IDENTIFICATION OF A PUTATIVE PPA-MLT-10 ORTHOLOG IN THE NEMATODE *PRISTIONCHUS PACIFICUS*

A thesis submitted in partial fulfillment of the requirements
For the degree of Master of Science in Biology

By

Maryn Cook

May 2014
The thesis of Maryn Cook is approved:

__________________________________________  ______________________________
David Bermudes, Ph.D.                          Date

__________________________________________  ______________________________
Rheem Medh, Ph.D.                              Date

__________________________________________  ______________________________
Cheryl Van Buskirk, Ph.D.                      Date

__________________________________________  ______________________________
Ray Hong, Ph.D., Chair                         Date

California State University, Northridge
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Signature Page</td>
<td>ii</td>
</tr>
<tr>
<td>List of Figures</td>
<td>iv</td>
</tr>
<tr>
<td>List of Tables</td>
<td>v</td>
</tr>
<tr>
<td>Abstract</td>
<td>vi</td>
</tr>
<tr>
<td>Chapter I: Introduction</td>
<td>1</td>
</tr>
<tr>
<td>Background</td>
<td>1</td>
</tr>
<tr>
<td>Recent and Relevant Evidence</td>
<td>2</td>
</tr>
<tr>
<td><em>Pristionchus pacificus</em></td>
<td>4</td>
</tr>
<tr>
<td>Insect Molting</td>
<td>6</td>
</tr>
<tr>
<td>Nematode Molting</td>
<td>12</td>
</tr>
<tr>
<td>Molting Behavior</td>
<td>15</td>
</tr>
<tr>
<td><em>mlt-10</em></td>
<td>16</td>
</tr>
<tr>
<td>Embryonic Molt</td>
<td>18</td>
</tr>
<tr>
<td>Hypothesis and Focus of Thesis</td>
<td>19</td>
</tr>
<tr>
<td>Chapter II: Methods and Materials</td>
<td>20</td>
</tr>
<tr>
<td>Chapter III: Results</td>
<td>29</td>
</tr>
<tr>
<td>Chapter IV: Discussion</td>
<td>57</td>
</tr>
<tr>
<td>References</td>
<td>66</td>
</tr>
</tbody>
</table>
List of Figures

| Figure 1.1 | Life cycles of *P. pacificus* and *C. elegans* |
| Figure 1.2 | Generalized structure of mature insect integument |
| Figure 1.3 | Ecdysteroid titers across *D. melanogaster* lifespan |
| Figure 1.4 | Schematic of the major changes in the insect integument during molting |
| Figure 1.5 | Generalized structure of dorso-ventral adult nematode integument |
| Figure 1.6 | *P. pacificus* J4 larva hermaphrodite undergoing molting |
| Figure 1.7 | Summary of signaling cascade of molting in insects, *C. elegans*, and *P. pacificus* |
| Figure 2.1 | Screenshot of BLAST results for Cel-mlt-10 query |
| Figure 2.2 | Cropped screenshot of BLAST results for Contig83-snap.49 query |
| Figure 2.3 | Screenshot of BLAST search for Contig83-snap.49 |
| Figure 2.4 | Primer strategy and denaturing gel electrophoresis of *Ppa-mlt-10* |
| Figure 2.5 | Gene structure of *Ppa-mlt-10* |
| Figure 2.6 | MLT-10 protein structures |
| Figure 2.7 | Multiple sequence alignment of selected MLT-10 proteins |
| Figure 2.8 | Proline-Rich C-terminal Domain of PPA-MLT-10 |
| Figure 2.9 | Phylogenetic tree positioning :PPA-MLT-10 with other MLT-10 orthologs |
| Figure 2.10 | Melting curves |
| Figure 2.11 | Diagram of primer design strategy for amplification of *Ppa-mlt-10* |
| Figure 2.12 | Result of qPCR experiment |
| Figure 2.13 | Screenshot of BLASTp query of proline-rich repeat domain |
| Figure 2.14 | Phylogram positioning the newly identified *P. pacificus* mltn genes |
| Figure 2.15 | Gel electrophoresis image of 5‘RACE of *Ppa-pnhr-2*genes |
| Figure 2.16 | Virtual clone of *Ppa-pnhr-2p::gfp* (MC16.1) |
| Figure 2.17 | Gel electrophoresis results of putative *Ppa-chs-1* expression |
| Figure 2.18 | Gel electrophoresis results of putative *Ppa-hch-1* expression |
| Figure 3.1 | Predicted 3-D ribbon models of PRR of PPA-MLT-10 using ITASSER |
| Figure 3.2 | Figure 3.3 Predicted 3-D structure of PPA-MLT-10 DUF using I-TASSER |
List of Tables

| Table 1.1 | Overview of insect integument components and function |
| Table 2.1 | List of Primers |
| Table 2.2 | Comparison of *Ppa-mlt-10* and *Cel-mlt-10* gene |
| Table 2.3 | Best BLASTp *P. pacificus* sequence |
| Table 2.4 | Putative *P. pacificus mltn* gene family and *C. elegans* paralogs |
ABSTRACT
IDENTIFICATION OF A PUTATIVE PPA-MLT-10 ORTHOLOG IN THE
NEMATODE PRISTIONCHUS PACIFICUS

By
Maryn Cook
Master of Science in Biology

Beetle-associated nematode Pristionchus pacificus (Diplogastridae) bears characteristics of both free-living and parasitic worms and is therefore a unique comparative system to study behavior and developmental regulation, including the regulation of molting. Molting is the regular and isochronous shedding of a cuticle or exoskeleton at the end of each juvenile or larval stage. Many nematodes, such as the widely-studied Caenorhabditis elegans (Rhabditidae), undergo four larval molts that divide five discreet developmental stages; however, in the nematode P. pacificus, only three larval molts are observed after hatching from eggs. Researchers hypothesize that rather than omitting a molt altogether, P. pacificus experience a molt prior to hatching known as the embryonic molt. Such an embryonic molt has been observed in other select Diplogastrids and is now a widely accepted developmental phenomenon defining the taxon. Another difference between P. pacificus and C. elegans is the presence of the insect ecdysone receptor homolog in its genome which suggests that the regulation of molting may differ within the phylum.

In an effort to study the developmental process and evolution of molting and genetically characterize the embryonic molt in P. pacificus, we draw analogy from the well-studied model organism C. elegans. Using standard reciprocal basic local
alignment search tools (TBLASTN, BLASTP, BLASTX), we show that the *P. pacificus* genome encodes the putative ortholog of *mlt-10*, a nematode-specific protein necessary for the animal to successfully remove the larval cuticle during molting. We performed a reverse transcription polymerase chain reaction (RT-PCR) experiment using *P. pacificus* cDNA to amplify the most similar translated nucleotide sequence and obtain the full-length mRNA sequence. We compared the cDNA sequence to the genomic sequence to identify intron/exon length and boundaries and to determine the start and stop of translation. To identify regions of sequence conservation we conducted a multiple sequence alignment of the predicted *P. pacificus* MLT-10 with MLT-10 from three *Caenorhabditis* species and select nematode species and performed phylogenetic analysis. Using quantitative PCR (qPCR) on cDNA obtained from synchronized and staged worm populations we obtained a preliminary developmental profile of putative *P. pacificus mlt-10* (*Ppa-mlt-10*). Using standard Blast queries of the conserved proline-rich-repeat C-terminal domain of MLT-10 followed by phylogenetic analysis, we show that the *P. pacificus* genome contains the *mltn* paralogous gene family and assign putative names to the newly identified genes.

Using the BLAST/RT-PCR procedure described above and in order to find genes expressed only during embryogenesis, we identified, amplified and sequenced putative *chs-1* and *hch-1* orthologs in the *P. pacificus* transcriptome. Finally, we have generated a transcriptional reporter of the *P. pacificus* ecdysone receptor suitable for microinjection. An attempt was made to separate an egg population into four discrete developmental stages based on egg mass using a simple sucrose gradient assay. The goal of this experiment was to obtain staged egg populations and measure embryonic expression of
Ppa-mlt-10 using qPCR. The assay failed to separate the eggs into the desired stages and instead produced a single heterogeneous population. We also pursued with unsuccessful outcome the RNA interference (feeding) of Ppa-pnhr-1 and Ppa-pnhr-2 and 5’RACE of Ppa-mlt-10 and Ppa-pnhr-1.

Together these experiments provide our laboratory a starting point from which the differences in the molting pathway within the nematode phylum can be established particularly as they pertain to endocrine signaling and the embryonic molt.
CHAPTER I
INTRODUCTION

Background

Phylum nematoda (roundworm) constitutes a large phylum of animals and encompasses nearly 20,000 species ranging from free-living to parasitic, terrestrial to marine, and innocuous to pathogenic. Nematodes possess a complete digestive system and a fluid-filled body cavity that facilitates efficient nutrient transfer and permits body wall muscles to contract without hindering digestion. The characteristic roundworm body surface is composed of a collagen-rich outer cuticle that must be periodically shed throughout the life cycle in a process called ecdysis (molting). Ecdysone, derived from the term “ecdysis” is a steroidal precursor that when bound to its receptor prompts molting in insects and defines the Ecdysozoa. The Ecdysozoa are a clade of molting invertebrates that includes nematodes, arthropods, and six smaller phyla. Ecdysis, the final step in the molting process, refers to the actual shedding of the old cuticle (exuvium). The process of molting has been extensively studied in insects and is well-characterized; however the process in nematodes remains elusive.

The assignment of nematodes to the Ecdysozoa clade was originally based on the common trait of molting with an underlying hypothesis that an ecdysteroid played a yet-to-be-determined role in nematode molting. Although endogenous ecdysteroids have been detected in the reproductive and body tissue of parasitic nematodes *Dirofia*ria *immitis* and *Ascaris* suum (Barker et al. 1991; Cleator et al. 1987), researchers have failed to identify an ecdysteroid in the most robustly studied nematode, free-living
Caenorhabditis elegans. In fact, studies of molting in C. elegans while supportive of a steroid-induced molting pathway, clearly point to an alternative pathway, one that does not include ecdysone.

Recent and Relevant Evidence

Teasing out the molting regulatory network in nematodes has been a relatively recent endeavor but ranks as one of great importance. Since parasitic nematodes are responsible for human disease and suffering, and the destruction and loss of agricultural crops worldwide, the development of pharmacological and biocidic agents targeting the molting process could have profound value.

In 2005, the most comprehensive analysis of molting in C. elegans to date was produced (Frand et al. 2005). Using genome-wide RNA interference, 159 genes essential for the successful removal of the cuticle during molting were identified. Annotated proteins (nuclear hormone receptors, transcription factors, cuticle collagens, proteases, for example) and novel genes [mlt-8 thru mlt-11 (MoLting defective)] were uncovered in the screen, many of which are conserved in nematodes and/or insects. In the first detailed proposed model for molting in C. elegans, Frand et al. (2005) describes the process as a periodic endocrine signaling cascade mediated by ligand-activated nuclear hormone receptors (nhr-23 and nhr-25) that results in the remodeling of the exoskeleton by the epithelial cells. Membrane signaling proteins (QUA-1, ACN-1, MLT-8, MLT-9) and extracellular components (MLT-10 and FBN-1) facilitate the generation of new cuticle, protease inhibitors (MLT-11 and BLI-3/5) prevent early release of old cuticle, and proteases (NAS-36/37) promote the actual ecdysis event (Frand et al. 2005).
The identification of the novel extracellular matrix component MLT-10 and subsequent discovery of a large nematode-specific (and putatively dispensable) \textit{mltn} gene family which contains 13 paralogs known as \textit{mltn-1} thru \textit{mltn-13} (MoLt TeN like) directed the focus to the characterization of the protein and its involvement in molting (Frand et al. 2005; Meli et al. 2010). The expression of a \textit{mlt-10} transcriptional reporter oscillates synchronously with the molting cycle and peaks just prior to ecdysis (Frand et al. 2005; Meli et al. 2010) Due to the positive correlation between \textit{mlt-10} expression and molting, a transgenic mutant containing an integrated \textit{mlt-10::gfp-PEST} fusion gene is now used as a positive fluorescence marker for molting in \textit{C. elegans} (Monsalve et al. 2011). It is for this purpose that we pursue the identification and characterization of \textit{mlt-10} in \textit{P. pacificus}.

The presence and function of \textit{nhr-23} and \textit{nhr-25}, orthologs of the insect ecdysone-induced early-late gene product DHR3 and the ecdyone-repressed bridge gene product Ftz-F1, respectively, suggests that molting in \textit{C. elegans} may rely on a steroid signal; however, the steroid does not appear to be ecdysone. Recently the presence of an ecdysone-signaling pathway has been proposed in beetle-associated nematode \textit{Pristionchus pacificus}. Homologs of the well-characterized arthropod ecdysone receptor subunits Ultraspiracle (USP) and Ecdysone Receptor (EcR) were identified in the \textit{P. pacificus} genome: \textit{Ppa-pnhr-1} and \textit{Ppa-pnhr-2} (Pristionchus Nuclear Hormone Receptor), respectively (Parihar et al 2010). In a pattern that strongly correlates with the molting schedule, mRNA expression of the genes, particularly \textit{Ppa-pnhr-1}, was found to oscillate across the lifespan of the animal with peak expression occurring during the intermolt period.
Additionally, ecdysteroids were found to play a role in the development of *P. pacificus*. In a series of experiments an exogenous ecdysteroid was applied to the culture media of synchronized *P. pacificus* J2 animals and scored development at 46 and 54 hours later (Manish Parihar, Master’s Thesis, 2008). Synchronized *C. elegans* L1 animals were used as a negative control. The results showed that normal larval development of *P. pacificus* was severely impeded in response to the exogenous ecdysteroid application— all *P. pacificus* animals arrested at the J2 larval stage; whereas *C. elegans* larvae developed normally to the adult stage (Manish Parihar, Master’s Thesis, 2008). Taken together the results suggest the presence of an ecdysone signaling pathway that affects development in *P. pacificus* while not in *C. elegans*. In an effort to understand molting and the embryonic molt in *P. pacificus* and compare molting across the nematode phylum, we have generated a *gfp* reporter construct to be used to assess spatial expression of the *Ppa-pnhr-2* promoter during embryonic and post-embryonic development.

*Pristionchus pacificus*

The *Pristionchus pacificus* adult is a ~1 mm self-fertilizing hermaphroditic roundworm that, in nature, resides predominantly on scarab beetles. *P. pacificus* displays a transparent, unsegmented, bilaterally symmetric body and radially symmetric head. *P. pacificus* has five pairs of autosomes and one pair of sex chromosomes existing in near macrosynteny with those of *C. elegans*. Like *C. elegans*, sex in *P. pacificus* is based on an XO sex determination system whereby hermaphrodites have two X chromosomes (XX) and males have one (X0). In nutritiously adequate environments spontaneous males occur at a very low rate (0.1%) as a result of a non-disjunction event of the X
chromosome (Sommer et al. 1996). However in starved populations large numbers of males can occur. The typical laboratory model has a 3.5 day life cycle (Figure 1.1) and feeds on an *Escherichia coli* (OP50) bacterial food source. The genome of *P. pacificus* was fully sequenced in 2008 and since it’s completion has been used as a satellite model organism and comparative system to *C. elegans* (Dieterich et al 2008).

![Figure 1.1 Life cycles of P. pacificus and C. elegans.](image)

Figure 1.1 Life cycles of *P. pacificus* and *C. elegans*. Like many nematodes, *P. pacificus* undergoes four molts during its life cycle: a single embryonic molt followed by three larval molts. Because the first molt takes place within the egg, *P. pacificus* hatches as a J2 animal. *C. elegans* experiences all four molts after hatching. Adapted from Hong and Sommer, 2006.

In favorable conditions the hermaphroditic adult lays eggs that progress through four juvenile stages, J1, J2, J3 and J4 (compared to L1, L2, L3, and L4 in *C. elegans*) before becoming adults. Each stage is punctuated by a molt. Total generation time in *P. pacificus* is 82 hours at 20 °C (compared to 75 hours in *C. elegans*) (Felix et al.1999) In unfavorable conditions such as crowding, extreme temperature, and food scarcity, *P. pacificus* J2 animals, like *C. elegans* L1 or L2 animals, can enter an alternative J3/L3
developmental stage known as dauer. Harsh environmental cues prompt the J2 individual to adopt this alternative dauer state by arresting feeding and reducing locomotion. The resistant dauer can exist in this state of stasis without feeding for months. Recovery can be induced by improving food availability, whereby the dauer individual undergoes the J3/J4 molt and develops normally to adulthood (Cassandra and Russell, 1975). Notably, in its natural environment and in a relationship known as necromeny, *P. pacificus* spends most of its life in the non-feeding dauer state, surviving on the body of the oriental beetle until the host dies. When the beetle dies, the dauer nematode reenters the normal life cycle while consuming the beetle carcass. However, under laboratory conditions, *P. pacificus* is free-living and therefore progresses through the normal life cycle that includes four molts/larval stages and a final adult stage.

*Insect Molting*

Most of what we know about molting comes from studies on the insect and fruit fly *D. melanogaster* from which parallels between the molting of insects and nematodes can be drawn. While nematodes maintain the worm-like (vermiform) body shape throughout the life cycle, insects experience distinct changes in body form as they develop from birth to adulthood. Despite this difference in metamorphogenesis, both phyla transition from one stage to the next by molting. After hatching *D. melanogaster* animals progress through multiple larval stages or instars, each of which is punctuated by a molt. Following a long state of inactivity as a pupa, the animal undergoes a final molt into adulthood. In a similar pattern, nematodes such as *P. pacificus* and *C. elegans* go
through four juvenile or larval stages, respectively, and four molts before reaching adulthood.

Insects (and nematodes) lack an endoskeleton; therefore, to allow for increased body size and the development of complex structures, the animals must shed an exoskeleton. The insect exoskeleton is made up of chitin, an extracellular chemical composite similar to collagen in vertebrates (and nematodes) and cellulose in plants, and provides the cuticle with support and protection (Figure 1.2).

![Figure 1.2 Generalized structure of mature insect integument. The insect cuticle is an extracellular matrix composed of chitin, protein, lipids and enzymes overlying a single epidermal layer. Adapted from Chapman, 2013.](image)

The insect cuticle is formed by a single layer of epithelial cells which deposit chitin and proteins to form an overlying extracellular cuticle (Page and Johnstone, 2007; Chapman, 1998; Cox et al 1981). Insects (and nematodes) are unable to synthesize sterols and thus depend upon dietary consumption of cholesterol for proper cuticular development (Monroe, 1960; Hieb and Rothstein, 1968). The composition and function of the insect integument is outlined in Table 1.1.
**Table 1.1** Overview of insect integument components and function

<table>
<thead>
<tr>
<th>Layer</th>
<th>Predominant Composition</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Epicuticle</td>
<td>Cement layer: Mucopolysaccharides and lipids</td>
<td>Protection of wax layer; not present in all insects</td>
</tr>
<tr>
<td></td>
<td>Wax layer: Hydrocarbons (grasshoppers); alcohols (caterpillars); free fatty acids (stone flies)</td>
<td>Waterproofing; intra/interspecific signaling</td>
</tr>
<tr>
<td></td>
<td>Epicuticle: Cuticulin (Highly polymerized lipids and proteins); tanned and untanned lipoproteins</td>
<td>Protection of procuticle from molting enzymes</td>
</tr>
<tr>
<td>Exocuticle</td>
<td>Anti-parallel/irregular chitin and sclerotized protein matrix</td>
<td>Mechanical rigidity; cuticle strength</td>
</tr>
<tr>
<td>Endocuticle</td>
<td>Uniform/parallel chitin and unsclerotized protein matrix</td>
<td>Cuticular strength</td>
</tr>
<tr>
<td>Epidermis</td>
<td>Epidermal cell: Short microvilli at apical membrane; extensive rough ER and Golgi complexes; pigment granules</td>
<td>Secretes cuticle, enzymes</td>
</tr>
<tr>
<td>Oenocyte</td>
<td>Large and polyploid; extensive endoplasmic reticulum</td>
<td>Lipid metabolism and synthesis of wax layer (hydrocarbons, sex pheromones)</td>
</tr>
<tr>
<td>Dermal gland</td>
<td>Single duct; microvilli – like structures at apical membrane</td>
<td>Production of cement layer</td>
</tr>
<tr>
<td>Pore canal</td>
<td>Cytoplasmic extensions of epidermal cells during/just after molt</td>
<td>Provides passage-way for epicuticle components</td>
</tr>
</tbody>
</table>

Data obtained from Chapman 2013
Molting in insects is a steroid-triggered behavioral process that strongly correlates with the rise and fall of the hormone ecdysone (Figure 1.3). The process begins when dietary cholesterol is converted to ecdysone and then modified into 20-hydroxyecdysone (20E) by the p450 enzymes (*spook, spookier, phantom, disembodied, shadow,* and *shade*) in the prothoracic gland of the animal (Gilbert et al. 2004; Warren et al. 2002). 20E is secreted into the target tissue via the hemolymph and enters the cuticular epidermal cells where it stimulates cell division and growth (Gilbert et al. 1999).

**Figure 1.3** Ecdysteroid titers across *D. melanogaster* lifespan. Pulses of 20-hydroxyecdysone stimulate holometabolous insect molting. Peak expression coincides with the intermolt phase while basal expression coincides with hatching or ecdysis. Adapted from Truman et al. 1994.

The increase in epithelial cell number is followed by apolysis or the separation of the overlying mature cuticle from the epidermal layer. The separation forms a space into which the epidermal cells secrete a gel containing attenuated proteases and chitinases. The epithelial cells secrete cuticulin which forms the new outer epicuticle. Cuticulin is composed of hardened lipoproteins and protects the underlying epidermal cells and newly
forming cuticle from degradation by the enzymes in the gel. Digestion of the old endocuticle occurs when the enzymes in the gel are activated. Most of the chitin and protein from the old endocuticle is reabsorbed by the epithelial cells and reused to form the new procuticle. The old exocuticle remains having been sclerotized during the animal’s previous molt. The oenocytes deposit a waxy layer on top of the new cuticle, and just before ecdysis the dermal cells secrete the cement layer. As a result of a surge in ecdysis-triggering hormone, the animal undergoes ecdysis whereby it loosens and sheds the old cuticle (Zitna et al 1996; Park et al 2002). When ecdysis occurs ecdysone titers are lowest, and within a few hours the newly formed cement layer and exocuticle undergo tanning (darkening) and sclerotization. Sclerotization is the cross-linking of the cuticle proteins and results in the hardening of the matrix. Synthesis of the unsclerotized endocuticle by the epidermal cells continues during the intermolt period until the next molt begins (Figure 1.4) (Chapman, 2013).
Figure 1.4 Schematic of the major changes in the insect integument during molting. The epidermal cells are stimulated by 20E, undergo cell division, and begin secreting components of the new procuticle. Following the cracking and shedding of the old cuticle, the new cuticle hardens, tans, and awaits the next molt. Adapted from Chapman 2013.
Pulses of the steroid 20E control molting and metamorphosis in insects via the ecdysone receptor: nuclear hormone receptor subunits Ultraspiracle (USP) and ecdysone receptor (EcR) (Koelle et al. 1992; Thomas et al. 1993; Yao et al. 1992). Nuclear hormone receptors are an evolutionarily conserved superfamily of transcription factors that when activated by a ligand, bind directly to specific genomic DNA sequences known as Hormonal Response Elements (HRE) (Schwabe and Rhodes, 1991). HRE reside upstream of target genes, usually other transcription factors, and regulate their expression. USP and EcR form a heterodimer and together with 20E make up the ecdysone receptor complex that binds directly to ecdysone response elements (ERE) (Thomas et al. 1993; Cherbas et al. 1990). As a result, a signaling cascade of gene activation occurs, beginning with early genes such as Broad Complex, E74 and E75 that encode transcription factors that both auto repress and simultaneously activate the late genes (Ashburner et al. 1974; Thummel, 1996).

**Nematode Molting**

Compared to insect molting, relatively less is known about the process in nematodes. Molting in nematodes such as *C. elegans*, while strongly suggestive of a nuclear hormone-mediated pathway, does not appear to involve ecdysone as the ligand. The process begins when the animal enters lethargus, a ~2 hour phase that, as the name implies, is characterized by a state of near inactivity. It is during lethargus that the new cuticle is synthesized from outermost layer to the innermost layer.

The nematode cuticle is synthesized by an underlying ectodermal cell layer which contains hypodermal cells that form discrete syncytia along the entire body surface and
specialized hypodermal cells called seam cells. Seam cells fuse during development and are located on the dorsal and ventral region of the animal’s hypodermis. The surface of the *C. elegans* (and *P. pacificus*) cuticle is characterized by regularly spaced circumferential corrugations called annuli which are present at all developmental stages. Longitudinally distending ridges called alae formed by the seam cells are present only during the L1 (J1), dauer, and adult stages (Cox et al. 1981; Singh and Suston 1978). The outermost layer of the adult *C. elegans* cuticle, the epicuticle, is composed of lipids and glycolipids, and is covered by a glycocalyx or surface coat secreted by gland cells (Nelson et al. 1983).

Below the epicuticle lies the cuticle which consists of a cortical, medial, and basal layer. The cortical layer contains highly cross-linked collagens and cuticulins secreted by the underlying hypodermis and is connected to the basal layer via the fluid-filled medial layer (Johnstone, 1994). The medial layer contains fibrous stems called struts which are located adjacent to each annuli and (notably) are only present during the adult stage (Riddle et al. 1991). The basal layer consists of two distinct strati: the outer, bi-layered collagenous layer that encircles the animal with circumferentially opposing angled fibers and an uncharacterized inner layer (Figure 1.5). Muscle contraction and motility is transmitted to the cuticle via filaments that connect the hypodermis to the body wall muscles (Francis and Waterston, 1991).
Studies of molting in *C. elegans* have identified two nuclear hormone receptors similar in principle to the ecdysone receptor found in *D. melanogaster*. *nhr-23* and *nhr-25* (*Nuclear Hormone Receptor*), orthologs of the ecdysone-induced early-late gene product DHR3 and the ecdyone-repressed bridge gene product Ftz-F1, respectively, affect the ability of the animal to undergo ecdysis (Kostrouchova et al. 1998; Sluder et al 1999; White et al. 1997; Broadus et al 1999). Loss of *nhr-23* and *nhr-25* function using RNA interference produces animals unable to completely disassociate from the old cuticle (the *mlt* phenotype) and/or bearing cuticular malformations (Kostrouchova et al. 1998; Asahani et al 2000). Both transcripts are expressed in the epidermis and oscillate in parallel with the molting cycle and synthesis of cuticular collagens (Kostrouchova et al. 1998; Kostrouchova et al. 2001; Gissendanner et al. 2004). Notably, the *nhr-23* gene
product binds HRE repeats in vitro in a pattern suggestive of heterodimeric partnering (Kostrouchova et al. 1998; Kostrouchova et al. 2001). Further, RNA interference of nhr-23 decreases reporter expression of an array of key molting epithelial or cuticular reporter genes (Kostrouchova et al. 1998; Asahani et al 2000; Frand et al. 2005). One such gene, mlt-10 is analyzed in this manuscript in depth.

Molting Behavior

Both insects and nematodes perform a sequence of stereotypical movements to loosen and discard the old cuticle. In insects, the movements are known as pre-ecdysis and ecdysis behavior. Pre-ecdysis behavior is characterized by regularly occurring bilateral abdominal compressions along the dorso-ventral axis. Ecdysis behavior is characterized by air-swallowing and rhythmic peristaltic contractions that start in the posterior abdominal segment and continue to the anterior segment. The entire process takes approximately an hour and results in the rupture and posterior removal of the old cuticle.

In nematodes like C. elegans and P. pacificus, each intermolt stage is followed by lethargus, an easily-observable, sleep-like behavioral state (Raizen et al. 2008). Animals undergoing lethargus are nearly inactive and unable to feed. During the first half of lethargus, in an unknown biochemical process, contact between the hypodermis and the overlying old cuticle is severed (apolysis) and the hypodermal cells begin synthesizing a new procuticle. Notably, during the L4/J4-to-adult molt, the old cuticle surrounding the anterior of the animal, referred to as a “buccal cap”, is particularly prominent and detectable using simple light microscope (Figure 1.6). During the second half of lethargus, the animal reverses and rotates longitudinally to loosen the old cuticle from the
new, and rupture of the cuticle surrounding the pharynx initiates ecdysis. For approximately 30 minutes the animal rocks back and forth until it ruptures the cuticle surrounding the head and crawls out as a next-stage larva (ecdysis) (Riddle et al. 1997; Singh and Soulston, 1978). The newly molted animal resumes normal feeding and locomotory behaviors and develops appropriately until the next juvenile/larval molt or adulthood.

**Figure 1.6** *P. pacificus* J4 larva hermaphrodite undergoing molt. During the J4/Adult molt, the old cuticle is particularly prominent and is referred to as the “buccal cap” (arrow).

*mlt-10*

MLT-10 (MoLTing defective) is a nematode-specific protein identified in a forward genetics screen for molting defective *C. elegans* animals (Frand et al 2005). The *mlt-10* transcript contains 8 exons and spans ~5 kilobases of Chromosome II; the protein consists of 690 amino acids (Wormbase.org). Protein sequence analysis of MLT-10 reveals a lysine rich domain of unknown function (DUF644) and a series of tandem proline-rich repeats (PRR) near the C-terminus (Meli et al. 2005). MLT-10 is conserved
among nematodes and a standard BLASTp search of the full-length sequence in NCBI identifies closest relatives *Caenorhabditis remanei* and *Caenorhabditis briggsae* as sharing 82% protein sequence identity (Altschul et al. 1990). Additionally, hypothetical protein orthologs of MLT-10 exist in parasitic nematodes *Wuchereria bancrofti* (40% identity) and *Brugia malayi* (39% identity) two causative agents of lymphatic filariasis in humans. A BLASTp search of the proline-rich repeat domain of MLT-10 (amino acids 514 to 690) significantly increases the identity scores of all aforementioned organisms. The importance of the uncharacterized PRR domain and its conserved role in nematodes will be discussed in Chapter III of this manuscript.

The *C. elegans* genome contains 13 *mlt-10* paralogs, *mltn-1* thru *mltn-13* (MoLt-TeN related) which appear as three gene clusters. All of the paralogs are expressed as transcripts; however they appear to be largely redundant as the individual knock-down of each paralog produces animals with a normal molting phenotype (Meli et al. 2005). The genomes of *C. remanei* and *C. briggsae* contain 11 and 13 putative *mltn* orthologs, respectively (Wormbase.org).

The loss or gain of *mlt-10* function affects the animal’s ability to properly shed the newly synthesized cuticle or develop to adulthood properly (Meli et al. 2010). The fluorescence of a *mlt-10* reporter tagged with a time-sensitive variant of *gfp* is first detected during late-embryogenesis as the L1 larva begins cuticle synthesis. The *gfp* signal localizes to the main body syncytia of the hypodermis approximately 3.5 hours prior to all four molts and increase in intensity for 3 hours during lethargus. Immediately following ecdysis, the signal drops to barely detectable levels where it remains until then next molting cycle occurs at which point the expression pattern repeats (or ceases when
the animal has reached adulthood) (Frand et al 2005; Meli et al. 2010). A MLT-
10::mCHERRY fusion protein was detected in secretory-like vesicles flanking the apical
membrane of hypodermal cells (Meli et al. 2010). Furthermore, knock-down of nhr-23 or
nhr-25 through RNA interference completely abrogates or drastically reduces mRNA
expression of the Cel-mlt-10p:gfp-PEST reporter, respectively (Frand et al. 2005; Meli et
al. 2010). Taken together the evidence suggests that MLT-10 is a conserved nematode-
specific secreted protein involved in the biosynthesis of new cuticle.

*Embryonic Molt*

*C. elegans* belongs to Diplogastroidea, a superfamily of nematodes characterized by,
among other traits, an embryonic molt. Unlike the nematode *C. elegans* which hatches as
an L1 (Larval 1) larva following an 18 hour embryogenesis, *P. pacificus* larvae undergo a
24 hour embryogenesis and hatch as J2 (Juvenile 2) individuals, having undergone a
J1/J2 molt while still inside the protective egg barrier (Fuerst von Lieven, 2005). Other
Diplogastroidea such as facultative insect parasite *Pristionchus iheritieri* and predatory
nematode *Butlerius degrissei* also molt a single time during embryogenesis prior to
hatching (Grootaert, 1976; Grootaert and Jaques, 1979).

The phenomenon of an embryonic molt is not exclusive to the nematode
Diplogastrids. In fact, it is well-known that many insects undergo an embryonic molt.
Insects such as grasshoppers, locusts, cockroaches, and beetles shed multiple cuticles
before hatching or concurrently with hatching (Schneider et al. 1957; Jones, 1956;
Bulliere et al. 1979; Micciarelli and Sbrenna, 1972). Additionally, ecdysone has been
found to trigger molting in the embryo in a similar cascade to the one found in the post-
embryonic development of insects. For example, ecdysteroid titers surge three times
during the first half of embryogenesis and a final time just before hatching of grasshopper
Locusta migratoria which results in the embryonic molt of first embryonic instar (E1)
into the pronymph and the embryonic molt of the 1st nymph cuticle, respectively (Truman
and Riddiford, 1999).

A schematic summarizing the process of molting as it pertains to the focus of this thesis is presented in Figure 1.7.
Figure 1.7 Summary of signaling cascade of molting in insects, *C. elegans*, and *P. pacificus*. A. Insect ecdysone receptor (EcR and USP) is bound by ecdysone and triggers the animal’s molting. B. *C. elegans* nhr-23 and nhr-25 (orthologs of DHR3 and FTZ-F1, respectively) affect expression of mlt-10 which results in the successful synthesis and removal of the cuticle. C. *P. pacificus* development is impeded by the application of exogenous ecdysone. Additionally the *P. pacificus* genome encodes the orthologs of EcR and USP (*Ppa-pnhr-1/2*) which oscillate with the molting schedule. It is unknown whether ecdysone is the ligand for the receptor and/or impacts molting. Additionally, it is unknown if *Ppa-pnhr-1* and *Ppa-pnhr-2* dimerize or control molting.
Hypothesis and Focus of Thesis

This thesis consists of the identification of Ppa-mlt-10, a putative mlt-10 ortholog in the P. pacificus genome. RT-PCR, sequence analysis, and a developmental expression profile of Ppa-mlt-10 performed by qPCR were used to test our hypothesis that Ppa-mlt-10 is the mlt-10 ortholog and provide the background data necessary for future studies aimed at elucidating the embryonic molt in P. pacificus. We also provide evidence that the P. pacificus genome encodes the mltn (MoLt TeN like) gene family and assign putative gene names based on phylogenetic positioning. We have also generated an ecdysone receptor fluorescence reporter, Ppa-pnhr-2p::gfp, suitable for microinjection into P. pacificus the results of which will enable our laboratory to carry-out spatial analysis of the gene and assess its relationship to molting. In that the signaling cascade that controls molting appears to differ within the nematode phylum, the current study reinforces our claim that P. pacificus is a viable comparative model to study molting in nematodes.
Chapter II

METHODS AND MATERIALS

*P. pacificus cultures*

*P. pacificus* (California strain—PS312) was used for all experiments. The animals were cultured at room temperature on standard nematode growth media (NGM) spiked with cholesterol and seeded with *E. coli* (OP50) as the nematode food source (Brenner, 1974; Stiernagle, 2006).

*Comparative Sequence Searches*

*Ppa-mlt-10, Ppa-hch-1, and Ppa-chs-1* were identified using the standard algorithm TBLASTN (Altschul et al. 1997) and the ‘HYBRID1 proteomics gene models, transcripts’ database to compare the MLT-10, HCH-1, and CHS-1 (*C. elegans*) protein sequences, respectively, to the Pristionchus.org library of predicted translated DNA sequences from *P. pacificus*. The most highly related predicted translated DNA sequences from the *P. pacificus* genome were compared to the GenBank library of all available protein sequences using the BlastX algorithm (Altschul et al. 1997).

The putative *Ppa-mltn* gene family was identified using the standard algorithm BlastP (Altschul et al. 1997) and the ‘HYBRID1 proteomics gene models, proteins’ database to compare the proline-rich repeat domain of PPA-MLT-10 (amino acids 597-776) to the Pristionchus.org library of predicted protein sequences from *P. pacificus*. The nine most highly related predicted proteins from the *P. pacificus* proteome were used in a phylogenetic analysis.
The genomic DNA of *Ppa-mlt-10* was identified using the BlastN (Altschul et. al. 1997) algorithm and the HYBRID1 Assembly (Sanger+454) contigs database to compare the putative cDNA *Ppa-mlt-10* sequence to the Pristionchus.org single contig, whole genome assembly. The two most highly related (and adjacent) genomic DNA sequences, were used to design nested upper and lower primers corresponding to the predicted 5’ and 3’ untranslated region of *Ppa-mlt-10* for a qualitative polymerase chain reaction experiment.

*Generation of full-length Ppa-mlt-10 by Reverse Transcription PCR (RT-PCR)*

Full-length cDNA of *Ppa-mlt-10* was generated using RNA obtained from a mixed-stage populations of worms and populations containing only eggs. RT-PCR reaction volumes contained 2X Apex PCR Master Mix, 0.2 μM of each primer, and deionized water for a total of 16 or 20 μL. The primary reaction contained 2 μL of template, and the secondary reaction contained 2 μL of a diluted (1:100) primary reaction as the template. Primers used were as follows: RHL294 (primary forward), RHL240 (primary reverse), RHL295 (secondary forward), RHL241 (secondary reverse). The conditions for the primary RT-PCR were as follows: 94°C for 3 minutes; 30 cycles of 92°C for 1:30 minutes, 55°C for 1:30 minutes, 72°C for 3 minutes; and a final elongation for 5 minutes. The conditions for the secondary RT-PCR were as follows: 94°C for 3 minutes; 35 cycles of 92°C for 1:30 minutes, 55°C for 1:30 minutes, and 72°C for 3 minutes; and a final elongation for 5 minutes. The secondary PCR product was purified using the DNA Clean and Concentrator kit (Zymo Research) according to the
manufacturer’s instructions and sequenced directly at Laragen (Culver City, CA) using the following primers: RHL345, RHL346, RHL347, MC016.

Comparative Sequence Analysis of MLT-10

Alignment of the following protein sequences was performed using the MUSCLE algorithm (Edgar, 2004) of MEGA5: Ppa-mlt-10, C. elegans (NP_493755.1), C. brenneri (XP_002648184.1), and C. remanei (XP_003097182.1) and best BLASTp sequences from the following species (as identified on wormbase.org): Ascaris suum (ASU_02697), Brugia malayi (Bm6119b and Bm2099b), Bursaphelenchus xylophilus (BUX.s01513 and BUX.s00725.1), Caenorhabditis angaria Cang_2012_03_13_001), Heterorhabditis bacteriophora Hba_12702), Loa loa (LOAG_11475), Meloidogyne hapla (MhA1_Contig438.frz3.g), and Strongyloides ratti (g6338_Sra). Similar protein sequences from Homo sapien (XP_005269394.1) and Tetraodon nigroviridis (gi 47221831).

Synchronization of nematode population

The worms from approximately 180 nearly starved mixed population plates were washed with deionized water and harvested in thirteen 15-mL polypropylene culture tubes and centrifuged at 2500 rpm for 3 minutes. The supernatant was discarded and 10 mL of a lysis solution (1 NaOH: 1 household bleach: 3 deionized water) was applied to the remaining worm pellet. The tubes were incubated at room temperature for 10 minutes and vortexed every two minutes for 10 seconds. Lysis of larvae was monitored using the dissecting microscope to ensure the remaining solution contained intact embryos. The
tubes were centrifuged again at 2500 rmp for 3 minutes. The lysis solution-supernatant was discarded and rinsed three times with deionized water to completely eliminate the lysis solution. The remaining embryos were suspended in 10 mL of S-basal (without a food source) and incubated at room temperature on a laboratory platform rocker for 24 hours. After 24 hours, the embryos had hatched and arrested development at the J2 stage. The 13 tubes containing the synchronously arrested J2 larvae were simultaneously spiked with 200 μL of liquid OP50 culture, lightly vortexed and centrifuged at 2500 rpm for 3 minutes. The tubes were aspirated to approximately 50 μL and lightly tapped to disturb the worm pellet. The remaining supernatant from each tube containing a concentrated J2 population was dropped onto 4-6 OP50-seeded NGM plates and incubated at room temperature. Staged samples were collected every four hours for 48 hours (from 0 to 48 hours) and prepared for RNA extraction.

**RNA Extraction**

To obtain the RNA content of each staged-sample, the synchronized worm populations were washed with deionized three times to completely remove the bacterial food source, suspended in 200 μL of TRizol and incubated at 4°C until all stages were gathered. The TRizol-suspended tubes were subjected to three freeze-thaw applications using liquid nitrogen and a 37°C water bath to ensure the cuticle is thoroughly broken. Total RNA from the worms was obtained using the in-column Direct-zol RNA MiniPrep kit (Zymo Research). The samples were subject to DNase treatment and the protocol from the kit was followed to obtain total RNA. The RNA concentration of each sample was measured using a standard nanodrop device. The samples were stored at -80°C. RNA
extraction of a mixed-stage population of animals and a population containing only embryos was performed using the same protocol.

**Complementary DNA Synthesis**

cDNA from 20 ng of total RNA for each synchronized worm sample was generated using the SuperScript VILO cDNA Synthesis kit (Life Technologies). The VILO kit is optimized for synthesizing first-strand cDNA used in real-time PCR experiments and uses random primers to prime the reverse transcription. cDNA of each staged worm sample was obtained following the manufacturer’s instructions and used in a real-time quantitative PCR experiment. Additionally, cDNA from total RNA of a mixed-stage population of animals and a population containing only embryos was synthesized using the same protocol.

**Comparative Sequence Analysis of mltn Gene Family**

A multiple protein sequence alignment of the putative *P. pacificus* mltn sequences and the proline-rich repeat domains of *mltn-1* thru *mltn-13* and their corresponding sequences in *C. remanei* and *C. briggsae* was produced with the MUSCLE algorithm in MEGA5.2 (default parameters). A phylogenetic tree was produced using the maximum-likelihood method (Jones-Taylor-Thornton model + r distribution) and 1000 bootstrap iterations.

**5’RACE of Ppa-pnhr-2**

To isolate the 5’ end of *Ppa-pnhr-2*—GenBank: GQ337701.1 (Parihar et. al. 2010), 6.3 ng of total RNA obtained from a mixed-stage population of worms was used in a 5’RACE (Rapid Amplification of cDNA Ends) experiment. The procedure was
performed using the First Choice RLM-RACE kit (Applied Biosystems) according to the manufacturer’s instructions. Reverse gene-specific primers RHL211 (outer) and RHL210 (inner) were used with the forward primers supplied by the kit to amplify the 5’ end of *Ppa-pnhr-2* in a nested PCR reaction. Five μL of the PCR product were visualized on a 2% denaturing agarose gel and the remaining product was purified using the DNA Clean and Concentrator kit (Zymo Research) according to the manufacturer’s instructions. The purified DNA was cloned into the pJET vector using the CloneJET PCR Cloning Kit (Thermo Scientific) according to the protocol supplied with the kit.

Competent bacterial cells (DH5α) were transformed with the vector according to standard cloning protocol. Confirmation of bacterial transformation was obtained by PCR and visualized using gel electrophoresis (2% agarose). A miniprep of the transformed bacteria was performed using the GeneJET Plasmid Miniprep Kit (Fermentas). A digest to confirm placement of the insert into the vector was performed using the BglII restriction enzyme. The plasmid (MC5.1) was sequenced at Laragen (Culver City, CA) using the primers supplied by the miniprep kit.

**Generation of transcriptional reporter *Ppa-pnhr::gfp***

To generate the transcriptional reporter *Ppa-pnhr-2::gfp*, the promoter insert on a diluted sample of plasmid MC5.1 was amplified using Phusion Taq Reaction Buffer and Phusion Taq Polymerase (Thermo Scientific), dNTP’s, and forward (RHL337) and reverse (RHL261) primers augmented with SalI and BamHI restriction sites, respectively. The PCR product was purified using the DNA Clean and Concentrator kit (Zymo Research) and digested with SalI and BamHI restriction enzymes. In parallel, the coding
sequence for gfp was cut out of a standard gfp vector using SalI and BamHI restriction enzymes. The denaturing gel containing the digested gfp product was purified using the ZymoClean Gel DNA Recovery kit (Zymo Research). Ligation of the Ppa-pnhr-2 promoter to the linearized gfp coding sequence was achieved using a standard laboratory T4 Ligase and buffer.

Competent bacterial cells (DH5α) were transformed with the Ppa-pnhr-2::gfp ligation product according to standard cloning protocol. Confirmation of bacterial transformation was obtained by PCR and visualized using gel electrophoresis. The miniprep of the transformed bacteria was performed using the GeneJET Plasmid Miniprep Kit (Fermentas). A digest to confirm placement of the insert into the vector was performed using the Xhol and EcoRI restriction enzymes. The plasmid (MC16.1) was sequenced at Laragen (Culver City, CA) using an internal primer (RHL338) to confirm proper ligation.

**Real-Time or quantitative PCR (qPCR)**

Quantitative PCR was performed on the Applied Biosystems 7300 Real-Time PCR thermocycler (Applied Biosystems) using the SYBR Green/ ROX qPCR Master Mix (Thermo Fisher). The master mix contains SYBER Green I intercalating dye, Taq DNA polymerase, dNTP’s and passive reference dye. A 96-well PCR plate was used to carry out the reaction. The qPCR reaction mixtures consisted of 9 μL of SYBR Green PCR Master Mix, 0.75 μL of forward and reverse primers, 1.0 μL of template cDNA in a total volume of 18 μL. Thermocycling was performed under the following conditions: 10 minutes of a polymerase activation stage at 90°C, followed by 40 cycles of 95°C for 15
seconds (denaturation), 55°C for 30 seconds (annealing), and 72°C for 30 seconds (elongation). The fluorescence was measured during the elongation step (72°C). A dissociation step which is used to establish a melting curve consisted of a single cycle of 95°C for 15 seconds, 60°C for 30 seconds, and 95°C for 15 seconds. The primers used to amplify *Ppa-mlt-10* were MC012 and MC013; the primers used to amplify *Ppa-β-tubulin* were AG11112 and AG11113; the primers used to amplify *Ppa-ama-1* were LS25505 and LS25504.

*Primers and primer design*

All primers used in the experiments described in this manuscript are outlined in Table 2.1.
### Table 2.1 List of Primers

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence (5’ to 3’)</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>RHL294</td>
<td>CCGAGTGGATCCTGGTCACATTTTC</td>
<td>Full-length <em>Ppa</em>-<em>mls</em>-10 forward primer (primary)</td>
</tr>
<tr>
<td>RHL295</td>
<td>ATGATCGATCGAGAGTCTGTCGC</td>
<td>Full-length <em>Ppa</em>-<em>mls</em>-10 forward primer (secondary)</td>
</tr>
<tr>
<td>RHL240</td>
<td>TGTAGGGCAGTTGGGAGAGCAGCAG</td>
<td>Full-length <em>Ppa</em>-<em>mls</em>-10 reverse primer (primary)</td>
</tr>
<tr>
<td>RHL241</td>
<td>AAGTGACAGAGTGCGCGAGCAACA</td>
<td>Full-length <em>Ppa</em>-<em>mls</em>-10 reverse primer (secondary)</td>
</tr>
<tr>
<td>MC016</td>
<td>CCTCTCCACCTGTCACTACAACCT</td>
<td>Full-length <em>Ppa</em>-<em>mls</em>-10 forward sequencing primer</td>
</tr>
<tr>
<td>RHL345</td>
<td>CCTTTACGGGTCTCCTTGGCATTGCTTGG</td>
<td>Full-length <em>Ppa</em>-<em>mls</em>-10 reverse sequencing primer</td>
</tr>
<tr>
<td>RHL346</td>
<td>CGCGATGGGGCATTGGCACTCATTTC</td>
<td>Full-length <em>Ppa</em>-<em>mls</em>-10 reverse sequencing primer</td>
</tr>
<tr>
<td>RHL348</td>
<td>CCTCTCGTCCCCTGATCCTTTTCTC</td>
<td>Full-length <em>Ppa</em>-<em>mls</em>-10 forward sequencing primer</td>
</tr>
<tr>
<td>MC012</td>
<td>GAAGCTCGTGCTGATGACGACGAG</td>
<td>rTPCR <em>Ppa</em>-<em>mls</em>-10 forward primer</td>
</tr>
<tr>
<td>MC013</td>
<td>CGATTCTCGATTCTGCAAGGAGAAG</td>
<td>rTPCR <em>Ppa</em>-<em>mls</em>-10 reverse primer</td>
</tr>
<tr>
<td>AG1112</td>
<td>CTGCGAGGAGGAACTGGAAC</td>
<td>rTPCR <em>Ppa</em>-β-tubulin forward primer</td>
</tr>
<tr>
<td>AG1113</td>
<td>GACGGTCAGAGAGAAGCTTAG</td>
<td>rTPCR <em>Ppa</em>-β-tubulin reverse primer</td>
</tr>
<tr>
<td>LS2505</td>
<td>CTGCAATGTCCAGTGGAAACGAC</td>
<td>rTPCR <em>Ppa</em>-amo-1 forward primer</td>
</tr>
<tr>
<td>LS2504</td>
<td>AATTTGACCCAGGGAGAACAC</td>
<td>rTPCR <em>Ppa</em>-amo-1 reverse primer</td>
</tr>
<tr>
<td>RHL337</td>
<td>atGTCGACGTCGCTCCGCGACAAGTGAG</td>
<td>Promoter of <em>Ppa</em>-phir-2 forward primer with Sall restriction site (bold)</td>
</tr>
<tr>
<td>RHL261</td>
<td>atGGAATCGCGAAGAGGGAAGGAGAAGGAGGAG</td>
<td>Promoter of <em>Ppa</em>-phir-2 reverse primer with BamHI restriction site (bold)</td>
</tr>
<tr>
<td>RHL338</td>
<td>GATAATGAGACGAGATGACGACAG</td>
<td><em>Ppa</em>-phir-2: gfp forward primer for fusion check</td>
</tr>
<tr>
<td>RHL211</td>
<td>GCATCGCTCAAGAGAGGAGGATTGAT</td>
<td>5’RACE <em>Ppa</em>-phir-2 reverse primer (secondary)</td>
</tr>
<tr>
<td>RHL210</td>
<td>ATGGCATGTACGAGGGATGTCTTG</td>
<td>5’RACE <em>Ppa</em>-phir-2 reverse primer (primary)</td>
</tr>
<tr>
<td>N/A</td>
<td>GCTGATGGGGGAGTGAAGAACACCTG</td>
<td>5’RACE reverse primer (primary) supplied by First Choice RLM-RACE kit</td>
</tr>
<tr>
<td>N/A</td>
<td>CGCGAGTCCCGAACCAGCTCGTGGCTTGGATG</td>
<td>5’RACE reverse primer (secondary) supplied by First Choice RLM-RACE kit</td>
</tr>
<tr>
<td>RHL304</td>
<td>GCGATGGATGAGGGAGATCGCG</td>
<td>Full-length <em>Ppa</em>-kels-1 (Contig 77.37) forward primer (primary)</td>
</tr>
<tr>
<td>RHL305</td>
<td>GATCGAGAGATCTGAGAAGGAG</td>
<td>Full-length <em>Ppa</em>-kels-1 (Contig 77.37) forward primer (secondary)</td>
</tr>
<tr>
<td>RHL306</td>
<td>GGTCAAGGAGAGATTTGAGCTCAC</td>
<td>Full-length <em>Ppa</em>-kels-1 (Contig 77.37) reverse primer (primary)</td>
</tr>
<tr>
<td>RHL307</td>
<td>CTCGAGAGATCTGTACCTAAC</td>
<td>Full-length <em>Ppa</em>-kels-1 (Contig 77.37) reverse primer (secondary)</td>
</tr>
<tr>
<td>RHL323</td>
<td>CTCGCTCCATGGTCACTACCTCAGT</td>
<td>Full-length <em>Ppa</em>-ksi-1 (Contig 139.18) forward primer (primary)</td>
</tr>
<tr>
<td>RHL324</td>
<td>CGGATTTGAGAAGATGGAGAGAGAAGGAG</td>
<td>Full-length <em>Ppa</em>-ksi-1 forward primer (secondary)</td>
</tr>
<tr>
<td>RHL325</td>
<td>CGATAGATGAGGAGAGTCTCGAGC</td>
<td>Full-length <em>Ppa</em>-ksi-1 (Contig 139.18) reverse primer (primary)</td>
</tr>
<tr>
<td>RHL326</td>
<td>CTCTGCTGGAGAGAGAATAAAC</td>
<td>Full-length <em>Ppa</em>-ksi-1 (Contig 139.18) reverse primer (secondary)</td>
</tr>
<tr>
<td>Ppa_SL1</td>
<td>GTTTTAATTACCCAAGGTGGAG</td>
<td><em>P. pacificus</em> splice leader 1 forward</td>
</tr>
<tr>
<td>Ppa_SL2b</td>
<td>TACCCAGTATCTCAGATTACGC</td>
<td><em>P. pacificus</em> splice leader 2b forward</td>
</tr>
</tbody>
</table>
Identification of Putative mlt-10 Gene in the *P. pacificus* Genome

To identify the *mlt-10* ortholog in *P. pacificus*, a bioinformatics technique commonly used in predicting putative orthologs known as a “reciprocal blast” was utilized. In a reciprocal blast, the sequence from a known gene or protein of interest is used to query the genome of an organism of interest for highly similar sequences. The sequence with the highest degree of similarity is used to back-query a set of genomes or the genome from which the known sequence originates. If the original gene or protein is retrieved, inference of putative orthology can be made.

First, the full-length protein sequence of *Caenorhabditis elegans* MLT-10 (CEL-MLT-10) was utilized to query translated sequences of the *P. pacificus* genome (pristionchus.org) using the basic local alignment search tool (BLAST). The output of such a search, informally referred to as a “snap number” identifies predicted open-frame coding DNA sequences in the *P. pacificus* genome with similarities to the query sequence using the Semi-HMM-based Nucleic Acid Parser (SNAP) gene-finder program (Korf 2004). The open-frame coding DNA sequence in the *P. pacificus* genome with the highest degree of similarity to *Cel-mlt-10* corresponds to a 2,331 base pair region, snap number Contig83-snap.49 (Figure 2.1).
Figure 2.1 Screenshot of TblastN results for Cel-mlt-10 query in pristionchus.org. Contig 83-snap.49 contains the *P. pacificus* coding sequence with the most significant similarity to *Cel-mlt-10*.

Second, the nucleotide sequence located on Contig83-snap.49 was translated and used to query all translated nucleotide sequences available in the National Center for Biotechnology Information (NCBI) database. In this search the *Cel-mlt-10* mRNA was identified as the sequence with the highest degree of similarity (Figure 2.2). This supports
the hypothesis that the genome of *P. pacificus* may contain the *Cel-mlt-10* ortholog and it is located on genomic region Contig83-snap.49.

**Figure 2.2** Cropped screenshot of BLAST results for Contig83-snap.49 query in the translated nucleotide NCBI database. The *Cel-mlt-10* mRNA is the sequence with the highest degree of similarity to the translated nucleotide sequence located on Contig83-snap.49 in the *P. pacificus* genome.
Amplification of Contig83-snap.49 mRNA using RT-RT-PCR

In order to amplify Contig83-snap.49 (putative Ppa-mlt-10) including the 5’ and 3’ ends, the Contig83-snap.49 sequence was used to query all genomic DNA sequences in the P. pacificus genome. The corresponding genomic DNA was found to be located on adjacent Contigs 83.29 and 83.30 (Figure 2.3) and the sequence was used as the template for a nested reverse transcription polymerase chain reaction (RT-PCR) experiment.

Figure 2.3 Screenshot of BLAST search for Contig83-snap.49 in the P. pacificus genome (pristionchus.org). The results indicate that the target is located on two adjacent Contigs, 83.29 and 83.30.

RT-PCR is a commonly used molecular biology technique employed to generate a measurable and homogeneous DNA product of a single or low-copy DNA sequence present within a heterogeneous cDNA sample. The thermocycling method subjects a sample containing the sequence template, DNA polymerase, dNTP’s and primers to a series (20 to 40 cycles) of heating and cooling treatments optimized for the amplification process. In brief and following a variable initialization step, the sample is heated to a temperature of 92-96 °C in order to denature all double-stranded DNA molecules into
single-stranded DNA. The sample is then cooled to 50-65 °C to facilitate the annealing of the primers and polymerase to the DNA template. The sample is then heated again to 72 °C to allow the extension of the DNA by recruiting complementary dNTP’s from the reaction mixture and yielding a newly synthesized double stranded DNA polymer. Ideally with each cycle, the number of target DNA molecules doubles, so that by the final extension step the sample contains an exponentially higher and measurable quantity of cDNA compared to the original sample.

The term “nested” refers to two separate RT-PCR experiments whereby the second reaction utilizes the product from the first PCR in a mixture with DNA polymerase, dNTP’s, and different primers. In that the primers of the second reaction contain sequences located 3’ inside of the primers from the first reaction, a nested RT-PCR enhances the specificity of the target amplification.

In order to ensure amplification of the entire transcript, including the start of translation and stop codon, nested primers were designed to anneal to sequences located in the predicted 5’ and 3’ un-translated regions of Contig83-snap.49 (Figure 2.4A). RT-PCR was performed using cDNA generated from an RNA sample isolated from a mixed stage population of *P. pacificus*. The product of the nested RT-PCR experiment was visualized on a denaturing agarose gel and represented a ~2.3 kilobase amplicon (Figure 2.4B).
Figure 2.4 Primer strategy and denaturing gel electrophoresis of Contig83-snap.49 (*Ppa-mlt-10*). A. The genomic DNA sequence was utilized in order to amplify the gene’s start and stop sites; therefore, the primer sequences corresponded to sequences located in the predicted 5’ and 3’ un-translated regions. B. The predicted 2,331 base amplicon was isolated from an RNA sample of mixed stage *P. pacificus* worms. Primers used are indicated below arrows.

Attempts at cloning the 2.3 kilobase PCR product were made however, only fragments of the transcript were recovered. Therefore, the PCR product was sequenced directly (Laragen) in a method known as “primer walking” and verified to match the target using gene editing software (Geneious). Additionally, a nested qRT-PCR was performed on cDNA generated from an RNA sample collected from an eggs-only population using the same primers and found to contain the same 2.3 kilobase target as the mixed stage population (Data not shown). This finding suggests that Contig83-snap.49 represents the *Cel-mlt-10* ortholog and is expressed as mRNA during embryogenesis and larval development. From this point, Contig83-snap.49 will be
referred to as \textit{Ppa-mlt-10}, per the standard gene-naming protocol adopted from the \textit{C. elegans} nomenclature scheme (Horvitz et. al. 1979).

\textit{Generation of Gene and Protein Structures of Putative \textit{P. pacificus} mlt-10}

In order to construct the gene structure (identify gene size, intron/exon boundaries, and exon number) of \textit{Ppa-mlt-10}, the predicted coding sequence of \textit{Ppa-mlt-10} was aligned with the corresponding genomic DNA sequence (Geneious). The putative \textit{Ppa-mlt-10} coding sequence contains 2,331 bases, spans a 5,802 base genomic DNA region and consists of 20 exons. The gene structure was visualized using a gene structure draw tool (www.bioinformatics.uni-muenster.de/tools/) (Figure 2.5).

\textbf{A} \quad \textit{Ppa-mlt-10}

\textbf{B} \quad \textit{Cel-mlt-10}

\textbf{Figure 2.5} Gene structures of \textit{mlt-10}. A. \textit{Ppa-mlt-10} structure was determined by comparison of cDNA sequence to \textit{P. pacificus} (California) genomic sequence. B. For comparison the \textit{mlt-10} gene structure (\textit{C. elegans}) was obtained from wormbase.org.

Our finding supports the evidence that \textit{P. pacificus} genes contain roughly twice the number of exons compared to \textit{C. elegans}—the coding sequence of \textit{Ppa-mlt-10} is 2.5 times longer than \textit{Cel-mlt-10} (Dieterich et al. 2008). Median \textit{Ppa-mlt-10} exon and intron
length are 47% and 65% shorter than *Cel-mlt-10*, respectively, compared to the global trend of 42% shorter exons and 37% longer introns (Dieterich et al. 2008). *Ppa-mlt-10* is a large gene—the length of the coding sequence is nearly four times greater than the median coding sequence length of all genes in the *P. pacificus* genome (Table 2.2).

### Table 2.2 Comparison of *Ppa-mlt-10* and *Cel-mlt-10* gene

<table>
<thead>
<tr>
<th>Gene Feature</th>
<th><em>P. pacificus</em></th>
<th><em>C. elegans</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of exons</td>
<td>20</td>
<td>8</td>
</tr>
<tr>
<td>Median exon length (bp)</td>
<td>94 (85)</td>
<td>178 (147)</td>
</tr>
<tr>
<td>Median intron length (bp)</td>
<td>168 (110)</td>
<td>477 (69)</td>
</tr>
<tr>
<td>Genomic length (bp)</td>
<td>5,802</td>
<td>5,062</td>
</tr>
<tr>
<td>Coding sequence length (bp)</td>
<td>2,331 (618)</td>
<td>2,073 (1,029)</td>
</tr>
</tbody>
</table>

The numbers in parenthesis represent the median length of the gene feature across the entire genome of each species (Dieterich et al. 2008).

In order to construct the protein structure of putative PPA-MLT-10, the translated sequence was aligned with MLT-10 (*C. elegans*) using the MUSCLE algorithm (default settings) in MEGA5.2 and compared to the structure of MLT-10 as identified by Meli et al. 2010. Patterns in the protein sequence were identified using PROSEARCH (workbench.sdsc.edu). The predicted PPA-MLT-10 protein consists of 776 amino acids, six Asparagine (N)-linked glycosylation sites, 17 protein-kinase C phosphorylation sites, 9 casein kinase II phosphorylation sites, and the conserved proline-rich repeat domain located from amino acid site 594 to the C-terminus (Figure 2.6).
Figure 2.6 MLT-10 protein structures (A) PPA-MLT-10. B) MLT-10. The DUF644 and serine-proline rich domains are highly conserved between the two nematodes. PPA-MLT-10 contains six N-linked glycosylation sites compared to four in C. elegans.

Multiple Sequence Alignment of MLT-10

The proline-rich region of putative PPA-MLT-10 protein was aligned using the Muscle (Edgar, 2004) algorithm in the open-source alignment program MEGA5.2 with protein sequences of three known MLT-10 proteins from the Caenorhabditis genus—C. elegans (NP_493755.1), C. brenneri (XP_002648184.1), and C. remanei (XP_003097182.1) and best BLASTp sequences from the following species (as identified on wormbase.org): Ascaris suum (ASU_02697), Brugia malayi (Bm6119b and Bm2099b), Bursaphelenchus xylophilus (BUX.s01513 and BUX.s00725.1), Caenorhabditis angaria Cang_2012_03_13_001), Heterorhabditis bacteriophora Hba_12702), Loa loa (LOAG_11475), Meloidogyne hapla (MhA1_Contig438.frz3.g), and Strongyloides ratti (g6338_Sra). Similar protein sequences from Homo sapien (XP_005269394.1) and Tetraodon nigroviridis (gi 47221831) were used as outgroups (Figure 2.7). PPA-MLT-10 shares 74% amino acid identity in the proline-rich-repeat region (Figure 2.8) and 39% amino acid identity in the DUF644 with MLT-10. Phylogenetic analysis was conducted using Maximum Likelihood (JTT plus r; 100 bootstrap replicates) (Figure 2.9).
Figure 2.7 Multiple Sequence Alignment of selected MLT-10 proteins. The proline-rich repeat C-terminal domain shows strong homology among select nematode species (Ppa: *P. pacificus*; Cel: *C. elegans*; Cre: *C. remanei*; Cbr: *C. brenneri*).
Figure 2.8 Proline-Rich C-terminal Domain of PPA-MLT-10 aligned with MLT-10. Three gaps were inserted in order to vertically align the repeats. There are 11 non-overlapping tandem proline repeats in the highly hydrophobic proline-rich region of Ppa-mlt-10.
Figure 2.9 Phylogenetic tree of positioning *P. pacificus* MLT-10 with other MLT-10 orthologs (known and hypothetical). The Maximum Likelihood phylogram (model GTR + gamma distribution) shows the evolutionary relationship of MLT-10 orthologs. Values on branches represent 100 replicates.
Developmental Expression Profile of Ppa-mlt-10

To generate a developmental expression profile of *Ppa-mlt-10* across the lifespan of *P. pacificus*, a real-time or quantitative polymerase chain reaction experiment (qPCR) was performed using cDNA obtained from synchronized RNA templates (see Methods). In brief, mixed-population NGM plates are bleached to obtain an isolated population of *P. pacificus* eggs. The eggs are allowed to hatch in the absence of nutrients where they develop to the J2 larval stage and synchronously undergo an arrest in development. The entire arrested population of J2’s is exposed to food and proceeds to develop normally. Every four hours for 48 hours (including an arrested J2 sample), whole animals are collected for their RNA content. Additionally, an egg RNA extraction was performed to be used as the endogenous control.

Due to its simplicity and efficiency, qPCR is currently the method of choice to measure RNA levels in a given sample. In order to run a successful qPCR and to obtain reliable results, three key elements are necessary: accurate Cycle threshold (Ct) measurements, stably expressed independent reference genes, and validation of a pure experimental sample.

To quantify the results of the qPCR, the comparative Cycle to Threshold (Ct) method (also known as the ΔΔCt method) was utilized. Generally speaking, the qPCR software quantifies the level of fluorescence, in this case, the SYBR Green fluorogenic dye, that logarithmically accumulates over a series of polymerase chain reaction cycles, usually 40, and determines the cycle threshold at which the fluorescence of the target exceeds the background fluorescence. A low Ct value (~≤29) indicates the sample contains higher amounts of target cDNA compared to a sample with a high Ct value.
which would indicate the sample contains little or no target cDNA. The comparative Ct method compares the Ct values of the target-of-interest to the Ct values of a control, either an untreated sample or RNA obtained from a specific stage. The Ct values of each staged experimental target and its corresponding control are then normalized to independent reference genes. The $\Delta\Delta$Ct calculation used in this experiment is as follows:

$$\Delta\Delta\text{Ct} = (\text{Ct}_{\text{target}} - \text{Ct}_{\text{reference gene}}) - (\text{Ct}_{\text{control}} - \text{Ct}_{\text{reference gene}})$$

The $\Delta\Delta$Ct values for ten technical replicates were standardized by log transformation to obtain fold change, averaged, and auto-scaled as described by Willems et. al. for the final expression profile of $Ppa-mlt-10$.

In order to choose a stably expressed reference or “house-keeping” gene for the qPCR, a simple qPCR was conducted using primers for $Ppa-y45F10D4$, $Ppa-cdc-42$, and $Ppa-ama-1$ on a mixed-stage cDNA template. In the only published paper that discusses the selection and validation of qPCR reference genes in $P.\text{pacificus}$, these transcripts were found to be the most stably expressed (Schuster and Sommer, 2012). The result of the qPCR identified $Ppa-ama-1$ as having the most robust expression under the tested conditions and was selected as a reference gene for the qPCR experiment. $Ppa-ama-1$ encodes the highly conserved large subunit of RNA polymerase II, shares 83% identity with $C.\text{elegans ama-1}$ and the Homo sapien homolog RPB1. $ama-1$ is commonly used as a house-keeping gene in qPCR experiments in $C.\text{elegans}$. Another reference gene, $Ppa-\beta$-tubulin was chosen based on its use as a reference gene in a qPCR experiment.
quantifying the expression of *Ppa-egl-4* (*P. pacificus* Egg-Laying defective) a protein kinase involved in insect pheromone attraction (Hong, Witte, and Sommer, 2008). A qPCR experiment using primers designed to amplify the most conserved region of *Ppa-β-tubulin* in developmentally synchronized populations of worms confirmed that the transcript is stably expressed across the lifespan of *P. pacificus*. (Data not shown).

To determine the purity of the sample, a melting curve was generated by the qPCR software. A melting curve is obtained by adding a dissociation step after each cycle and visualizing the temperature at which 50% of the double-stranded DNA becomes single-stranded. A single distinct peak indicates the sample is pure and contains a single product, whereas more than one peak suggests the sample is contaminated with multiple products such as off-target cDNA, primer-dimers, or genomic DNA, in addition to the intended target. The presence of single and distinct peaks for *Ppa-mlt-10*, *Ppa-β-tubulin*, and *Ppa-ama-1* in the melting curve analyses indicates the RNA samples are pure and contain the intended PCR product (Figure 2.10)

![Melting curves](image)

**Figure 2.10** Melting curves *Ppa-β-tubulin* (A); and *Ppa-mlt-10* (B)
To increase the specificity of the amplification of *Ppa-mlt-10* and to decrease the likelihood of amplifying off-target cDNA or gDNA, forward and reverse primers that span two adjacent exons were designed (Figure 2.11). Primers MC012 (forward) and MC013 (reverse) were used to amplify *Ppa-mlt-10* in the qPCR experiment.

**Figure 2.11** Diagram of primer design strategy for amplification of *Ppa-mlt-10* using qPCR. Primers (arrows) that span two adjacent exons increase the likelihood of amplifying the target rather than off-target cDNA or gDNA.

Since the data reflect the relative expression of *Ppa-mlt-10* for a single biological sample representing one biological condition or synchronization event, statistical analysis cannot be performed. Generally speaking a minimum of three biological replicates is necessary to determine the mathematical significance of relative gene expression in qPCR experiments in which case a student T-test can be performed (Goni et. al. 2009).

Relative expression of *Ppa-mlt-10* was found to oscillate across the lifespan of *P. pacificus* with a greater than 3000-fold increase in peak expression occurring immediately prior to each molting event, particularly the J2/J3 molt and J3/J4 molt.
Conversely, expression of \textit{Ppa-mlt-10} drops dramatically at the 24 hour time point (Figure 2.12).

\textbf{Figure 2.12} Result of qPCR experiment. A. Larval expression of \textit{Ppa-mlt-10}. Transcript expression levels were normalized against \textit{Ppa-\beta-tubulin} and \textit{Ppa-ama-1} for a total of 10 technical replicates. Error bars denote the standard error. B. Timeline of synchronized J2 \textit{P. pacificus} animals following exposure to food (adapted from Parihar et. al. 2010).

This result is consistent with the observation by Frand et. al. 2005 of a 100-fold increase in \textit{gfp} fluorescence of transgenic animals [\textit{Ex(Cel-mlt-10::gfp pest)}] prior to each molt compared to the absence of \textit{gfp} expression at the intermolt period. Together with the multiple sequence alignment of \textit{Ppa-mlt-10} with three \textit{Caenorhabditis mlt-10} proteins, phylogenetic analysis, and protein domain comparative, these findings suggest
that *Ppa-mlt-10* is differentially expressed throughout the larval development of *P. pacificus* and represents the ortholog of *Cel-mlt-10*.

*Identification of Ppa-mltn Family of Paralogs*

To identify the *mltn* paralogous gene family in *P. pacificus*, the proline-rich repeat domain of putative PPA-MLT-10 (amino acids 597-776) was used to query translated sequences in the *P. pacificus* genome (pristionchus.org) using BLASTp (Figure 2.13). Generally speaking gene families arise from a series of duplication, transposition, or unequal crossover events of a gene or genes.
Figure 2.1 Screenshot of BLASTp query of proline-rich repeat domain of putative PPA-MLT-10. As expected the query retrieved Contig83-snap.49 as the best match. Sequences with scores >50 were identified as putative *P. pacificus* mltn genes. The red arrow indicates the cutoff point at which significance was deemed too low.
Because the majority of the *P. pacificus* genome has not yet been annotated or mapped, the genomic location of a sequence identified by the Contig-snap is arbitrary and remains so until experiment proves otherwise. Using the proline-rich repeat domain rather than the entire *Ppa-mlt-10* protein sequence produced a series of individual significant Contig-snap sequences rather than multiple tandem significant sequences that may or may not be genomically linked (See Figure 2.1 for example). Nine sequences with significant similarities to the proline-rich repeat region of *Ppa-mlt-10* (Contig83-snap.49) were identified as putative *Ppa-mltn* (*Pristionchus pacificus* MoLt Ten-like) paralogs (Table 2.3). The cutoff point at which significance was deemed too low was determined by the following criteria: 1. Score <50; 2. E value > -10; and 3. Lack of proline repeat consensus SPX₄P.

**Table 2.3 Best mltn BLASTp *P. pacificus* sequences**

<table>
<thead>
<tr>
<th><em>P. pacificus</em> Genomic Location</th>
<th>Score (bits)</th>
<th>E value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contig83-snap.49</td>
<td>350</td>
<td>1e-97</td>
</tr>
<tr>
<td>Contig97-snap.14</td>
<td>202</td>
<td>7e-53</td>
</tr>
<tr>
<td>Contig4-snap.488</td>
<td>133</td>
<td>4e-32</td>
</tr>
<tr>
<td>Contig43-snap.178</td>
<td>131</td>
<td>2e-31</td>
</tr>
<tr>
<td>Contig30-snap.11</td>
<td>119</td>
<td>8e-28</td>
</tr>
<tr>
<td>Contig75-snap.126</td>
<td>105</td>
<td>9e-24</td>
</tr>
<tr>
<td>Contig37-snap.66</td>
<td>95</td>
<td>1e-20</td>
</tr>
<tr>
<td>Contig50-snap.85</td>
<td>86</td>
<td>7e-18</td>
</tr>
<tr>
<td>Contig45-snap.90</td>
<td>82</td>
<td>1e-16</td>
</tr>
<tr>
<td>Contig78-snap.37</td>
<td>65</td>
<td>1e-11</td>
</tr>
</tbody>
</table>
A multiple protein sequence alignment of the translated putative *P. pacificus* MLTN sequences with the proline-rich repeat domains of MLTN-1 thru 13 (wormbase.org) starting with the first SPX₄P consensus was produced with the ClustalW algorithm in MEGA5.2 (default parameters). MLT-10 of *C. elegans*, *C. remanei*, *C. briggsae*, and *P. pacificus* were used as outgroups. A phylogenetic tree was produced using the maximum-likelihood method (Jones-Taylor-Thornton model + Γ distribution) and 100 bootstrap iterations (Figure 2.14)
Figure 2.14 Phylogram of the newly identified *P. pacificus* MLT-10 and MLTN proteins. The tree shows the evolutionary relationship of the putative PPA-MLTN sequences with known *C. elegans* mltn family. The proline-repeat domains of MLT-10 *C. elegans* (Cel), *C. remanei* (Cre), *P. pacificus* (Ppa), and *C. briggsae* (Cbr) were used as the outgroups. Bootstrap values are from 100 iterations.
Based on the topology of the tree and a distance matrix, the most similar C. elegans MLTN sequences were identified and putative gene names were assigned where possible.

Table 2.4 Putative *P. pacificus* *mltn* gene family and *C. elegans* paralogs

<table>
<thead>
<tr>
<th><em>P. pacificus</em> Genomic Location</th>
<th><em>P. pacificus</em> <em>mltn</em> gene name</th>
<th>Most similar <em>C. elegans</em> <em>mltn</em> gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contig83-snap.49</td>
<td><em>Ppa-mlt-10</em></td>
<td><em>mlt-10</em></td>
</tr>
<tr>
<td>Contig97-snap.14</td>
<td><em>Ppa-mltn-3</em></td>
<td><em>mltn-4</em></td>
</tr>
<tr>
<td>Contig4-snap.488</td>
<td><em>Ppa-mltn-8</em></td>
<td><em>mltn-12</em></td>
</tr>
<tr>
<td>Contig43-snap.178</td>
<td>Unresolved</td>
<td><em>mltn-9</em></td>
</tr>
<tr>
<td>Contig30-snap.11</td>
<td><em>Ppa-mltn-12</em></td>
<td><em>mltn-12</em></td>
</tr>
<tr>
<td>Contig75-snap.126</td>
<td>Unresolved</td>
<td><em>mltn-9</em></td>
</tr>
<tr>
<td>Contig37-snap.66</td>
<td>Unresolved</td>
<td><em>mltn-13</em></td>
</tr>
<tr>
<td>Contig50-snap.85</td>
<td>Unresolved</td>
<td><em>mltn-12</em></td>
</tr>
<tr>
<td>Contig45-snap.90</td>
<td>Unresolved</td>
<td><em>mltn-13</em></td>
</tr>
<tr>
<td>Contig78-snap.37</td>
<td>Unresolved</td>
<td>None</td>
</tr>
</tbody>
</table>

*Identification of the 5’ End of Ppa-pnhr-2 Using RLM-R.A.C.E.*

In order to confirm the 5’ end of the mRNA *Ppa-pnhr-2* transcript, RNA Ligase Mediated Rapid Amplification of CDNA Ends (RLM-RACE) (Ambion) was performed on RNA obtained from a mixed-stage population of worms. The RLM-RACE protocol differs from the classic RACE in that the polymerase chain reaction-based technique amplifies only full-length, capped mRNA rather than any cDNA template which may or may not contain the 5’ end of the target mRNA. In a series of steps, the RLM-RACE method requires total RNA to be treated with calf-intestine phosphatase and tobacco acid pyrophosphatase to remove the 5’ free phosphate group and cap from all precursor RNA molecules, respectively. Next a 5’RACE Adaptor is added to the 5’end of the altered
RNA using a T4 ligase and the entire treated RNA sample is subject to standard random decamer reverse transcription. Finally, two gene-specific reverse (antisense) primers are used with two forward (sense) primers corresponding to the 5’ adapter sequence in a two-step nested PCR experiment. The PCR product is visualized on an agarose gel and can be cloned or sequenced directly. The result of our RLM-RACE experiment confirms that the PCR product contains the 5’ end of *Ppa-pnhr-2* (Figure 2.15).

![Gel electrophoresis image of 5’RACE of *Ppa-pnhr-2*. The lane containing the inner positive template PCR product indicates the presence of the desired product. Outer +/- lanes contain the PCR product of the outer nested PCR reaction. The inner -T lane contains the inner reaction without the cDNA template.](Image)

Initially, a classic RACE technique was utilized to obtain the 5’ end of *Ppa-pnhr-2*; however this technique did not yield a positive result. The classic RACE method when used with nematodes such as *P. pacificus* and *C. elegans*, subjects untreated total or Poly(A)-selected RNA to a PCR experiment using gene-specific reverse primers with a forward primer corresponding to a highly conserved 22-nucleotide spliced leader (SL) sequence. More than half of all *C. elegans* mRNAs are trans-spliced at the 5’ end to one of ten spliced leaders, the most common of which is SL1 and the same has been found to be true for *P. pacificus* (Blumenthal, 1995; Lee and Sommer, 2003). The use of one of
the conserved spliced leader sequences as the forward primer in a 5’RACE experiment is common and often elucidates the 5’ ends of trans-spliced transcripts. In two separate RACE experiments on RNA obtained from mixed stage *P. pacificus* worms, primers corresponding to the SL1 and SL2 sequences were used as the forward primers along with reverse *Ppa-pnhr*-2-specific primers. The PCR products were visualized on denaturing agarose gels, the results of which suggest a heterogeneous population of non-specific amplicons (Data not shown).

*Generation of Transcriptional Reporter Ppa-pnhr-2p::gfp*

In order to analyze promoter activity of *Ppa-pnhr*-2, a micro-injectable transcriptional reporter was created using traditional molecular biology cloning technique (See Methods). The promoter region of *Ppa-pnhr*-2, ~1.8 kilo bases upstream of the start of translation, was ligated to the coding sequence of the 5’ end of green fluorescent protein (gfp) to generate the *Ppa-pnhr*2p::gfp construct (Figure 2.16). In nematodes such as *P. pacificus*, microinjection is used to visualize promoter activity of a gene of interest. An injection mixture which consists of the transcriptional reporter construct (in this case, *Ppa-pnhr*-2p::gfp) *P. pacificus* genomic DNA, and a co-injection marker, is injected into the distal tip gonad of individual worms. The transcriptional reporter may exist in the nascent embryo as an extra-chromosomal array or be heritable in subsequent generations. The F2 generation of the injected individual is analyzed for gfp expression. Stable lines bearing the extra-chromosomal array can be obtained through application of gamma or UV irradiation to the mutant strain, microparticle gene bombardment, or coinjection with a single-strand oligonucleotide followed by outcrossing with wild-type males.
Initially, the less labor-intensive protein fusion technique was attempted to create a gfp reporter construct suitable for microinjection. This PCR-based protocol is commonly used in C. elegans with great success and utilizes three separate qPCR experiments. First, the gfp coding sequence is amplified from a standard Fire gfp vector. Concurrently, the promoter region for a gene of interest is amplified from genomic DNA using a sequence-specific 5’ primer and a 3’ sequence-specific primer containing a 24 nucleotide overhang complementary to the standard Fire gfp vector. Lastly, the two PCR
products are utilized in a final PCR experiment which anneals the two templates together at the 24 nucleotide overhang and amplifies the single promoter::gfp product (Hobart 2002). This method while successful in producing the initial PCR products did not yield a final fusion product, as visualized on an agarose gel (Data not shown).

Identification of chs-1 and hch-1 orthologs in P. pacificus using RT-PCR

In order to identify genes in P. pacificus that may be exclusively expressed during embryogenesis for use as early and late embryonic markers, sequences from two C. elegans proteins were used to query translated nucleotide sequences in the P. pacificus genome using a reciprocal BLAST: chs-1(CHitin Synthase) and hch-1 (defective HatCHing). chs-1 is required for a number of early embryonic processes, specifically chitin and eggshell synthesis, meiotic accuracy, and establishment of the anterior/posterior orientation in the newly fertilized zygote (Veronico et al. 2001; Johnston et al. 2006). hch-1 is expressed during late-embryogenesis and required for normal hatching in C. elegans (Hishida et al. 1996). Our results show that while the genome of P. pacificus contains putative orthologs for chs-1 and hch-1, neither transcript is expressed solely during embryogenesis thus neither of these genes is useful as an embryonic marker (Figures 2.17 and 2.18).
Figure 2.17 Gel electrophoresis results of putative *Ppa-chs-1* expression in egg and J4 populations. A. Expression of *Ppa-chs-1* using cDNA obtained from egg population and generated using reverse transcriptase (Lane 1), cDNA obtained from egg population without reverse transcriptase (Lane 2), cDNA obtained from J4 population generated using reverse transcriptase (Lane 3), cDNA obtained from J4 population without reverse transcriptase (Lane 4), *P. pacificus* genomic DNA. The predicted length of the target is 774 base pair (cDNA) and 839 (gDNA). B. The quality of the cDNA used in the RT-PCR was verified using the well-characterized *Ppa-obi-1* transcript.
**Figure 2.18** Gel electrophoresis results of putative *Ppa-hch-1* expression in egg and J4 populations. A. Expression of *Ppa-hch-1* using cDNA obtained from egg population and generated using reverse transcriptase. B. Expression of *Ppa-hch-1* using cDNA obtained from a mixed stage population (Lane 1) and J4 population (Lane 2). NOTE: The mixed stage cDNA visualized in Lane 1 may have been of poor quality; therefore no reaction is observed to have taken place.
Molting in nematodes is the regular and repeated shedding of an exoskeleton or cuticle throughout the lifespan of the worm. In most circumstances such as in the nematode *Caenorhabditis elegans*, the worm undergoes four molts during larval development. However, it is hypothesized that the beetle-associated nematode *Pristionchus pacificus* undergoes three larval molts and one molt prior to hatching known as the embryonic molt. The embryonic molt, while a generally accepted phenomenon in some nematodes, remains an imprecise observation. Our laboratory seeks to characterize this proposed heterochronic shift in development in *P. pacificus* beyond simple imaging using the tools of comparative genomics and molecular biology. Furthermore and drawing analogy from the molting of insects, we wish to elucidate the *P. pacificus* molting pathway. In that the genome of *P. pacificus* encodes the homologs of the well-characterized insect ecdysone receptor (*Ppa-pnhr-1* and *Ppa-pnhr-2*), we wish to explore and characterize a putative ecdysone signaling pathway in *P. pacificus*. The necromenic nematode and laboratory-amenable *P. pacificus*, while not a parasite categorically, shares many characteristics with its parasitic counterparts (ecdysone sensitivity, nematode predation, embryonic molt, etc.) and thus presents as a useful laboratory tool to study processes that may have global application.

In this thesis we propose that the genome of *P. pacificus* encodes the ortholog of *mlt-10*, and we therefore refer to the gene that encodes this protein as *Ppa-mlt-10*. Reciprocal BLAST analysis and multiple protein sequence alignment of putative PPA-
MLT-10 with MLT-10 of three *Caenorhabditis* species (*C. elegans*, *C. briggsae*, and *C. remanei*) indicates the proline-rich repeat C-terminal domain shares high amino acid identity. Additionally, regions of conservation are detected throughout the DUF644. The result of a preliminary qPCR experiment measuring the expression pattern of *Ppa-mlt-10* across the lifespan of the animal suggests that *Ppa-mlt-10* is expressed as mRNA and vacillates synchronously with the first two of three larval ecdyses. These results are similar to the developmental profile of *mlt-10* in *C. elegans* and support the present hypothesis.

We also show that the genome of *P. pacificus* encodes the putative *mltn* gene family. Comparative sequence analysis of the proline-rich repeat C-terminal domain of nine putative *P. pacificus* *mltn* proteins with the *mltn* gene family of *C. elegans* positions the newly identified *P. pacificus* in a Maximum Likelihood tree. We identified the 5’UTR of *Ppa-pnhr-2* using 5’RACE, the results of which were used to generate the transcriptional reporter construct *Ppa-pnhr-2::gfp*. Finally, we identified putative *hch-1* and *chs-1* orthologs in *P. pacificus* using a comparative BLAST approach; however, based on the results of an RT-PCR experiment we rejected our hypothesis that these transcripts were expressed exclusively during embryogenesis.

These results mark a starting point from which to study molting in *P. pacificus*. For instance a *Ppa-mlt-10* transcriptional reporter (*Ppa-mlt-10::gfp*) could be generated by cloning the promoter region of *Ppa-mlt-10* in front of the coding sequence of *gfp*. The purified fluorescence construct is co-injected into the young adult animal and could be used to visualize the first molt that occurs inside the *P. pacificus* egg. Furthermore, a transgenic line bearing a stably expressed *Ppa-mlt-10::gfp* could be produced using
standard gamma or UV irradiation protocols. Such a transgenic strain can be used as a control for molting in many experiments.

The *Ppa-pnhr-2::gfp* transcriptional reporter described in this manuscript, while useful in identifying the spatial dynamics of promoter activity, is not useful in identifying the developmental expression pattern of *Ppa-pnhr-2*, or any promoter for that matter. It is well-known that wild type *gfp* is highly stable and has a half-life of ~26 hours in mouse cells (Corish and Tyler-Smith, 1999). In order to circumvent this obstacle, a PEST (proline, glutamic acid, serine, and threonine) sequence is often inserted at the C-terminus immediately prior to the stop codon of *gfp*. The PEST sequence is a proteolytic signal sequence and marks the product for degradation, reducing the half-life of *gfp* to 9.8 hours (Corish and Tyler-Smith, 1999). Currently, the *gfp*-PEST fusion is commonly used in reporter assays in *C. elegans*.

Given the goals of our laboratory, to elucidate the dynamic regulation of molting in *P. pacificus* with particular attention to the embryonic molt, *Ppa-mlt-10::gfp-PEST* and *Ppa-pnhr-2::gfp-PEST* transcriptional reporter would be more suitable.

Embryogenesis of *P. pacificus* spans ~24 hours and the first molt is predicted to take place ~7 hours prior to hatching. This leaves a narrow window in which to visualize a fluorescence reporter. In simple terms, the less stable *gfp*-PEST would be more useful. Notably, an N-terminal fusion of a mouse cyclin destruction box to *gfp* combined with the C-terminal PEST sequence further reduces the half-life of *gfp* to 5.5 hours in mouse cells (Corish and Tyler-Smith, 1999). Adopting this *gfp* cloning technique could prove to be a useful laboratory tool.
Of interest to the author is the highly conserved proline-rich region of putative PPA-MLT-10. The tandemly repeated region of PPA-MLT-10 is a 179-residue domain at the C-terminus (aa598 to aa776). In total this region contains 33 prolines, compared to 26 prolines within residues 1 to 597. There are 11 non-overlapping, irregularly spaced repeats consisting of 7 amino acids with consensus sequence SPX₄P although there are 17 repeats in total. This may allow for pleomorphy in the protein’s conformation. Each proline is always preceded by a serine and the general consensus sequence ends with a proline (Figure 3.1). A TBlastN query identifies many predicted coding sequences containing (SPX₄P)ₙ. Sequences found in *Plasmodium falciparum* (XM 002808631.1) and *Anaplasma phagocytophilum* (CP006618.1) share 53% and 45% identity with the proline-rich C-terminal region of *Ppa-mlt-10*.

This discrete domain can be assigned to the broader protein category known as Proline Rich Regions (PRR). The proline amino acid is unique in its conformational rigidity and without venturing beyond the scope of this discussion is often positioned at the N-terminal region of a helix (Williamson, 1994). In fact, overall PRR’s are highly unmalleable and are associated with rapid binding (Williamson, 1994). Some well-known proteins with tandemly repeated proline rich motifs are mucin, elastin, gluten, and collagen, to name a few. Some contain proline repeats that span the entire amino acid while some contain one or two small domains.

Three-dimensional protein modelling of the proline-rich domain of PPA-MLT-10 using the freeware I-TASSER (http://zhanglab.ccmb.med.umich.edu/I-TASSER/) predicted the tertiary structure as a left-handed quadrilateral β-helix known as a pentapeptide repeat protein (PRP) (Zhang, 2008) (Figure 3.2). PRP’s are a vast family of
proteins found in prokaryotes and eukaryotes, the function of which remains unknown. Many bacterial PRPs appear to confer resistance to fluoroquinolones by inhibiting DNA gyrase (Vetting et. al. 2006). An Arabidopsis PRP found in the thylakoid membrane is involved with glycolipid localization and assembly (Kieselbach et. al. 1998).

Figure 3.2 Predicted 3-D ribbon models of PRR of PPA-MLT-10 using ITASSER. A longitudinal view of the predicted three dimensional ribbon model of proline-rich repeat domain of PPA-MLT-10; C-score: -1.11 (A); cross section (B) The C-score is used to assess the accuracy of the predicted model. C scores range from -5 (low confidence) to 2 (high confidence).

Three-dimensional protein modelling of the DUF644 of putative PPA-MLT-10 using I-TASSER predicted the tertiary structure found in Figure 3.3. (Zhang, 2008). The structural analog to the DUF644, i.e. the known Protein Data Bank crystallized structure with the highest similarity to the query, was identified as BBA73, an outer surface lipoprotein found in the spirochete bacteria Borrellia burgdorferi. In order to confirm the prediction results, we would need to crystallize the PPA-MLT-10 protein.
Figure 3.3 Predicted 3-D structure of putative PPA-MLT-10 DUF using I-TASSER. The three dimensional model of the DUF644 of PPA-MLT-10 was generated using the template of the most similar structural analog, outer surface lipoprotein BBA73 from *Borrelia burgdorferi* (C-score: -3.17).

The Maximum Likelihood phlyogram of the *mltn* gene family implies an expansion of the basic SPX₄P protein module through a series of duplication events prior to the speciation of the *P. pacificus* and *Caenorhabditis mlt-10* proteins as we know them today. The most common ancestor of all sequences analyzed contained the truncated SPX₄P consensus, and it appears that as the SPX₄P modules duplicated and increased in number, the C-terminal domain became the functional unit that characterizes *mlt-10*. In fact, as previously mentioned RNA interference of each individual *mltn* gene indicates the transcripts are redundant and produces no observable function (Meli et al. 2010). Based on this, we hypothesize that the putative *P. pacificus mltn* genes identified here (with the exception of Contig97-snap.14) are also redundant. According to the tree topology, *Ppa-mlt-10* recently underwent a gene duplication event to produce the sequence located on Contig97-snap.14. Our experiments show that Contig97-snap.14 is
expressed as mRNA (data not shown); however further experimental evidence is needed to determine its function, necessity, and relationship to molting. Phylogenetic analysis beyond what is mentioned here surpasses the scope of this thesis.

Taken together our literature review, experimental results and 3-D models generated by I-TASSER suggest that PPA-MLT-10 may be an extracellular matrix lipoprotein. During molting, such a molecule may be transported in secretory vesicles to the apical membrane of nematode hypodermal cells and secreted into the extracellular space. The molecule may exist as a monomer or as an oligomeric quaternary helical complex as visualized in the ITASSER 3-D protein model. In such a case, PPA-MLT-10 may serve to protect the underlying procuticle from degradation by molting enzymes as some kind of lubricant or gel. Interestingly, the sequence most similar to MLT-10 in humans is a predicted Mucin Isoform 4 (using BLASTp) that contains a C-terminal domain consisting of SPX2P consensus repeats. In that mucins are proteins secreted by epithelial cells, heavily glycosylated, and serve a protective role to the underlying epidermis, a functional homology could be assigned to MLT-10 and is worth investigating.

Many mucins are pathogen-binding and protect the underlying epidermis from infection and in theory mlt-10 may bind directly to bacterial or fungal pathogens. *Staphylococcus aureus, Pseudomonas aeruginosa* and *Salmonella enterica* have been found to deleteriously infect *C. elegans*; therefore, it is reasonable to propose that these microbes also infect *P. pacificus*. To test the hypothesis that Ppa-mlt-10 directly binds to a pathogenic bacterial cell such as the Gram-positive *S. aureus*, a relatively straightforward assay can be performed. A *P. pacificus* strain containing a PPA-MLT-
10:mCHERRY translational fusion protein is reared on the pathogenic food source rather than *E. coli* and subjected to a Gram-stain (followed by a counterstain to detect Gram-negative bacteria such as *P. aeruginosa* or *S. enterica*). Overlying images containing the red fluoresce and gram stain can be analyzed for proximity and localization of binding.

Here we report the first *mlt-10* ortholog to be identified in quasi-parasitic nematode *Pristionchus pacificus*. In that parasitic nematodes are deleterious to animal and plant populations, the *P. pacificus* model provides the opportunity to study the important developmental process of molting. Interruption of larval molting could have profound pharmacological effects. Our findings deepen our understanding of the complex mechanisms controlling ecdysis and confer a broader understanding of the relationship between all molting animals.
REFERENCES


