appears in Mount 7 (characteristic of genes enriched in the germline, in oocytes, during meiosis and during mitosis).

Further work will be necessary to clone and sequence full transcripts produced from the F48E3.8 and C54G7.3 genes since sequence analysis of full transcripts will predict the complete peptide sequences of each isoform. In this work, I have confirmed selected exon-intron boundaries for the C54G7.3 gene (**Figure 3.1**).

The currently assigned transcripts of the genes are not likely to reflect the full scope of transcript diversity, since amplification across internal regions of the genes return a larger number of products than would be obtained if transcript

numbers were accurate (**Figure 3.4** and **Figure 3.8**). Various molecular cues suggest that in C54G7.3 an internal promoter may be used to drive a truncated transcript that has not been previously detected (**Figure 3.5** and **Figure 3.7**). However, my attempts to confirm the presence of "late" SL1-spliced transcript, driven from the internal promoter, have not yielded a result when tested with a semi-nested PCR approach (**Figure 3.8**). Attempts to use a Northern Blot to address the questions of increased transcript number were inconclusive since no PDA-encoding transcript was detected at all (**Figure 3.10**).

The Choi and Newman (2006) analysis, which elucidated the SL1 splicing of "late" transcripts produced from internal promoters in genes containing large introns, did not include reference to the genes F48E3.8 and C54G7.3 since no SL1-led transcripts have been previously identified for these genes. One gene that was included in their analysis was *tropomyosin-1* (*tmy-1*) which utilizes a two-promoter system and drives transcripts from an internal promoter. Comparing the *tmy-1* gene structure to either F48E3.8 or C54G7.3 shows some interesting correlates. The two promoter system is used to drive tissue-specific transcripts of *tmy-1*: the tropomyosin gene is expressed in body-wall muscle, but the internal promoter is used to produce transcripts (encoding isoforms III and IV) in the pharynx and intestine (Anyanful *et al.*, 2001). While my analysis has focused on the expression and activity of F48E3.8 in the pharynx, it has recently been reported that F48E3.8 is expressed in the body-wall muscle and, in fact, the sub-cellular localization of the protein is similar to that of TMY-1 based on their analyses (Meissner *et al.*, 2011). F48E3.8 is expressed both in the pharynx and

body-wall muscle, then, and further work to determine if F48E3.8 utilizes a two-promoter system may be warranted. I have presented results showing the presence of putative PHA-4 binding sites within the ~ 1 kb region immediately preceding the translation initation site for F48E3.8 transcript b. Six potential PHA-4 binding sites are found here, including one element (TGTTTAT) which reads as a TRTTKRY element in both directions. The internal promoter of *tmy-1* is responsive to the PHA-4 transcription factor (Anokye-Danso *et al.*, 2008), drives both embryonic and post-hatching pharyngeal expression (Anyanful et al., 2001) and includes these bi-directional elements since the internal promoter includes a TGTTTGT sequence (Anyanful *et al.*, 2001).

While I noted no *bi-directional* TRTTKRY elements in the canonical promoters of F48E3.8 and C54G7.3, I found one such putative PHA-4 binding site within the F48E3.8 large intron and four instances of these *bi-directional* TRTTKRY elements in the C54G7.3 large intron. In C54G7.3, two of these elements are TGTTTGT sequences. This specific sequence has been described as one of the TRTTKRY sequences with relatively strong to moderate PHA-4 binding activity (Raharjo *et al.*, 2010). This suggests that the putative internal promoter of the C54G7.3 gene may have high affinity for the PHA-4 transcription factor consistent with other evidence suggesting the region may include a promoter.

My analysis of the C54G7.3 canonical and putative internal promoters using the Promoter 2.0 Prediction Software confirm the likelihood of a functional promoter in the region of the canonical promoter of this gene, but not in the large

intron of the gene. In general, software designed to predict eukaryotic promoters are currently considered unreliable owing to the diversity of sequences in eukaryotic promoters and the lack of empirical evidence on characterized eukaryotic promoters from which to build consensus sequences that drive this type of software. The *C. elegans* promoter database (CEPDB) is one repository for sequence information on identified *C. elegans* promoters. A canonical promoter for C54G7.3 is listed, and the reported sequence does correspond to the published genetic sequence preceding the C54G7.3 ORF (**Table 3.1**). The source of this information is not listed and my attempts to contact the webmaster have been unanswered. Experimental evidence demonstrating the presence of a functional internal C54G7.3 promoter may be obtained by generating a GFP reporter construct for this gene driven by this putative regulatory region.

I used an RT-PCR time course to investigate gene expression dynamics for the F48E3.8 and C54G7.3 genes with the best results presented when sufficiently dense populations of worms are grown in solid media at 25°C and RNA extracts are DNAse-treated prior to cDNA synthesis (**Figure 3.12B**). In these experiments, I could not re-create the punctuated bursts of expression described for the *chs-2* gene when analyzed by Veronico *et al.* (2001). However, using *col-12* and *dpy-13* we observe the cyclical antiphase waves of expression that have been described for the genes by Johnstone and Barry (1996). Multiple trials of this experiment – in my hands – suggest that *ama-1* levels are not consistent throughout development and as such this gene may not be the ideal positive control against which to normalize the expression of other genes. Based

on literature reviews, I have been unable to find any gene which has been established as an ideal positive control for this type of experiment. The absence of a solid positive control makes it difficult to draw conclusions about whether the changes in expression pattern for any of the genes investigated are a result of *bona fide* biological changes or simply the result of limited template. Underscoring the possibility of the latter concern, peaks for most of the genes are overlapping. Nonetheless, differences are seen in the peaks of expression between the two collagens *col-12* and *dpy-13* suggesting that all variation cannot solely be attributed to template variation. *col-12* and *dpy-13*, then, may be of some limited utility in drawing conclusions on timing, so the collagens may be used as rough markers for the presence of different developmental stages in these time courses. While genes like *col-12* and *dpy-13* may be used as markers for the various stages it might also be possible to observe waves of expression of the transcription factor *lin-29* (or other heterochronic genes) and to apply its expression to analyses of this type.

The PFam analysis of the F48E3.8 and C54G7.3 isoforms returns an output with a remarkably low number of functional domains. The only reproducibly detected domains are those of the PDA (**Figure 2.2 and Figure 2S.1**) and a number of repeats of an EB module domain. According to PFam annotations, the EB domain (which is found in many *C. elegans* proteins) has no known function; the domain is known to include 8 cysteine residues that are predicted to form four disulphide bridges but the functional significance of these residues has not been confirmed. One notable difference between nematode PDA

proteins and some insect homologs is the absence of the LDLa receptor domain that is found in Groups 1 and 2 of the insect PDA proteins (Dixit *et al.*, 2008; Luschnig *et al.*, 2006; Wang *et al.*, 2006). The PFam analysis of nematode PDAs returns one discernible chitin-binding domain (CBD) that is found in both predicted isoforms of C54G7.3 but none in any isoforms of F48E3.8. The CBD may be dispensible for substrate interaction: deacetylation of a substrate may not require binding it; however, the domain is present in 4 of the 5 groups of chitin deacetylases found in insects. It is possible that nematode PDAs function without independent substrate binding domains or that inaccuracy in the exon-intron boundaries yield protein sequences with interrupted (and hence unrecognizable) chitin-binding domains. It is also possible that, in the absence of empirical data on nematode chitin-binding domain sequences, the current PFam algorithms fail to accurately detect chitin-binding domains that may have unique, nematode-specific attributes. Empirical work will be necessary to determine whether isoforms of F48E3.8 and C54G7.3 bind to chitin *in vitro* and may be assessed by passing crude protein extracts through chitin-packed columns followed by elution with SDS and subsequent protein identification. Establishing the chitin-binding properties of these enzymes will further elucidate the mechanism of action of the nematode PDAs and, coupled with sequence analysis, may provide the information to enhance the predictive capacity of bioinformatic tools designed to help domain prediction.

Results from my SignalP 4.0 analysis suggests that the two predicted isoforms of C54G7.3 may have signal peptides and, as such, may be secreted

through the classical secretory pathway. In F48E3.8, the protein isoform corresponding to transcript c is found to have a high probability of secretion via a non-classical pathway (**Figure 3.2**). Empirical evidence exists that F48E3.8 is likely to be secreted. Meissner *et al.* (2011), using GFP tagging of proteins expressed in embryonic body-wall muscle, showed that F48E3.8 protein is localized to the endoplasmic reticulum and as such, despite the lack of a signal peptide, the protein appears to enter the pathway that positions it as an enzyme that is secreted or that functions in the ER or Golgi (where it may work to alter carbohydrates during the assembly of other macromolecules that are secreted). A large-scale attempt to identify N-linked glycoproteins produced in *C. elegans* has been conducted (Kaji *et al.*, 2007). In this analysis, F48E3.8 protein is identified as a protein that is N-glycosylated and specifically its carbohydrate moiety has been shown to include mannose. Protein sequences were analyzed using mass-spectrometry and the peptide sequence for F48E3.8 is common to all three isoforms that would be derived from the three currently annotated transcripts. Kaji *et al.* (2007) note that a large proportion of the N-glycosylated proteins identified in their study lack signal peptides, suggesting that many proteins in *C. elegans* are secreted via non-classical mechanisms. The non-classical secretion of numerous proteins has also been described in *B. malayi* where many proteins detected in the excretory-secretory protein (ESP) catalog lack a signal peptide (Moreno and Geary, 2008). Detecting the specific processing and localization of F48E3.8 and C54G7.3 would require the creation and microscopic analysis of tagged versions of these proteins. Secretion from the cell is the presumed fate for

these proteins if they mediate any of the functions outlined in **Table 2.4**. I have proposed that plant parasitic nematodes may secrete polysaccharide deacetylase enzymes as virulence factors that aid in tissue penetration and demonstrated that that one homolog of the *C. elegans* F48E3.8 is *Meloidogyne incognita* Msp9 (Heustis *et al.*, 2012). Msp9 transcript is found in the *M. incognita* pharyngeal gland cells and it encodes a signal peptide (Huang *et al.*, 2003).

In all our analyses, we have focused on the somatic expression and function of two nematode polysaccharide deacetylases in *C. elegans*, but we have not directly investigated the possibility that these genes have a role in the nematode germline. In sheathed filarial parasites like *B. malayi*, where the chitinous eggshell is remodeled into a microfilarial sheath (**Figure 1.2**, **Table 2.4**), polysaccharide deacetylases could mediate the biochemical transition necessary to remodel this otherwise rigid structure. Even in the absence of such gross changes in morphology, deacetylation of the chitinous eggshell may act in coordinate function with chitinases to alter the embryonic covering in preparation for hatching. In a large-scale screen using microarrays to identify genes enriched in the *C. elegans* germline, F48E3.8 was identified as a sperm-enriched transcript – a finding confirmed by an RT-PCR assay (Reinke *et al.*, 2000). Intriguingly, information available from the Biological General Repository for Interaction Datasets (BioGRID) lists F48E3.8 as interacting with K08E3.6 (cyk-4) (**Table 3.1**). *cyk-4* (*cytokinesis-4*) encodes a Rho GAP that is involved in establishing the anterior-posterior axis of the early embryo and in forming the microtubules that

function in the central spindles during anaphase and which polarize the foregut epithelium (Wormbase WS227).

Efforts to further characterize the F48E3.8 and C54G7.3 genes during *C. elegans* development will benefit from loss-of-function alleles for the genes; there are currently no available worm strains carrying useful alleles for either. While two alleles of the C54G7.3 gene have been generated – gk2686 (a silent mutation) and gk1581 (a missense mutation) – they are available in two strains that carry many mutations: the gk2686 allele is present in the VC2452 strain and the gk1581 allele is present in the VC1923 strain which carry 224 and 323 other mutations respectively. These strains were generated in large mutagenesis screens designed to study the biases in generating mutant alleles when using different mutagens to induce mutations that are identified through deep sequencing (Flibotte *et al.*, 2010), and therefore are not useful for characterizing these two individual genes.

FIGURES

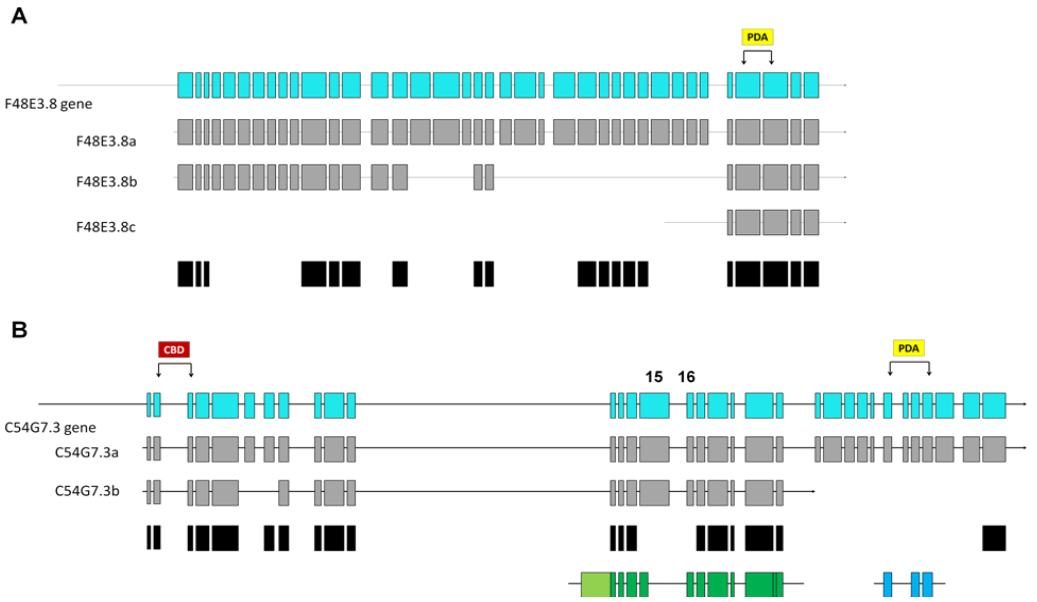


Figure 3.1. Current annotations of gene structure and predicted transcripts of F48E3.8 and C54G7.3 including exon-intron boundaries. Gene maps showing predicted exon-intron boundaries in the genes (connected turquoise boxes), predicted transcripts (connected grey boxes) and a summary of information regarding any exons with known expression in any transcript (black boxes) are presented here, based on information available on Wormbase (WS227). [This image puts the genes in the correct orientation with 5' to the left.] The region originally predicted to encode the PDA domain is highlighted with brackets above each gene. **(A)** For F48E3.8, 38 exons are found across the ORF of the gene and three transcripts (a,b,c) have been assigned using information available from ESTs. Multiple cDNA clones exist for fragments of the gene and the exons with confirmed expression by ESTs for these clones are depicted here using black boxes below the ORF to signify confirmed expression. The three predicted transcripts of F48E3.8 are predicted to encode a PDA domain. **(B)** For C54G7.3, 28 exons are found across the ORF of the gene and two transcripts (a,b) have been assigned. Multiple cDNA clones exist for fragments of the gene and the exons with expression confirmed by ESTs for these clones are depicted here using black boxes below the ORF to signify expression. Transcript a (but not b) is predicted to encode the PDA domain. The exon-intron boundaries confirmed by sequencing the yk1620b03 clone is presented here (connected green boxes); boundaries confirmed through a cloned amplimer of the PDA-encoding region (connected blue boxes) are also presented here.

Figure 3.2. F48E3.8 and C54G7.3 isoforms based on transcript predictions. This comparison of isoforms from F48E3.8 and C54G7.3 present the peptides by scale. Black solid box borders signify isoforms whose length are based on fully sequenced cDNA; black hatched box borders signify isoforms whose length are based on transcripts that have not been fully sequenced. Red hatched boxes are hypothetical isoforms. The PDA domain, if present in an isoform, is depicted by a yellow box. Signal peptides are depicted with blue boxes at the start of the peptide. **(A)** F48E3.8 isoforms A-C all lack signal peptides for classical secretion, but isoform C is predicted to contain a non-classical secretion signal (S). A predicted non-classical secretion signal is not found in isoform A. **(B)** C54G7.3 isoforms A and B share an N-terminal region predicted to encode a signal peptide for classical secretion. The hypothetical isoform of C54G7.3, which would be produced through activity of an internal promoter, does not have a signal peptide and also does not bear non-classical secretion signals.

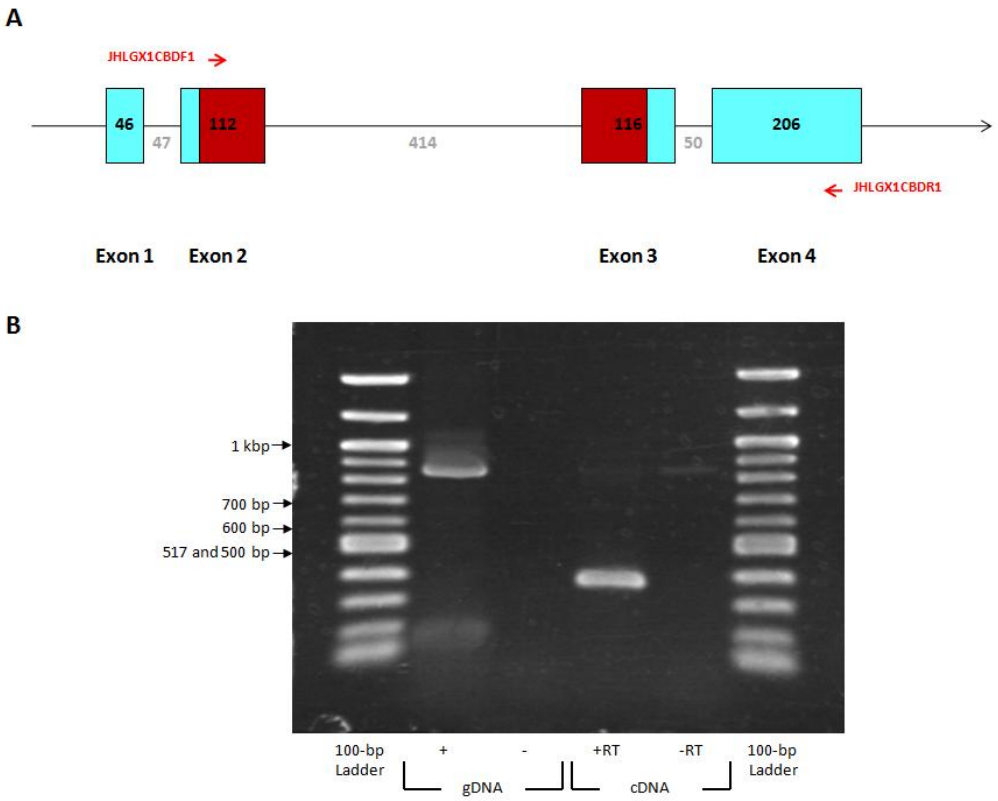


Figure 3.3. C54G7.3CBD primer specificity and analysis of amplimer size. PCR was performed using the JHLGX1CBDF1 and JHLGX1CBDR1 primers. **(A)** Schematic depicting annealing sites for these primers relative to the region originally predicted to encode the CBD domain (burgundy). **(B)** PCR was performed using these primers under the following conditions: Stage 1: 94°C, 3 minutes; Stage 2: [94°C, 1 minute; 56°C, 1 minute; 72°C, 3 minutes] x30; Stage 3:72°C, 10 minutes. Amplimers are predicted to be 836 bp (gDNA) and 372 bp (cDNA).

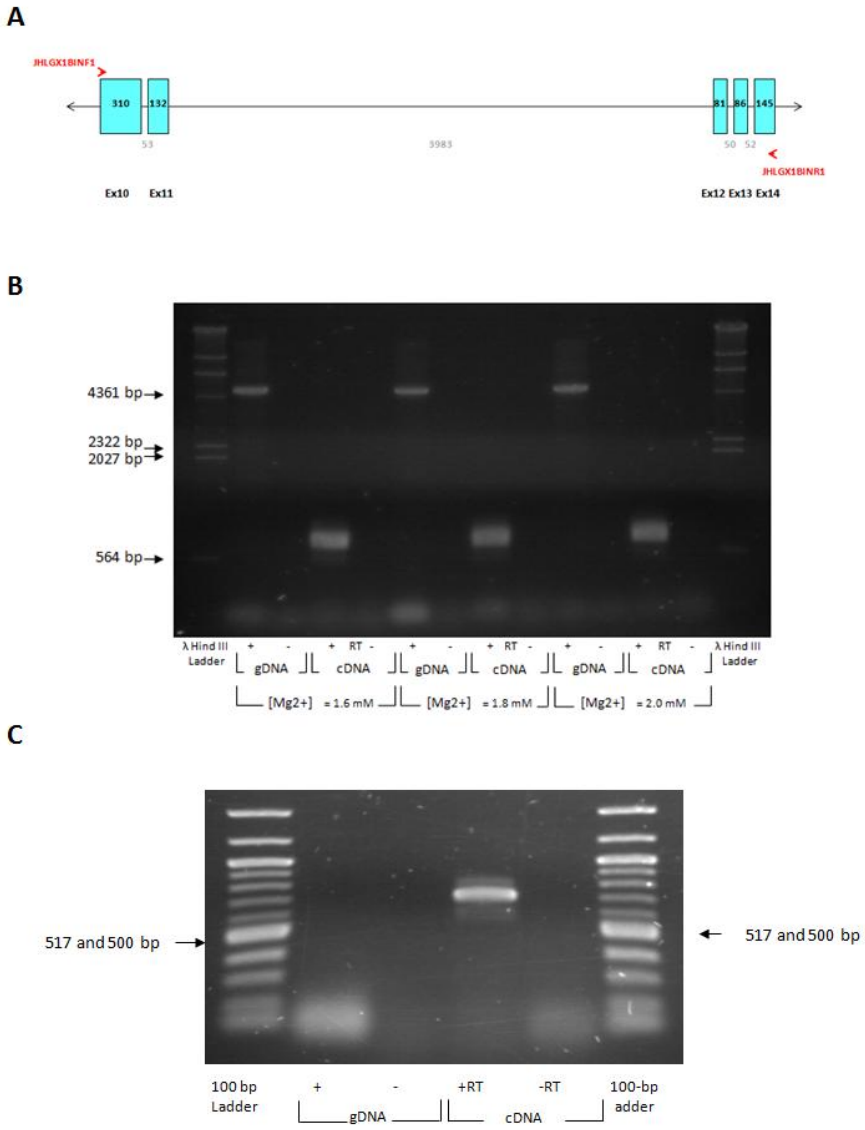


Figure 3.4. PCR generates amplimers across the large predicted intron of C54G7.3 and suggests greater transcript complexity. PCR was performed using the JHLGX1BINF1 and JHLGX1BINR1 primers. **(A)** Schematic depicting annealing sites for these primers. PCR was performed under the following conditions: Stage 1: 94°C, 3 minutes; Stage 2: [94°C, 1 minute; 62.4°C, 1 minute; 72°C, 3 minutes] x30; Stage 3:72°C, 10 minutes. **(B)** Using three different concentrations of Mg^{2+} with Elongase enzyme mix, the appropriately-sized amplimers were obtained to match predictions of 4813 bp (gDNA) and 675 bp (cDNA). **(C)** When Taq Polymerase is used, no amplimer is detected from gDNA and additional bands, which may reflect alternative splicing, are obtained from cDNA. Primer sequences are listed in **Appendix A**.

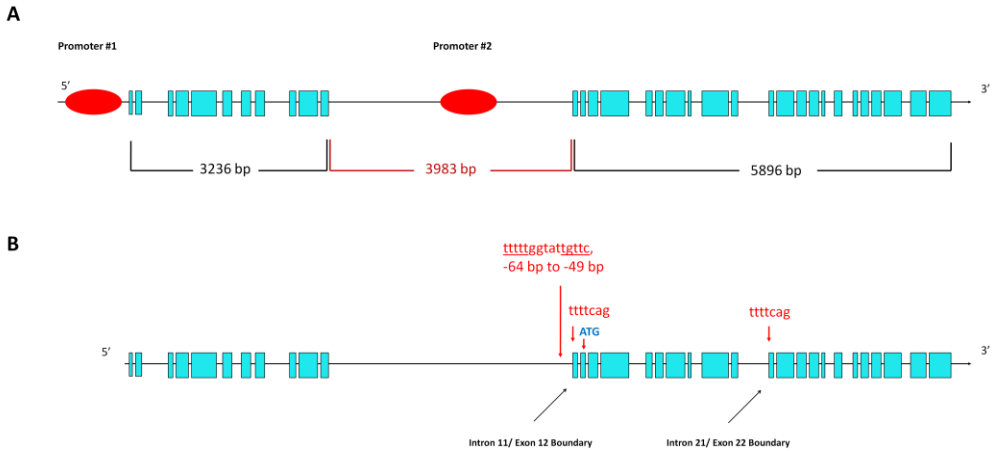
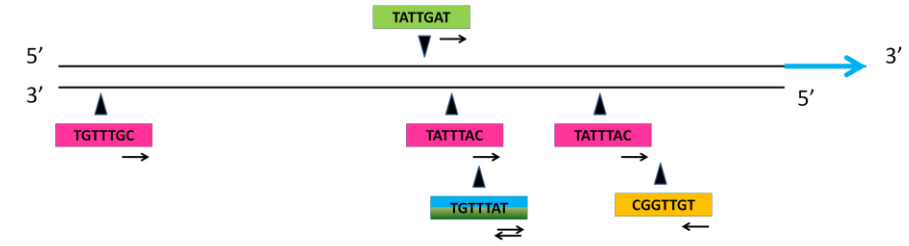
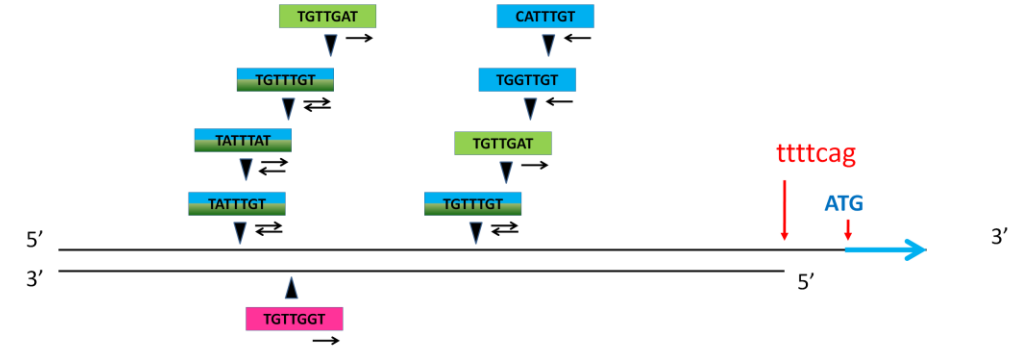


Figure 3.5. Key features of the C54G7.3 gene suggest an internal promoter. The C54G7.3 gene contains 33 exons that, when compared to transcripts a or b sequence, are separated by a large ~4 kb intron. 11 exons appear before this large intron. **(A)** This gene structure is reminiscent of the gene structure elucidated by Newman and Choi (2006): a canonical promoter region is found upstream of the gene and drives transcripts starting with exons as early as Exon 1 (Promoter #1) and can also drive expression initiating with downstream exons (Promoter #2). For C54G7.3, this promoter region is hypothetical and no evidence has established its functionality as a promoter or the presence of SL1-led "late" transcripts that are produced from this promoter. (B) Various attributes at the boundary between predicted intron 11 and exon 12 suggest the transcript produced could interact with the SL1-leader sequence carrying snRNP that allows nascent transcripts to be trans-spliced. This includes the UC-rich docking site (approximately -50 bases from the ATG-encoding sequence) where the snRNP sits and the frequently observed SL1 acceptor site sequence (UUUUCAG). An in-frame ATG would initiate translation of a peptide sequence that maintains an in-frame PDA domain.



~ terminal 1kb of F48E3.8 large intron



~ terminal 1kb of C54G7.3 large intron

Figure 3.6. Analysis of putative PHA-4 binding sites in large introns of the F48E3.8 and C54G7.3 genes. PHA-4 interaction in a promoter region is mediated by TRTTKRY elements. Six of these elements are found within the ~1 kb region at the end of the predicted large intron of F48E3.8 and as many as nine of these elements are found within the ~1 kb region at the end of the predicted large introns of C54G7.3. The TRTTKRY elements can sit on either DNA strand and read in either direction. It is noteworthy that *bi-directional* PHA-4 recognition sites are observed here.

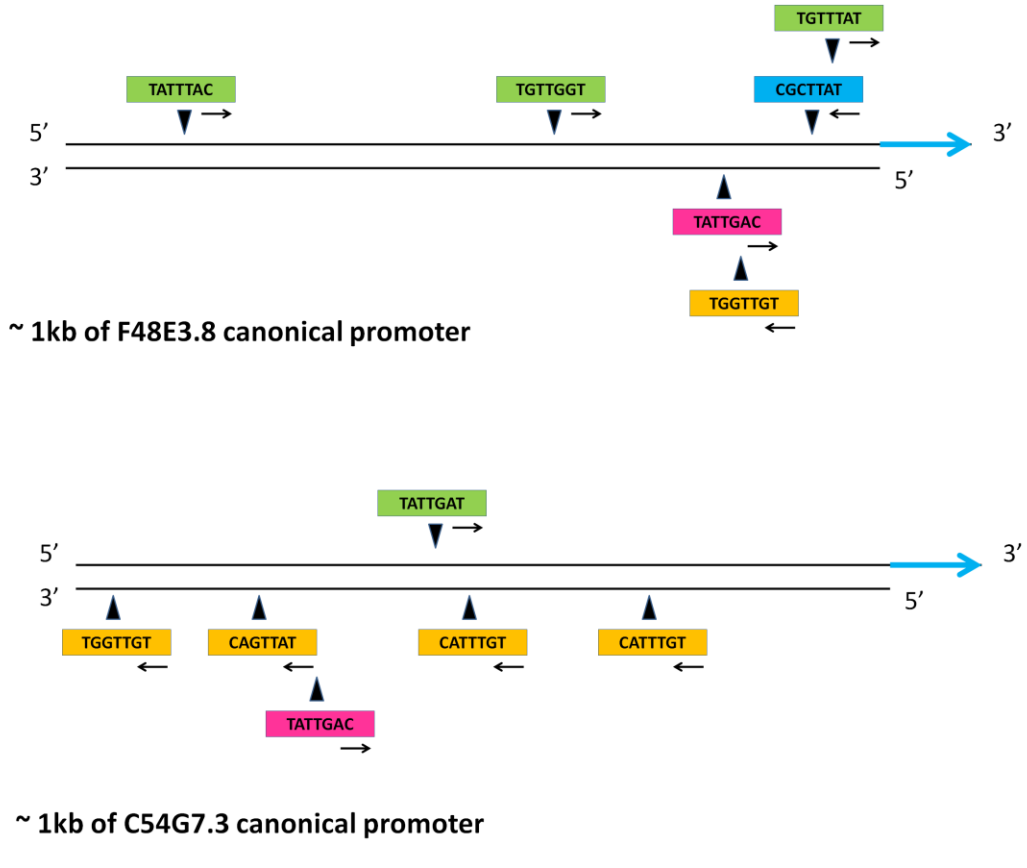


Figure 3.7. Analysis of putative PHA-4 binding sites in the canonical promoters of the F48E3.8 and C54G7.3 genes. Six putative PHA-4 binding sites are found in the canonical promoter of F48E3.8 and similarly six putative PHA-4 binding sites are found in the canonical promoter of C54G7.3. (I have identified an additional PHA-4 binding site in the C54G7.3 promoter as five were previously reported by Gaudet and Mango (2002).

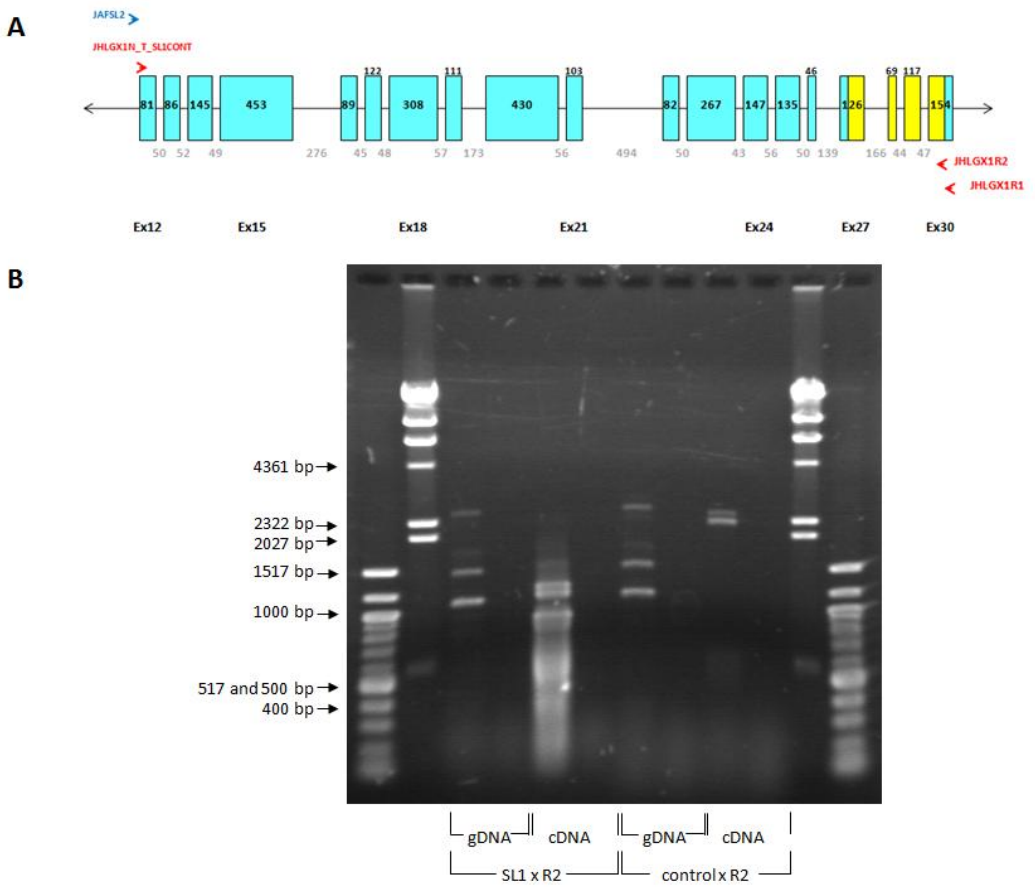


Figure 3.8. A semi-nested SL1 splice leader PCR fails to identify a discernible novel transcript produced from the putative internal promoter. (A) The region targeted for amplification initiates immediately downstream of the predicted large intron and extends through the PDA-encoding domain (yellow). **(B)** The semi-nested PCR consisted of two 30-cycle reactions. In step one, JHLGX1R1 was used as a reverse primer and in step two JHLGX1R2 was used as the reverse primer. The JAFSL2 primer, including the SL1 sequence (*SL1 in [b] of figure*), was used in attempts to isolate an SL1-led transcript. The JHLGX1NEWTRANSL1CONT primer was used in a control reaction designed to amplify starting with the sequence immediately 3' of the likely SL1 sequence, and previously identified as a predicted exon (*control in figure*). In each 30-cycle reaction, the following PCR conditions were used: Stage 1: 94°C, 3 minutes; Stage 2: [94°C, 1 minute; 47.6°C, 1 minute; 72°C, 3 minutes] x2; Stage 3: [94°C, 1 minute; 60.9°C, 1 minute; 72°C, 3 minutes] x28; Stage 4:72°C, 10 minutes. Amplimers are predicted to be approximately 4912 bp (gDNA) and 3017 bp (cDNA) among the primer combinations. Primer sequences are listed in **Appendix A**. The smears observed after the second 30-cycle round of PCR (using R2 in this round) are likely the result of SL1 binding to numerous transcripts edited with the SL1 sequence.

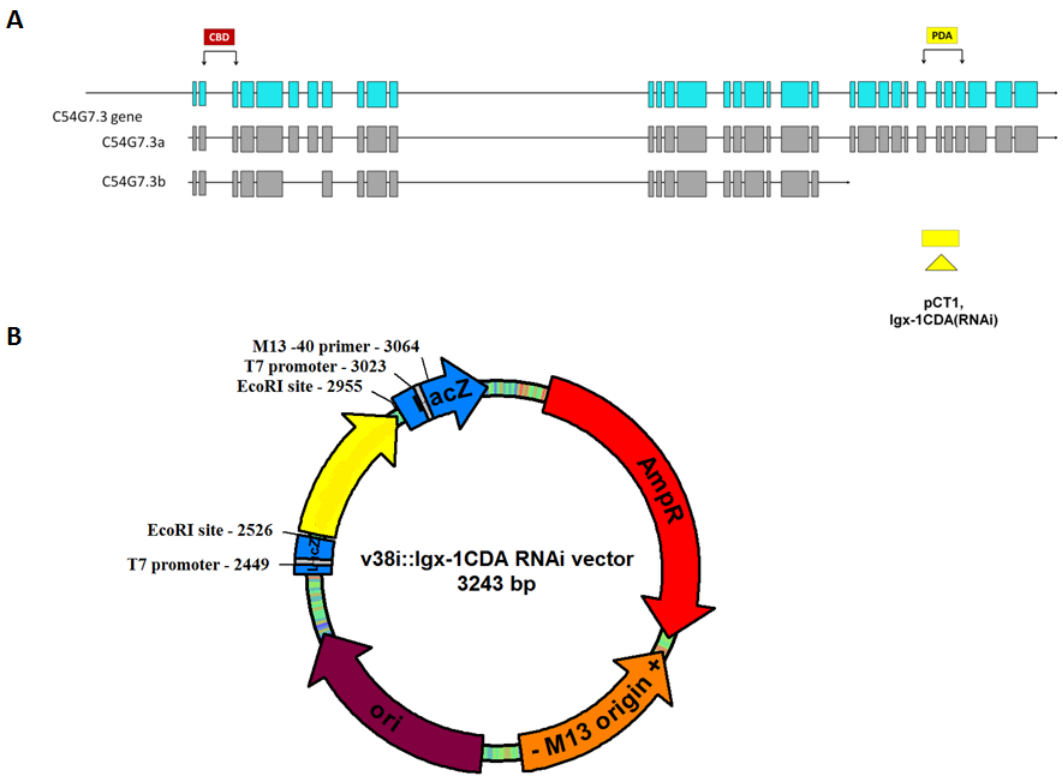


Figure 3.9. A construct designed to generate anti-sense probe targeting the C54G7.3PDA in Northern Blot analysis. (A) The region of the C54G7.3, represented in the cloned cDNA, used to generate the construct is represented by a yellow box below its corresponding region in the C54G7.3 gene. (B) EcoRI sites were used to clone a fragment of C54G7.3 cDNA encoding the PDA domain inserting it into the Litmus v38i vector as shown. This construct was linearized and used to generate ssRNA probe for use in Northern Blots.

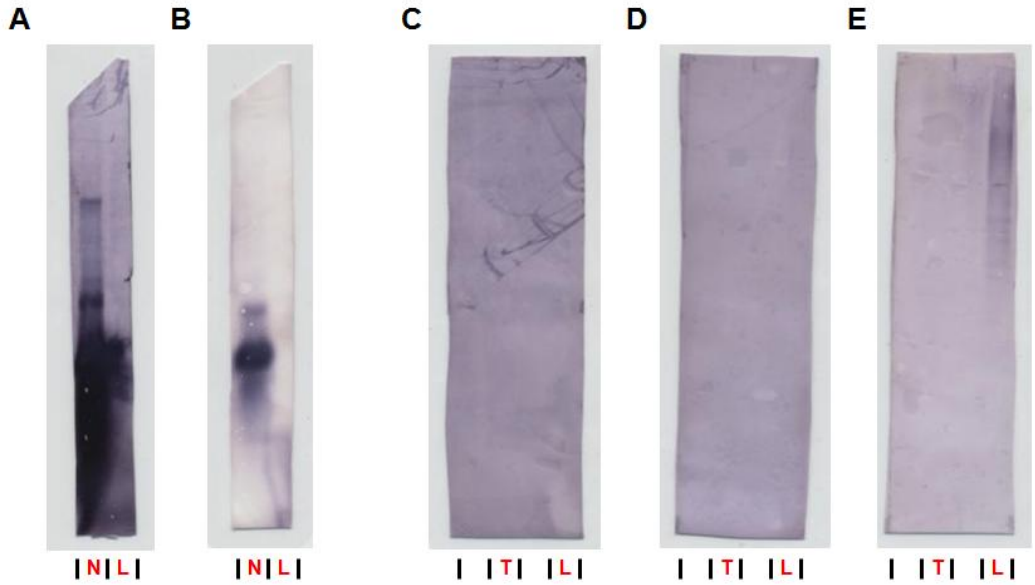


Figure 3.10. The C54G7.3PDA domain-encoding region is not detected in Northern Blots using ssRNA probe to detect transcripts present in representative mixed-stage RNA samples from *wild-type* N2 worms. Northern Blots for *Neo* RNA (control provided my manufacturer) using a *Neo* antisense probe (A,B) and for the PDA-encoding domain of C54G7.3 using a C54G7.3PDA antisense probe (C,D,E) are shown. Blots strips correspond to 2 or 4 lanes as demarcated below the images with RNA in each lane as follows: control *Neo* RNA, N; target mixed-stage *wild-type* N2 RNA, T; and ladder, L. **(A)** *Neo* RNA can de detected after standard pre-hyb and hyb conditions followed by incubation in detection buffer for 16 hours. Extensive staining is observed in the lane labeled [N], appearing to spill over into the ladder RNA [L]. **(B)** When incubation in the detection buffer is reduced to 2 hours, *Neo* RNA is still detectable but the effects of overexposure are dramatically reduced. **(C)** No message is detected using C54G7.3PDA antisense probe when 2.5 µg total mixed stage RNA is run and probed [T]. In this trial, incubation in detection buffer was allowed for 16 hours. **(D)** Increasing the total RNA per lane to 15 µg (a six-fold increase) and allowing incubation in detection buffer for 72 hours still does not distinguish target RNA in the lane labeled [T]. **(E)** A six-fold increase in RNA per lane with 6X probe and incubation in detection buffer for 24 hours also does not permit detection of any transcripts [T]. However, the increase in probe results in non-specific binding and detection of the RNA ladder [L].

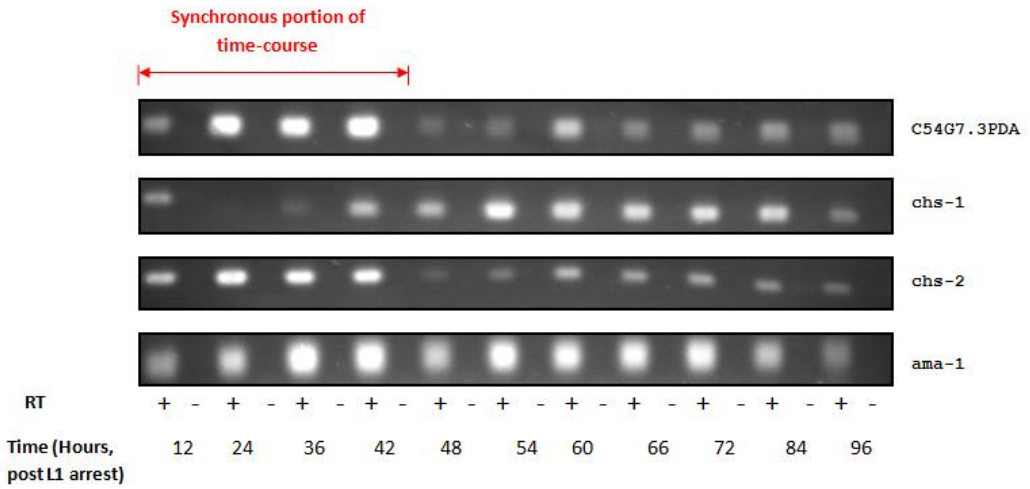


Figure 3.11. Developmental time course of C54G7.3 gene expression in wild-type *C. elegans*, harvested from liquid culture, using RT-PCR. A monoxenic worm culture was bleach-synchronized and L1 worms allowed to develop axenically for 16 hours, leading to synchronous L1 arrest. Thereafter, worms were grown in the presence of OP50 *E. coli*. Samples of culture were taken at the increments outlined, worms purified by sucrose flotation and stored in the presence of Trizol at -80°C. Thawed worms were homogenized and RNA extracted for use in synthesizing first-strand cDNA using RT (+) with no RT controls (-), when 1 µg RNA (with no DNAse treatment) was used for a 20-µL synthesis reaction diluted for a final volume of 60 µL to serve as template. PCR was conducted for segments of the genes of interest *chs-1*, *chs-2* and C54G7.3 (amplifying the predicted PDA domain, C54G7.3PDA using JHLGX1F1 and JHLGX1R1 primers). A segment of the gene *ama-1* was amplified as a positive control. All PCR reactions were conducted using temperatures specific to the annealing conditions and 30 cycles of amplification. 10 µL of a 50 µL PCR reaction are loaded to each lane of this 1% agarose gel.

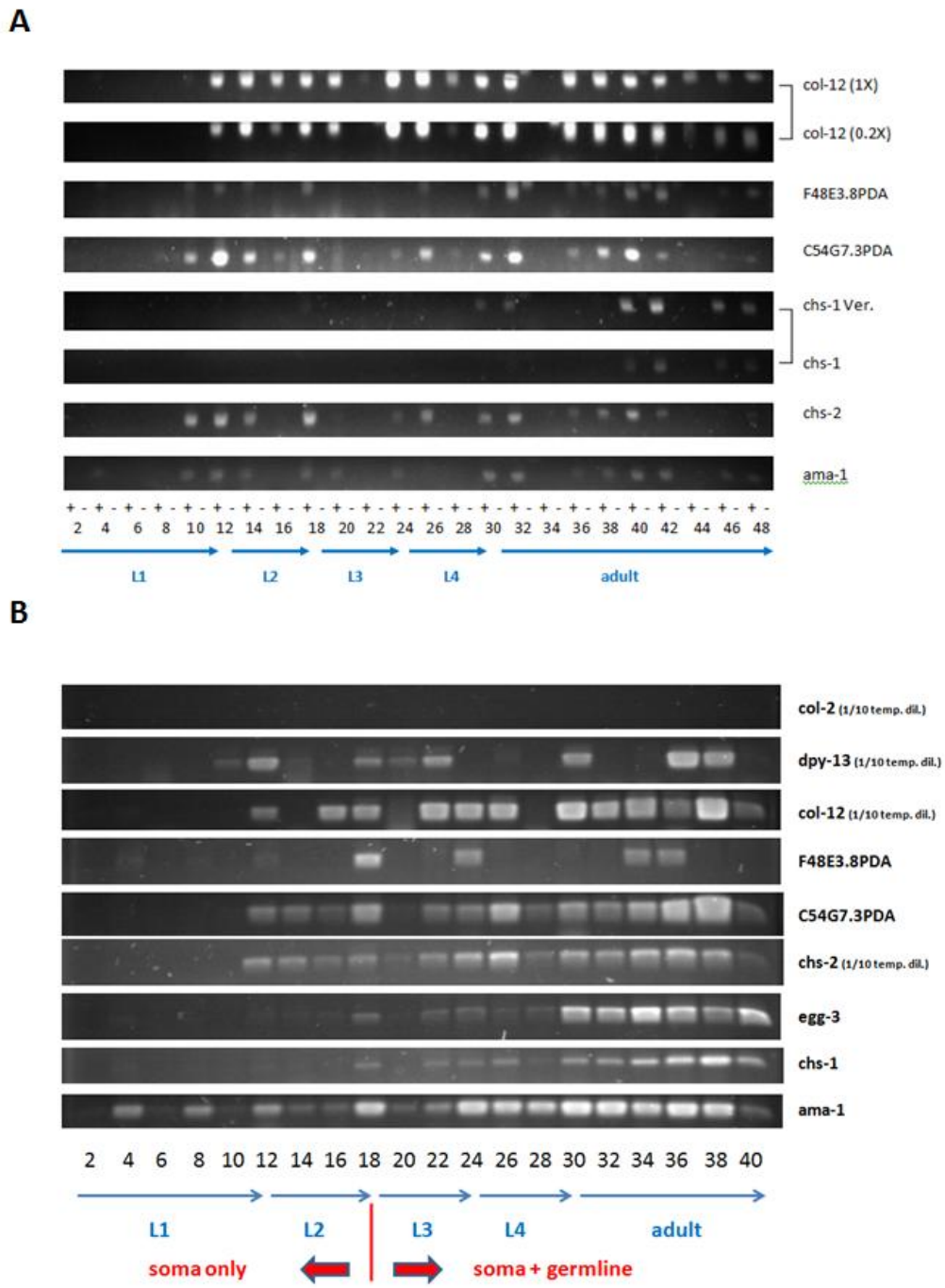


Figure 3.12. Developmental time courses of F48E3.8 and C54G7.3 gene expression in wild-type *C. elegans*, harvested from solid media, using RT-PCR. Synchronous post-hatching populations of the *C.* elegans wild-type N2 strain grown on solid media were used to extract RNA at 2-hr increments, representing expression from all stages of development (L1 to adult) before and after the presence of an F1 generation resulting from fertilization. At each time point, worms were purified by sucrose flotation before being stored at -80°C. First strand cDNA (+) was generated from each RNA extract (except No RT controls, -) and used as a template for 30 cycles

of PCR. The expression of *ama-1* (the large subunit of RNA polymerase II), *chs-1* (the "eggshell" chitin synthase), *chs-2* (the "pharyngeal" chitin synthase), *lgx-1*/C54G7.3, F48E3.8, and three previously studied collagens (*col*-12, *dpy-13* and *col*-2) were studied. *ama-1* serves as a positive control for extraction efficiency. *col-12* shows periodic bursts in expression and acts as a control for timing, but has high expression levels and requires less template for detection (0.2X). Each amplimer effectively reflects expression in 0.1 µg total RNA. **(A)** For this trial, 12,000 worms were grown on plates to harvest L1 and L2 larvae; 6,000 worms were grown on plates to harvest L3 and L4 larvae; 3,000 worms were grown on plates to harvest young adults, and, 1,500 worms were grown on plates to harvest older (>40-hr post-synchronization) adults. RNA extracts were not DNAse-treated and not standardized although the amount of RNA was quantified (230 – 1470 µg/mL in the samples). 2 µL of this RNA was used in a 20 µL synthesis reaction that was diluted to 50 µL upon completion. **(B)** For this trial, 25,000 worms were grown on plates to harvest L1 and L2 larvae; 12,500 worms were grown on plates to harvest L3 and L4 larvae; 6,250 worms were grown on plates to harvest adults. RNA extracts were DNAsed treated and standardized to 200 µg/mL before 10 µL was used for a 40 µL synthesis reaction that was ultimately diluted to 100 µL.

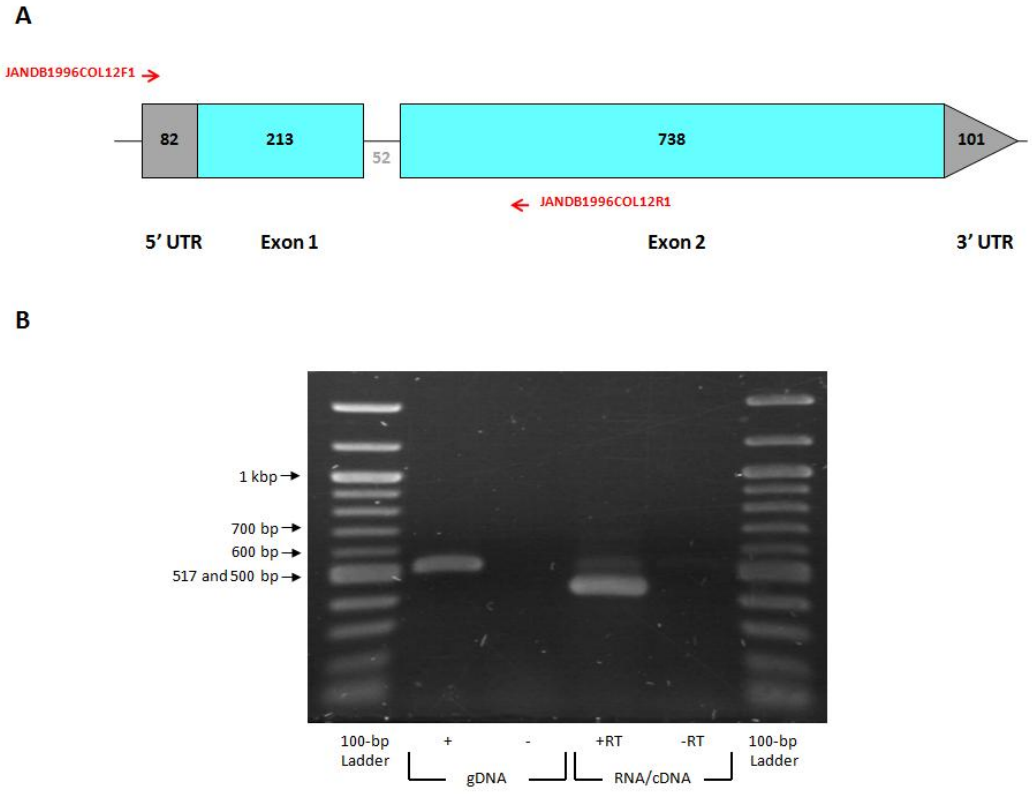


Figure 3.13. ***col-12*** **primer specificity and analysis of amplimer size.** PCR was performed using the JANDB1996COL12F1 and JANDB1996COL12R1 primers, described by Johnstone and Barry (1996). **(A)** Schematic depicting annealing sites for these primers. **(B)** PCR was performed using these primers under the following conditions: Stage 1: 94°C, 3 minutes; Stage 2: [94°C, 1 minute; 54.0°C, 1 minute; 72°C, 3 minutes] x30; Stage 3:72°C, 10 minutes. Amplimers are predicted to be 512 bp (gDNA) and 460 bp (cDNA). These primer sequences are listed in **Appendix A**.

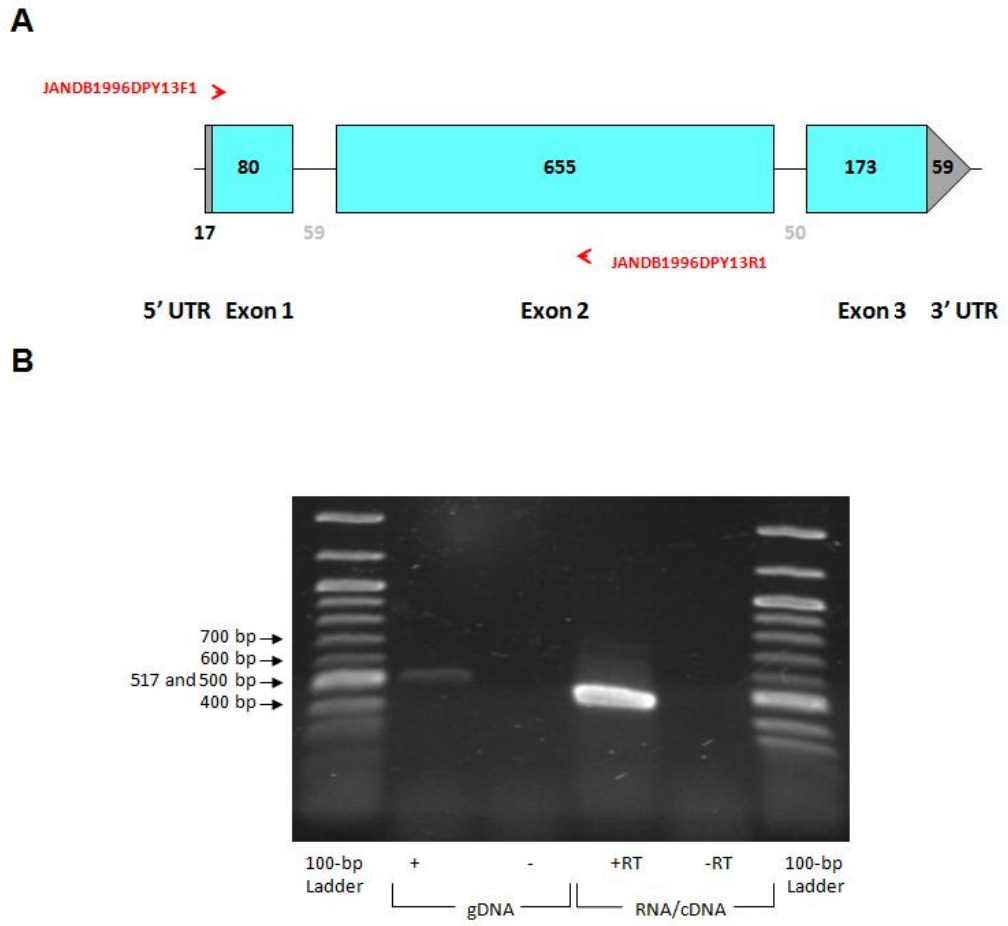


Figure 3.14. ***dpy-13*** **primer specificity and analysis of amplimer size.** PCR was performed using the JANDB1996DPY13F1 and JANDB1996DPY13R1 primers, described by Johnstone and Barry (1996). **(A)** Schematic depicting annealing sites for these primers. **(B)** PCR was performed using these primers under the following conditions: Stage 1: 94°C, 3 minutes; Stage 2: [94°C, 1 minute; 54.0°C, 1 minute; 72°C, 3 minutes] x30; Stage 3:72°C, 10 minutes. Amplimers are predicted to be 517 bp (gDNA) and 458 bp (cDNA). These primer sequences are listed in **Appendix A**.

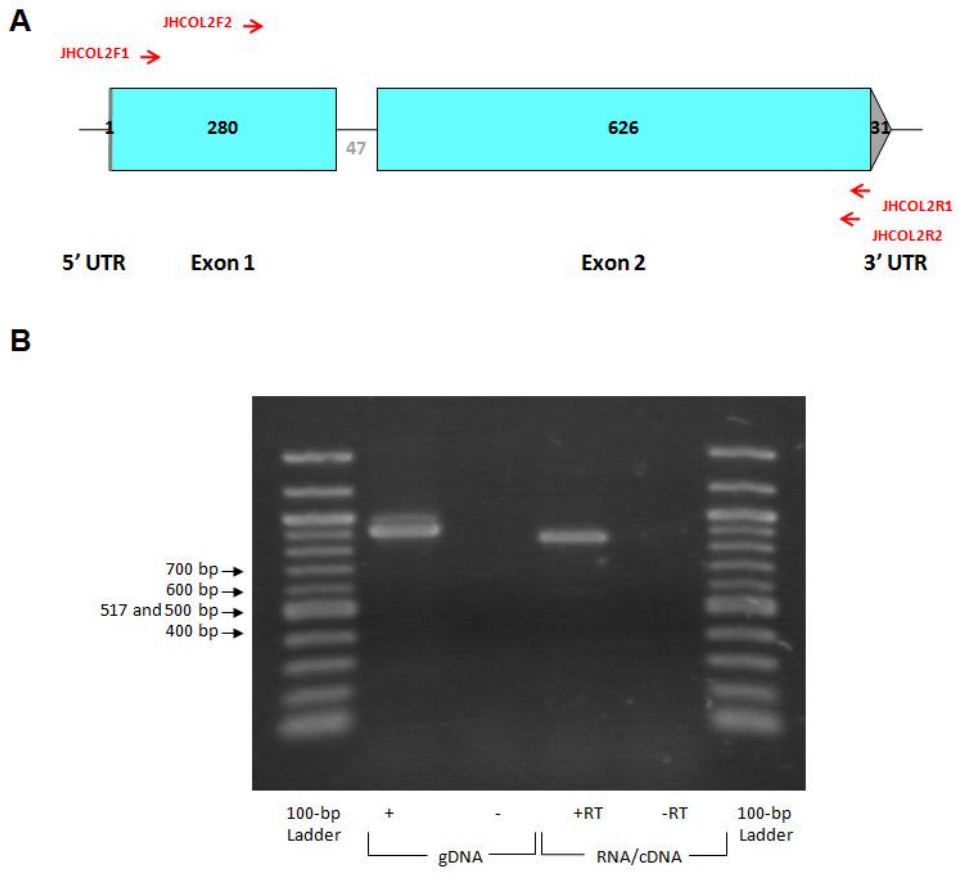


Figure 3.15. ***col-2*** **primer specificity and analysis of amplimer size.** PCR was performed using JHCOL2F1 and JHCOL2R2 primers. **(A)** Schematic depicting annealing sites for these primers. **(B)** PCR was performed using these primers under the following conditions: Stage 1: 94°C, 3 minutes; Stage 2: [94°C, 1 minute; 61.5°C, 1 minute; 72°C, 3 minutes] x30; Stage 3:72°C, 10 minutes. [N.B. In subsequent experiments, the annealing temperature was changed to 61.1°C.] While the amplimers of the *col-2* gene obtained using the JHCOL2F1 and JHCOL2R1 primers are observed at the predicted sizes of 918 bp (gDNA) and 871 bp (cDNA), numerous additional bands are observed at lower intensity from both templates. These additional bands may be explained by cross-reactivity with one of the many other collagen in *C. elegans*. In particular, BLAST analysis suggests that the JHCOL2F1 primer has a lower (but notable) complementarity to the *C. elegans col-92* gene. Primer sequences are listed in **Appendix A**.

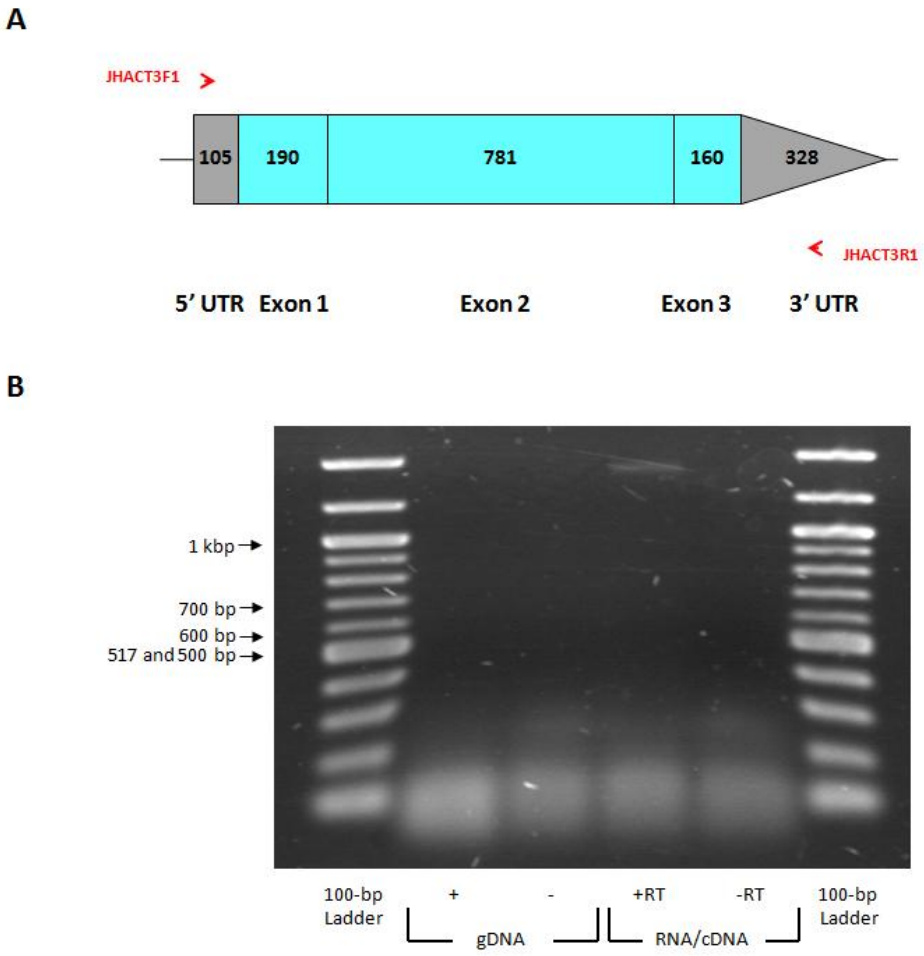


Figure 3.16. ***act-3*** **primer specificity and analysis of amplimer size.** PCR was performed using the JHACT3F1 and JHACT3R1 primers. **(A)** Schematic depicting annealing sites for these primers, in the 5' and 3' UTRs. **(B)** PCR was performed using these primers under the following conditions: Stage 1: 94°C, 3 minutes; Stage 2: [94°C, 1 minute; 60.0°C, 1 minute; 72°C, 3 minutes] x30; Stage 3:72°C, 10 minutes. An amplimer of 1395 bp was predicted to be generated from cDNA and is faintly visible from +RT reactions here. No amplimer was expected from gDNA.

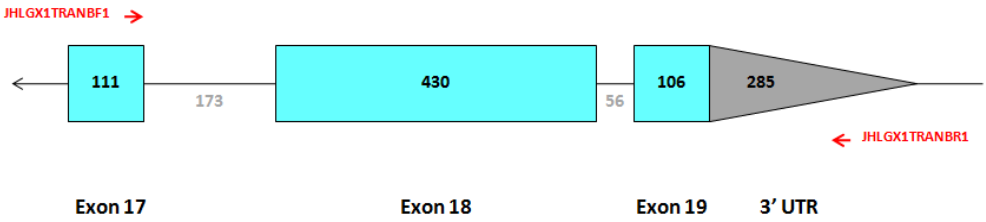


Figure 3.17. C54G7.3 transcript b primer annealing sites. Schematic depicting annealing sites for the JHLGX1TRANBF1 and JHLGX1TRANBR1 primers in transcript b.

TABLES

Table 3.1. Unpublished references to the F48E3.8 and C54G7.3 genes

Web Source	Methods & Results
C. elegans Promoter Database (CEPDB) *The website lists promoter sequences for a number of different genes.* General Homepage: http://rulai.cshl.edu/cgi-bin/CEPDB/home.cgi	Website includes a listing for a promoter for the C54G7.3/*lgx-1* gene. The listed sequence does correspond with the canonical promoter of *lgx-1* based on my cross-checking. Specific URL: http://rulai.cshl.edu/cgi-bin/CEPDB/AcePerl-1.87/seqget The sequence listed in the database : tcgtatgcaaaccaaatttacccttgtctctgaaaaaaattgtcaagttttaagatttgagcttcttga catattattattttggaataatgcatagtttttctcaatacacttggcgaaaaccttaaagactcccca ttcaaactatctgacttatcggttaccactcatttgaccggtgctctttttaatgtgatccgccctattg cacaagtacaagtcatcttgaggcacagccgagatttgtaaacatgttgagtatgcgcaccgtaa aatgacttcaatagtacttgatactcatgcacttgcttccctttttcaattcgttgtgacttgaccacaa catctcacttccgtatttgaaattgtgctcagaagtagaattactccttattattatttttatagatttatt gtaaactgcaataccctattaaaaaatttcaggctcgccaacgacagaagatctagcgggaaca atcagcttcaacaacataacgatgcgagtgcttctcgctatttcagttatctatcacataactatagg tgagttttgtagtaatttattttaacaatgcacattaaatttttagctcaacaa
www.microRNA.org	cel-miR-231 targets F48E3.8 *No information on experimental verification or method is provided with this information.*
Biological General Repository for Interaction Datasets (BioGRID) General Homepage: thebiogrid.org	CYK-4 (K08E3.6) interacts directly with F48E3.8 Specific URL: http://thebiogrid.org/41980/summary/ caenorhabditis-elegans/cyk-4.html
Nematode Expression Database (NextDB) General Homepage: http://nematode.lab.nig.ac.jp/	Clone 272e1 provides expression for F48E3.8 and shows faint expression in the presumptive pharyngeal primordium of embryos http://nematode.lab.nig.ac.jp/db2/ShowCloneInfo.php?clone=272e1 Clone 208c10 provides expression for C54G7.3 and shows pharyngeal staining in all stages, although intensity is lower in embryos http://nematode.lab.nig.ac.jp/db2/ShowCloneInfo.php?clone=208c10

Chapter 4

Functional and Biochemical Characterization of *C. elegans* Chitin Synthases

ABSTRACT

UDP-N-acetylglucosamine is converted to chitin through the action of chitin synthases. In nematodes, the two chitin synthase genes *chs-1* and *chs-2* have been characterized and they are known to function non-redundantly to deposit eggshell chitin and pharyngeal chitin respectively. Their temporal expression patterns are consistent with these functions. In this chapter, I report the use of the RT-PCR-based temporal analysis of gene expression patterns to better understand the onset of *chs-1* expression in the germline and to determine whether either of the two genes is expressed in other somatic structures, in the dauer larvae, or in the post-molting adult. I employ a radioactive substrate incorporation assay to study whether properties of the chitin synthase enzymes in *C.elegans* are similar to the biochemical properties reported in studies from other species. Since most other studies have focused on the more abundant germline isoform CHS-1, I take advantage of germline-ablated strains of *C. elegans* to additionally investigate the properties of the somatic isoform CHS-2. I also discuss attempts to use knockouts of the *chs-1* gene to address our unpublished results that chitin may be present in the *C. elegans* vulva.

INTRODUCTION

Chitin synthases polymerize chitin, β-1,4-linked-N-acetylglucosamine, using the precursor UDP-GlcNAc. Genes encoding chitin synthases (CHS) have been identified in many nematode species. A single chitin synthase gene was first detected in *Meloidogyne artiella* using Southern Blot analysis (Veronico *et al.*,

2001) and single chitin synthases were identified in the two filarial nematodes *B. malayi* and *D. immitis* along with their two homologs in *C. elegans* (Harris and Fuhrman, 2002; Harris *et al.*, 2000). Subsequent work led to the identification of an additional *B. malayi* chitin synthase homolog along with two from the filarial parasite *O. volvulus* (Foster *et al.*, 2005). It is now a widely regarded fact that all nematodes possess two *chs* genes. In nematodes where the presence of two *chs* genes has been established, and where the expression and function of these genes has been investigated, it is now known that the genes do not function redundantly. The CHS-1 protein polymerizes chitin necessary for the chitinous eggshell and the CHS-2 protein produces chitin necessary to line the pharynx.

chs-1 is expressed in the nematode germline. The orthologs *Bm-chs-1* and *Ov-chs-1* are expressed in gravid and non-gravid female adults as determined using RT-PCR and *in situ* hybridization (Harris and Fuhrman, 2002; Harris *et al.*, 2000). The *Bm*-CHS-1 protein is also detected in oocytes, embryos and microfilariae in *B. malayi* (Harris *et al.*, 2000). Work in *C. elegans* has established non-redundant patterns of expression and function for the two nematode chitin synthases. When studied in a synchronous population of worms, expression of the *Ce-chs-1* begins at a low level following the L2/L3 molt and gradually increases to its most abundant levels by the start of adulthood (Veronico *et al.*, 2001). This is consistent with expression during the production of oocytes in L4 worms which continues throughout adulthood as oocytes are continuously generated. Attempts to drive GFP expression under the control of the *Ce-chs-1* promoter have not been successful (Zhang *et al.*, 2005); however, GFP-tagged

Ce-CHS-1 protein is detected in oocytes prior to fertilization (Maruyama *et al.*, 2007). Loss of *chs-1* activity via RNAi results in a reduction in brood size for *C. elegans* (Fanelli *et al.*, 2005) owing to a disruption of eggshell integrity and consequently the inability to support the development of embryos (Zhang *et al.*, 2005). In *M. artiella*, the single *chs* gene that was first described appears to be more closely related to *Ce-chs-1* than to *Ce-chs-2* and RNAi against *Ma-chs-1* delays hatching of juveniles and causes a reduction in eggshell chitin measurable by staining with calcofluor. A knockout of the *chs-1* gene has been generated in *C. elegans*. Homozygotes of the *chs-1* deletion (*chs-1Δ*) are sterile owing to improperly formed eggshells and failed embryonic development (Johnston *et al.*, 2006). The CHS-1 protein, therefore, is termed the "eggshell" chitin synthase.

In *C. elegans*, the *chs-2* gene is expressed in punctuated bursts prior to each of four molts, suggesting that the gene is expressed to re-synthesize somatic chitin lost with each molt (Veronico *et al.*, 2001). When the *chs-2* promoter is used to drive expression of GFP (Zhang *et al.*, 2005) or a β-galactosidase-GFP fusion protein (Veronico *et al.*, 2001), expression of the gene is observed in the *C. elegans* pharynx prior to the L1/L2 molt. The spatial and temporal expression pattern for *Ce-chs-2* suggest that it is the gene responsible for producing pharyngeal chitin, and specifically that it replaces the lining of the pharynx which is shed along with the cuticle at each larval molt. Loss of *chs-2* (using RNAi by feeding) in *C. elegans* results in a delay in development (Fanelli *et al.*, 2005). When gene expression is disrupted using RNAi by injection, *C. elegans* larvae arrest at the L1/L2 molt, consistent with the gene's role in replacing the

pharyngeal lining (Zhang *et al.*, 2005). The CHS-2 protein, then, is termed the "somatic" chitin synthase.

Fanelli *et al.* (2005) note that in experiments targeting *Ma-chs-1*, *Ce-chs-1* and *Ce-chs-2* using RNAi, the effects can be attributed directly to a loss of the specific target transcripts since transcript abundance decreases, when measured using qRT-PCR.

While much of the work on the nematode chitin synthases has focused on the functional role of the two chitin synthases *chs-1* and *chs-2* and the molecular properties associated with their regulation, only limited information regarding the biochemical properties of the CHS-1 and CHS-2 isoforms has been obtained. The chitin synthase enzymes are multi-pass transmembrane proteins that create a channel through which chitin, polymerized by the cytosolic-facing catalytic domain, is extruded into the extracellular space. Much of our understanding of the kinetics and mechanism of enzyme function comes from parasitic nematodes. In particular, the pig intestinal parasite *Ascaris suum* has provided much of the biochemical data (Dubinsky *et al.*, 1986a; Dubinsky *et al.*, 1986b); the large size of the worm is ideal for the isolation of abundant protein quantities. Subsequent work has elucidated the properties of the chitin synthase activity in the plant parasite *Meloidogyne artiella* (Fanelli *et al.*, 2005). In these studies, crude protein extracts are used as the source for the chitin synthase enzymes when substrate incorporation (using radioactive labels) is used to measure the catalytic properties that are normalized to time and total protein content. Importantly, both studies focus on the "eggshell" chitin synthase CHS-1 since this isoform is far more

abundant in protein extracts, and chitin is produced at far greater quantities in the germline than in the soma (**Table 2.3**). Dubinksy *et al.* (1986) investigated the effects of co-factors, trypsin treatment and substrate availability (among other variables) and Fanelli *et al.* (2005) conducted similar experiments using *M. artiella* egg extracts to investigate the effects of trypsin treatment and the presence of nikkomycin, an inhibitor of chitin synthases that was previously characterized in insects.

The activity of CHS isoforms results in the deposition of chitin in the two well-established sites. Evidence of chitin in the nematode eggshell was first provided using the nematode *Aspiculuris tetraptera* and observed with the aid of electron microscopy (Wharton, 1979) and its importance as a protective covering, by virtue of its biochemical properties, was quickly established (Wharton, 1980). Subsequent work has shown that chitin is present in very diverse nematodes including *Heterodera glycines* (Burgwyn *et al.*, 2003). The *M. artiella* eggshell stains vividly with calcoflour (a dye that binds to many highly repetitive, large molecular weight carbohydrates including cellulose) and that staining can be ablated through RNAi against the *Ma-chs-1* gene (Fanelli *et al.*, 2005), supporting the presence of chitin (and not another carbohydrate) as the epitope in the nematode eggshell.

In some filarial parasites, the chitinous eggshell undergoes a dynamic transformation that leads to its elongation into a protective sheath around the microfilariae. There is no established report of additional chitin synthase activity to create this structure. This structural transformation in the sheathed filarial

parasites was observed for *B. pahangi* (Rogers *et al.*, 1976), *Acanthocheilonema viteae* (Ellis *et al.*, 1978), *Litomosoides carinii* and *B. malayi* (Schraermeyer *et al.*, 1987).

The presence of chitin in the nematode pharynx was also first demonstrated using a parasitic roundworm. Using electron microscopy, chitin was observed in the pharynx of *Oesophagostomum dentatum* (Neuhaus *et al.*, 1996). The localization of chitin in both the eggshell and pharynx have been more recently confirmed in *C. elegans* using highly specific chitin-binding probes from bacterial chitinases (Zhang *et al.*, 2005).

A cohesive framework for the non-redundant use of *chs-1* and *chs-2* genes to regulate chitin deposition in all previously documented localizations of the carbohydrate has been developed. Here, I report attempts to use methods previously applied to the study of the chitin synthases to further explain expression patterns for the genes and determine whether these expression patterns give insight into novel functions. I discuss attempts to use substrate incorporation assays to study the properties of the *C. elegans* chitin synthase isoforms, using germline-ablated strains to isolate proteins extracts where CHS-2 activity could be studied in the absence of CHS-1. Finally, I report our attempts to further characterize a novel pattern of chitin deposition at the *C. elegans* vulva, use a *chs-1* knockout to ask whether ablation of the gene disrupts the staining pattern we see in the hermaphrodite sex structure.

MATERIALS AND METHODS

Strains

The wild-type *C. elegans* Bristol N2 strain, the SS104 strain (carrying the temperature-sensitive *glp-4* allele, bn2), the NL2099 strain (carrying the *rrf-3* allele, pk1426), the JK560 strain (carrying the *fog-1* allele, q253), the MT1007 strain (carrying a missense *lin-24* allele, n432), the BA17 strain (carrying the *fem-1* allele, hc17) and the RB1189 strain (carrying a *chs-1* partial deletion, ok1120). The *chs-1* deletion in RB1189 was generated by the *C. elegans* Gene Knockout Consortium but sequenced and characterized in Zhang *et al.* (2005).

Worm Growth and Sample

Worm growth and sampling of worms were performed as previously described [Chapter 2]. Attempts to synchronize and harvest worms from the MT1007 strain carrying a *lin-24* mutant allele are summarized in **Table 4.1**.

RT-PCR

RT-PCR was performed as previously described (Chapter 2 and Chapter 3). Additional information on all primer sequences used in these studies is available in **Appendix A**.

RNAi

A *chs-1* amplimer generated using the primers JHCHS1F1 and JHCHS1R1 was cloned to T vector before liberation with BamHI and EcoRI ends

and sublcloning to the Litmus v28i Vector (NEB) (**Figure 4.3**). The presence of *chs-1* sequence in the Litmus 28i vector was confirmed through sequencing prior to using this construct for the generation of dsRNA.

RNAi experiments were performed as described previously (Chapter 2).

Chitin Synthase Assays

Chitin synthase assays were performed as previously described by Dubinsky *et al.* (2009).

Chitin Detection

Without sorting, worms from the RB1189 strain were purified by sucrose flotation, fixed in ethanol and stained with Wheat Germ Agglutinin at 2 µg/mL in PBS.

RESULTS AND DISCUSSION

Germline expression patterns. Veronico et al. (2001) established an RT-PCR-based time course to investigate the expression pattern of the *C. elegans* chitin synthases *chs-1* and *chs-2*. In their results, *chs-1* expression is seen beginning at 18 hours, the time designated at the L2/L3 molt. Since oocytes do not become visible in these population until L4s (that are first seen at 24 hours), the early burst of *chs-1* expression was not fully consistent with a germline expression pattern. We reasoned that germline expression could begin earlier than the onset of oocyte production or that *chs-1* expression was associated with

activity in a different tissue or the alternative dauer life stage (if growth conditions during their experiment were not ideal). I confirmed that *chs-1* expression could be seen as early as the 18-hr time point when the experiment was performed with worms grown on solid media (**Figure 3.12B**). This *chs-1* expression was not associated with dauer larvae since food was not depleted and since *col-2*, a collagen highly upregulated in dauer larvae (Cox and Hirsh, 1985), was not detected in this trial (**Figure 3.12B**). Recent work has shown that the CHS-1 protein found at the surface of *C. elegans* oocytes is associated with the EGG-3 protein (Maruyama et al., 2007). In order to determine whether the early *chs-1* expression was associated with a pre-oocyte germline, I assayed for *egg-3* expression in the same time course. Like *chs-1*, *egg-3* expression begins after the L2/L3 molt and suggests that active expression of oocyte genes may begin this early. This conclusion is confirmed by data available from the Nematode Expression Database (NEXTDB, http://nematode.lab.nig.ac.jp/) which shows that *chs-1* expression measured by *in situ* hybridization is seen in the gonad arms in L3 stages, before the production of oocytes (cosmid T25G3 clone 270c1, http://nematode.lab.nig.ac.jp/db2/ShowCloneInfo.php?clone=270c1).

Expression in dauer larvae. I made attempts to synchronize N2 *wild-type* worms under conditions of limited food to generate a dauer time course that would permit an analysis of chitin synthase (and chitin deacetylase) expression during dauer formation. Attempts at this time course did not yield enough RNA for reliable RT-PCR experiments. Recently, a genome-wide analysis of gene expression upon dauer entry has been published, with results made accessible

through an online repository (Jeong *et al.*, 2009). Findings from these experiments show a robust upregulation of the *chs-1* gene in dauer larvae (http://dauerdb.org/view_ma2.php?mode=search&search_cls=total&txtSearch=T25G3.2). Two possibilities follow from this observation: either *chs-1* expression is tied to the generation of a novel chitinous dauer structure, or, initiation of certain steps in germline development occur in preparation for the return to normal growth when mating would be favorable. Although no conclusions can be drawn about the localization and function of this expression, further work will be necessary to determine whether *chs-1* is upregulated to create a novel chitinous structure in the dauer. The accuracy of the Dauer Metabolic Database in providing useful information regarding the upregulation of genes in dauer larvae can be assessed by looking at the expression patterns presented for other genes which have previously been shown, through empirical work, to be upregulated in the dauer. In that respect, *col-2* may serve as a positive control for interpreting data from the database. *col-2* expression is elevated upon entry to dauer (http://dauerdb.org/view_ma2.php?mode=search&search_cls=Gene_Title&txtSearch=col-2). The possibility that CHS-1 may function outside the germline, and perhaps in the dauer larvae, is reinforced by recent observations from *B. malayi*. Bennuru *et al.* (2011), in a stage-specific analysis of the *B. malayi* proteome, have reported that Bml_35215 (CHS-1) is detected in various life stages. *Bm*CHS-1 was detected in uterine microfilariae (UTMF), released microfilariae (mf), infective L3 larvae and in adults (both male and female). The prevalence of this protein in various life stages and in males is surprising given the gene expression

and protein function characterized for CHS-1 in *C. elegans*. The presence of the *Bm*CHS-1 protein in infective L3 larvae again opens the possibility that CHS-1 may be utilized for a process distinct from the presence in oocytes and, as such, in the soma. However, the key to understanding the significance of this CHS-1 protein presence (and the burst of *chs-1* expression in *C. elegans* dauer larvae) will be to study germline development in the dauer, asking whether the normal process of germline development begins in both L3 and L3d in *C. elegans.*

The dauer cuticle (presumably optimized to reduce desiccation, among other things, in worms facing adverse conditions) and a pharyngeal plug are two places where chitin may be present but not previously identified. It is noteworthy that *pha-4* expression is upregulated in dauer larvae, its levels gradually rising during the transition to the dauer state and peaking at 72-hr when one of the methods for generating dauer are employed (http://dauerdb.org/view_ma2.php?mode=search&search_cls=Gene_Title&txtSearch=pha-4). Since I have identified numerous putative PHA-4 binding sites in the upstream regulatory regions of both *chs-1* and *chs-2*, it is possible that the PHA-4 transcription factor may activate one or both of these enzymes leading to the synthesis of chitin in the pharynx (**Figure 4.1**).

Vulva morphogenesis. Unpublished results from our lab, suggest that chitin is present as a lining to the *C. elegans* vulva. I attempted to support these findings using the RT-PCR time course. The MT1007 strain carries a *lin-24* mutation that stochastically disrupts vulva formation in some worms from a population of the worms grown under normal conditions. I also attempted to

synchronize this strain to ask whether loss of the vulval cells resulted in observable differences in the patterns of *chs-1* or *chs-2* expression. Efforts to synchronize these worms failed, however, since the absence of the vulva generates a "bag of worms" phenotype. In this case, large numbers of progeny hatch inside a parental hermaphrodite rather than being held or released with a chitinous protective covering. During normal synchronization, bleach treatment destroys parental tissue and releases the embryos that are covered with a chitinous eggshell. In the "bag of worms", the F1 are not protected by an eggshell upon bleach treatment and cannot be synchronized. Efforts to modify the protocol to permit synchronization of *lin-24* worms were unsuccessful (**Table 4.1**).

Trace levels of *chs-1* expression are detected in males of *B. malayi* and *D. immitis* when assayed using RT-PCR (Harris and Fuhrman, 2002; Harris *et al.*, 2000) and the *Bm*-CHS-1 protein has been detected in adult males (Bennuru *et al.*, 2011). Since males lack oocytes – the only previously established site for *chs-1* expression – it is possible that CHS-1 could have an as-yet uncharacterized function, like acting at the sex structures. The *in situ* hybridization data available for the *chs-1* gene (NEXTDB, cosmid T25G3 clone 270c1) does not reveal expression at the vulva in any life stage. It is noteworthy, however, that *chs-2* expression by these *in situs* is not limited to the pharynx, the only previously established site for *chs-2* expression. *chs-2* is also detected in the region of the developing vulva and in a large region on the dorsal side of L3 and L4 worms (cosmid F48A11 clone 316g4, http://nematode.lab.nig.ac.jp/db2/ShowCloneInfo.php?clone=316g4). We cannot

rule out the possibility that non-specific binding is being observed here, but these results may suggest additional somatic expression of *chs-2* in, among other places, the developing sex structures. The chitin we observe at the vulva may be deposited by *chs-2* and not *chs-1*.

Post-molting adult growth. In their analysis of the temporal expression of various cuticular collagens, Johnstone and Barry (1996) note that in addition to peaks of expression during each larval stage – necessary for the creation of the cuticle that will replaced at each molt – an extra burst of expression is seen for some (but not all) collagens, such as *col-12* and *sqt-1*. Although no functional work has elucidated a role for these genes in the adult, the authors suggest they may be expressed for post-molting growth since the adult increases significantly in size and its cuticle is enlarged accordingly. No studies have previously explored the role of the chitin synthases in post-molting growth, despite the fact that the pharynx – which is lined with chitin – increases in length in adults, like the cuticle. Chitin synthases may also be reactivated in response to damage. In additional trials of the RT-PCR time course analysis, the SS104 strain (carrying a *glp-4* mutant allele) was grown under non-permissive conditions in liquid culture. These conditions restrict production of the germline. I observed *chs-2* expression in adult stages of the worms (**Figure 2S.12**) and reasoned that *chs-2* expression may continue in adult worms, contrary to results presented by Veronico *et al.* (2001), to enable growth or repair of the adult pharynx. I made various attempts at a detailed expression time course similar to those reported in **Figure 3.11** or **Figure 3.12**, using germline-ablated SS104 worms. However, I was unable to

establish populations of high enough density to harvest enough worms for sufficient RNA yield. In order to confirm adult expression of *chs-2*, I used the JK560 strain grown under non-permissive conditions such that the *fog-1* mutant allele limits production of sperm but not oocytes. Again, *chs-2* expression was observed (**Figure 4.2**) supporting an adult expression of the *chs-2* gene. However, when additional samples from the same time course as those depicted in **Figure 2S.12** were used for RNA extraction *including a DNAse treatment to more accurately quantify the nucleic acids*, higher levels of *chs-1* were observed (**Figure 2S.13**). The conspicuous levels of *chs-1* suggest that germline restriction is not fully penetrant and the presence of *chs-2* in the presumptive germline-restricted post-molting adults of the SS104 strain (**Figure 2S.12**) may reflect expression from F1 worms. Further work will be necessary to investigate the possibility that *chs-2* is reactivated in adults.

Overall assessment of the RT-PCR time course method. In our temporal analysis of gene expression, we have been unable to re-create the punctuated burst of activity prior to each molt which has been previously reported for the *chs-2* gene (Veronico *et al*., 2001). Our results for *chs-2*, as well as the two chitin deacetylases which we have recently identified, are more similar to the expression patterns observed for the collagens (including *col-12*) when this method was first employed to study their temporal expression patterns (Johnstone and Barry, 1996). Similar to the results for collagens, we find that all the genes implicated in somatic chitin metabolism appears to have cyclical waves of intensity that appear to follow at intervals coincident with the molting pattern of

four larval stages preceding adulthood (**Figure 3.12**). It remains possible that these genes do, in fact, have discrete expression patterns that are irresolvable within this population-based approach as explained by Johnstone and Barry (2006). It has been previously hypothesized that all collagens (including *col-12*) may be targets of one the terminal terminal transcription factors of the heterochronic gene pathway which regulates cyclical waves of gene expression during larval molting. One such transcription factor is LIN-29 (Liu *et al.*, 1995). It should be interesting to determine, then, whether *chs-2* (and perhaps the chitin deacetylases F48E3.8 and C54G7.3) is a target of this regulatory pathway. Previously established methods exist to determine whether a transcription factor such as LIN-29 binds to specific regulatory regions. It should be possible to determine whether amplified ~2 kb regions of the *chs-2* promoter are bound by LIN-29 causing reduced mobility in gels when analyzed *in vitro*. An alternative approach would show interaction between this DNA segment and a putative transcription factor in a yeast one-hybrid system. Our efforts to better characterize the role of the *chs-2* gene during development, including attempts to understand novel functions, will benefit from the development of temperature-sensitive alleles of the gene.

Biochemical properties of the CHS isoforms. I have been unable to detect chitin synthase activity in *wild-type*, *glp-1* (lacking sperm and eggs) and *fem-1* (lacking sperm) worms: although I have detection of radioactive substrate in my trials, the calculated values are never higher than matched no substrate controls. If these trials had provided positive results, it would enable to us to

assign levels of activity to the combined activity of CHS-1 and CHS-2 (*wild-type*), to look at the activity of the CHS-1 enzyme in pre-fertilized oocytes along with that of CHS-2 (*fem-1*) and to look at the activity of CHS-2 alone. Hence, we would be able to deduce the contributions of the pre-fertilization and post-fertilization CHS-1 enzymes and also the somatic CHS-2.

In studies of chitin synthase enzymes obtained in crude protein extracts from *A. suum*, the highest level of activity detected (under a variety of conditions) is 0.025 pmol GlcNAc incorporated/min/mg protein (Dubinsky *et al.*, 1986a). Fanelli *et al.* (2005) report much higher values from a similar approach applied in *M. artiella*, observing activity within the range of 180 – 5800 pmol GlcNAc incorporated/min/mg protein. The latter results are an order of magnitude greater than those first reported from *A. suum*. However, Fanelli *et al.* (2005) never detail the inclusion of matched no substrate controls and it may be necessary to re-visit the accuracy of these values. In both *A. suum* and *M. artiella*, trypsin treatment of the crude protein extracts was found to increase the activity of the enzymes using this assay. This is taken as a conclusion that the enzymes exist as zymogens that are cleaved and activated by trypsin treatment. While this is a reasonable conclusion it is not exhaustive and trypsin treatment may also destroy any proteins that associate with the CHS proteins. Trypsin treatment may activate the CHS enzymes by destroying an inhibiting protein rather than altering the form of the CHS peptide sequence directly.

Bennuru *et al.* (2011) identified CHS-2 protein in the UTMF but not other life stages (mf, infective L3 larvae, adult female, adult male). *Bm*CHS-2 is

presumed to mediate pharyngeal chitin deposition in *B. malayi*. If the protein is present only briefly – to synthesize a new pharyngeal lining prior to each molt before it is rapidly degraded – it may explain the the failure to detect CHS-2 protein in the stage-specific *B. malayi* proteomic analysis. Alternatively, this failure at detection may simply reflect the fact that the protein is present at very low levels (even if it has a longer lifetime) or that it may be difficult to isolate as a membrane-embedded protein.

Detection of chitin at the* C. elegans *vulva. Previous work from our lab has shown that various stains that bind to chitin in the *C. elegans* eggshell and pharynx, also label the *C. elegans* vulva. This observation suggests that chitin may line this structure. As an additional step to confirm the presence of chitin at the vulva, we asked whether ablation of one or both *chs-1*or *chs-2*, using RNAi, could disrupt the staining pattern. A *chs-1* amplimer was successfully cloned to the Litmus v28i vector enabling me to generate dsRNA that could be used to silence the *chs-1* gene (**Figure 4.3**). Similar attempts to clone a *chs-2* fragment were unsuccessful.

When parental worms are fed *E. coli* expressing *chs-1* dsRNA, brood size is diminished (**Figure 4.4**). This result is similar to the results reported for *Ce-chs-1* RNAi by microinjection of dsRNA to the gonads (Zhang *et al.*, 2005). This construct can be used then to silence *chs-1* expression to ask whether the gene has an important role in deposting chitin at the vulva.

A knockout (KO) has been generated for *C. elegans chs-1* but not *chs-2*. This KO has been instrumental in dissecting the role of CHS-1 in synthesizing

eggshell chitin and, as such in providing an protective barrier (Zhang *et al.*, 2005). The region of the CHS-1 protein removed is summarized in **Figure 4.5**. The *chs-1*Δ deletion allele is carried in a strain against a lethal balancer chromosome. In populations of the RB1189 strain, *chs-1*+/*chs-1*+ worms are homozygous for this lethal chromosome and are continuously removed from the population. Remaining worms are heterozygous *chs-1*+/*chs-1*Δ or homozygous *chs-1* Δ /*chs-1*Δ. In heterozygotes, the single copy of *chs-1*+ is haplosufficient and progeny with normal eggshells are formed (**Figure 4.6A**) while the homozygous null mutants cannot form progeny with normal eggshells so that they are sterile (**Figure 4.6C**). The absence of eggs allows us to determine which worms are fully *chs-1 null* and, as such, are the worms whose vulval staining pattern are revealing about the putative role for of CHS-1 in vulval chitin. We find that for both heterozygotes and homozygotes of the *chs-1 null*, the staining pattern with Wheat Germ Agglutinin is maintained. Hence, if chitin is present at the vulva, its localization cannot be attributed to CHS-1 alone (**Figure 4.6B,D**).

FIGURES

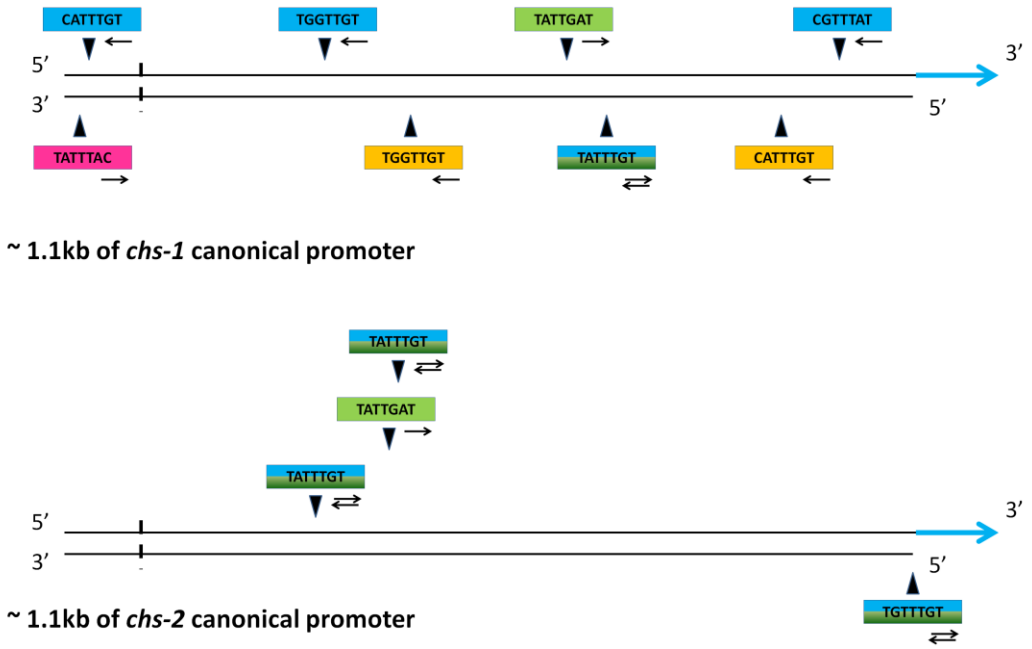


Figure 4.1. Analysis of putative PHA-4 binding sites in the canonical promoters of the *chs-1* and *chs-2* genes. Six putative PHA-4 binding sites are observed within the terminal ~1 kb sequence of the canonical *chs-1* promoter while four PHA-4 binding sites are observed within the terminal ~1 kb sequence of the canonical *chs-2* promoter. Bi-directional PHA-4 recognition sites are seen in the promoters for both these genes.

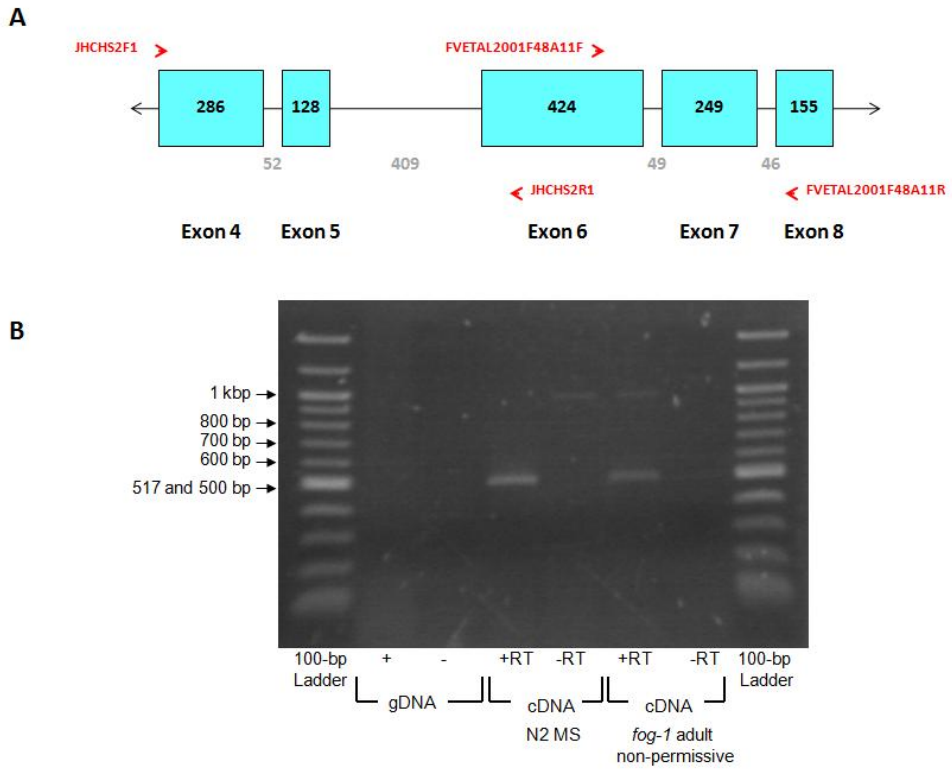


Figure 4.2. *chs-2* is detected in adults of the JK560 strain which produce oocytes but not sperm. 2 µg of total mixed-stage JK560 RNA (without DNAse treatment) was used for a 40 µL synthesis reaction that was ultimately diluted to 100 µL with cDNA representing expression in a strain producing oocytes but no sperm (and hence not undergoing self fertilization). The expression of *chs-2* in post-molting adult worms was tested using PCR with this cDNA template. **(A)** Schematic showing annealing sites for the JHCHSF1 and JHCHSR1 primers that were used to test for expression. **(B)** *chs-2* amplimers of the appropriate size are obtained from N2 MS *wild-type* cDNA and from the *fog-1* adults. The latter suggests that when an F1 generation is not being produced, *chs-2* expression is seen in post-molting adults and suggests expression in the parental generation. No amplimer is detected from gDNA, but this likely reflects a problem with the template, since gDNAa amplimers are also detected from the two cDNA samples (since no DNAse treatment was done on the isolated RNA).

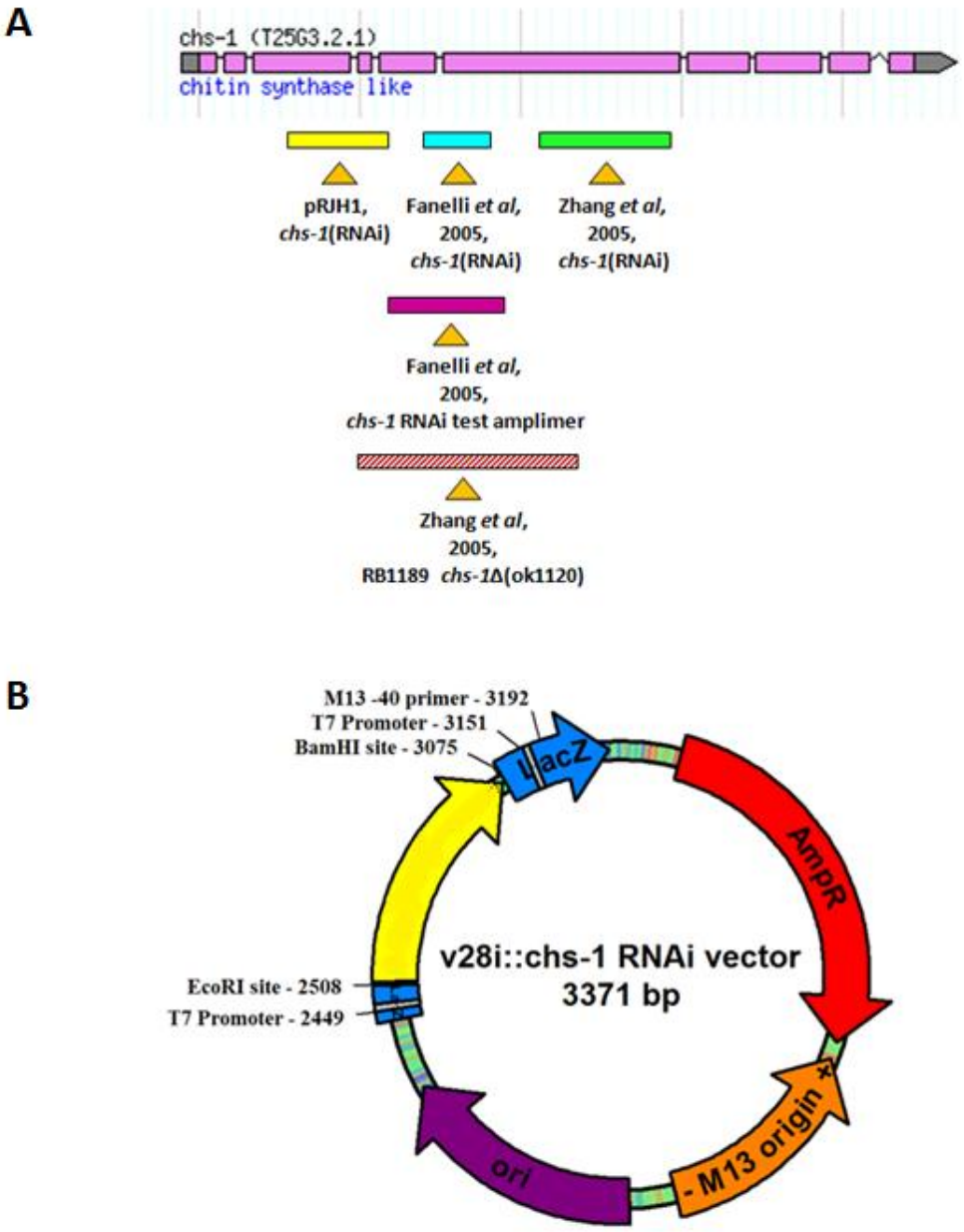


Figure 4.3. A construct bearing a *chs-1* insert in the Litmus 28i vector can be used to generate ssRNA probe targeting the product of this gene *in vivo*. (A) Various regions of the *chs-1* gene have been targeted in previously published RT-PCR and RNAi experiments and with the generation of the knockout. cDNA from the *chs-1* gene was amplified for the generation of a construct (pRJH1) and the corresponding genetic region is shown. **(B)** This cDNA fragment was cloned to a Litmus v28i vector and is depicted as a yellow insert. The presence of the *chs-1* insert in the vector was confirmed through sequencing.

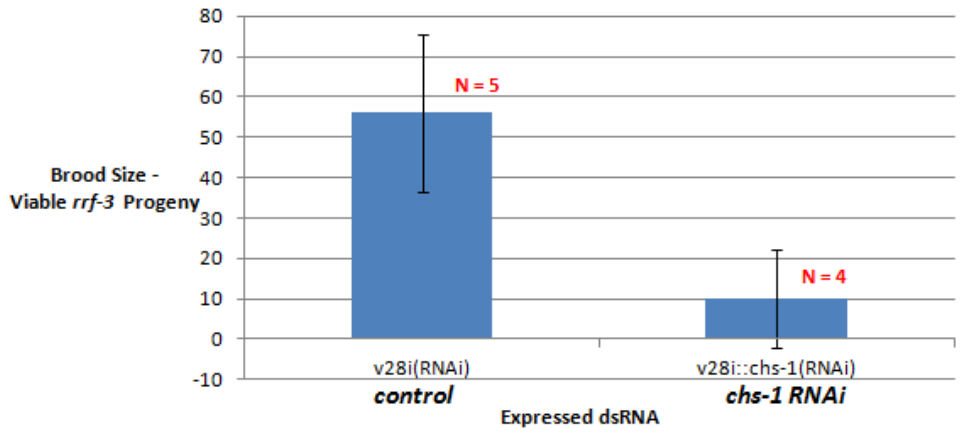


Figure 4.4. dsRNA introduced by feeding disrupts the *chs-1* gene in the NL2099 strain resulting in a decrease in brood size. Feeding dsRNA targeting the *chs-1* gene in parental worms of RNAi-hypersensitive *rrf-3* worms results in a significant reduction in brood size. Control worms receive dsRNA that would be generated from the v28i construct that bears no insert from the *chs-1* gene.

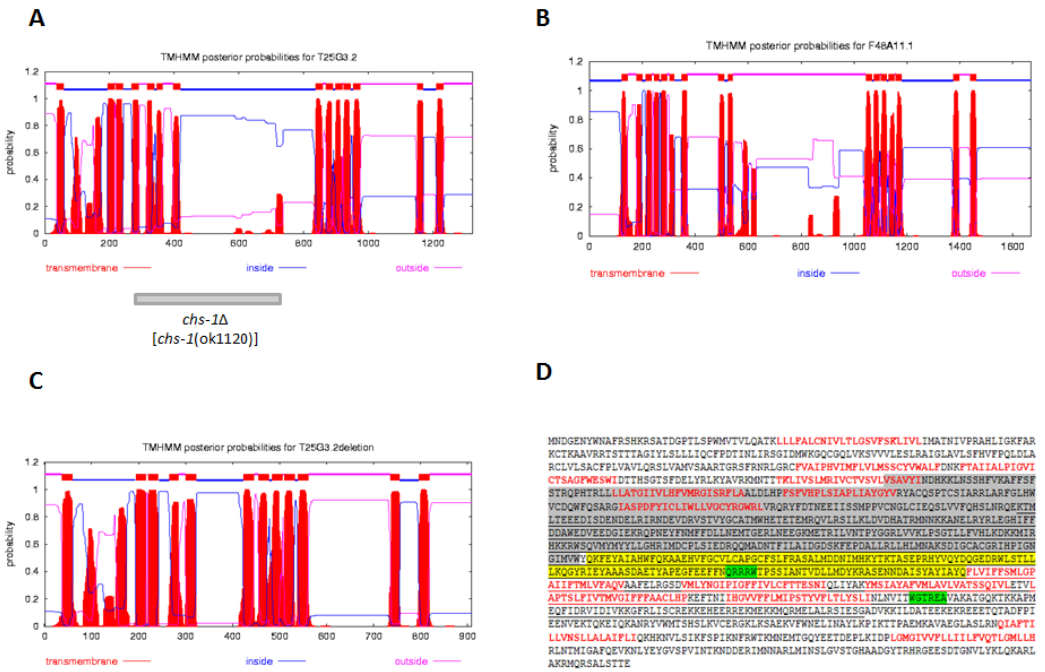


Figure 4.5. The CHS-1 and CHS-2 proteins are multipass transmembrane proteins with a cytosolic-facing catalytic site and the ok1120 deletion allele of *chs-1* produces a non-functional protein. (A) Hydrophobicity plot for the *C. elegans* CHS-1 protein. Multiple transmembrane domains are shown as red peaks. The RB1189 strain carries a deletion allele of the *chs-1* gene. The region of the protein corresponding to the in-frame deletion is highlighted by the grey box below. **(B) A** hydrophobicity plot showing the similar topology observed for the *C. elegans* CHS-2 protein. **(C)** Hydrophobicity plot of the protein sequence that should be generated for the *chs-1* deletion. **(D)** The CHS-1 protein sequence is presented. Residues in red font are thirteen transmembrane sequences equivalent to the peaks in the CHS-1 hydrophobicity plot in [A]. The sequence deleted in the ok1120 allele is highlighted in a grey box (*chs-1*Δ in [A] and is distinct from the catalytic domain that is highlighted in yellow. Two stretches of residues have been noted as strongly conserved when compared to those of other species: QRRRW is highly conserved among fungi and animals while WGTREA is conserved among animals. The *chs-1*Δ does not directly remove the catalytic domain but is likely to produce a protein isoform with major misfolding (that may also be misdirected in the cellular excretory pathway) which explains the loss of activity for the ok1120 allele.

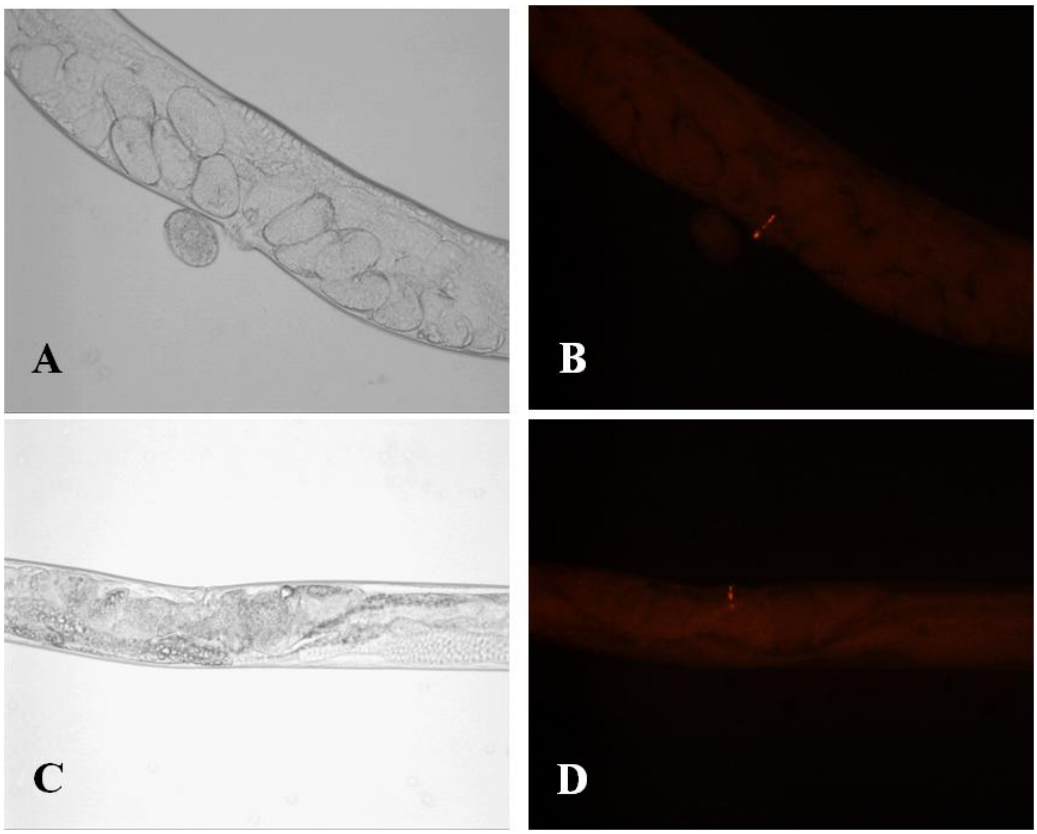


Figure 4.6. Ablation of CHS-1 activity in homozygous RB1189 worms does not disrupt chitin staining at the vulva. Worms were purified from liquid culture using sucrose flotation, fixed using ethanol and stained with Wheat Germ Agglutinin at 2 μg/mL in PBS. No attempt was made to distinguish worms on the basis of genotype prior to staining. The worms analyzed by microscopy were either *chs-1Δ* homozygotes or heterozygotes while chitin synthase +/+ worms would be eliminated by homozygousity of a larval lethal allele carried on the GFP-encoding balancer chromosome. The worms showing no GFP in the pharynx were homozygous for *chs-1* loss and the worms used as the basis for conclusions on the role of the gene at the *C. elegans* vulva.

TABLES

Table 4.1. Attempts to synchronize a strain carrying a *lin-24* mutant allele

Protocol	Observation at 0 hrs	Observation at 16 hours
Normal Procedure for worms that do not display a " bagging" phenotype		
Monoxenically-grown worm culture treated with bleach solution (3 minutes), washed, resuspended without food	All worms dead Eggs visible in suspension or in carcasses	Synchronized, arrested culture of L1 worms
Modified protocols		
Monoxenically-grown worm culture treated with bleach solution (**5 minutes**), washed and resuspended without food	All worms dead (desired)	All worms still dead (not desired)
Monoxenically-grown worm culture treated with bleach solution (3 minutes), **then sonicated**, washed and resuspended without food	All worms dead (not desired)	All worms still dead (not desired)
Monoxenically-grown worm culture treated with bleach solution, **then immediately sonicated**, washed and resuspended without food	Not all worms dead (not desired)	

Table 4.2. Unpublished references to the *chs-1*, *chs-2* and *egg-3* genes.

Web Source	Methods & Results
Nematode Expression Database (NextDB) General Homepage: http://nematode.lab.nig.ac.jp/	Clone 270c1 provides expression information for *chs-1* which is evidently expressed in the gonads beginning in L4 and continuing to adult worms. It is unclear whether the staining can be seen as early as L3s. This data is available at http://nematode.lab.nig.ac.jp/db2/ShowCloneInfo.php?clone=270c1 Clone 316g4 provides expression for *chs-2* which stains the pharynx, even in embryo; also stains much more than pharynx. This data is available at http://nematode.lab.nig.ac.jp/db2/ShowCloneInfo.php?clone=316g4 Clone 117h10 provides expression information for F44F4.2 and show expression in the germline beginning as early as the worms labeled L3-L4. This data is available at http://nematode.lab.nig.ac.jp/db2/ShowCloneInfo.php?clone=117h10

STATEMENT OF AUTHOR CONTRIBUTION

RJH completed all experiments described in this Chapter with the exception of staining for RB1189 mutants (**Figure 4.6**) that was done by William Einhorn.

Chapter 5

Chitin Synthases and Chitin Deacetylases:

Recent Advances and Future Avenues of Research

Perspective: Chitin Synthases in Nematodes

Roles for the two nematode chitin synthases, *chs-1* and *chs-2*, have been well-established and efforts to understand the molecular and functional properties of these genes have provided many insights into their temporal and spatial regulation. Information on the biochemical properties of the isoforms encoded by these genes is more limited and future efforts should be directed at a deeper understanding of these properties for both the germline and somatic isoforms. Post-translational regulation of these enzymes is, in particular, poorly understood. Recent findings suggest it may be an important feature of chitin synthase regulation and these findings warrant some discussion.

Maruyama *et al.* (2007) demonstrated the physical interaction between EGG-3, a sperm receptor on the surface of the oocyte, and CHS-1. In their model, CHS-1 exists within the membranes of vesicles present in the oocyte cortex. CHS-1 is associated with EGG-3 that is positioned in the membrane. Sperm interaction with EGG-3 causes endocytosis of this receptor and the fusion of CHS-1 to the plasma membrane. CHS-1 is regulated by its presence in a complex. These authors provide a response to Dubinsky et al. (1986) who formulated a hypothesis that the sperm may be activators of chitin synthases. EGG-3, then, is a mediator of this interaction.

Recent work in fungi suggest that chitin synthases and chitin deacetylases interact directly within complexes at the cell membrane/ECM (Gilbert *et al.*, 2012). Here, the authors propose that chitin is synthesized and secreted into the ECM where the close association with CDA proteins allow for chitosan synthesis.

If chitin synthases commonly function in association with other proteins, it may also be worthwhile to revisit some of the early work on CHS activation through trypsin treatment (Dubinsky *et al.*, 1986a; Fanelli *et al.*, 2005). The authors suggest that the increase in activity seen with trypsin treatment suggests that the protein may exist as a zymogen (a precursor of a functional enzyme that may become activated by the removal of a segment of the protein sequence). However, just as trypsin treatment may cleave a zymogen it can also destroy additional proteins that interact with and inhibit the activity of the chitin synthase enzyme. Hence, the mode by which trypsin cleavage contributes to increased activity measured by *in vitro* assays will need further investigation.

Perspective: Chitin Deacetylases in Nematodes

Heustis *et al.* (2012) provides the first comprehensive identification, systemization and characterization of nematode polysaccharide deacetylases. We have evaluated the developmental importance of these genes (**Figure 2.4**) and presented a framework for understanding various possible activities of polysaccharide deacetylases that may be important for the development of free-living and parasitic roundworms (**Table 2.4**). This summary also provides a useful framework for testing various roles *in vivo*.

In our initial attempts to dissect the role of chitin deacetylation, we have focused on the implications for pharyngeal morphogenesis. The first metazoan chitin deacetylases were characterized in *D. melanogaster*, where *serp* and *verm* were found to regulate growth of the embryonic trachea by deacetylation of the

chitinous lining there (Luschnig *et al.*, 2006; Wang *et al.*, 2006). Our attention has been drawn to the question of regulating growth in epithelial tubes, and potential role for chitin deacetylases in regulating this process, due to the structural analogies that can be observed between the *Drosophila* embryonic trachea and the nematode pharynx – although the latter is of far greater complexity. Much work will need to be done to address the extent of direct physical interactions between chitosan and the surrounding ECM or cell membranes of underlying tissue in the pharynx to determine how deacetylation could support structural integrity. The creation of free amine groups opens various possibilities for covalent or electrostatic interactions with surrounding material (**Figure 5.1**). It is possible that deacetylation of the pharyngeal chitin may limit longitudinal growth of the pharynx (as reported in *D. melanogaster*) just as much as it could facilitate coordinated extension of the chitinous lining with its underlying pharyngeal myoepithelial cells (**Figure 5.2**).

These models, presented in **Table 2.4**, though, are not all mutually exclusive. Deacetylation of the pharyngeal lining in *C. elegans* may be integrally tied to pharyngeal morphogenesis but may also be a protective measure to maintain the integrity of the pharyngeal lining against exogenous chitinases. Parasitic roundworms encounter chitinases produced by the innate immune response of both plants and animals. Free-living bacteriovore nematodes like *C. elegans* are exposed to the chitinases produced by their food source.

Given the important role these enzymes can play during nematode development, their activity is bound to be tightly regulated and future efforts can

aim to address the mechanisms for this control. In insects, some but not all CDAs have an LDLa receptor domain present within the protein sequence. *Tc*CDA1-3 all bear this domain (Dixit *et al.*, 2008) and are the enzymes shown to function in the exoskeleton where they help with cuticular molting as *T. castaneum* progress through pupal and larval stages (Arakane *et al.*, 2009). The LDLa receptor domain can bind LDLa molecules that carry small ligands, including hormones. It is possible that this domain allows CDAs in insects to be responsive to insect hormones that regulate molting. In contrast, no LDLa receptor domain has been detected in any nematode PDA sequence (**Figure 2S.1**). If hormone regulation is a feature of the post-translational regulation of the insect PDAs that regulates their activity in time, it encourages the question of temporal regulation in nematode PDAs.

Assessing Prospects for Targeting Chitin Metabolism in Nematodes

Chitinase activity in nematodes was first detected in *Onchocerca gibsoni* (Gooday *et al.*, 1988). The authors were also the first to note the possibility of targeting enzymes of the chitin metabolism pathway for the development of selective interventions against plant and vertebrate parasites. They noted that (secreted) chitinases were likely to be more viable targets than (membrane-embedded) chitin synthases since targeting the latter requires small molecules that can permeate the cell membrane. Subsequent work, though, has shown the chitinases are a large family of enzymes that demonstrate spatiotemporal specialization during nematode development. Moreover, chitinases are produced

in plants and vertebrates and are important components of the innate immune response.

Chitin deacetylases may prove to be better therapeutic targets: these enzymes exhibit low-copy number with one or two genomic homologs seen in the *Nematoda* [Heustis *et. al*., 2012; Chapter 2], they are likely to be secreted proteins [Chapter 3], and are expressed in various stages of development [Chapter 3]. If these enzymes are multi-substrate – acting on various targets *in vivo* – it also suggests that disruption of these enzymes may affect multiple developmental processes and the cumulative effect may result in more rapid parasite loss.

While chitin is not synthesized in all organisms, its importance in biology is underscored in the evolutionary processes that have driven all organisms to produce chitinases. Bacteria employ chitinases for nutrition, as these enable their ability to use chitin as an energy source. Plants produce proteins with chitin-binding properties, such as the lectin Wheat Germ Agglutinin, that can act as a "plant antibodies" when they facilitate the agglutination of foreign cells. Plants also produce chitinases that act as innate immune elicitors that may serve protective functions against their nematode parasites and insect pests. Vertebrates similarly utilize chitinases produced by macrophages as part of the innate immune response (Boot *et al*., 1995). Chitinases are also produced in human gastric juice (Paoletti *et al.*, 2007).

Some evidence suggests that genetic attributes of human hosts may determine their susceptibility to infection by filarial parasites as well as their response to established infections, and again the evolutionary interaction with

chitin metabolism at its interface, can be observed. Human genetic polymorphisms in the gene for the phagocyte specific chitotriosidase CHIT1, a component of the human innate immune respone - as well as in the promoter of the gene encoding the mannose-binding protein MBL2 - has been linked to susceptibility to infection by *W. bancrofti* (Choi *et al.*, 2001; Meyrowitsch *et al.*, 2010).

FIGURES

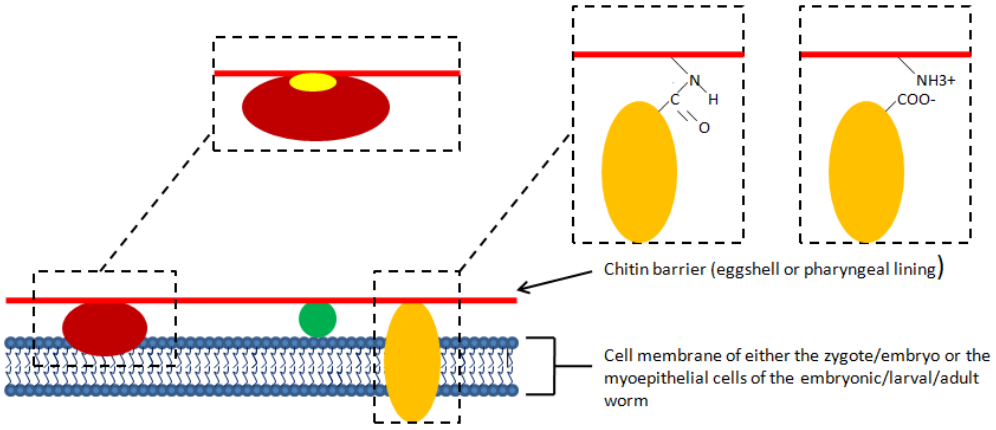


Figure 5.1. The relevance of chitin and chitosan in nematode development. The model depicts a hypothetical lipid bi-layer that is the cell membrane of an epithelial cell. Some proteins are embedded transmembrane proteins (orange oval) or peripheral proteins associated with one membrane, here shown on the extracellular surface (burgundy oval or green spheres). Chitin-binding domains (CBD, yellow) permit non-covalent interactions with chitinous substrates. One or more CBD can be found in any of these types of proteins, but is shown here in only one example for simplicity. Deacetylation of chitin creates reactive amine groups that can form covalent or ionic linkages with nearby molecules including macromolecules at the surface of the underlying epithelium of the nematode embryo or pharynx. One example of such a macromolecule is shown here as a transmembrane protein, although the other classes of proteins can also similarly be linked to chitosan.

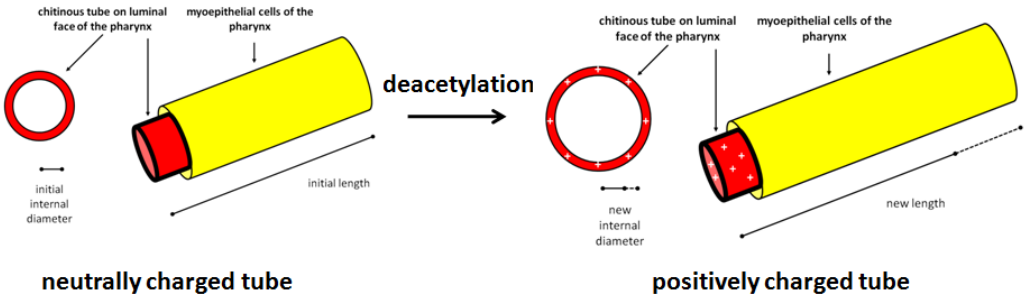


Figure 5.2. Chitin Deacetylation during Pharyngeal Morphogenesis. Deacetylation leads to an accumulation of positive charge along the surface of the chitinous tube lining the pharynx. This may result in longitudinal or radial expansion of a positively charged tube (mediated by new physical and chemical properties of the tube or based on new chemical bonds formed between the free amine groups and the adjacent myoepithelial ECM and cell membranes). Alternatively, deacetylation may limit growth as it has been reported to do in the *Drosophila* embryonic trachea.

Appendix

LIST OF PRIMERS

During primer design, all sequences are analyzed used the BLASTn program to ensure that they do not have a reasonable likelihood of annealing to other non-target sequences from *C. elegans* and also from the *E. coli* that are likely to provide the largest source of contaminating DNA/RNA obtained from harvested worms.

PRIMER NAME	PRIMER SEQUENCE AND NOTES
JHLGX1F1	5'-AAG ACG AGC TCC GGG TGT CTC CGC-3'
	Entire primer anneals with *predicted* Exon 27 of C54G7.3/*lgx-1* Red bases form an intrinsic SacI restriction site
JHLGX1F2	5'-CCT CAG GAT CCG TGT CTC CGC CCA GAT GAA ACA CC-3'
	Blue bases anneal to *predicted* Exon 27 of C54G7.3/*lgx-1* Red bases form a BamHI restriction site
JHLGX1R1	5'-TGT GAG ATC TCA CCT CCT AGT GCA AGT TGG GGC G-3'
	Blue bases anneal to *predicted* Exon 30 of C54G7.3/*lgx-1* Red bases form a BglII restriction site
JHLGX1R2	5'-CGC TAG AAT TCT CCT CTT CAG GAG CTC CAC CGA AC-3'
	Blue bases anneal to *predicted* Exon 30 of C54G7.3/*lgx-1* Red bases form a EcoRI restriction site
C54G7.3PDA PRIMER DESIGN NOTES ARE INCLUDED IN PRIMER BINDER #1	
JHLGX1CBDF1	5'-GAG TTA GAT GGA CTT CTC ACA GCG G-3'
	Entire primer anneals within *predicted* Exon 2 of C54G7.3/*lgx-1*
JHLGX1CBDR1	5'-GAA AAA TTG AGC TGG TCC ATC AAC GG-3'
	Entire primer anneals within *predicted* Exon 4 of C54G7.3/*lgx-1*
C54G7.3CBD PRIMER DESIGN NOTES ARE INCLUDED IN PRIMER BINDER #1	
JHF48E3_8PDAF1	5'-CTT CTG ACC GAA TGC CCG AGA GAC GG-3'
	Entire primer anneals with *predicted* Exon 34 of F48E3.8

JHF48E3_8PDAF2	5'-CAT GCA GAC TTC CCA GTT GCT TCT GCA CTA G-3'
	Entire primer anneals with *predicted* Exon 34 of F48E3.8
JHF48E3_8PDAF3	5'-CCT TCA AGA CTT CTG ACC GAA TGC CCG-3'
	Entire primers anneals with *predicted* Exon 34 of F48E3.8
JHF48E3_8PDAR1	5'-GTT TGT GGC CAA TAT GGG CTG CTG ACC-3'
	Entire primer anneals with *predicted* Exon 36 of F48E3.8
JHF48E3_8PDAR2	5'-TCT CGC TCT TCC ACT GAA ACT GCA CGT CG-3'
	Entire primer anneals with *predicted* Exon 38 of F48E3.8
JHF48E3_8PDAR3	5'-GAT TTG GGA CAA ACG CCA CAC ATT CGG-3'
	Entire primer anneals with *predicted* Exon 38 of F48E3.8
F48E3.8PDA PRIMER DESIGN NOTES ARE INCLUDED IN PRIMER BINDER #2	
JHAMAF1	5'-TTC CAA GCG CCG CTG CGC ATT GTC TC-3'
	Taken from Johnstone and Barry (1996) Entire primer anneals within Exon 1 of *ama-1*
JHAMA1R1	5'-CAG AAT TTC CAG CAC TCG AGG AGC GGA-3'
	Taken from Johnstone and Barry (1996) Entire primer anneals within Exon 3 of *ama-1*
AMA-1 PRIMER ANALYSIS NOTES ARE INCLUDED IN PRIMER BINDER #2	
JHACT3F1	5'-TCG TCT TCA CAT CTT GCC ACA GTT GCC-3'
	Blue bases anneal to the 5' UTR of *act-3*
JHACT3R1	5'-GTG GGG TCT TCT TAT GTG GCA GGA TCG G-3'
	Blue bases anneal to the 3' UTR of *act-3*
ACT-3 PRIMER DESIGN NOTES ARE INCLUDED IN PRIMER BINDER #2	
JHCHS1F1	5'-TCT TGG ATC CTC TCG CCA GAT GCC TTG TTC TCA G-3'
	Blue bases anneal within Exon 3 of T25G3.2/*chs-1* Red bases form a BamHI restriction site
JHCHS1R1	5'-ACA AGA ATT CGC CAA GAA GCG TGA AAT CCC TCG C-3'
	Blue bases anneal to Exon 5 of T25G3.2/*chs-1*

	Red bases form an EcoRI restriction site
FVETAL2001T25G3F	5'-GAA GAT GAA CAC CAC CAC G-3'
	Blue bases anneal to Exon 3 of T25G3.2/*chs-1* Primers used in Veronico *et al.* (2001), obtained through personal comm.
FVETAL2001T25G3R	5'-CAA TCT CCA TCC TCT ATA GC-3'
	Blue Bases anneal to Exon 5 of T25G3.2/*chs-1* Primers used in Veronico *et al.* (2001), obtained through personal comm..
CHS-1 PRIMER DESIGN AND ANALYSIS NOTES ARE INCLUDED IN PRIMER BINDER #3	
JHCHS2F1	5'-ACA AGG ATT CAC CAT GTT CCG AAA CGT GAA ACG C-3'
	Blue bases anneal to Exon 4 of F48A11.1/*chs-2* Underlined bases were intended to be a BamHI restriction site
JHCHS2R1	5'-ACA AGA ATT CAA GCA GCA GAC AAC GTA GAT GCA G-3'
	Blue bases anneal to Exon 6 of F48A11.1/*chs-2* Red bases form an EcoRI restriction site
FVETAL2001F48A11F	5'-GCT GAT GAA GCT GAG CAA G-3'
	Blue bases anneal to Exon 6 of F48A11.1/*chs-2* Primers used in Veronico *et al.* (2001), obtained through personal comm.
FVETAL2001F48A11R	5'-GCA AGC GAC ATT GAT GCA C-3'
	Blue bases anneal to Exon 8 of F48A11.1/*chs-2* Primers used in Veronico *et al.* (2001), obtained through personal comm.
CHS-2 PRIMER DESIGN AND ANALYSIS NOTES ARE INCLUDED IN PRIMER BINDER #3	
JHEGG3F1	5'-GCG CAC CTC TGA CAG TCA TCT TCC GC-3'
	Blue bases anneal to Exon 1 of F44F4.2/*egg-3*
JHEGG3R1	5'-AGG TGG CAA CTG AAC TTC CCC GGC-3'
	Blue bases anneal to Exon 4 of F44F4.2/*egg-3*
EGG-3 PRIMER DESIGN NOTES ARE INCLUDED IN PRIMER BINDER #3	

JANDB1996COL12F1	5'-ACT TGG CTT CTA AAG TCC AGT GAC A-3'
	Taken from Johnstone and Barry (1996) Blue bases anneal in the 5'UTR of the single *col-12* transcript
JANDB1996COL12R1	5'-TCC GCA TGA GCA GCA TGA TCC TCC A-3'
	Taken from Johnstone and Barry (1996) Blue bases anneal to Exon 2 of *col-12*
COL-12 PRIMER ANALYSIS NOTES ARE INCLUDED IN PRIMER BINDER #3	
JANDB1996DPY13F1	5'-ATG GAC ATT GAC ACT AAA ATC AAG GCC-3'
	Blue bases anneal within Exon 1 of *dpy-13*
JANDB1996DPY13R1	5'-TTG TGG GCA TGG CTT GCA TGG TGG TGG G-3'
	Blue bases anneal within Exon 2 of *dpy-13*
DPY-13 PRIMER DESIGN NOTES ARE INCLUDED IN PRIMER BINDER #3	
JHCOL2F1	5'-GCC TAT TCG GCA GTC ACC TTC TCG G-3'
	Blue bases anneal within Exon 1 of *col-2*
JHCOL2F2	5'-CTG CTC GTG ATA TCT GGA GTG AGG TGC-3'
	Blue bases anneal with Exon 1 of *col-2*
JHCOL2R1	5'-AAA TTT AGC GGC GTC GGG TTC CGT C-3'
	Blue bases anneal to a region spanning Exon 2 and the 3' UTR
JHCOL2R2	5'-GTC GGG TTC CGT CCT CAA AGA AGA CTC C-3'
	Blue bases anneal with Exon 2 of *col-2*
COL-2 PRIMER DESIGN NOTES ARE INCLUDED IN PRIMER BINDER #4	
JHCOL19F1	5'- GGG CAA GCT CAT TGT GGT TGG ATC C-3'
	Blue bases anneal to Exon 1 of *col-19*
JHCOL19F2	5'- TTG GAT CCT GCG GAG TAC TTG TGT GCG-3'
	Blue bases anneal to Exon 1 of *col-19*
JHCOL19R1	5'- TGT AAG CTG CAC GAG ATG GGC ATG G-3'
	Blue bases anneal to Exon 2 of *col-19*
JHCOL19R2	5'- CGG CAT CCT CTC CCT TTT GTC CTG G-3'

	Blue bases anneal to Exon 2 of *col-19*
COL-19 PRIMER DESIGN NOTES ARE INCLUDED IN PRIMER BINDER #4	
JHLGX1BINF1	5'-CCA AAC AAC TCG CCC ACC GAT GCC-3'
	Entire primer anneals within *predicted* Exon 10 of C54G7.3/*lgx-1*
JHLGX1BINR1	5'-GGA GCT TCC GTG ATG AAT GGC TGA GC-3'
	Entire primer anneals within *predicted* Exon 14 of C54G7.3/*lgx-1*
LGX-1 BIG INTRON PRIMER DESIGN NOTES ARE INCLUDED IN PRIMER BINDER #4	
JHLGX1TRANBF1	5'- GTT GAT AAC TCT GGA ACA TGT GAG CCG C-3'
	Entire primer anneals within *predicted* Exon 17 of C54G7.3/*lgx-1*
JHLGX1TRANBR1	5'-GGG AGA GAA AGG CAA AAC GGG AAA TTG GG-3'
	Entire primer anneals within the 3' UTR of C54G7.3/*lgx-1*b
LGX-1 TRANSCRIPT B 3'-UTR PRIMER DESIGN NOTES ARE INCLUDED IN PRIMER BINDER #4	
JAFSL2 [*N.B. This is a SLI primer]	5'-GCC GCT GCA GGT TTA ATT ACC CAA GTT TGA G-3'
	Blue bases anneal as the SL1 splice leader sequence Red bases are the first 5 of 6 bases of a PstI restriction site
JHLGX1NEWTRANSL1CONT	5'-ATA CGG GAT CCC AAG TGA GCC AGT AAC ATC-3'
	Blue bases anneal *predicted* Exon 12 of C54G7.3/*lgx-1* Red bases form a BamHI restriction site
SL1 & CONTROL (FOR LATE TRANSCRIPT TEST) PRIMER DESIGN NOTES ARE INCLUDED IN PRIMER BINDER #4	

During primer design, all sequences are analyzed used the BLASTn program to ensure that they do not have a reasonable likelihood of annealing to other non-target sequences from *C. elegans* and also from the *E. coli* that are likely to provide the largest source of contaminating DNA/RNA obtained from harvested worms.

References

REFERENCES

Anokye-Danso, F., Anyanful, A., Sakube, Y., Kagawa, H., 2008. Transcription factors GATA/ELT-2 and forkhead/HNF-3/PHA-4 regulate the tropomyosin gene expression in the pharynx and intestine of *Caenorhabditis elegans*. *J Mol Biol*. 379**,** 201-11.

Anyanful, A., Sakube, Y., Takuwa, K., Kagawa, H., 2001. The third and fourth tropomyosin isoforms of *Caenorhabditis elegans* are expressed in the pharynx and intestines and are essential for development and morphology. *J Mol Biol*. 313**,** 525-37.

Arakane, Y., Dixit, R., Begum, K., Park, Y., Specht, C. A., Merzendorfer, H., Kramer, K. J., Muthukrishnan, S., Beeman, R. W., 2009. Analysis of functions of the chitin deacetylase gene family in *Tribolium castaneum*. *Insect Biochem Mol Biol*. 39**,** 355-65.

Baker, L. G., Specht, C. A., Donlin, M. J., Lodge, J. K., 2007. Chitosan, the deacetylated form of chitin, is necessary for cell wall integrity in *Cryptococcus neoformans*. *Eukaryot Cell*. 6**,** 855-67.

Banks, I. R., Specht, C. A., Donlin, M. J., Gerik, K. J., Levitz, S. M., Lodge, J. K., 2005. A chitin synthase and its regulator protein are critical for chitosan production and growth of the fungal pathogen *Cryptococcus neoformans*. *Eukaryot Cell*. 4**,** 1902-12.

Bendtsen, J. D., Jensen, L. J., Blom, N., Von Heijne, G., Brunak, S., 2004. Feature-based prediction of non-classical and leaderless protein secretion. *Protein Eng Des Sel*. 17**,** 349-56.

Bendtsen, J. D., Kiemer, L., Fausboll, A., Brunak, S., 2005. Non-classical protein secretion in bacteria. *BMC Microbiol*. 5**,** 58.

Bennuru, S., Meng, Z., Ribeiro, J. M., Semnani, R. T., Ghedin, E., Chan, K., Lucas, D. A., Veenstra, T. D., Nutman, T. B., 2011. Stage-specific proteomic expression patterns of the human filarial parasite *Brugia malayi* and its endosymbiont *Wolbachia*. *Proc Natl Acad Sci* U S A. 108**,** 9649-54.

Berninsone, P., Hirschberg, C. B., 1998. Heparan sulfate/heparin N-deacetylase/N-sulfotransferase. The N-sulfotransferase activity domain is at the carboxyl half of the holoenzyme. *J Biol Chem*. 273**,** 25556-9.

Blair, D. E., Hekmat, O., Schuttelkopf, A. W., Shrestha, B., Tokuyasu, K., Withers, S. G., van Aalten, D. M., 2006. Structure and mechanism of chitin deacetylase from the fungal pathogen *Colletotrichum lindemuthianum*. *Biochemistry*. 45**,** 9416-26.

Blair, D. E., Schuttelkopf, A. W., MacRae, J. I., van Aalten, D. M., 2005. Structure and metal-dependent mechanism of peptidoglycan deacetylase, a streptococcal virulence factor. *Proc Natl Acad Sci* U S A. 102**,** 15429-34.

Blair, D. E., van Aalten, D. M., 2004. Structures of *Bacillus subtilis* PdaA, a family 4 carbohydrate esterase, and a complex with N-acetyl-glucosamine. *FEBS Lett*. 570**,** 13-9.

Blaxter, M. L., De Ley, P., Garey, J. R., Liu, L. X., Scheldeman, P., Vierstraete, A., Vanfleteren, J. R., Mackey, L. Y., Dorris, M., Frisse, L. M., Vida, J. T., Thomas, W. K., 1998. A molecular evolutionary framework for the phylum *Nematoda*. *Nature*. 392**,** 71-5.

Blumenthal, T., Trans-splicing and operons (June 25, 2005), *WormBook*, ed. The *C. elegans* Research Community, Workbook, doi/10.1895/wormbook.1.5.1, http:www.wormbook.org.

Blumenthal, T., Evans, D., Link, C. D., Guffanti, A., Lawson, D., Thierry-Mieg, J., Thierry-Mieg, D., Chiu, W. L., Duke, K., Kiraly, M., Kim, S. K., 2002. A global analysis of *Caenorhabditis elegans* operons. *Nature*. 417**,** 851-4.

Boot, R. G., Renkema, G. H., Strijland, A., van Zonneveld, A. J., Aerts, J. M., 1995. Cloning of a cDNA encoding chitotriosidase, a human chitinase produced by macrophages. *J Biol Chem*. 270**,** 26252-6.

Bulik, D. A., Olczak, M., Lucero, H. A., Osmond, B. C., Robbins, P. W., Specht, C. A., 2003. Chitin synthesis in *Saccharomyces cerevisiae* in response to supplementation of growth medium with glucosamine and cell wall stress. *Eukaryot Cell*. 2**,** 886-900.

Burgwyn, B., Nagel, B., Ryerse, J., Bolla, R. I., 2003. *Heterodera glycines*: eggshell ultrastructure and histochemical localization of chitinous components. *Exp Parasitol*. 104**,** 47-53.

Caufrier, F., Martinou, A., Dupont, C., Bouriotis, V., 2003. Carbohydrate esterase family 4 enzymes: substrate specificity. *Carbohydr Res*. 338**,** 687-92.

Choi, E. H., Zimmerman, P. A., Foster, C. B., Zhu, S., Kumaraswami, V., Nutman, T. B., Chanock, S. J., 2001. Genetic polymorphisms in molecules of innate immunity and susceptibility to infection with *Wuchereria bancrofti* in South India. *Genes Immun*. 2**,** 248-53.

Choi, J., Newman, A. P., 2006. A two-promoter system of gene expression in *C. elegans*. *Dev Biol*. 296**,** 537-44.

Cleves, A. E., Kelly, R. B., 1996. Rehearsing the ABCs. Protein translocation. *Curr Biol*. 6**,** 276-8.

Consortium, T. C. e. S., 1998. Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science*. 282**,** 2012-8.

Cox, G. N., Hirsh, D., 1985. Stage-specific patterns of collagen gene expression during development of *Caenorhabditis elegans*. *Mol Cell Biol*. 5**,** 363-72.

Das, S., Van Dellen, K., Bulik, D., Magnelli, P., Cui, J., Head, J., Robbins, P. W., Samuelson, J., 2006. The cyst wall of *Entamoeba invadens* contains chitosan (deacetylated chitin). *Mol Biochem Parasitol*. 148**,** 86-92.

Dixit, R., Arakane, Y., Specht, C. A., Richard, C., Kramer, K. J., Beeman, R. W., Muthukrishnan, S., 2008. Domain organization and phylogenetic analysis of proteins from the chitin deacetylase gene family of *Tribolium castaneum* and three other species of insects. *Insect Biochem Mol Biol*. 38**,** 440-51.

Dubinsky, P., Rybos, M., Turcekova, L., 1986a. Properties and localization of chitin synthase in *Ascaris suum* eggs. *Parasitology*. 92 (Pt 1)**,** 219-25.

Dubinsky, P., Rybos, M., Turcekova, L., Ossikovski, E., 1986b. Chitin synthesis in zygotes of *Ascaris suum*. *J Helminthol*. 60**,** 187-92.

Ellis, D. S., Rogers, R., Bianco, A. E., Denham, D. A., 1978. Intrauterine development of the microfilariae of *Dipetalonema viteae*. *J Helminthol*. 52**,** 7-10.

Fanelli, E., Di Vito, M., Jones, J. T., De Giorgi, C., 2005. Analysis of chitin synthase function in a plant parasitic nematode, *Meloidogyne artiellia*, using RNAi. *Gene*. 349**,** 87-95.

Flibotte, S., Edgley, M. L., Chaudhry, I., Taylor, J., Neil, S. E., Rogula, A., Zapf, R., Hirst, M., Butterfield, Y., Jones, S. J., Marra, M. A., Barstead, R. J., Moerman, D. G., 2010. Whole-genome profiling of mutagenesis in *Caenorhabditis elegans*. *Genetics*. 185**,** 431-41.

Foster, J. M., Zhang, Y., Kumar, S., Carlow, C. K., 2005. Parasitic nematodes have two distinct chitin synthases. *Mol Biochem Parasitol*. 142**,** 126-32.

Franks, D. M., Izumikawa, T., Kitagawa, H., Sugahara, K., Okkema, P. G., 2006. *C. elegans* pharyngeal morphogenesis requires both *de novo* synthesis of pyrimidines and synthesis of heparan sulfate proteoglycans. *Dev Biol*. 296**,** 409-20.

Gaudet, J., Mango, S. E., 2002. Regulation of organogenesis by the *Caenorhabditis elegans* FoxA protein PHA-4. *Science*. 295**,** 821-5.

Gaudet, J., Muttumu, S., Horner, M., Mango, S. E., 2004. Whole-genome analysis of temporal gene expression during foregut development. *PLoS Biol*. 2**,** e352.

Ghedin, E., Wang, S., Spiro, D., Caler, E., Zhao, Q., Crabtree, J., Allen, J. E., Delcher, A. L., Guiliano, D. B., Miranda-Saavedra, D., Angiuoli, S. V., Creasy, T., Amedeo, P., Haas, B., El-Sayed, N. M., Wortman, J. R., Feldblyum, T., Tallon, L., Schatz, M., Shumway, M., Koo, H., Salzberg, S. L., Schobel, S., Pertea, M., Pop, M., White, O., Barton, G. J., Carlow, C. K., Crawford, M. J., Daub, J., Dimmic, M. W., Estes, C. F., Foster, J. M., Ganatra, M., Gregory, W. F., Johnson, N. M., Jin, J., Komuniecki, R., Korf, I., Kumar, S., Laney, S., Li, B. W., Li, W., Lindblom, T. H., Lustigman, S., Ma, D., Maina, C. V., Martin, D. M., McCarter, J. P., McReynolds, L., Mitreva, M., Nutman, T. B., Parkinson, J., Peregrin-Alvarez, J. M., Poole, C., Ren, Q., Saunders, L., Sluder, A. E., Smith, K., Stanke, M., Unnasch, T. R., Ware, J., Wei, A. D., Weil, G., Williams, D. J., Zhang, Y., Williams, S. A., Fraser-Liggett, C., Slatko, B., Blaxter, M. L., Scott, A. L., 2007. Draft genome of the filarial nematode parasite *Brugia malayi*. *Science*. 317**,** 1756-60.

Gilbert, N. M., Baker, L. G., Specht, C. A., Lodge, J. K., 2012. A Glycosylphosphatidylinositol Anchor Is Required for Membrane Localization but Dispensable for Cell Wall Association of Chitin Deacetylase 2 in *Cryptococcus neoformans*. *MBio*. 3.

Gooday, G. W., Brydon, L. J., Chappell, L. H., 1988. Chitinase in female *Onchocerca gibsoni* and its inhibition by allosamidin. *Mol Biochem Parasitol*. 29**,** 223-5.

Graber, J. H., Salisbury, J., Hutchins, L. N., Blumenthal, T., 2007. *C. elegans* sequences that control trans-splicing and operon pre-mRNA processing. *RNA*. 13**,** 1409-26.

Guo, W., Li, G., Pang, Y., Wang, P., 2005. A novel chitin-binding protein identified from the peritrophic membrane of the cabbage looper, *Trichoplusia ni*. *Insect Biochem Mol Biol*. 35**,** 1224-34.

Harris, M. T., Fuhrman, J. A., 2002. Structure and expression of chitin synthase in the parasitic nematode *Dirofilaria immitis*. *Mol Biochem Parasitol*. 122**,** 231-4.

Harris, M. T., Lai, K., Arnold, K., Martinez, H. F., Specht, C. A., Fuhrman, J. A., 2000. Chitin synthase in the filarial parasite, *Brugia malayi*. *Mol Biochem Parasitol*. 111**,** 351-62.

Heustis, R. J., Ng, N. K., Brand, K. J., Rogers, M. C., Le, L. T., Specht, C. A., Fuhrman, J. A., 2012. Pharyngeal Polysaccharide Deacetylases Affect Development in the Nematode *C. elegans* and Deacetylate Chitin *In Vitro*. *PLoS One*. 7**,** e40426.

Huang, G., Gao, B., Maier, T., Allen, R., Davis, E. L., Baum, T. J., Hussey, R. S., 2003. A profile of putative parasitism genes expressed in the esophageal gland cells of the root-knot nematode *Meloidogyne incognita*. *Mol Plant Microbe Interact*. 16**,** 376-81.

Jeong, P. Y., Kwon, M. S., Joo, H. J., Paik, Y. K., 2009. Molecular time-course and the metabolic basis of entry into dauer in *Caenorhabditis elegans*. *PLoS One*. 4**,** e4162.

John, M., Rohrig, H., Schmidt, J., Wieneke, U., Schell, J., 1993. *Rhizobium* NodB protein involved in nodulation signal synthesis is a chitooligosaccharide deacetylase. *Proc Natl Acad Sci* U S A. 90**,** 625-9.

Johnston, W. L., Krizus, A., Dennis, J. W., 2006. The eggshell is required for meiotic fidelity, polar-body extrusion and polarization of the *C. elegans* embryo. *BMC Biol*. 4**,** 35.

Johnstone, I. L., Barry, J. D., 1996. Temporal reiteration of a precise gene expression pattern during nematode development. *EMBO J*. 15**,** 3633-9.

Kaji, H., Kamiie, J., Kawakami, H., Kido, K., Yamauchi, Y., Shinkawa, T., Taoka, M., Takahashi, N., Isobe, T., 2007. Proteomics reveals N-linked glycoprotein diversity in *Caenorhabditis elegans* and suggests an atypical translocation mechanism for integral membrane proteins. *Mol Cell Proteomics*. 6**,** 2100-9.

Kamath, R. S., Ahringer, J., 2003. Genome-wide RNAi screening in *Caenorhabditis elegans*. *Methods*. 30**,** 313-21.

Kim, S. K., Lund, J., Kiraly, M., Duke, K., Jiang, M., Stuart, J. M., Eizinger, A., Wylie, B. N., Davidson, G. S., 2001. A gene expression map for *Caenorhabditis elegans*. *Science*. 293**,** 2087-92.

Kumar, S., Tamura, K., Nei, M., 2004. MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief Bioinform*. 5**,** 150-63.

Liu, Z., Kirch, S., Ambros, V., 1995. The *Caenorhabditis elegans* heterochronic gene pathway controls stage-specific transcription of collagen genes. *Development*. 121**,** 2471-8.

Luschnig, S., Batz, T., Armbruster, K., Krasnow, M. A., 2006. *serpentine* and *vermiform* encode matrix proteins with chitin binding and deacetylation domains that limit tracheal tube length in *Drosophila*. *Curr Biol*. 16**,** 186-94.

Maruyama, R., Velarde, N. V., Klancer, R., Gordon, S., Kadandale, P., Parry, J. M., Hang, J. S., Rubin, J., Stewart-Michaelis, A., Schweinsberg, P., Grant, B. D., Piano, F., Sugimoto, A., Singson, A., 2007. EGG-3 regulates cell-surface and cortex rearrangements during egg activation in *Caenorhabditis elegans*. *Curr Biol*. 17**,** 1555-60.

Meissner, B., Rogalski, T., Viveiros, R., Warner, A., Plastino, L., Lorch, A., Granger, L., Segalat, L., Moerman, D. G., 2011. Determining the sub-cellular localization of proteins within *Caenorhabditis elegans* body wall muscle. *PLoS One*. 6**,** e19937.

Meyrowitsch, D. W., Simonsen, P. E., Garred, P., Dalgaard, M., Magesa, S. M., Alifrangis, M., 2010. Association between mannose-binding lectin polymorphisms and *Wuchereria bancrofti* infection in two communities in North-Eastern Tanzania. *Am J Trop Med Hyg*. 82**,** 115-20.

Mitreva, M., Jasmer, D. P., Zarlenga, D. S., Wang, Z., Abubucker, S., Martin, J., Taylor, C. M., Yin, Y., Fulton, L., Minx, P., Yang, S. P., Warren, W. C., Fulton, R. S., Bhonagiri, V., Zhang, X., Hallsworth-Pepin, K., Clifton, S. W., McCarter, J. P., Appleton, J., Mardis, E. R., Wilson, R. K., 2011. The draft genome of the parasitic nematode *Trichinella spiralis*. *Nat Genet*. 43**,** 228-35.

Moreno, Y., Geary, T. G., 2008. Stage- and gender-specific proteomic analysis of *Brugia malayi* excretory-secretory products. *PLoS Negl Trop Dis*. 2**,** e326.

Neuhaus, B., Bresciani, J., Peters, W., 1996. Ultrastructure of the Pharyngeal Cuticle and Lectin Labelling with Wheat Germ Agglutinin-gold Conjugate Indicating Chitin in the Pharyngeal Cuticle of *Oesophagostomum dentatum* (Strongylida, Nematoda). *Acta Zoologica*. 78**,** 205-13.

Paoletti, M. G., Norberto, L., Damini, R., Musumeci, S., 2007. Human gastric juice contains chitinase that can degrade chitin. *Ann Nutr Metab*. 51**,** 244-51.

Raharjo, W. H., Logan, B. C., Wen, S., Kalb, J. M., Gaudet, J., 2010. *In vitro* and *in vivo* characterization of *Caenorhabditis elegans* PHA-4/FoxA response elements. *Dev Dyn*. 239**,** 2219-32.

Rastogi, S., Borgo, B., Fox, P., Pazdernik, N., Mardis, E., Kohara, Y., Havranek, J., Schedl, T., *glp-4* encodes the valyl amino-acyl tRNA synthetase VARS-2. International Worm Meeting Abstract Booklet, Vol. 2011, Los Angeles, CA, USA, 2011.

Reinke, V., Smith, H. E., Nance, J., Wang, J., Van Doren, C., Begley, R., Jones, S. J., Davis, E. B., Scherer, S., Ward, S., Kim, S. K., 2000. A global profile of germline gene expression in *C. elegans*. *Mol Cell*. 6**,** 605-16.

Rogers, R., Ellis, D. S., Denham, D. A., 1976. Studies with *Brugia pahangi*. 14. Intrauterine development of the microfilaria and a comparison with other filarial species. *J Helminthol*. 50**,** 251-7.

Schraermeyer, U., Peters, W., Zahner, H., 1987. Formation by the uterus of a peripheral layer of the sheath in microfilariae of *Litomosoides carinii* and *Brugia malayi*. *Parasitol Res*. 73**,** 557-64.

Simmer, F., Tijsterman, M., Parrish, S., Koushika, S. P., Nonet, M. L., Fire, A., Ahringer, J., Plasterk, R. H., 2002. Loss of the putative RNA-directed RNA polymerase RRF-3 makes *C. elegans* hypersensitive to RNAi. *Curr Biol*. 12**,** 1317-9.

Thein, M. C., McCormack, G., Winter, A. D., Johnstone, I. L., Shoemaker, C. B., Page, A. P., 2003. *Caenorhabditis elegans* exoskeleton collagen COL-19: an adult-specific marker for collagen modification and assembly, and the analysis of organismal morphology. *Dev Dyn*. 226**,** 523-39.

Trudel, J., Asselin, A., 1990. Detection of chitin deacetylase activity after polyacrylamide gel electrophoresis. *Anal Biochem*. 189**,** 249-53.

Veronico, P., Gray, L. J., Jones, J. T., Bazzicalupo, P., Arbucci, S., Cortese, M. R., Di Vito, M., De Giorgi, C., 2001. Nematode chitin synthases: gene structure, expression and function in *Caenorhabditis elegans* and the plant parasitic nematode Meloidogyne artiellia. *Mol Genet Genomics*. 266**,** 28-34.

Wang, S., Jayaram, S. A., Hemphala, J., Senti, K. A., Tsarouhas, V., Jin, H., Samakovlis, C., 2006. Septate-junction-dependent luminal deposition of chitin deacetylases restricts tube elongation in the *Drosophila* trachea. *Curr Biol*. 16**,** 180-5.

Wharton, D., 1980. Nematode egg-shells. *Parasitology*. 81**,** 447-63.

Wharton, D. A., 1979. Oogenesis and egg-shell formation in *Aspiculuris tetraptera* Schulz (Nematoda: Oxyuroidea). *Parasitology*. 78**,** 131-43.

Wood, W. B. (Ed.) 1988. The Nematode *Caenorhabditis elegans*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Zhang, Y., Foster, J. M., Nelson, L. S., Ma, D., Carlow, C. K., 2005. The chitin synthase genes *chs-1* and *chs-2* are essential for *C. elegans* development and responsible for chitin deposition in the eggshell and pharynx, respectively. *Dev Biol*. 285**,** 330-9.