THE FUNCTION OF THE TRANSCRIPTION FACTOR HLH-6 IN THE ESOPHAGEAL GLAND CELLS OF ROOT KNOT NEMATODES (*MELOIDOGYNE*)

by

Royden A. Loewen

A thesis submitted to the Faculty of Graduate Studies of The University of Manitoba In partial fulfillment of the requirements for the degree of

Master of Science

Department of Biological Sciences

University of Manitoba

Winnipeg, MB

Abstract

Root-knot nematodes (RKN) are plant parasitic nematodes that cause massive crop loss. RKN are adept at invading and feeding on plant tissue with secretions they release from their gland cells. The proteins within these secretions, called the "secretome," have been studied intensely using transcriptomics and proteomics. In our laboratory, the highly derived gland cells of RKN are being studied by comparing glandular development in the model RKNs *Meloidogyne hapla* and *Meloidogyne incognita* to the bacterial feeding, model nematode, *Caenorhabditis elegans*. In the current study, pharyngeal and gland important transcription factors, PHA-4 and HLH-6, respectively, were examined for their role in transcriptional regulation of parasite-specific genes. Homologues of these transcription factors were identified, cloned and sequenced from *M. incognita* strain Que1. Phylogenetic analyses confirmed the homologues and revealed a new PPN-specific gene, *hlh-6-like,* not present in the free-dwelling or animal parasitic nematodes examined. Bioinformatic comparisons of the homologues revealed conservation of the DNA-binding motifs of Ce-HLH-6 and Mi-HLH-6 proteins, as well as conservation of their promoter regions. While HLH-6 antibodies proved inconclusive, expression assays revealed expectant levels of PHA-4 and HLH-6 in *Meloidogyne* as compared to *C. elegans*. This conservation supports the hypothesis that Mi-HLH-6 plays an important role in gland cell development and function in *M. incognita* and other RKN. In identifying genes essential for parasitism, such as those regulating the function of the gland cells and their secretions, we are providing new targets for knockdown during RKN infection. The future goal of this research is to develop effective control mechanisms against parasite specific targets to inhibit this pest in the agricultural setting.

Acknowledgements

I thank Dr. Jay Kormish for mentoring me through the last several years, guiding me through the many ups and downs of research, and teaching me the discipline of science. I thank my committee members Dr. Mario Tenuta, and Dr. Sylvie Renault, for their help and support throughout my studies. I also want to thank Dr. Mehrdad Mahran and Jennifer Francisco for their help in initiating the root knot nematode populations. Finally I want to thank all of my lab mates, Shinhye, Oiza, Max, Old Stephanie, New Stephanie, Monty, Mackenzie, Emma, and Patrycja, for their irreplaceable help and friendship.

Dedication

For my family,

and their unwavering support.

Table of Contents

List of tables

List of figures

List of copyrighted material used in this study

Section	Reference	License ID
Figure 1.1	Williamson & Gleason, 2003	3852681240886

1. Introduction

1.1 Plant Parasitic Nematodes

Plant parasitic nematodes (PPN) are roundworms that infect the roots of many agricultural plants, reducing total worldwide crop yields by over $100 billion annually (McCarter, 2008). PPN include root knot nematodes (RKN, *Meloidogyne* spp.) and cyst nematodes (CN, *Heterodera* and *Globodera* spp.), and are capable of infecting virtually all crop plants (Bird & Opperman, 2009). Current methods to control RKN are insufficient: crop rotation has limited value since RKN can survive in a dauer-like state for up to nine years and some species can have a broad host infectivity; broadly toxic nematicides have been banned; various RKN species have adapted to overcome plant resistance mechanisms; and affected agricultural land quarantine is sometimes the only option for infection outbreaks (Agri Food Canada, 2012; Brito et al., 2015; McCarter, 2008).

The most agriculturally damaging PPN include the RKNs *Meloidogyne incognita, M. hapla* and the CNs *Globodera pallida* and *Heterodera glycines. M. incognita* is the most important RKN in its ability to infect a wide range of plants and habitats, and is especially prevalent in tropical regions, and the southern United States. *M. hapla* is the "Northern RKN", and is an issue in northern United States, southern Canada, and northern Europe (Taylor & Sasser, 1978). *H. glycines* is possibly the most pertinent PPN to Manitoban agriculture as it is steadily expanding northward following the production of soybeans (Tenuta & Tenuta 2014). *G. pallida*, the potato CN, is the only CN with a fully sequenced genome (Cotton et al., 2014).

Several PPNs including most of those mentioned above, have had their genomes sequenced recently, and using this new data, major research is currently being conducted

to find mechanisms to control RKN in the agricultural setting (Abad et al., 2008; Cotton et al., 2014; Fosu-Nyarko & Jones, 2015; Lunt et al., 2014; Opperman et al., 2008).

RKN infect plants in their J2 larval stage, entering at the elongation zone of the root and burrowing into the zone of root differentiation. Here they insert their protrusable stylet into a plant cell and, using glandular secretions, dedifferentiate plant cells into giant cells which the developing nematode then feeds on (Bellafiore et al., 2008)(see Figure 1). The invasion of the plant root requires secretions from the worm's three gland cells to degrade plant cell walls, suppress plant immune responses, and induce the formation of giant multinucleate feeding cells (Goverse & Smant, 2014).

1.2 Host-Parasite Interaction

The host parasite interaction is facilitated by the "secretome": the collection of proteins comprising the nematode secretions. Most of the secretions come primarily from the nematode's three gland cells: two sub-ventral esophageal gland cells (SvGs) used during initial migration and infection within the plant root, and one dorsal esophageal gland cell (DG) which is the most active gland cell beginning two days after infection and produces secretions aiding in giant cell maintenance (see Figure 1.2; Quentin et al., 2013). All three gland cells extend long processes anteriorly where they release secretions into the esophageal lumen. The DG process empties in the procorpus near the base of the stylet; the SvGs release their secretions in the metacorpus at the base of the metacorpal pump (Eisenback, 1989). A limited number of secretions also come from the chemosensory amphids and the hypodermis (Bellafiore et al., 2008; Haegeman et al., 2012).

The secretome is made up of "effectors", which include proteins and small molecules capable of altering plant cells. Initial studies of the secretome looked at the

transcriptome of the gland cells due to its relative accessibility, and thus, the ability to identify glandular secretions quickly (Huang et al., 2003). Later, methods were developed to isolate the minute secretions of the worm which allowed researchers to begin to determine the components of the secretome through proteomics (Bellafiore et al., 2008). In this way researchers isolated the proteins of the secretome and used tandem mass spectrometry coupled to liquid chromatography (LC-MS-MS) to identify them. This method exposed and expanded on the three essential classes of effector molecules: plant cell wall modification, immune interaction and suppression, and feeding cell formation and maintenance.

Cell wall modification is an important function of the secretome, allowing the worm to move through root tissue, and aid in the restructuring of plant cell walls during formation of giant cells. The presence of cell wall degrading carbohydrate-active enzymes (CAZymes) in all PPNs studied thus far is therefore unsurprising (Bird & Opperman, 2009). Through the study of its genome, 61 CAZymes were found in *M. incognita* (Abad et al., 2008) including cellulases, pectate lyase and polygalacturonase to break down pectate, cylanase to break down xylan, and endo-1, 4-betagalactosidase to break down pectin. In addition to these hydrolytic enzymes, expansins were found in the secretory product, and it has been proposed that they disrupt the non-covalent bonding between the polysaccharide chains of the cell wall, enabling greater accessibility to the CAZymes (Abad et al., 2008). Besides simple degradation of the cell wall, the secretome may play a role in predigesting molecules of the plant for ingestion in the worm. Invertases found in the secretions break down sucrose into glucose and fructose, rather than aiding in cell wall degradation as the other CAZymes seem to do. Invertases may be used to pre-digest plant cell components before consuming them (Abad et al., 2008).

When PPN invade plant roots the plant immune system is activated, and PPN must quickly suppress the response in order to survive. The immune response is triggered in the plant by the presence of pathogens. PPN face a unique challenge in suppressing the immune response of plants throughout the six-week portion of their life cycle that they remain inside the plant root tissue (Haegeman et al., 2012). The first response of plants to invasion is the recognition of pathogen associated molecular patterns (PAMPs). This response is called PAMP-triggered immunity (PTI) and includes production and release of reactive oxygen species (ROS) that work to degrade nematode tissue, as well as strengthen the cell walls of the area of infection with the help of lignification and callose deposits (Goverse & Smant, 2014). Through this mechanism the plant attempts to destroy the nematode and contain it from further migration (Quentin et al., 2013). Although this response is seen in plants after PPN invasion, known PAMPs have not been detected in the worm, though they may yet be found in the secretome (Quentin et al., 2013). To combat ROS the nematode releases a variety of antioxidant proteins in the secretome to metabolize ROS, including glutathione peroxidase, peroxiredoxin, and superoxide dismutase (Quentin et al., 2013).

Secondarily to PTI, effector triggered immunity (ETI) occurs. In this process the plant recognizes effector molecules, or the effector molecule's detrimental effects on plant cells, and invokes localized cell death. This is known as the hypersensitive response, which the PPN secretome appears to be generally effective in preventing (Haegeman et al., 2012). From transcriptomic analysis of infected plant cells researchers found that plant genes involved in defense are down regulated early in the plant-nematode interaction (Damiani et al., 2012). Chorismate mutase of the secretome prevents host defense triggering, and Hs10A06, and Hs4F01 are effectors that disrupt

plant defense signaling pathways, thereby protecting invading worms from the immune response. As well, calreticulin has been shown to suppress PTI. Using its secretome, PPN effectively silence the plant immune response, maintaining their own health, and keeping their feeding cells alive.

There are, however, two effector proteins – CG1 and MAP-1 – found in the PPN secretome, that activate the resistance gene *Mi-1* (Hewezi & Baum, 2012). The *Mi-1* gene protects tomato and related plants from *Meloidogyne* (Goggin et al., 2006; Haegeman et al., 2012). But *Mi-1* is one of the few resistance methods that effectively work against RKN, and only protects tomato plants from some species of *Meloidogyne* (Williamson & Kumar, 2006).

The establishment and maintenance of feeding cells, or "giant cells," as they are called due to their large size relative to other plant cells, is also an important function of the PPN's secretome. In RKN, giant cells are formed through cellullar hypertrophy and repeated nuclear division without cytokinesis. In contrast, the CN's feeding cells are a syncytium formed through breakdown of plant cell walls and cellular fusion (Williamson & Kumar, 2006). It is these giant cells that give rise to the characteristic gall seen on the plant root. A number of proteins have been found that are suggested to play a role in giant cell formation and maintenance. Included amongst these are secreted proteins that have on them a nuclear localization signal (NLS), suggesting that once inside the plant cell they are directed to the nucleus where they can directly affect change on plant cell gene expression. Bellafiore and colleagues found 26 such proteins in the secretome, and another 40 proteins appeared to have nucleotide binding ability (Bellafiore et al., 2008). Collectively amongst these 66 proteins were found helicases, RNA and DNA binding proteins, histones, and a nuclear assembly protein. These proteins are the most likely to

be able to enter the plant cell and effect gene expression.

There are a number of specific proteins suggested to affect giant cell formation (Bellafiore et al., 2008). CDC48 induces cell proliferation, SKP1 encourages cell cycle progression, TCTPs (translationally controlled tumour protein) increase cell growth and prevent apoptosis of the cell. CDPK is essential in plant cell elongation and has been found in the secretome, and CPI-2 and API-2 inhibit programmed cell death. These proteins therefore collectively encourage hypertrophy while preventing the cell from apoptotic pathways, thereby creating a viable giant cell from which RKN can receive all of its nutrients.

1.3 PPN Genomics

New advances in our understanding of PPN come from genomics. The field of genomics has exploded in the last decade with ever faster and cheaper sequencing technologies (Buermans & den Dunnen, 2014). *C. elegans* was the first multicellular organism to be sequenced and since then many other nematodes have had their genome's sequenced (The *C. elegans* Genome Sequencing Consortium 1998). Several nematode species, which are included in later analyses within this thesis, are highlighted below, and were chosen because they represent important models in their ecological niche, and a diverse range of life-history survival strategies, including plant (clade IV), human (clade III), and insect parasites (clade V), as well as free-dwelling taxa (clade V)(Blaxter et al., 1998).

Several PPN species, all of clade IV, have now had their genomes sequenced and were included for phylogenetic analysis. *M. incognita* is known as the southern root-knot nematode and is capable of infecting virtually all crops; in contrast *M. hapla* is known as the northern root-knot nematode and is somewhat less common, however it is important

as a model genetic species (Opperman et al., 2008). Unlike *M. incognita*, *M. hapla* reproduction involves meiosis, thus allowing for the development of genetic linkage maps (Abad et al., 2008; Liu et al., 2007). *M. floridensis* is found in Florida and is very closely related to *M. incognita*, and is also capable of sexual reproduction (Lunt et al., 2014). *Globodera pallida* is the only CN with a genome sequenced thus far, and is a major pathogen of potatoes (Cotton et al., 2014). *Brugia malayi* of clade III is a model filarial parasite, capable of infecting humans and other animals (Ghedin et al., 2007). Several nematodes of clade V were chosen in addition to *C. elegans*, including the free-dwelling *C. briggsae*, the omnivorous *Pristionchus pacificus*, and the entemopathogenic *Heterohabditis bacteriophora* (Bai et al., 2013; Dieterich et al., 2008; Stein et al., 2003). This new collection of genomic data is being used to understand the evolution of parasitism in PPN.

Analyses of these genomes reveal strong reductions in the RKN genomes; indeed *M. hapla* at 54Mbp is the smallest metazoan genome yet discovered. Specifically, reduction has occurred in areas of self-protection such as lost antibacterial and antifungal genes, as most of their lifecycle is spent inside the relative safety and shelter of the plant root (Abad et al., 2008). Unsurprisingly, RKN show gene expansion in areas of parasitism, such as in secretory genes involved in cell wall modification (Bird & Opperman, 2009). As noted earlier, CAZymes have been found in the secretome of *Meloidogyne*. The presence of CAZyme genes in metazoans is extremely rare, and therefore it was initially suggested that these genes could have come from symbiotic bacteria living within RKN, similar to how bacteria are used to degrade cellulose in other metazoans. This theory was disproven upon close analysis of a CAZyme, endogluconase, which showed a polyadenalated 3’ end of the cDNA, a feature of eukaryotic gene

expresssion. Genomic analysis of both *M. incognita* and *M. hapla* revealed CAZymes that had more in common with bacterial genes than metazoan genes, but that these genes were in fact mapped conclusively within the RKN chromosomes (Abad et al., 2008; Opperman et al., 2008). It is now assumed that the CAZymes were acquired in RKN through horizontal gene transfer (HGT). HGT theory suggests genes have been incorporated from bacteria or fungi living a similar plant pathogenic - or plant associated - lifestyle into RKN genomes through some mechanism of gene transfer (Haegeman et al., 2011). Multiple HGT events have been proposed to account for the host of cell-wall degrading enzymes present in the secretome (Abad et al., 2008). Plant parasitism has evolved separately at least three times in nematodes, and each case has evidence of HGT occurring.

Comparative genomics also showed many similar genes between the filarial parasite *B. malayi*, and the PPN *M. incognita's* secretome genes; this was somewhat surprising since they are considered to have diverged evolutionarily up to a billion years ago (Blaxter et al., 1998). 26 out of 80 known *B. malayi* secreted proteins were also found in the *M. incognita* secretome. This set of homologous genes between species long separated in evolution, suggests they are highly important in the parasitic interaction; the conserved set of genes are thought to aid primarily in host immune suppression, revealing the importance of such genes in successful parasitism (Bellafiore et al., 2008; Bird & Opperman, 2009). Further studies in comparative genomics will reveal conserved sequences necessary for the different styles of parasitism employed by the different species of PPN (Haegeman et al., 2012).

1.4 Pharyngeal glands of C. elegans

C. elegans is one of the most well studied organisms and its well-annotated genome provides a useful model for comparative studies with other organisms, especially with other nematodes, including PPN. Despite being separated by 100s of millions of years of evolution there are many basic genetic similarities between *Caenorhabditis* of clade V, and *Meloidogyne* of clade IV (Blaxter et al., 1998). Besides obvious morphological similarities such as their basic body plan and organization, they have similarities on the genome level such as the use of splice leaders to control operons (Guiliano & Blaxter, 2006). Because of *C. elegans'* close relation to PPN, it is being used increasingly as a model in anthelminthic discovery (Burns et al., 2015). For these same reasons it will be a useful model for understanding the evolution of parasitism in PPN, and in finding and studying parasite specific genes integral to PPN's success.

In *C. elegans* the alimentary system has been used to study how transcription factors (TF) affect organ development (Kormish et al., 2010). In respect to the current study, it is useful to note the genetic pathways activating pharyngeal and gland cell specification and function have already been elucidated in *C. elegans*, and can be compared to *Meloidogyne* to establish similarities and differences in PPN.

One of the more complex organs in *C. elegans* is the pharynx. This neuromuscular organ initiates digestion as it pumps food into the intestine of the nematode. The pharynx includes 95 cells composed of seven different types: muscle, neuron, marginal, arcade, gland, epithelial and valve cells (Altun & Hall, 2009). These tissues work together to coordinate the pumping of the pharynx, the physical breakdown of food, and the efficient passage of food into the intestine. Various secretions present in

the pharynx are produced and secreted by five pharyngeal gland cells, in comparison to *Meloidogyne's* three (Smit et al., 2008).

The *C. elegans* transcription factor PHA-4 is considered to be a master regulator of pharynx development. PHA-4 has been found to be necessary in all pharyngeal cells for specifying their fate as pharyngeal and is also required for subtype specification of the diverse types of cell that make the organ (Gaudet & Mango, 2002; Horner et al., 1998; Kalb et al., 1998; Mango et al., 1994). For example, PHA-4 works with CEH-22 in the development of the pharyngeal muscle cells, and is required to work in conjunction with HLH-6 for proper gland cell development and function to occur (Okkema et al., 1997; Smit et al., 2008).

PHA-4 is a forkhead transcription factor which is highly conserved, and found in the regulation of foregut development of *Drosophila* (*forkhead*) and mammals (FoxA/HNF3) (Zorn & Wells, 2010). Forkheads have approximately 100 amino acids, and a conserved DNA-binding domain (DBD) consisting of a helix-turn-helix motif with loops on either side (W. H. Raharjo, et al., 2010). See Figure 1.3 for the conserved amino acids of the forkhead DBD.

PHA-4 was initially revealed in *C. elegans* as the factor binding the consensus sequence found in a large group of putative pharyngeal genes. The sequence that PHA-4 binds is TRTTKRY (R = A/G, K = T/G, Y = T/C). The requirement of PHA-4 throughout development was established by observing *pha-4* temperature sensitive mutants. In this genetic background, any move from a permissive to a restrictive temperature proved to be lethal, thus showing the necessity of PHA-4 throughout *C. elegans'* development. Substantial lethality occurred as well if the shift to restrictive temperature happened after pharyngeal development completed. These experiments

showed that PHA-4 is necessary throughout development and for the mature function of the pharynx as well. (Gaudet & Mango, 2002)

Gaudet and Mango in 2002 demonstrated that variations in the TRTTKRY sequence and flanking regions affected the binding affinity of PHA-4. *In vitro* assays have shown that different TRTTKRY sequences affect the binding affinity of PHA-4, but that the sequences flanking the TRTTKRY core also affect the affinity level of PHA-4. By mutating sequences in embryos to have higher levels of affinity for PHA-4, an earlier onset of expression was observed *in vivo* for the modified sequence. In similar fashion, gene expression was delayed in onset if the *cis*-acting factor was mutated to have lower binding affinity to PHA-4. These differing levels of affinity allow a mechanism whereby PHA-4 can activate different pharyngeal genes during different stages of development depending upon its affinity for a *cis*-regulatory element. When PHA-4 levels are lower they will only initiate genes with high affinity for PHA-4, whereas genes with lower levels of affinity will be activated when PHA-4 levels are higher. Though high affinity for PHA-4 is necessary for early activation, it is not sufficient, and therefore other activators are required to work with PHA-4 to activate the various pharyngeal genes (Gaudet & Mango, 2002; Kalb et al., 1998)

HLH-6 is one of the transcription factors that work with PHA-4 to regulate development in the pharynx. HLH-6 is specific to gland cells, and is necessary for their proper development and function (Smit et al., 2008). As seen in Figure 1.4, the pharynx has five gland cells located in the posterior bulb: two anterior gland cells (g1AR, and g1AL), a posterior dorsal cell (g1P) and two posterior ventral cells (g2R and g2L)(Sulston et al., 1983); all of the gland cells have long extensions reaching down to various points along the pharynx through which they release secretions. The g1P

extension travels from the posterior end of the terminal bulb, through the isthmus, the metacorpus, and empties into the procorpus at the anterior end of the pharynx near the buccal cavity. The g1As empty into the pharyngeal lumen near the metacorpus, and the g2s release into the terminal bulb (Altun & Hall, 2009).

HLH-6 is a basic helix-loop-helix (bHLH) transcription factor. bHLH factors play important developmental roles in yeast to humans, in many cell types including neuron, muscle, and gland; they are divided into several classes based on tissue specificity, dimerization partner, and DNA-binding specificity (Jones, 2004; Massari & Murre, 2000). HLH-6 is a class II bHLH protein, meaning it is not broadly expressed, and likely functions with a class I bHLH as a heterodimer (Grove et al., 2009). The binding of HLH-6 with HLH-2 has been shown through a yeast two-hybrid assay but the two transcription factors acting together as a heterodimer has not been shown to bind DNA in an electrophoretic mobility shift assay (Grove et al., 2009; Smit et al., 2008). As seen in Figure 1.3, the conserved motif of bHLH is approximately 60 amino acids long and has two essential domains. The amino-terminal domain, called the basic domain, binds DNA, and the carboxy-terminal domain is the helix-loop-helix domain which dimerizes with another bHLH protein homo- or hetero-dimerically (Jones, 2004). The basic domain binds DNA at E-boxes (Enhancer-box) of the canonical sequence: CANNTG (Seipel et al., 2004). A bHLH factor is required for gland development across many organisms in metazoan including nematodes (*hlh-6),* cnidarians (*ash-2),* and mammals (*Sgn-1)* (Seipel et al., 2004; Smit et al., 2008; Yoshida et al., 2001).

HLH-6 was found in a similar fashion to PHA-4. Smit et al. 2008 began with a list of probable gland specific genes (Ao et al., 2004). Many of the genes were *phat* genes (pharyngeal gland toxin related), which are thought to be secretory in nature due to

their signal peptide. Four of the genes were confirmed with GFP/YFP transcriptional reporters. Three showed GFP expression in all of the gland cells, but *phat-5* was only expressed in g1AR and g1AL. From the initial set of genes a consensus sequence was searched for in their promoter regions that might bind a master regulator TF. From this search the *cis* factor PGM1 (Pharyngeal gland motif 1) was found in most of the gland-specific genes. When PGM1 was mutated, expression in the glands of the GFP/YFP constructs was greatly reduced or gone entirely except in the gene *Y8A9A.2*. (Smit et al., 2008)

Examination of PGM1 revealed E-boxes that are known binding sites for basic helix-loop-helix (bHLH) transcription factors. Mutation of the E-box eliminated glandular expression of reporter constructs demonstrating the necessity of this site for gland expression. The E-box is not the only required element for gland expression since mutations in flanking sequences to the E-box also reduced expression. Therefore it was concluded that an extended sequence following the E-box including CAnvTGhdYMAAY (where V = A, C or G, H = A, C or T, D=A, G or T, M = A or C, and Y = C or T) was necessary and sufficient for glandular expression. Indeed when three tandem copies of PGM1 were placed upstream of a promoterless GFP reporter, the elements were shown to be sufficient for gland-specific expression. A later paper revealed a slightly less stringent PGM1 as CAnnTGnnYMAAY or CAnnTGnYMAAY (Ghai et al., 2012). PGM1 is an important *cis* acting element in gland specific expression. (Smit et al., 2008)

A candidate approach was used to identify the transcription factor binding to the PGM1 regulatory element (Smit et. Al. 2008). bHLHs are known to bind E-boxes as heterodimers of class I and II molecules (Grove et al., 2009; Massari & Murre, 2000).

HLH-2 is a ubiquitous class I bHLH present in *C. elegans*, and HLH-6 is the only class II bHLH found in gland cells. Therefore HLH-6 was proposed as the gland specific *trans* element responsible for binding to PGM1. Indeed when either PGM1 or HLH-6 was knocked out, GFP/YFP reporter data confirmed reduction of expression for both experimental conditions (Smit *et al.* 2008).

hlh-6 is required for gland cell development and function. In a predicted null mutation for *hlh-6 (tm299),* most of the gland cells still develop, though in 84% (n=90) of the cases the g2 cells do not develop. In *hlh-6* mutants 32% (n=105) of the nematodes die in the L1 larvae stage from starvation. Those that did grow into adulthood also showed starvation defects such as longer growing periods and smaller brood sizes. Feeding mutants a less sticky bacteria (HB101) than that used in standard *C. elegans* cultures (OP50) rescued this phenotype. Because less sticky bacteria rescued the mutants it was suggested that the function of the glandular secretions controlled by *hlh-6* is to line the lumen of the pharynx and allow for a smooth passage of food. (Smit et al., 2008) To assign just how much gland cell functionality is lost in *hlh-6* mutants, the glands were entirely removed through genetic ablation, in order to compare phenotypes. The phenotype of nematodes with no gland cells, and therefore zero gland function, is the same as *hlh-6* mutants, suggesting that despite the fact that there are gland cells still present in *hlh-6* mutants, the function of these cells is considerably limited. Therefore *hlh-6* is important in the differentiation of the gland cells, though not the only factor involved as many of the gland cells still develop, and it is necessary for function of the gland cells. *hlh-6* is activated by gland specific genes (Raharjo & Gaudet, 2007), and in turn activates a battery of gland specific genes including the secreted *phat* proteins that aid in feeding. (Smit et al., 2008)

To learn more about how gland cells develop, researchers looked for the genetic factors activating *hlh-6.* An important study in 2007 revealed three distinct promoter elements contributing to the activation of *hlh-6,* including a PHA-4 binding site (PBS), and two previously unclassified elements called HRL1 (*hlh-6* candidate Regulatory element) and HRL2 (Raharjo & Gaudet, 2007).

In order to find these important regulatory factors, researchers deleted different portions of the promoter region up to 1175bp upstream of *hlh-6* to find which areas were necessary for *hlh-6* expression. Deleting anything upstream of -747bp had virtually no effect. Deleting areas from -747 to -241 had a very small loss of expression, and deleting the segment downstream of -241 resulted in a nearly complete loss of expression. In the -747 to -241 region several E-boxes were found, showing that *hlh-6* likely auto regulates its own expression. This also explains the slight loss of function when that region is knocked out. Once activated, *hlh-6* can maintain its signal, but requires the actions of the promoters downstream of -241 for initial activation. (Raharjo & Gaudet, 2007)

The area downstream of -241 was called the minimal promoter, and contained three distinctive regions A, B, and C, and one further region with a PBS. When any one of these regions was knocked out, significant loss of expression of *hlh-6* occurred. When the region from *C. elegans* was compared to *C. briggsae*, two conserved sequences were observed. One in the B region (HRL1b), and one in the C region (HRL2c). Though there was no significant conservation seen in region A, *C. elegans* had an identical HRL2 sequence in A as in C, thus it was named HRL2a. (Raharjo & Gaudet, 2007)
Transgenic animals with GFP transcriptional reporters containing the various enhancer elements with a minimal promoter revealed their roles in *hlh-6* expression. None of HRL1, HRL2, or PBS can initiate expression on their own. PBS combined with HRL2b

activates weak non-pharyngeal expression, as does HRL2b with HRL1c. PBS combined with HRL2c activates expression in gland, and non-gland cells, but all are descendants of the MS blastomere. An enhancer construct combining PBS, HRL2c, and HRL1b showed strong gland specific expression, and some weak expression outside of the pharynx. Thus it was proposed that the overlap of PHA-4 and the transcription factor binding HRL2 in the posterior pharynx activated *hlh-6* in cells there, and the activity of HRL1 repressed activation of *hlh-6* in non-secretory cells, thus limiting *hlh-6* to gland cells. (Raharjo & Gaudet, 2007)

A more recent study has revealed LAG-1 as the transcription factor binding to HRL1 to restrict *hlh-6* expression to gland cells. The connection between LAG-1 and HLR1 has been shown as LAG-1 is able to bind HRL1 *in vitro*, and a previously known LAG-1 binding site (LBS) is able to effectively replace the HRL1 binding site *in vivo* (Ghai & Gaudet, 2008).

LAG-1 is a member of the CSL (*C*BF, *S*uppressor of Hairless, *L*AG-1) family of proteins controlled through Notch signaling. In canonical Notch signaling, CSL is a transcription factor that represses a gene unless Notch signaling is active in which case the CSL becomes an activator. However, despite its similarity to CSL, LAG-1 has only been shown to act as a repressor in conjunction with HRL1 in gland development, not as an activator, and is therefore non-canonical. As mentioned previously HRL1 appears to function to restrict *hlh-6* expression outside of gland cells. Indeed, if the three HRL1 binding sites found in the promoter of *hlh-6* (HRL1b, HRL1c, and HRL1d) are knocked out, *hlh-6* gland expression is not effected, but ectopic expression of HLH-6 occurs. Sensibly, removing LAG-1 produces the same ectopic expression of HLH-6. (Ghai & Gaudet, 2008)

Thus, HLH-6 is an important transcription factor regulating gland cell development and function but is far from the whole picture, as its removal does not remove all gland development, or all gland secretions. The upstream regulators of HLH-6: PHA-4, LAG-1, and the unknown factors binding HRL2, and HRL3, may have a greater role in turning on other regulators in the glands that are yet to be uncovered. Researchers recently looked for factors involved in gland development and regulation aside from *hlh-6*, and found a number of genes not under HLH-6 control. Some genes they found to be regulated by the previously mentioned *cis*-regulatory element HRL3 including *nas-12* and *Y8A9A.2*. Other genes like *ZK596.1*, *scl-3*, *wrt-3*, and *Y76B12C.3* were found that were both HLH-6, and HRL3 independent. (Ghai et al., 2012)

These genes were found from looking through a list of probable gland specific genes, and finding those that did not have the HLH-6 binding site, PGM1. From an initial list of 47 genes, researchers tested ten of them with GFP reporters. By knocking out PHA-4, they confirmed that all of these gland specific genes had a PBS element, as all gland and pharynx specific genes have thus far been found to contain. One of the ten genes, a phat-like protein, was PGM1 dependent, and *B0280.7* was HLH-6 dependent without containing a PGM1 site. This suggests that HLH-6 controls some gland specific genes both directly and indirectly. Two others, *ZK596.1*, and *scl-3*, showed somewhat diminished expression when HLH-6 was knocked out showing a partial dependence on HLH-6. Four more analyzed genes showed no dependence on HLH-6, these genes were *wrt-3*, *nas-12*, *Y8A9A.2*, and *Y76B12C.3*. (Ghai et al., 2012)

The promoters of *nas-12*, and *Y8A9A.2* were examined further in an attempt to identify gland specific non-*hlh-6* regulatory regions. A series of 5’ deletions in the promoters of each gene were carried out to find functional regions. A minimal effective promoter

fragment of 132bp was found, and was compared to similar sequences from other *Caenorhabditis* species. Several conserved regions were found; one was a PBS, and another was HRL3. PBS and HRL3 are also found upstream of *hlh-6* and this suggests that the same genes that turn on the important regulator and gland specific transcription factor HLH-6, also turn on a suite of other genes confined to gland cells (Ghai et al., 2012).

As seen in the differences between the subventral gland cells (SvG) and dorsal gland cells (DG) in PPN, the subtypes of gland cells in *C. elegans* have slightly different functions (Ghai et al., 2012). For example *Y8A9A.2* is only expressed in g1P, and the g2s, not in the g1A cells. This is despite having PBS, and HRL3 present in its regulator region just like many other genes expressed throughout all gland cells have. Through comparison with other *Caenorhabditis* species, the regulatory region of *Y8A9A.2* had several conserved sequences. One of them, CR3 (conserved region 3), was an NHR-like site, found to be bound by NHR-48. *nhr-48* was a previously unclassified NHR gene. NHR-48 functioning through CR3 effectively repressed expression of Y8A9A.2 in g1A cells. (Ghai et al., 2012)

phat-5, on the other hand, was found to be expressed only in g1A cells, the reciprocal of the *Y8A9A.2* pattern. In a similar analysis to find its conserved promoter regions, the region CR1-phat-5 was found to be necessary and sufficient for repressing *phat-5* in g2s, and sufficient but not necessary in g1P. Thus far a *trans*-factor acting on CR1-phat-5 has not been found (Ghai *et al.* 2012). The pharyngeal gland cells have many layers of developmental and functional regulation, including gland subtype regulation factors.

1.5 Objectives

The aim of this research is to identify *pha-4* and *hlh-6* homologues within RKN and to elucidate their role within development and function in pharyngeal and gland development. Comparative expression analysis of identified homologues will serve as a foundation for understanding when and where these genes are active in *Meloidogyne* when compared to *C. elegans.* Such an analysis can lead to the discovery of highly conserved functions of these transcription factors in pharyngeal development as well as elucidate novel functions that have been acquired to accommodate the parasitic versus free-living lifestyle of these two species. The following subaims were explored in the context of this thesis:

1) Which of the identified basic helix-loop-helix and forkhead proteins in the genomes of *Meloidogyne* spp, are the most likely true homologues of HLH-6 and PHA-4? Phylogenetic analyses using Clustal Omega, PAUP4.0, and MEGA6 were carried out on the DNA binding domains of HLH-6 and PHA-4 and their related proteins in the nematodes *C. elegans, C. briggsae, P. pacificus, H. bacteriophora, B. malayi, M. hapla, M. incognita, M. floridensis,* and *G. pallida*. Maximum parsimony, and maximum likelihood trees were constructed and tested for significance using bootstrap values.

2) Is *Meloidogyne hlh-6* gene under the same regulatory control as *C. elegans hlh-6*? Homology synteny, transcription factor and enhancer element searches where performed on the upstream region of the *Meloidogyne hlh-6* gene.

3) Where and when are the HLH-6 and PHA-4 homologues functioning during *Meloidogyne* development and transitions through the parasitic lifestyle? To detect protein expression patterns antibodies were developed against CeHLH-6 and MhHLH-6 and tested for specific immunoreactivity in Western analysis and whole mount immunohistochemistry. In parallel, mRNA for *pha-4, hlh-6,* and an *hlh-6-like* protein, were tested for their transcript levels throughout the life-cycle of *Meloidogyne.* The cDNA of *pha-4* and *hlh-6* homologues, as well as *hlh-6-like* were cloned from a Quebec *Meloidogyne* isolate and sequenced. The transcript levels of *pha-4, hlh-6,*and the *hlh-6-like* genes were semi-quantitatively tested in various *Meloidogyne* life stages to elucidate when their respective transcription levels were most active.

1.6 Significance

Finding new control strategies of PPN in the agricultural setting is of upmost importance for global food security (Fosu-Nyarko & Jones, 2015). An estimated 125 billion dollars of global agricultural output is lost due to PPN (Chitwood, 2003). Original sources of nematode control include carbamates and organophosphates, which are now banned due to their non-specific neurotoxic effects. Bromomethane was also used to control PPN and has been banned because of its role in ozone depletion (McCarter, 2008). PPN have also shown the ability to quickly evolve and thereby develop the ability to infect plants that were previously resistant (Li et al., 2011). Thus the dearth of nematode-specific pesticides, and the resistance breaking habits of PPN must be addressed to meet the ever increasing demand in agricultural output in a sustainable and eco-friendly way.

RNA interference (RNAi) functioning through transgenic plants has the potential to create crop plants capable of resisting invasion by PPN. RNAi is the process by which dsRNA within cells creates a knockdown effect specific to that sequence. The effectiveness of RNAi was first shown in *C. elegans*, and has proven to be an invaluable tool in the study of genetics (Fire et al., 1998). Studies have proven that RNAi is capable of knocking down genes in PPN, and proteins of the secretome have been targeted in this way with varying degrees of success (Huang et al., 2006; Quentin et al., 2013).

To find a robust and PPN-specific target for RNAi, essential to parasitism, I propose to look for a master regulator of gland function in PPN. HLH-6 may be such a target, as it has the role of controlling glandular development and function within the non-parasitic nematode *C. elegans*. If HLH-6 plays a similar role in *Meloidogyne* as it does in *C. elegans*, it could be that many of the genes of the esophageal glands, including those of the secretome, are under its transcriptional control. Knocking down HLH-6 in *Meloidogyne* via RNAi could therefore be a highly effective way of controlling the pest in agricultural crops.

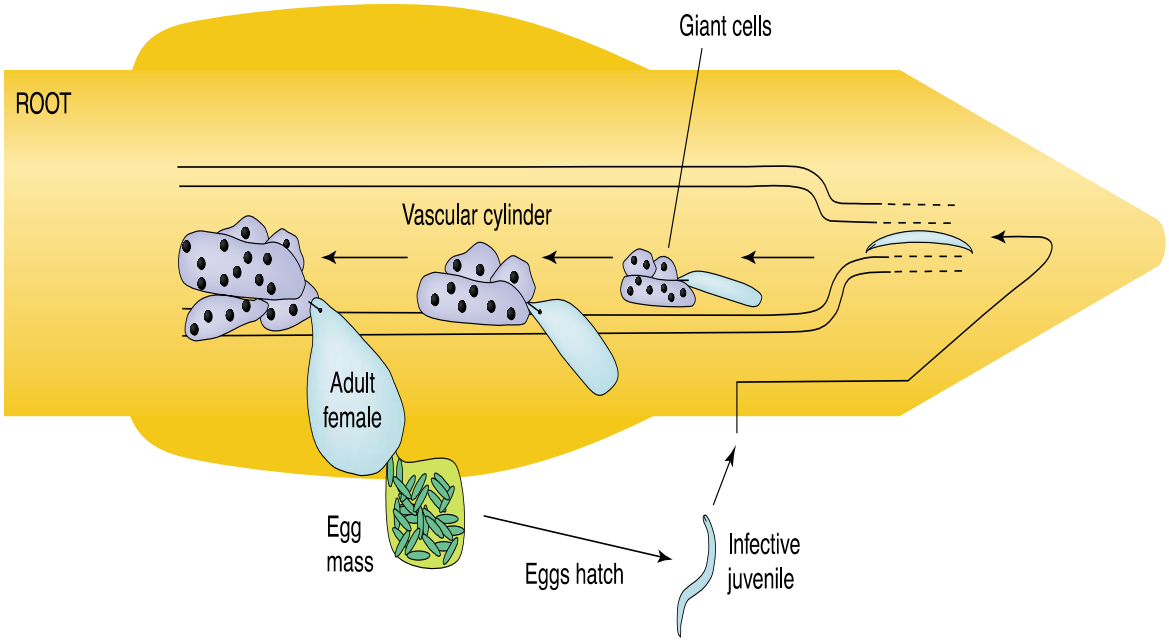


Figure 1.1 The life-cycle of Root Knot Nematodes. Second stage larval nematodes (J2s) hatch from their eggs on the outside of the root into the soil, from there they penetrate the root at the zone of elongation, and migrate through the plant tissue into the differentiation zone of the root. The nematodes become sendentary, and use their secretions to create dedifferentiated plant giant cells which they then feed on. The nematode goes through several molts until reaching adulthood, at which point the female lays her eggs on the surface of the root. Used with permission from Williamson & Gleason (2003).

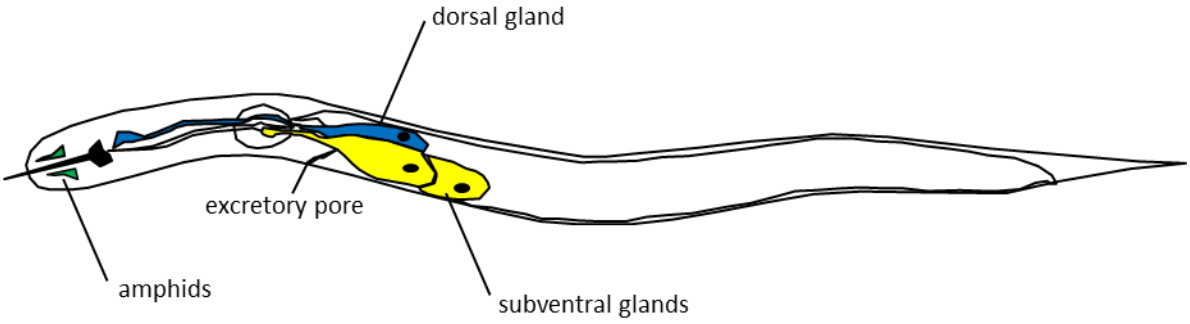
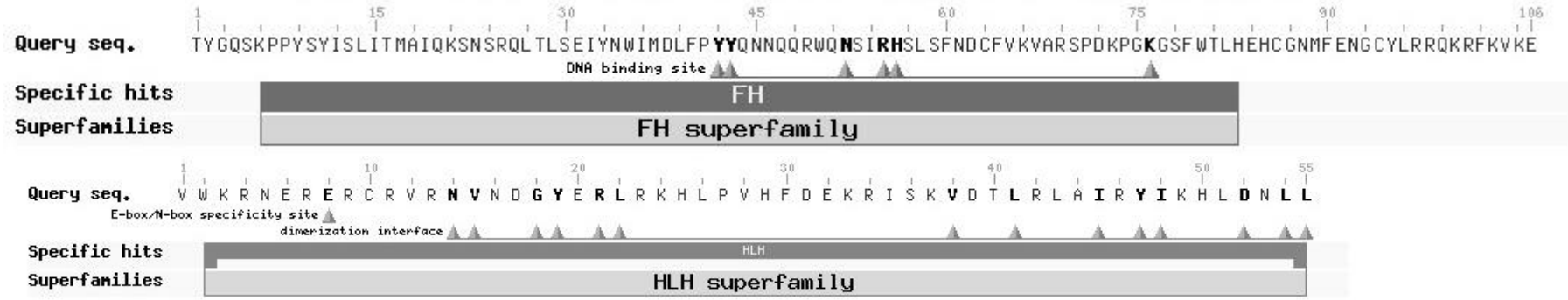


Figure 1.2 Glands and other secretory organs of a typical plant-parasitic nematode. Anterior is to the left. Amphids (green) located at the anterior tip of the worm release some minor secretions. There are three esophageal gland cells: the single dorsal gland (blue) is slightly anterior to the subventral glands (yellow) and is active later in the sedentary stage of the worm; the two subventral gland cells are most active during the migratory life-cycle stage. Modified from Haegeman et al. (2012).

A)



B)

```
PHA-4          TYGQSKPPYSYISLITMAIQKSNSRQLTLSEIYNWIMDLFPYYQNNQQR----WQNSIRHSLSFNDCFVKVARSPD-----KPGKGSFWTLHEHCGNMFENGCYLRRQKRFKVKE
Cb-PHA-4       TYGQSKPPYSYISLITMAIQKSATRQLTLSEIYNWIMELFPYYQNHQQR----WQNSIRHSLSFNDCFVKVARSPD-----KPGKGSFWTLHEHCGNMFENGCYLRRQKRFKVKE
Hb-PHA-4       NYGHAKPPYSYISLITMAIQQSPTHKLTLSEIYNWIMELFPYYRQNQQRSGRGWQNSIRHSLSFNDCFVKLARTPD-----KPGKGSFWTLHELCGNMFENGCYLRRQKRFKVKE
Pp-PHA-4       NYGSSKPPFSYISLITMAIQRSETHMMTLNEIYTWIMTQFPYYRQNQQR----WQNSIRHSLSFNDCFVKVPRTPD-----KPGKGSFWTLHSLCGNMFENGCYLRRQKRFKVRE
Mh-PHA-4       NFGPNKPPYSYISLISMAIQQSSRKMCTLSEIYNFIMEYFEYYKNHNQRC--SWQNSIRHSLSFNDCFVKVPRTPDRIKNCRPGKGSFWTLHALCGDMFENGCTLRRQKRFKLKD
Mf-PHA-4       SYGPNKPPYSYISLISMAIQQSSRKMCTLSEIYNFIMAYFEYYKNHNQRC--SWQNSIRHSLSFNDCFVKVPRTPDRIKNCRPGKGSFWTLHALCGDMFENGCTLRRQKRFKLSK
Que1-PHA-4     SYGPNKPPYSYISLISMAIQQSSRKMCTLSEIYNFIMAYFEYYKNHNQRC--SWQNSIRHSLSFNDCFVKVPRTPDRIKNCRPGKGSFWTLHALCGDMFENGCTLRRQKRFKLSK
Gp-PHA-4       NFGPNKPPYSYISLISMAIQQSSRKMCTLSEIYNFIMEYFDYYKNHQQRC--SWQNSIRHSLSFNDCFVKVPRTPD-----RPGKGSFWTLHELCGDMFENGCTLRRQKRFKLAH
Bm-PHA-4       KIGNAKPPFSYISLITMAIQRSESRMLTLSEIYQFIMDNYAYYRQNQQRSA-GWQNSIRHSLSFNDCFVKVPRTPD-----KPGKGSFWTLHEDCGNMFENGCYLRRQKRFKTND
consensus       yG  KPPySYISLItMAIQqSesrmlTLsEIYnfIMd f YYrnhqQR    WQNSIRHSLSFNDCFVKvpRtPD     kPGKGSFWTLHelCGnMFENGCyLRRQKRFKvke

HLH-6          SVWKRNERERCRVRNVNDGYERLRKHLPVHFDEKRISKVDTLRLAIRYIKHLDNLL
Cbr-HLH-6      SVWKRNERERCRVRNVNDGYEKLRRHLPVHFDEKRISKVDTLRLAIRYIRHLDNLL
Hb-HLH-6       SVWKRNERERFRVRCVNNGYDSLRRHLPVSFSDKRISKVDTLRLAIRYIRHLEAVL
Pp-HLH-6       SVWKRNERERYRVRCVNEGYETLRHHLPLSSSEKRISKVDTLRLAVKYVRHLEFVL
Bm-HLH-6       TVWKRNERERIRVRCVNDGYERLRDHLPLTESDRRISKVDTLRLAIRYIRHLEALL
Mh-HLH-6       TVWKRNERERIRVREVNNGYTNLKSKLPLTENDRRISKVDTLRLAIAYIHHLENLI
Mf-HLH-6-1     TVWKRNERERIRVRQVNDGYTRLKNTLPLTENDRRISKVDTLRLAIAYIHHLENLI
Mi-HLH-6-1     TVWKRNERERIRVRQVNDGYTRLKNTLPLTENDRRISKVDTLRLAIAYIHHLENLI
Mi-HLH-6-2     TVWKRNERERIRVRQVNDGYTRLKNTLPLTENDRRISKVDTLRLAIAYIHHLENLI
Mf-HLH-6-2     TVWKRNERERIRVRQVNDGYTRLKNTLPLTENDRRISKVDTLRLAIAYIHHLENLI
Gp-HLH-6       TVWRRNERERQRVQQLNFGYTHLRERLPLIDRDRRMSKVDSLRLAIAYIRHLELLL
consensus      tVWkRNERERlRVrnvNdGYtrLr hLPlte drRiSKVDtLRLAiaYikHLenll
```

Figure 1.3 A) Conserved amino acid residues from Forkhead, and HLH superfamilies from NCBI (Coordinators, 2016). B) DBDs of CePHA-4, and CeHLH-6 aligned with homologous proteins from several model nematode species: *Globodera pallida* (Gp), *Meloidogyne hapla* (Mh), *Meloidogyne incognita* (Mi), *Meloidogyne floridensis* (Mf), *Pristionchus pacificus* (Pp), *Heterohabditis bacteriophora* (Hb), *Caenorhabditis elegans*, *Caenorhabditis briggsae* (Cb), and *Brugia malayi* (Bm). Consensus line shows conserved residues of respective superfamilies underlined.

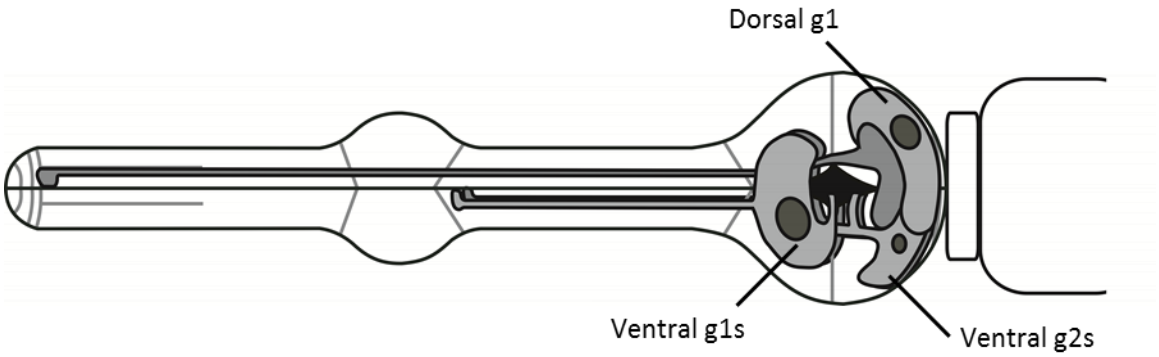


Figure 1.4 Pharynx of *C. elegans* showing placement of the five gland cells. Anterior is to the left. The gland cell bodies are in the posterior bulb of the pharynx, and extensions of the gland cells empty their secretions throughout the pharynx at the points marked with black arrows.

2. Materials and Methods

2.1 Nematode husbandry

2.1a - *Caenorhabditis spp.* Strain Husbandry

C. elegans strain N2(wild type) was maintained on NGM plates seeded with *Escherichia coli* strain OP50 at 15 or 20°C using standard methods (Brenner, 1974); strain GD251 (*rol-6*(*e187*) *hlh-6*(*tm299*) *unc-4*(*e120*) II; *pcxIs* [*pha-4*::GFP::His2B *rol-6(e187)*]) was maintained on *E. coli* strain HB101 and kept at room temperature. To prepare large quantities of *C. elegans* protein, liquid cultures were used. 1L of autoclaved liquid 2XYT (16g tryptone, 10g yeast extract, 5g NaCl, 1mL 1M NaOH) was inoculated with OP50 from a previously streaked plate, and grown overnight at 37°C. 1L of NGM broth (3g NaCl, 5g Tryptone, 25ml KPO_4(1M), 1ml $CaCl_2$(1M), 1ml $MgSO_4$(1M), 1ml cholesterol (5mg/ml)) was made, 200mL were added to four separate 250mL flasks with 25mL of OP50. *C. elegans* were collected from 20 recently starved plates by washing them off with M9, these were added to the 2XYT and OP50 flasks. The flasks were incubated on an Excella E1 Platform Shaker at minimal RPM at room temperature for five to ten days, fed with fresh bacteria as necessary, before the worms were collected for protein isolation.

2.1b - *Meloidogyne spp.* Strain Husbandry

Meloidogyne strains were grown on Rutgers VFA #4050 (Tomato Growers Supply Company) tomato plants or Tenderlong Imperator carrot. Seeds were bleached ten minutes in 5% sodium hypochlorite, and rinsed with distilled water five times. The seeds

were planted in Sunshine Mix #1 soil (Planet Natural) and covered in a single layer of vermiculite (ProMix) in seedling trays, and grown in a Conviron A1000 growth chamber under eighteen hours of light per day at 22°C. Plants were fertilized with Miracle-Gro Water Soluble All Purpose Plant Food as directed by manufacturer once a week, and watered every other day as needed. Two-four weeks after seeding, the plants were replanted in larger six inch wide pots, in a mixture of three parts sand (Quikrete Play Sand), and one part soil. Infected plants were kept in a designated room under eighteen hours of T5 full spectrum light, and watered every other day or as needed. Soil, sand and vermiculite were autoclaved before use.

2.2 Genotyping of GD251 and *Meloidogyne spp.*

Genotypes were verified using PCR and gel electrophoresis. Template DNA from *C. elegans* was collected by lysing ten adult worms in a PCR tube of 20μL of single worm lysis buffer (50mM KCl, 2.5mM $MgCl_2$, 10mM Tris HCl pH 8.3, 0.45% Tween-20, 0.45% NP-40, 0.01% gelatin, 2% 20mg/mL proteinase K), each tube was frozen for fifteen minutes at -80°C, then using a thermocycler heated to 60°C for 60 minutes, and 95°C for fifteen minutes. This protocol was adapted from Williams *et al.* (1992).

Genomic DNA from *Meloidogyne* was isolated using Gene JET Genomic DNA Purification Kit (Thermo Scientific) as per kit instructions with the following modifications. Twelve adults were extracted from root tissue using needles and forceps, placed in a 1.5mL centrifuge tube, and flash frozen in liquid nitrogen in the kit digestion solution. One nematode would be ideal, however more nematodes were necessary in order to collect enough DNA. Prior to extraction the nematode/digestion solution was

ground while frozen with a 1.5mL fitted plastic pestle. After three repeated freeze-grinds, fresh proteinase K (same as original volume) was added to the solution, and vortexed for five minutes. The homogenized tissue was then taken forward as the sample in the regular kit protocol, except that in order to obtain a more concentrated DNA sample 50μL of elution buffer rather than 200μL was used to elute the genomic DNA.

2.3 Cloning and Sequencing

2.3a - RNA Isolation and cDNA Synthesis from three *Meloidogyne* Life-Stages

RNA was collected from *Meloidogyne* egg masses, J2s, and adult females. For each extraction approximately 150 egg masses, 5000 J2s, or 100 adults were used to collect a reasonable amount of RNA for analysis, which was about 20-500ng/μL RNA per 10μL of eluent. Individual egg masses were picked from infected roots using needles and forceps under a dissecting microscope and collected in glass well slides in water. Egg masses were bleached (20μL 5% hypochlorite in 300μl water) for about five minutes to dissolve gel matrix and clean the eggs. BSA was added to quench the bleach, and the collected embryos were washed several times with water. Some egg masses were allowed to hatch in water at room temperature in the dark for collection of J2s, and adult females were dissected from roots using needles and forceps. RNA extraction was conducted preferentially immediately following sample collection, barring that, samples were stored in 20μL of RNA-later (Quiagen) in microcentrifuge tubes at -80°C until RNA extraction.

RNA was isolated using the RNeasy microkit (Quiagen) with the following adaptations. Various life-cycle samples were brought up to 50μL in Buffer RLT in microcentrifuge tubes, and frozen in liquid nitrogen, then ground with a mortar while the

sample thawed. This freeze-grind process was repeated a total of five times before adding another 300μL of Buffer RLT, and vortexing the sample for five minutes. The samples thus homogenized were taken through the regular kit protocol, with the exception that the 80% ethanol wash was conducted twice instead of once. Eluted samples were quantitated on a NanoDrop 2000c (Thermo Scientific).

RNA samples were tested for quality before cDNA synthesis using an RNA 6000 pico chip on an Agilent BioAnalyzer. RNA was then used to synthesize cDNA using BioRad iScript Select cDNA Synthesis Kit using oligo(dT) primers as per manufacturer's instructions. The cDNA was then stored at -80°C until further use including cloning and expression analysis.

2.3b - Cloning and Sequencing HLH-6, HLH-6-LIKE, and PHA-4 from *Meloidogyne*

Three predicted *Meloidogyne* genes: *Mh-pha-4*(MhA1_Contig1988.frz3.gene1), *Mh-hlh-6*(MhA1_Contig384.frz3.gene12), and *Mh-hlh-6-like*(MhA1_Contig809.frz3.gene10) from *Meloidogyne* strain Que1 were cloned for sequencing using Invitrogen Topo XL PCR Cloning Kit. Primers internal to the predicted genes were developed using Primer3 software (Koressaar & Remm, 2007; Untergasser et al., 2012); primers developed from the start codon, or in reverse from the stop codon, were chosen manually and given a nucleotide length closest to a 60°C melting temperature (estimated melting temperature = (C+G)*4 + (T+A)*2). Splice leader primers were also tested as the forward primer upstream of the ATG. See Figure 2.1 for primers used to amplify each fragment.

cDNA fragments were amplified by PCR using long PCR kit from Thermo Scientific, following the manufacturer's protocol. Of the 25μL product, 20μL was loaded into a 1% agarose gel with crystal violet, and 2μL was added to 8μL of H_2O and run in parallel on a 1% agarose gel with SYBR Gold. As SYBR Gold enables greater resolution of band size but interferes with later cloning reactions, the SYBR Gold gel was used to verify the size of the amplicon, while the DNA was purified from the crystal violet gel using the Thermo Scientific GeneJET Gel Extraction Kit. The purified amplicons were adenylated by adding 1.5μL Taq polymerase buffer, 1.5μL dNTPs, 1μL Taq polymerase to 10μL of gel purified DNA and incubated for 20 minutes at 72°C. 4μL of purified amplicon with added 3' A-overhang was added to 1μL of TOPO-XL vector, mixed gently, and incubated for 30 minutes at room temperature. 1μL of 6x Topo Cloning Stop Solution was then mixed in, briefly centrifuged, and placed on ice. 3μL from cloned vector solution was added to 50μL One Shot TOP10 Chemically Competent *E. coli* cells, incubated on ice for 30 minutes, heat-shocked at 42°C for 30 seconds without shaking, and incubated on ice for two minutes. 250μL of S.O.C. medium was added, the vial was capped tightly, and shook horizontally at 125 RPM at 37°C for 90 minutes. A 1:4 dilution, and an undilute sample of these transformed cells were spread separately on pre-warmed 2XYT + kanamycin (50μg/mL) plates, and incubated overnight at 37°C.

Colonies from the plates were then selected by screening with M13F and M13R primers supplied in the kit to test for the presence of the correctly sized insert. Positively screened colonies were grown up overnight in 2XYT + kanamycin(50μg/mL) broth at 37°C shaking at 150 RPM, and DNA was collected from them using a mini-prep kit (Thermo Scientific). Inserts were tested for orientation using an M13F or M13R primer

and an internal primer to the cloned sequence as seen in Table 2.1. 8μL of positively verified template DNA from mini prep (150-400 ng/μL) and 4μL of each necessary primer (2μM) were added to wells in a 96 well plate, and sent to Manitoba Institute of Cell Biology (MICB) for sequencing. A sample of transformed *E. coli* was stored in 15% glycerol stocks at -80°C.

2.4 Phylogenetics and Sequence Homology

2.4a - DNA Binding Domain Alignments

DNA binding domains (DBD) as predicted by Pfam from Ce-HLH-6, and Ce-PHA-4 were used as queries in BLAST searches for homologues in a series of nematode species (Finn et al., 2015; Lipman, 1990). These species included other Clade V nematodes *Caenorhabditis briggsae, Heterohabditis bacteriophora, Pristionchus pacificus,* Clade IV nematodes *Globodera pallida, Meloidogyne hapla, Meloidogyne incognita, Meloidogyne floridensis,* and Clade III nematode *Brugia malayi.* BLAST hits with an e-value of 0.05 or less were kept for further analysis.

2.4b - Maximum Parsimony Analysis using PAUP4.0

To generate maximum parsimony tree, the DBD lists in two separate files (HLH-6 and other bHLH related proteins, and PHA-4 and other forkhead related proteins), were run through PAUP 4.0 (Swofford, 2002). DBD lists were reformatted with Clustal Omega by outputting the file as Nexus output, the file was saved as a txt file through Word, and opened in PAUP. Parsimony was selected under analysis, and 1,000,000 fast-heuristic bootstraps were chosen, other settings were left as default and the program was

run. The tree was viewed by selecting print bootstrap consensus tree, and saved at this point as a PDF for later editing.

2.4c - Maximum Likelihood Analysis using MEGA 6

The bHLH and forkhead proteins were also analyzed using MEGA6 (Tamura et al., 2013). DBD sequences were pasted into alignment tool in FASTA format, and aligned using Clustal Omega, this alignment was exported, and opened up under the Models tab in order to find the best protein model which was LG+G. The alignment was then opened in the Phylogeny tab. Under phylogeny settings Maximum Likelihood was selected, bootstrap 10000, and LG+G as the protein model, other settings were left as default. Program instructions were followed from Hall (2013).

2.4d - Glandular Gene Promoter Region Homology

Dot plots were created using the YASS genomic similarity search tool (http://bioinfo.lifl.fr/yass/) on default settings, comparing the promoters of *Ce-hlh-6* with *Cb-hlh-6, Mh-hlh-6,* and *Mh-hlh-6.* Promoters were defined as 1500 bp upstream of the start codon of the gene of interest, or up until the next gene upstream if fewer than 1500 bp away. HRLs (*hlh-6* Regulatory eLement) were searched for in *Cb-hlh-6* and all of the PPN *hlh-6* promoters using YASS including HRL2a(AATAAATA), HRL1b(GTCGGAAG), HRL1c(GTGGCTA), HRL2c(AATAAATA), HRL3(TGCACAG) and HRL1d(GTGGATA) (Ghai et al., 2012; Raharjo & Gaudet, 2007). Hits were included if they had at least 4/7, or 5/8 conserved nucleotides.

PGM1 (Pharyngeal Gland Motif 1) was also searched for in the promoter regions of *hlh-6* genes. PGM1 has been found previously to have a consensus sequence of CAnvTGhdYMAAY (Stringent PGM1), and CAnnTGnnYMAAY or CAnnTGnYMAAY (Variable PGM1) (N= A, C, G, or T, V=A, C or G, H=A, C or T, D=A, G or T, M=A or C, and Y=C or T) (Ghai et al., 2012; Smit et al., 2008). These sequences were searched for using the program DNA Pattern Find (bioinformatics.org).

PBS (PHA-4 Binding Sequence) elements were searched for in the promoter regions of *hlh-6* genes and putative gland specific genes using the PWM (Positional Weight Matrix) for PBS downloaded from the JASPAR motif database as seen in Figure 2.2 (Mathelier et al., 2014). The PWM was used in the program Motif Matcher (https://users.soe.ucsc.edu/~kent/improbizer/motifMatcher.html) to search for probable PBS elements.

The improbizer program (https://users.soe.ucsc.edu/~kent/improbizer/improbizer.html) was used to search for regulatory elements in the *hlh-6-like* genes. 1500bp of upstream region from the ATG were used from *Mh-hlh-6-like, Mf-hlh-6-like, Mi-hlh-6-like-1, Mi-hlh-6-like-2,* and *Gp-hlh-6-like*. One motif was searched for occurring once per sequence with an initial motif size of six. Other settings were left as default. Ten control runs, which used randomized sequence of the input genes, were performed to demonstrate the original motif found on actual sequence was significant. This method was adapted from Smit et al. (2008).

2.5 Antibody Development and Use

2.5a - Antibody development

The predicted CeHLH-6 and MhHLH-6 proteins had a predicted mass of 30.8kD and 24.4kD respectively as determined by ExPASy Compute pI/Mw Tool (Gasteiger et al., 2005). Antibodies were developed by using GenScript's silver polyexpress package, which included peptide selection, peptide synthesis, pre-immune serum, inoculation of two rabbits, and antibody purification. Peptides for antigenicity were selected outside of the DBD region and within areas of hydrophobicity to ensure binding of the antibody to proteins in their natural folded state, or DNA-bound proteins. In this way antibodies were developed that could be used for immunohistochemistry and chromatin immunoprecipitation (ChIP). Epitopes were chosen in the N-terminus region, and another in the C-terminus region in each Ce-HLH-6 (Ce-HLH-6-1: CQSQPKPSSKASLDT; CeHLH-6-2: CFNGFQEESEGNI) and Mh-HLH-6 (Mh-HLH-6-1: EKGIKRKRGDEIDQC; Mh-HLH-6-2: MGECSSGENTPPPN). GenScript synthesized 2 peptides (13-15 amino acids long) for each Ce-HLH-6 and Mh-HLH-6, and ensured a cysteine residue was on one end to enable conjugation of the peptide to the adjuvant KLH. KLH ensures a strong immune response from the rabbits. Mh-HLH-6-2 had an internal cysteine residue and therefore KLH was conjugated to the N-terminal amide group instead. For each antigen, two rabbits were injected. After a 6-week immunization period the antibodies were affinity purified from the anti-serum of the rabbits, tested for an ELISA titer of at least 1:64000, and shipped to our lab. One other antibody used was the monoclonal antibody MH27 which binds adherens junctions in *C. elegans*. It was

generated from mice cells, and was provided by the Gaudet lab (Francis, R., Waterston, 1991).

2.5b - Protein Sample Preparation

Liquid cultures of worms were settled on ice for one hour, supernatant removed, and worms were collected from the bottom of Erlenmeyer flasks into 50 mL conical tubes. Tubes were then spun at 1000g for three minutes, the supernatant removed, pellet resuspended in 40 mL of M9, and this wash was repeated two times in order to remove as much bacteria as possible. The pellet of worms was then resuspended in Adjusted Nuclear Buffer (ANB: 10mM Tris HCl (pH 8), 80mM KCl, 2mM EDTA, 0.1% NP-40, and 1 tablet of EDTA-free protease inhibitor (Roche)/ 10mL solution). The original Nuclear Buffer recipe was from Zanin et al., 2011.

Ceramic mortar and pestles were autoclaved, and pre-cooled using liquid nitrogen. The nematode/ANB suspension was dripped into liquid nitrogen in the mortar. Frozen balls of worms (garbonzos) were ground thoroughly, nitrogen was used to re-cool sample and mortar and pestle as necessary. Ground frozen powder was then poured into 50mL conical tube and stored at -80°C, or used immediately for protein purification.

Crude nuclear protein lysates were generated as follows. Ground nematodes were thawed on ice and spun at room temperature in ANB at 100g for three minutes to pellet large debris, and supernatant was removed from the pellet and re-spun at 2000g for fifteen minutes to pellet nuclear portion of extract. The supernatant (cytoplasmic fraction) was then removed from the pellet (nuclear fraction), and all samples frozen at -80°C until next step. Adapted from Shakes et al. 2012.

2.5c - Protein Sample Quantitation

Protein extracts were quantified using a RC DC Protein Assay Kit (Bio Rad) as per manufacturer's instructions. The kit follows the principles of the Lowry Assay (1951) The Lowry is a colourometric assay whereby protein isolates react with copper tartrate solution and Folin reagent to produce a blue colour with a maximum absorbance at 750nm. For the RC DC assay 12.5µL protein fraction was added to12.5µL 2x dye-free SDS-gel loading buffer (GLB: 100mM Tris pH 6.8, 3.2% SDS, 16% glycerol, 40mM DTT (dithiothreitol) and pipetted vigorously. This protein in GLB solution was used as the sample following the regular kit instructions. Optical density readings were taken on a FlexStation 3 microplate reader. Bovine serum albumin (BSA) was used to generate a standard curve, from which protein samples were compared and concentration values interpolated.

2.5d - SDS-PAGE and Electrotransfer

The SDS-PAGE was performed as adapted from Green and Sambrook (2012). A 10% acrylamide resolving gel was prepared (1.9mL H_2O, 1.3mL Tris (1.5M, pH 8.8), 0.05ml SDS (10%), 1.7mL 30% acrylamide mix (AccuGel), 0.05mL 10% ammonium persulfate, and 2µl TEMED) in Erlenmeyer flask on ice, the solution was then quickly poured using a Pasteur pipette between glass plates leaving space at the top for stacking gel (2.1cm from the top of the glass). 100% ethanol was added to overlay the resolving gel and the gel polymerized in a vertical position at room temperature for 45 minutes. The ethanol was poured off and the top of the gel was rinsed with distilled H_2O. A 5% stacking acrylamide gel (0.68mL H_2O, 0.13mL 1.0M Tris (pH 6.8), 0.01mL 10% SDS,

0.17mL 30% acrylamide mix (AccuGel), 0.01mL Ammonium persulfate (10%), and 1μL TEMED) was then prepared in a similar manner and pipetted on top of the resolving gel, a five or ten well comb of 1mm or 1.5mm thickness was then added and the gel polymerized over 45 minutes at room temp in a vertical position. The comb was then removed, and the wells were washed with distilled water. The gel was mounted in BioRad electrophoresis apparatus and 1L Tris-Glycine electrophoresis buffer (3g Tris Base, 14.4g Glycine, 1g SDS (pH 8.3)) was added to top and bottom reservoirs.

While the stacking gel polymerized, 10μL of 2x SDS gel-loading buffer (with 0.04% bromophenol blue) was added to 10μL of protein sample of known concentration and heated to 85°C for two minutes. The sample was added to each well in a predetermined order, including a Kaleidoscope Ladder (Bio-Rad) of known molecular weights. The sample volume to add was dependent on comb size, for example 20μl of sample was added to a well in a ten-well comb of 1mm thickness. The electrophoresis apparatus was then attached to a power supply, and 8V/cm was applied to the gel until the dye front moved past the resolving gel, at which point the voltage was increased to 15V/cm, and the gel was run until the dye front reached the bottom of the glass plates.

Following electrophoresis, glass plates and gel were removed from the electrophoresis apparatus, immersed in 4°C transfer buffer (5.82g Tris base, 2.93g glycine, 3.75mL 10% SDS, 20% methanol, 1L H_2O), and the glass plates were removed with wedges. A PVDF membrane (Bio-Rad) pre-cut to the dimensions of the gel was soaked in methanol for five seconds, then moved to transfer buffer along with the gel, filter paper, and fiber pads and equilibrated for fifteen minutes. These materials were then assembled into a gel cassette in the following order: black cassette holder, fiber pad, filter

paper, gel, PVDF membrane, second filter paper, second fiber pad, clear cassette holder. Once the correct order was ensured, the cassette was closed firmly and placed in cassette module in electrotransfer tank, along with an ice pack. The tank was filled till the blotting line with cooled transfer buffer, a stir bar was added, and placed on stir plate, and placed in a larger tank, which surrounded the transfer tank that was then filled with ice to prevent overheating. The transfer was run at 100V for 1-1.5 hours.

2.5e - Western Blotting

After electrotransfer, the membrane was removed from transfer apparatus and rinsed with distilled H_2O. The membrane was incubated in 20mL of blocking buffer (5% Milk Powder (Carnation) in TBST (10mL 1M Tris-HCl (pH 7.4), 9g NaCl, 0.2mL 100% Tween 20, 1L H_2O) overnight at 4°C on a rocking platform. Membrane was then rinsed briefly rinsed in TBST at room temperature, and immersed in 10mL TBST with primary antibody (Ce-HLH-6-1 1:5000; Ce-HLH-6-2 1:5000; pre-immune serum 1:1000) overnight at 4°C. The membrane was then rinsed for five minutes six times in 30ml of TBST at room temperature, and incubated with the secondary antibody (1:5000). The secondary antibody was HRP-conjugated goat anti-rabbit (Bio-Rad). To test antigen specificity, the peptide that was used to generate the antibody was incubated with the primary antibody to show the peptide's ability to specifically bind the antibody. The peptide was added at 10x the concentration of the primary antibody.

The membrane was rinsed for five minutes six times in 30mL of TBST at room temperature, placed on a plastic sheet protector protein side up, and covered with 500µL of freshly prepared ECL solution (250µL solution A, 250µl solution B (Bio-Rad)),

covered with a second sheet protector, and incubated for five minutes with minimal light exposure. Residual moisture was wiped away with a Kimwipe, and the sheet protectors with the membrane in them were taped into a film cassette. In the dark room, film was placed over the membrane and exposed for 30 seconds, and another exposure for three minutes, and developed using an automated developer. If the bands were unclear longer exposure times of fifteen minutes and up to several hours were used to get a stronger signal. As a consistency control, membranes were re-blocked overnight, and treated to the same process as above using tubulin (provided by the Gaudet lab) as the primary antibody (1:1000), and HRP (Bio-Rad) goat anti mouse (1:5000) as the secondary.

2.5f - Immunohistochemistry of Second Stage Larval Nematodes

L2s of *C. elegans*, and J2s of *Meloidogyne* were stained with their respective HLH-6 antibodies: Ce-HLH-6-1, Ce-HLH-6-2, Mh-HLH-6-1, and Mh-HLH-6-2 and the adherens junctions antibody AJM-1. *C. elegans* L2s, and *Meloidogyne* J2s were collected from their respective growth media into glass well slides and rinsed several times with M9. To increase permeabilization nematodes were sedated with levamisole (1:30), and cut in half using needles under a dissecting scope. Several hundred severed nematodes were placed with minimal volume onto a polylysine slide (ThermoScientific). A long cover slip was placed on the sample perpendicular to the slide, excess moisture was wicked away from edges using filter paper, the location of the cover slip was gently outlined using a diamond pen. To freeze crack the nematodes the slide was then flash frozen in liquid nitrogen for ten seconds, removed, and the cover slip quickly flicked off.

Methanol fixation was then carried out by placing slides in -20°C methanol for ten minutes, then -20°C acetone for ten minutes. The samples were rehydrated with an ethanol series of 95%, 70%, 50%, 30% ethanol in PBS (8g NaCl, 0.2g KCl, 1.44g Na_2HPO_4, 0.24g KH_2PO_4,1L H_2O (pH 7.2)), followed by 100% PBS. Slides were immersed for two minutes in each step of the series at room temperature.

The slides were then washed three times in PBS for five minutes each, excess liquid was removed with Kimwipe, and a border was drawn around the sample using a hydrophobic pen. The sample was blocked using 500µL of 20% fetal calf serum, in a humidity chamber for 90 minutes at 4°C. The block was removed and the sample was covered with 50µL of primary antibody (All HLH-6s 1:100; MH27 1:500) with cover slip over top and incubated overnight at 4°C. Slide was then washed three times for five minutes each in PBS, and covered with 50µL of Alexa Fluor secondary antibody (Life Technologies). Secondary antibody dilutions were as follows: 1:2000 Alex Fluor 488 goat anti rabbit for polyclonal antibodies, and 1:2000 Alexa Fluor 568 goat anti mouse for monoclonal antibodies. The sample was incubated at room temperature for one hour with minimal light exposure, then washed three times for five minutes each in PBS. A drop of antifade mounting solution (Life Technologies) was added before placing a coverslip over the sample and sealing it with clear nail polish. The samples were viewed under Zeiss Axio Imager Z1 epiflourescent microscope with rhodamine and GFP filters. Pictures were taken at 400x magnification, and photos were edited with ImageJ (http://rsb.info.nih.gov/ij/) (Schneider et al., 2012).

2.6 Semi-Quantitative Expression Analysis

Semi-quantitative RT-PCR analysis was performed on MhHLH-6, MhHLH-6-like, MhPHA-4, and MhPMP-3 transcripts. Primers were developed using Primer3 software to produce amplicons in length from 65-175bp, with melting temperatures around 58°C, and where possible to be spanning exons in order to prevent contaminating genomic DNA from interfering (See Table 2.2)(Green & Sambrook, 2012; Koressaar & Remm 2007; Untergasser et al. 2012). Template came from *Meloidogyne* strain Que1 cDNA collected variously from egg masses, pre-parasitic J2s, or adult female worms as described earlier, kept at a concentration of 10ng/µL. 20µL PCR reaction included 9.6µL H_2O, 4µl 5x Phire-II buffer (Thermo Fisher Scientific), 2µl 2mM dNTPs, 1µL each 10µM forward and reverse primers, 2µL template cDNA, 0.4µl Phire Hot Start II DNA Polymerase (Thermo Fisher Scientific). A no-sample water control was included in every run. PCR reactants were overlaid with 10µL of mineral oil to prevent evaporation during removal cycles. Samples were heated to 98°C for 30 seconds, followed by 40 cycles of 98°C for five seconds, 56-60°C (depending on the primers) for five seconds, 72°C for ten seconds, then given a one time final extension of 72°C for one minute. PCR samples were removed after two minute pauses in the cycle at the extension phase of PCR (72°C) at cycles 32, 34, and 36. Each cycle sample was quantified on an Agilent DNA1000 chip on Agilent Bio Analyzer following manufacturer's instructions.

a) *Mh-pha-4* (MhA1_Contig1988.frz3.gene1)

ggtttaattacccaagtttaag
oJDK0288-SL1.1-F-1
atgatatcatgggaagttgtagaagaagtaagacaacaaatacaacaacaacaagaaaag

caacaacaacaacagcaacaatttatttcttctcttgatatttcttcaacaaattcttta

acaggaggaatgtcttcaacaacaacaacaacagattatctaaatgaccaacaacaatgg

aataatatttatcaaaataataattcttcaccaaattcttcttccacaatttcattaatt

aaccaaaaacccttcattaactaattctttatattcttccccaatttatcctccacaatat

gcttctttaaatggttttcaaaataatctttatattccttcttcttcccctttcttcttct

tcaaatccacaaatttgtcaatatatacaacaacaacaattattagaacaaaataattct

ttaaatttatcatatacttctgcttttcaatcagcaattaattctgctacaaataattta

aattcttcaattagaggggagaagaacaaaaatatattgaaataaataaacaattaaaa

aatagtaaaaatcagtcaacaaataatcaaattggagattctaataataatcaaagtcca

acaataaatactatggaattaactgttcatgaaatgcaaaaaattaaaaatcaaggaaat

tttggtccaaatAAACCTCCTTATTCTTATATTTCTTTAATTTCTATGGCTATACAACAA

AGTGACCGAAAAATGTGTACATTGAGTGAAATTTATAATTTTATAATGGAATACTTTGAA

TATTACAAAAATCATCGTTGTTCTTGGCAAAATTCAATAAGACATTCTTTATCTTTTAAT

GATTGTTTTGTTAAAGTACCTCGAACACCTGATAGAATAAAAAATTGCAGACCAGGCAAA
oJDK0167-F-2
GGTTCTTTTTGGACATTACATGCATTGTGTGGAGATATGTTCGAAAATGGATGTTTTTTG
oJDK0168-R-1
AGAAGACAAAAACGTtttaaattaaaagataaagaaaaaccagaaaggcaaagaagaact

aaaaatatttcgaataaattaactttaacagaaagaaaattacaaaaacaacaaaaacaa

atgattaattctgcttcttctattttattagaacaaaaaccaaaaagggaaatagaatct

tcaatagatttaaaattagaacctttaattcaaacagatataggactacgactactacaa

ttaataacagaagaaataaataataataataatttaatagatattattggggatggaaat

aattcttccaataaagaaaaacatttaaataataatttaattcaattaataaggccatca

gcagaaaatttaaataataatttaattttaccttcacaatttaataataattttgaatta

tctcatttaacaacacaatgcccttcaagttcttctggaattttacaaggaaattctaat

ttatgtggatctcctgttatttcttctattggaggggacaaatgttatcatcattatct

tctgcttattgtcatcaatttccaaattcaacaacagctatagtccatccaacaataaca

caacaatttccccatcaattttatacttcaaataattctgatttagataattgttcaagt

cctcaatcaacaaataataatttaaattcacaacaattaacatgccttaattcttctgga

ccttttttaaatattaatttaaatataaatcaaaatttaattgattatcaaccaataatt

cctcctaattttccaccagaatattcaatgatttatggaggaccagcacagattaataca
oJDK0174-R-2
atttataatggaaagttagttttttttaaatatttttgtttaatgtctttaaatattttt

ttttaa

b) *Mh-hlh-6* (MhA1_Contig384.frz3.gene12)

ggtttaattacccaatttgag
SL1.2 oJDK0289-F-1
atgaacaataatccaataaatgaccaatcagtccctttaacaatagatattctcaccaat

gaacaacaaatgcttctagaacgtcaacaagcattatgtgtagctgctgctcaggctgtc

aattttgctaatataaaatatggaagcttctgcctttcaaagaagcagcagtgtaatagtt

aatccaccaattcaagaaaaaggaattaaaaggaaaagaggagatgaaatcgaccaacaa

gaaatacaacaaaaaatgcctaatatgggttgcgggatgcgggaaccaaaagaagttgta

ggtgtaaatcgttttattagtccttttcggaagaggctaaaataccgttgccacacgaa

ttgaatgagttaagcacacatggtacaaccgtttggaaaCGCAATGAAAGGGAAAGAATT

AGAGTTCGAGAAGTTAACAATGGATATACGAATTTGAAAAGCAAGTTGCCTTTGACTGAA

AACGATCGACGTCTTAGCAAAGTTGATACTCTTCGCCTAGCTATTGCATATATTCATCAT
oJDK0307-F -2 oJDK0170-R-1
TTGGAAaatttaattaatgaaggagtaaatcatttgattgagtgccaatgttttaattat

gcgatgggggaatgcagcagtggagaaaatacaccaccaccaaattaa
oJDK0216-R-2

c) *Que1-hlh-6*

ggtttaattacccaatttgagtttacaatgaataataatccaataaacgaccaatcacaa

cccttaactatagatattctcactaatgaacaacaacaaatgcttctagaacgtcaacaa

gcattgtgtgtagctgcggcgcaaaccattaattttgctaatataaatatagatgcttct

gcttttcaaagaagcaacagcgtaatagtagttccacctagccaagaaaaagggactaaa

agaaaaagaaaagatgataatgggcaacaagaattacagcaaaaaatgcctaaaatgaca

acacctgttaaaagcccaaaagaggttttaggtgtaaatcgttttattagtcccttctct

gaagaggctaaaataccgttGCCGCATGAATTGAATGAATTGAGCACACACGGAACTACC
oJDK0331-F
GTCTGGAAGCGTAATGAACGAGAAAGGATTCGCGTCAGACAAGTTAATGATGGATATACG

AGGTTGATAAACGATCGACGTCTTAGCAAAGTTGATACGCTTCGCCTAGCTATTGCATAT
oJDK0332-R
ATTCACCATTTGGAAAacttaattaatgaaggggttaatcatctaatcgagtgccaatgt

tttaattatgtaatgggagaatgtagcagtgctgaaaatacaccaccaccaaattaa

d) *Mh-hlh-6-like* (MhA1_Contig809.frz3.gene10)

ggtttaattacccaagtttaag
SL1.1 oJDK0288-F-1
atgtttgaatgttcttcttcttcatcttggctttctaatcaacaaaataattcttcatat

tcttcatcattttcccctattaattcttctaatttatatcaatcaatatctttaaataat

aattctaacaatttttatttaaataatacaaaaataaatttaaattcaaaacaacaaaaa

ttaaataaaaatggaaaagaaaaacaaaaaaatattttatcaaaaagaaaaagaaataat

aaaaataatttaaatgttggaattaaatttggaggtaaaaagccacatcaaGTTGCTAGA

AGAAATGAAAGAGAACGTAAAAGAGTTCAACAAGTAAATGATGGATATGAAAAATTAGCA

AATACTTTAAATAATTTCGAACCTATTTGCAATGAAAGAAAATTAACAA**AAGCAGAAACA**
oJKD0301-R-1
TTAAAAACAGCAATTTTATATATTAAACATTTAGAAGAATTATTAaaacaacaacaacct
oJDK0319-F-2
ttaaaaataaatttagaaaaacaaaataaaattaaaaatgaaaataatattgaaaattca

aataatttttatcagaatttaaattcttctcaagaaaatcaaaaagaacagcaattctta

acaaaacattcaccaactaattttatttattcacaatcaataaatacaaattataataat

acaaataataataatatttatttaaataataattcatttcaaaattgtttatttccttct
oJDK0320-R-2
aaacaaaaataa

Figure 2.1 Primers used for cloning are underlined in predicted sequence of a) *Mh-pha-4* (strain VW9), b) *Mh-hlh-6* (strain VW9), c) *Que-1-hlh-6*, and d) *Mh-hlh-6-like* (strain VW9). VW9 is the published *M. hapla* genome sequence (Opperman et al., 2008), and Que1 is a Canadian regional isolate of *Meloidogyne* from Quebec which was used for all molecular work in this report. DNA binding domains are capitalized; primers are underlined with primer name underneath with F or R included in name to denote directionality, and a number to show pairs. Primer is in bold text if overlapping another.

Table 2.1 Cloning Primers					
Target	Fragment region	Primers (oJDK)	Tm°C	Amplicon (bp)	Nested
Mh-pha-4	5'	288: GGTTTAATTACCCAAGTTTAAG 168: CCATTTTCGAACATATCTCCAC	56	904	no
	mid	167: AAATTGCAGACCAGGCAAAG 174: TTCTGGTGGAAAATTAGGAGGA	54	923	yes
Mh-hlh-6	5'	289: GGTTTAATTACCCAATTTGAG 170: TGCAATAGCTAGGCGAAGAG	56	509	yes
	mid	331: TGAGCACACACGGAACTACC 332: GCTAGGCGAAGCGTATCAAC	58	130	no
	3'	307: AACGATCGACGTCTTAGCAA 216: TTAATTTGGTGGTGGTGTATTTT	56	168	no
Mh-hlh-6-like	5'	288: GGTTTAATTACCCAAGTTTAAG 301: TGCTGTTTTTAATGTTTCTGCTTT	54	360	no
	mid	319: AAGCAGAAACATTAAAAACAGCAA 320: GAAGGAAATAAACAATTTTGAAATGAA	58	310	no
TOPO-XL-Vector	plasmid insert	M13F: GTAAAACGACGGCCAG M13R: CAGGAAACAGCTATGAC	54	variable	no

Table 2.2 Semi-Quantitative Expression Primers				
Transcript	Primers (oJDK)	Tm°C	cDNA (bp)	Genomic (bp)
MHPHA-4	167: AAATTGCAGACCAGGCAAAG 168: CCATTTTCGAACATATCTCCAC	58	69	69
MHHLH-6	307: AACGATCGACGTCTTAGCAA 216: TTAATTTGGTGGTGGTGTATTTT	56	168	168
MHHLH-6-LIKE	300: AAAAGAGTTCAACAAGTAAATGATGG 301: TGCTGTTTTTAATGTTTCTGCTTT	60	114	163
PMP-3	211: TTTGGTGGAGTTACCCATCG 212: GGAGGACGTTCGGTCTCTAA	60	86	125

```
A 585 534 537  41   0 862 1010 1010   0 1010
C 71    0   0   0 407 148    0    0 623    0
G 0   471  79 969   0   0    0    0   0    0
T 354   5 394   0 603   0    0    0 387    0
```

Figure 2.2 Positional Weight Matrix of PBS from JASPAR database shown here, was used to search the upstream regions of gland specific genes in *C. elegans* and PPN for PHA-4 binding sites. The above sequence extends the canonical Fox-A binding site from TRTTKRY, to TRTTKRYHHD (R = A or T, K = C or T, Y = G or T, H = A, T, or C, D = A, T, or G

3. Results

3.1 RKN Husbandry

To expand nematode cultures quickly, soil with *Meloidogyne* was added to that around the roots of the plants at the time they were transplanted. Isogenic strains were developed by removing single egg masses from infected roots, and transferring them individually onto new uninfected plants. Three strains of *Meloidogyne*: Quebec1 (Que1), Quebec2 (Que2), and Ontario1 (Ont1) were raised on tomato plants in this way. The two Quebec strains were isolated from carrot plants in Ste-Cothilde-de-Chateauguay Quebec, and after being multiplied several generations on tomatoes, were sent to us by Dr. Benjamin Mimee. The Ontario strain was isolated near Guelph in Ontario, and was provided by Dr. Mary Ruth McDonald. The strains were first reared in Soil Science at the University of Manitoba, then passed on to our lab. Ontario1 was grown on both tomatoes and carrots. Of the strains, Que1 was found to propagate the most quickly, and was therefore used for all of the molecular work in this report.

3.2 Genotyping

Initially it was assumed based on the reports of the nematologists who sent the strains to our lab, that strains Que1, Que2, and Ont1 were all *M. hapla.* As research progressed this was shown to be incorrect and that in fact Que1 and Que2 were *M. incognita* and only strain Ont1 was *M. hapla*. The *Meloidogyne* DNA was tested with various primers to differentiate *Meloidogyne* species. Primers JMV1, JMV-hapla and JMV-tropical (oJDK0226, 0228, and 0229) were used to differentiate between temperate and tropical species. JMV1 was paired with JMV-hapla for *M. hapla* (440 bp), and with

JMV-tropical for tropical RKN such as *M. incognita* (549 bp). These primers exploit length polymorphisms of the intergenic spacer (IGS) region in closely related species to distinguish one from the other (Wishart et al., 2002). As can be seen in Figure 3.1 strain Ont1 tested positive for *M. hapla*, and Que1 tested positive for a tropical RKN species.

3.3 Homologues of HLH-6 and PHA-4 in RKN

Previous work has shown that PHA-4 and HLH-6 are involved in the developmental pathway of pharyngeal gland cells in the free-dwelling nematode *C. elegans*, and it has been proposed that this pathway may also exist in plant parasitic nematodes (PPN) (Gaudet & Mango, 2002; Smit et al., 2008). Other research has been conducted proving the usefulness of *C. elegans* as a model in studying PPN (Burns et al., 2015). The following are the results of experiments attempting to study and define the glandular developmental pathway in the PPN *M. incognita.*

Homologues of HLH-6 and PHA-4 were found amongst various nematode species. To ensure their accuracy, their DBDs were compared phylogenetically, and to deduce functionality, the promoter regions of *hlh-6* were compared for similarities.

The full *C. elegans* amino acid sequence of HLH-6 and PHA-4 were trimmed down to their DNA binding domains (DBD) using Pfam as a guide to where their active sites were (Finn et al., 2015). The genes were slightly extended from their Pfam predicted DBDs by extending the amino acid sequence three amino acids towards the N-terminus, and three amino acids toward the C-terminus in HLH-6, and in PHA-4 the DBD was extended to include five more amino acids towards the N-terminus, and fifteen more amino acids towards the C-terminus. These additional regions were included because

they included conserved residues that extended directly from the Pfam predicted DBD. These regions of conservation can be seen in the CLUSTAL Omega alignments in Figure 3.2.

Homologues of HLH-6 and PHA-4 were searched for by BLASTing their (DBD), against clade V nematodes: *C. elegans, C. briggsae, Pristionchus pacificus, Heterohabditis bacteriophora,* clade IV nematodes: *M. hapla, M. incognita, M. floridensis, Globodera pallida,* and clade III nematode: *Brugia malayi*. A wide range of nematode genomes were included in the BLAST search in order to understand how gland development is evolving in nematodes of diverse feeding habits. *C. elegans* and *C. briggsae* are closely related free-dwelling bacteriovores (Stein et al., 2003). *P. pacificus* lives in association with beetles, and feeds on bacteria, fungus, and other nematodes (Sommer, 2006). *H. bacteriophora* is an insect parasite, and *M. hapla, M. incognita, M. floridensis,* and *G. pallida* are plant parasitic nematodes (Abad et al., 2008; Cotton et al., 2014; Lunt et al., 2014; Opperman et al., 2008). *B. malayi* is a filarial nematode, capable of parasitizing humans (Ghedin et al., 2007).

C. elegans HLH-6 and PHA-4 DBDs were used as queries in a BLAST search, and hits were included for further analysis if they had an e-value of 0.05 or less. A list was thus assembled of possible homologues and closely related proteins of HLH-6 and PHA-4 in a cross section of nematode species. Some homologues were missing in the newly annotated genomes, or were only partial hits; envisaging these as mis-predicted splice sites due to their lacking cDNA data, the more likely complete gene was found manually by using tBLASTn. The complete DBD of a related species was used as the query in a tBLASTn search on ParaSite BLAST page against the species, which had an

incomplete or missing DBD. The top hit often revealed a more complete gene than the original prediction, through new or extended exons. Thus some of the genes used in the phylogenetic analysis were predicted based on homology to *C. elegans* genes. See Figure 3.3 for an example of a gene re-predicted *via* comparison to the *C. elegans* genome. The genes that were re-predicted from their original Worm Base conformation are noted in appendix Table 6.1.

As research progressed, RKN sequences of HLH-6, HLH-6-LIKE, and PHA-4, became available from cDNA sequence data from our own strains of *Meloidogyne*, and thus more accurate predictions could be made for these genes. In some cases it appeared that the predicted peptide sequence for these RKN proteins was incorrect. As seen in Figure 3.4, the original Mh-PHA-4 DBD compared to the actual DBD as translated from Que1 sequence data has its splice sites mis-predicted. These were adjusted with Que1 sequence data, and thus the accuracy of the phylogenetic analysis improved.

The DBDs were aligned with T-Coffee and presented using Boxshade from ExPASy as seen in Figure 3.5 (Artimo et al., 2012). The closely related bHLH factors include the *C. elegans* proteins HLH-1, HLH-2, HLH-4, HLH-3, and HLH-14 seen in Figure 3.5A. HLH-1 is the Myo-D homologue, the only myogenic regulatory factor (MRF) in *C. elegans;* it is essential for body wall muscle development (Krause et al., 1990). HLH-1 was used as the out-group in the phylogenetic tree, as it is the only class II bHLH protein included in this analysis (Massari & Murre, 2000). HLH-2 is a class I bHLH, and is expressed in embryos in neuronal cells, and after embryogenesis is broadly expressed and present in many structures and cell types of the worm including pharynx, neurons, muscle, and vulval cells. HLH-2 dimerizes with many bHLH's including HLH-

3, and HLH-14 in neuronal development (Krause et al., 1997). HLH-2 is a possible binding partner of HLH-6; it has been shown to bind HLH-6 in a yeast two-hybrid assay, though not in an electrophoretic mobility shift assay (Grove et al., 2009; Smit et al., 2008). It is possible that HLH-2 and HLH-6 bind PGM1 together, but require a third factor to complete a stable complex(Smit et al., 2008).

The forkhead transcription factors included for analysis are the *C. elegans* proteins LIN-31, FKH-2, FKH-9, and DAF-16 aligned in Figure 3.5B. LIN-31 is a forkhead-like transcription factor that negatively regulates vulval development (Howard & Sundaram, 2002). FKH-2 is a forkhead protein like PHA-4; it is important in the development of chemosensory neurons, and has been shown to positively regulate a number of other genes (Molin et al., 2000). FKH-9 is a forkhead with an unknown function in *C. elegans* (Hope et al., 2003). DAF-16 is the only Forkhead O (FOXO) homologue in *C. elegans*, and targets the insulin/IGF-1 signaling which in turn helps regulate major processes such as longevity and metabolism (Kwon et al., 2010). These genes, which are closely related to PHA-4 and HLH-6, help contextualize the phylogenetic trees, and show how the TF families are evolving amongst nematode species.

Specific homologues were verified using MP and ML phylogenetic analyses that included related bHLH and forkhead proteins. The phylogenetic trees were developed using the conserved DBD region of each protein. For alignments of each TF and their homologues separately, see appendix Table 6.1. The species specific full gene names (searchable on Worm Base) of the homologues of HLH-6 and PHA-4 are listed in Table 6.1, along with other closely related genes that were initially considered as possible

homologues. Using this list of genes a phylogenetic analysis was conducted for forkhead TFs including PHA-4, and bHLH TFs including HLH-6.

These collections of homologues were analyzed with maximum parsimony (MP) and maximum likelihood (ML) phylogenetic trees after aligning the sequences with Clustal Omega. MP and ML are both character based, tree-searching methods, similar to Bayesian Inference. These tree-searching methods begin with a random tree established from Clustal alignments, and search for simpler trees based on scores of evolutionary divergence. Trees with lower scores, and therefore lower levels of evolution are kept, until the tree with the lowest possible score is found. Phylogenetic analyses like MP and ML use amino acid replacement matrices to estimate evolutionary distances between sequences. MP uses a general matrix to calculate the minimum number of changes to generate the branch order on a tree. ML uses a more intricate model, which notes which amino acids are being exchanged in the sequence. ML estimates the substitution rate of any amino acid to any other amino acid, taking into account the physical properties of the amino acids that are changed. The LG matrix developed by Le and Gascuel was used for the analysis of the data sets analyzed in this report as recommended by MEGA version 6 (2008).

MP is a slightly older and faster program than ML, but ML has been shown to be generally more accurate (Hall, 2011). It has been argued that MP is the more accurate method when rates of evolution are non homogenous across the sequences (Kolaczkowski & Thornton, 2004). However, the sequences included within the data set analyzed here are all from the DBD, that is the active site, of the proteins, and therefore more likely to be evolving at a similar rate, than sequences from disparate regions of a

protein (Hall, 2011). Based on this assumption, the ML tree will be more accurate than MP.

The MP tree as seen in Figure 3.6, shows that most of the bHLH proteins group together into clades with their predicted homologues. A clade in this sense are all of the descendants of an interior node, and taxa are the branch ends of the tree which represent the existing sequences used in the analysis (Hall, 2011). The MP tree for bHLH proteins shows the HLH-6 taxa grouping together, and clearly places the predicted PPN proteins predicted by BLAST analysis with their *Caenorhabditis* homologues. The MP consensus tree collapses branches with less than 50% support from other equally parsimonious trees. This lack of uncertainty is clearly indicated in the many polytomies.

The internal node, representing the most recent common ancestor, of the HLH-6 clade, had a bootstrap value of only 58. For each node on the trees, a bootstrap score of 70 or over correlates with a significance level of 95%; values under 65 are considered unresolved (Hillis and Bull, 1993). The HLH-4 clade, and the HLH-1 clade grouped together with clarity. The HLH-4 clade has a bootstrap value of 77, and HLH-1 of 99. The HLH-14, and HLH-3 predicted homologues are shown to intermingle with each other; both of these TFs are involved in neuronal development (Grove et al., 2009; Krause et al., 1997). The HLH-2 clade, the only class I bHLH protein included, was the out-group. This tree helped prove the RKN HLH-6s found through BLAST searches were the most likely homologues to their non-parasitic counterparts. However because the HLH-6 clade had a relatively low bootstrap value, further validation was pursued.

An interesting result from this phylogenetic analysis of bHLH proteins was the discovery of a new gene group, one that grouped separately from the other classes of *C.*

elegans TFs in the MP consensus tree, and yet consistently grouped closely with the HLH-6 proteins. Because the new gene aligned most closely to HLH-6 it was called HLH-6-LIKE. This new gene group was found exclusively in clade V nematodes, that is, the plant parasites. HLH-6-LIKE is most likely either a diverged form of HLH-6, or the missing HLH-3 or HLH-14 gene. There was a very high degree of conservation within the HLH-6-LIKE DBD as seen in Figure 3.5A, especially amongst the RKNs. Because of this ambiguous gene group near HLH-6, the data were analyzed further with the more accurate but time consuming ML phylogenetic method

To probe the data further a ML tree was made using the same set of protein sequences from Figure 3.5A but run using the ML method on the program MEGA version 6 (Tamura et al., 2013). These results can be seen in Figure 3.7A. The bHLH tree revealed greater strength throughout the clades of regulatory genes, and the new unclassified PPN gene HLH-6-LIKE placed closest to HLH-6. There are strong bootstrap values in the ML tree for the internal nodes for the transcription factor families of HLH-6 (85), HLH-4 (90), HLH-1 (98), and HLH-2 (98) suggesting strong conservation of these TFs across the nematode phylum. The HLH-1 clade was especially strong and had very little divergence in any of the species. Homologues for *hlh-6* were verified for each species analyzed, including *M. hapla*: *MhA1_Contig384.frz3.gene12* (*Mh-hlh-6*), *M. incognita*: *Minc00130*, and *Minc07019* (*Mi-hlh-6-1, Mi-hlh-6-2*), and *M. floridensis*: *maker-nMf.1.1.scaf00640-augustus-gene-0.13* (*Mf-hlh-6-1*), and *maker-nMf.1.1.scaf11064-augustus-gene-0.3* (*Mf-hlh-6-2*). Note that *M. incognita* and *M. floridensis* often have paralogues of genes due to the hybrid nature of its genome.

The gene list for both phylogenetic analyses had some notable absences; some homologues of the bHLH proteins could not be found. The Myo-D homologue HLH-1, an important gene in muscle development, could not be found in *G. pallida*. However, orthologues of HLH-1 were found in other clade IV nematodes (Krause et al., 1990). Other proteins that were absent from analysis were the homologues of the out-group class I bHLH protein HLH-2 in *M. incognita,* and *G. pallida*. HLH-2 is important as the most likely heterodimer partner for HLH-6 (Grove et al., 2009).

HLH-14 and HLH-3 homologues may not exist in PPN. The conservation of HLH-14 and HLH-3 homologues are much less strong, the likely homologues of these genes do not group neatly within their families as the other bHLH proteins do. Notably absent from the HLH-14 and HLH-3 groups are any clade IV homologues. Because this result is consistent in all of the PPN species tested, it may be that HLH-3 and HLH-14 orthologues do not exist or are highly diverged in clade IV nematodes.

The ML tree of bHLH DBDs shows increased certainty in the placement of the HLH-6-like proteins. They are placed near the HLH-6 family, and are predicted to branch from the HLH-6 line. The history of the HLH-6-LIKE group is unclear, but it is likely a plant parasite-specific TF.

Most of the forkhead proteins were shown to group together into predicted orthologous groups, as seen in the MP phylogenetic tree in Figure 3.6B. The PHA-4 clade grouped together with a bootstrap value of 81 thus helping prove our homologues are likely correct. The other FoxA clades like LIN-31, FKH-2, and FKH-9 grouped together with significant bootstrap values. Only Bm-LIN-31 fell out of its predicted

place, branching from FKH-2 descent rather than LIN-31. DAF-16, the only FoxO TF forms the out-group in this tree.

A ML tree was also compiled from the forkhead genes to give robustness to homology proof. FKH-9 has a family node value of 95, DAF-16 a value of 99, and FKH-2 a value of 80. Within the LIN-31 family the predicted *B. malayi* homologue Bm-scaffold1191:-660:1444:1 again appears to be an outlier, grouping more strongly with FKH-2 family. *B. malayi* is the only clade III nematode included in the analysis, and it is therefore not surprising that this species sometimes acts as an outlier in the trees. The LIN-31 family, not including the *B. malayi* homologue, has an internal node bootstrap value of 95.

The PHA-4 clade has a bootstrap value of 78. The initial Worm Base predicted homologue of *pha-4* was *Minc12066* and appeared truncated. It did not align well with other RKN *pha-4* homologues. When it was included in a phylogenetic analysis it grouped in the PHA-4 clade but the bootstrap value was 67, and of low statistical significance (not shown). When cDNA data for Mi-PHA-4 from Que1 strain was used instead to create the translated DBD of the transcription factor, the PHA-4 clade held together with a significant bootstrap value of 76. *MhA1_Contig1988.frz3.gene1* (*Mh-pha-4*) is the most likely *M. hapla* homologue of PHA-4, and *maker-nMf.1.1.scaf03071-snap-gene-0.6* (*Mf-pha-4*) is the probable *M. floridensis* homologue.

Some genes are missing from the forkhead transcription factor trees including homologues of FKH-2 in *M. floridensis,* and the DAF-16 homologue in *M. incognita, P. pacificus,* and *H. bacteriophora*. Just as *B. malayi* of clade III tends to act as an outlier in each family of proteins, clade IV nematodes (*M. incognita, M. hapla, M. floridensis, G.*

pallida), and clade V nematodes (*C. elegans*, *C. briggsae, H. bacteriophora, P. pacificus*) group together. This is generally seen across all four phylogenetic trees presented here.

3.2 Gland Specific Promoters

3.2a - The promoters of HLH-6 targets

The promoters of some RKN secretome genes share similarities to the *phat* genes produced in *C. elegans'* glands. The promoters of gland specific genes under HLH-6 and PHA-4 control are well characterized in *C. elegans*. Therefore in addition to the search for direct homologues of the TFs HLH-6 and PHA-4, the promoter regions of secretome genes - possible HLH-6 targets in PPN - were analyzed for similarities to HLH-6 targets in *C. elegans*. These similarities help prove further homology in the gland development genetic pathways between the nematode clades. Similar regulators could mean similar functionality, as a specific set of regulatory elements upstream of a gene allows for specific temporal and spatial expression patterns. In adapting to parasitism PPN secretome genes may well have evolved to be controlled by the existing transcription factors controlling gland secretions that is seen in *C. elegans*.

An important conserved sequence in the upstream region of HLH-6 and many gland specific genes in *C. elegans* is PGM1 (pharyngeal gland motif 1) defined as CAnvTGhdYMAAY (V=A, C or G, H=A, C or T, D=A, G or T, M=A or C, and Y=C or T) (Smit et al., 2008). The sequence was later redefined less stringently as CAnnTGnnYMAAY or CAnnTGnYMAAY (Ghai et al., 2012). I am calling the original PGM1 “stringent-PGM1”, and the redefined PGM1 “variable-PGM1”.

Some RKN secretome genes contain the PGM1 enhancer. To test the model that gland specific proteins in PPN are under transcriptional control of HLH-6 as they are in *C. elegans*, well-studied secretome proteins were analyzed for the presence of the HLH-6 binding site, PGM1, in their upstream regions (within 1500 bp of start codon). Table 3.1A shows gland specific genes from *C. elegans* and the location and type of PGM1 in their promoter regions. These sites were found initially by searching for a shared motif among gland specific gene promoter regions, and then tested for functionality (Smit et al., 2008). The genes used to find the motif were *hlh-6, B0507.1, phat-1, phat-2, phat-3, phat-4, phat-5, phat-6,* and *dod-6.* Most of the PGM1s are within 500 bp of the start codon, only the PGM1 ahead of *phat-6* is substantially further away at -1815 bp; its PGM1 is in the coding region of the gene upstream of it. This hit is much less likely to be true, as all of the other PGM1s in the other PHAT proteins are within 250 bp of the start codon. Several of the other *phat* genes were experimentally determined, using transcriptional reporters, to be PGM1 dependent (Smit et al., 2008). However *phat-6* was not verified to be PGM1 dependent experimentally, and the fact that its PGM1 is so far away in another gene, suggests it may not in fact be under HLH-6 regulation.

PGM1 was also searched for in RKN as seen in Table 3.1B. Six well-studied gland gene were analyzed in *M. hapla* and *M. incognita* homologues. Homologues of *M. hapla* and *M. incognita* were used as these are the two most well studied RKN. However where *M. incognita* data was missing, homologues from the very closely related RKN, *M. floridensis*, were used as a substitute. Twelve secretome genes were examined in total: *Mh-crt, Mi-crt, Mh8D05, Mf-8D05, Mh-eng-1, Mi-eng-1, Minc03866, Mha1_contig203.frz3.gene9, Mh-D15, Mi-D15, Mh-eff-1,* and *Mi-eff-1*. These specific

genes were looked at because they are all verified genes of the secretome, that are produced specifically in either the SvG or DG cells. Four of the gene pairs were SvG specific, two were DG specific. Representative effectors of both the SvG and DG were shown to be positive for the presence of PGM1. In total six of the twelve secretome genes had a PGM1 element in their upstream regions including: *mh-crt, mh-eng-1, mi-eng-1, mha1_contig203.frz3.gene9, mi-d15,* and *mf-eff-1*. *mh-crt* is a calreticulin gene, predicted to be involved in suppressing the plant immune system (Jaouannet, Magliano, et al., 2012). *mh-eng-1,* and *mi-eng-1* are endogluconases, presumably used to degrade plant cell walls during invasion, and giant cell formation by PPN (Rosso et al., 1999). *mha1_contig203.frz3.gene9* has unknown function, but has been shown to be an important protein in the secretome through RNAi knockdown experiments (Danchin et al., 2013). *mi-d15* and *mf-eff-1* have nuclear localization signals, and are therefore thought to localize to plant cell nuclei where they could help reprogram the plant cell as a feeding structure for the nematode (Jaouannet, Perfus-Barbeoch, et al., 2012; Lin et al., 2012). Strangely, in only one homologue pair was PGM1-positive in both genes. Four of the six positives were in *M. hapla*. This small data set lends weight to the theory of HLH-6 controlling at least a sub-set of gland specific genes in *Meloidogyne* through PGM1, as it has been shown to do in *C. elegans*.

Gland specific genes from *Meloidogyne,* including *hlh-6* homologues, and secretome genes, maintain PHA-4 binding sites (PBS) in their promoter regions. PHA-4 binds PBS, which has the canonical sequence TRTTKRY (R = A/G, K = T/G, Y = T/C) and is found upstream of all pharynx specific genes in *C. elegans* including gland specific genes. Several candidate PBS sites were found in all upstream regions of *Meloidogyne*

gland specific genes, just as they had been noted in *C. elegans* genes. See table 3.1B for the number of PBS sites per promoter region for each gene analyzed. Although many candidates were located, it is likely that only a few are legitimate binding sites of PHA-4. For example in the *C. elegans phat-5* gene there are fully eight TRTTKRY elements present in upstream sequence, but only two of them were found to be active (Smit et al., 2008). This is because other *cis*-factors are involved in its binding process.

A more rigorous PBS algorithm, "PBS-stringent", uses a PWM from JASPAR for PHA-4 binding sites to probe promoter regions of gland specific genes via the online program Motif Matcher; PBS-stringent follows the sequence TRTTKRYHHD (R = A or G, K = G or T, Y = C or T ,H = A, C, or T, D = A, T, or G). This search showed many fewer possible binding sites than the canonical sequence, but at least one stringent-PBS element was located upstream of each PPN gland specific gene as seen in Table 3.1B. This does not prove that these are necessarily active PHA-4 binding sites, but they are candidates for such activity. An embryo specific PBS element, "PBS-embryo", was looked for as well. This novel PBS was uncovered from a ChIP-seq search for DNA regions of PHA-4 activity, and does not contain the canonical TRTTKRY sequence (Zhong et al., 2010). Rather the binding sequence is GAGAGAS (S = G or C). This sequence appears less frequently in the gland specific genes than the sequence containing TRTTKRY, and many of the secretome genes do not have it. This correlates with the model that secretome genes are used later in life stages for invasion of plant tissue, not for embryo development.

3.2b - The promoters of *hlh-6* and *hlh-6-like*

In addition to the promoters of HLH-6 target genes, the promoter of *hlh-6* itself has been well studied. The promoter of *hlh-6* contains several important regulatory elements, which were searched for in PPN *hlh-6* and *hlh-6-like* genes. The characterized regulators of the *hlh-6* promoter include the aforementioned PGM1, and PBS, and several HRL (Hlh-6 Regulatory eLement) elements. The HRL elements are HRL1b(GTCGGAAG), HRL1c(GTGGCTA) HRL1d(GTGGATA), HRL2a(AATAAATA), HRL2c(AATAAATA), and HRL3(TGCACAG). The transcription factor LAG-1 binds to HRL1, but the factors binding HRL2 and HRL3 are unknown (Ghai & Gaudet, 2008).

The *hlh-6* genes analyzed from PPN included *Mh-hlh-6*, *Mf-hlh-6-1*, and *Gp-hlh-6*. Each of these genes had a stringent PGM1 maintained in its promoter region as can be seen in Table3.1. *Mi-hlh-6* could not be analyzed because the upstream region is absent in available sequence data. In contrast all of the PPN *hlh-6-like* genes including *Mh-hlh-6-like, Mf-hlh-6-like*, and *Gp-hlh-6-like* came up negative for any type of PGM1 in their promoter regions. *hlh-6-like* is therefore less likely under the regulatory control of HLH-6.

To search for HRL elements the genomic similarity search tool YASS was used. YASS allowed for the direct comparison of the *C. elegans hlh-6* promoter with the other *hlh-6* or *hlh-6-like* genes. These promoters were compared between *C. elegans* and *C. briggsae* as a positive control of the method. This can be seen in Figure 3.8A where PGM1, HRL1, and HRL2 show up as predicted conserved regions. When the *C. elegans hlh-6* promoter was compared with the *M. hapla hlh-6* promoter, several copies of PGM1

and HRL1 and HRL2 were also noted. When the *M. hapla hlh-6-like* promoter was compared to *C. elegans*', no promoters were noted, and the only HRLs were relatively distant from the ATG start codon. While the *C. elegans* HRLs fall within the first 250 bp upstream of the ATG, three out of four of the discovered HRLs upstream of *Mh-hlh-6-like* are over 1000 bp away. As has been noted earlier, HRL1 and HRL2 are often in close proximity to one another, and this is also seen frequently in these dot plots (Ghai et al., 2012; I. Raharjo & Gaudet, 2007). Also notable in the dot plots are regions of conservation besides the PGM1 and HRLs, but these did not match known functional regions of the *C. elegans hlh-6* promoter.

A compilation of these searches is given in Figure 3.9, showing the entire promoter of the *C. elegans* and *C. briggsae hlh-6* genes, along with the PPN *hlh-6,* and *hlh-6-like* genes for which upstream data was available. The *C. elegans* promoter has its functional elements near its ATG. The *hlh-6* minimal promoter, as described by Raharjo et al. *et al.* includes the HRL elements and the PBS, and includes the first 241 bp upstream of the ATG start codon (I. Raharjo & Gaudet, 2007). Further upstream is the autoregulatory region of the gene that includes 2 PGM1 elements. The *C. briggsae* version of the promoter appears basically conserved relative to *C. elegans.* The *Mh-hlh-6* promoter also maintains most of the same elements as *hlh-6,* however they are further away from the ATG, with most of the enhancers falling over 500 bp away. The *Mf-hlh-6* promoter has at least one copy of all of the elements seen in *C. elegans,* but they are noticeably more spread out. The *Mf-hlh-6-2* promoter only has HRL elements, and is missing PGM1, and PBS elements. This promoter is the shortest one examined, as there was only data for 674 bp in the upstream region, after which the sequence was missing.

Therefore the absence of these elements is not necessarily meaningful. *Gp-hlh-6* has the PGM1 and PBS, but only one HRL element in the entire 1500 bp upstream sequence.

The last five promoters pictured in Figure 3.9, are from the *hlh-6-like* genes found in PPN. These genes do not appear to have as much conservation in their promoters as the *hlh-6* like genes from PPN. There were no PGM1 elements found in any of them, and the HRL elements appear quite dispersed. Thinking that the *hlh-6-like* promoters were under different regulatory control than *hlh-6* genes, the five *hlh-6-like* promoters were analyzed for a conserved element between them. Using the Improbizer program to search for shared sequence motifs in the same way that Smit *et al.* did to find PGM1, a strong candidate sequence for a functional element was found (2008). The consensus 17 base pair sequence of this new region is TTCMSCCGSGMKMKCCA (M = A or C, S = C or G, K = G or T), and the exact sequences and their locations are shown in Figure 3.9B. The transcription factor that could bind this sequence is unknown.

3.3 Antibodies

Injecting transgenes is not possible in *Meloidogyne*, and therefore protein function was explored with the use of antibodies. From the alignments and phylogenetic trees previously discussed *Mh-pha-4(MhA1_Contig1988.frz3.gene1),* and *Mh-hlh-6(MhA1_Contig384.frz3.gene12)* were predicted to be the homologues of PHA-4 and HLH-6 in *M. hapla*. Antibodies to use in immunohistochemistry and chromatin immunoprecipitation (ChIP) were developed against these TF proteins as described in methods section 2.3. The locations of these antibodies on their respective proteins are noted in Figure 3.2. During ChIP assays the transcription factor's DNA binding sites

must be actively binding the DNA, and thus the epitopes were chosen outside of this region. The antibodies were tested for function and specificity on Western blots, and used in an immunhistochemistry assay to test for tissue localization of HLH-6 in J2 *Meloidogyne* and L2 *C. elegans.* The second larval stage was used for both species because J2s are the only free-dwelling soil stage in *Meloidogyne*, and therefore would be the most accessible for collection.

Through several Western blots, the antibodies developed against HLH-6 in *C. elegans,* Ce-HLH-6-1 and Ce-HLH-6-2, were shown to be non-specific. Strain GD251, a mutant for *hlh-6* due to mutation *tm299,* was used as a negative control for the Ce-HLH-6 antibodies. *tm299* is a 595 bp deletion which begins in the first exon and ends one bp from the end of the second intron, resulting in a frameshift mutation, and therefore is a probable null (Smit et al., 2008). To verify the *hlh-6(tm299)* deletion mutation in strain GD251, primers oJDK0114 and oJDK0115 were used to produce an amplicon (196 bp) showing the truncated gene in strain GD251, compared to wild type strain N2 (821 bp). The *hlh-6(tm299)* mutant strain was a gift from the Gaudet lab, who in turn received it from the Mitani lab (2000).

HLH-6 is a transcription factor, and is only expressed in the gland cells; in order to amplify the signal of HLH-6, efforts were made to isolate a nuclear fraction of proteins to compare to a cytoplasmic fraction. HLH-6 has a molecular weight of 30.8 kD and the antibodies were expected to bind to HLH-6 at this level. A band was noted around 31 kD on the positive wild type control lane as seen in Figure 3.10 (grey arrow). However this same band was noted in the negative control *hlh-6(-),* the null mutant for *hlh-6*. Therefore the product bound here was not HLH-6. If the antibody were binding HLH-6 and

something else of similar molecular weight, we might expect to see a heavier band in the wild type lane than the *hlh-6(-)* lane. As can be seen by comparing lanes three and four, and eight and nine, the bands binding a protein at around 32 kD are of similar strength, suggesting that the antibody is not binding to HLH-6, or only very weakly. Instead the antibody is binding a protein of similar weight, but one that is present in both the wild type nematode and the *hlh-6* mutant. An anti-tubulin antibody (50 kD) was used to show that there was no loading difference in protein concentration from lane to lane. There was a much higher concentration of tubulin in the cytoplasmic fraction than the nuclear fraction; this was expected as tubulin is a cytoplasmic protein. There is also no difference between cytoplasmic and nuclear extracts in the 31 kD bands, though we would expect to see HLH-6, a TF, at higher levels in the nuclear extract than cytoplasmic.

Finally, lanes two, six, seven, and ten tested the pre-immune (PI) serum against *C. elegans* proteins and found antibodies in the serum capable of binding non-specific *C. elegans* proteins. The PI serum was a negative control, and should not have shown activity against *C. elegans* proteins unless the rabbit had been previously exposed to *C. elegans.* These results suggest the antibodies CeHLH-6-1 and CeHLH-6-2 were not specific or effective against HLH-6 in *C. elegans* protein extracts. *Meloidogyne* was not tested because J2s could not be collected in high enough concentrations to extract measurable amounts of proteins from. However they were tested in immunohistochemistry.

Adherens junctions were stained in *C. elegans* embryos and L2s using the monoclonal antibody MH27; this was done to show the context of the worm, and as a positive control for the staining procedure (Francis, R., Waterston, 1991). MH27 was also

tested for cross reactivity in *Meloidogyne* J2s. L2s of *C. elegans*, and J2s of *Meloidogyne* were also stained with their respective HLH-6 antibodies: Ce-HLH-6-1, Ce-HLH-6-2, Mh-HLH-6-1, and Mh-HLH-6-2 to test for tissue localization in the animals. The second larval stage of both nematodes was used because this life-stage is the most easily accessible life-stage in RKN. It is the stage when J2s have hatched from their eggs and are free-dwelling in the soil.

The antibody MH27, which binds to the AJM-1 protein in apical junctions, was found to be effective in *C. elegans* but did not show cross-reactivity in RKN (Bossinger et al., 2001). CeHLH-6-1, CeHLH-6-2, MhHLH-6-1, and MhHLH-6-2 were not shown to work in *C. elegans* N2 or *Meloidogyne* strain Que1. As seen in figure 3.11, MH27 did indeed stain the adherens junctions, which appear to be encircling the epithelial cells of the J2 nematode. MH27 was also tested for possible cross reactivity in *Meloidogyne*, however it was shown to be ineffective in strain Que1 as seen in 3.7B.

The HLH-6 antibodies were tested against their respective species however none of the antibodies were effective in binding their targets. Ce-HLH-6-1 and Ce-HLH-6-2 antibodies did not stain gland cells specifically as seen in Figures 3.11C-D. In the instances where structures do light up within the terminal bulb of the pharynx, there is significant non-specific binding occurring in many other areas in the worm, and five glandular nuclei are not seen as would be predicted. See Figure 1.3B for *C. elegans* glandular positions.

The Mh-HLH-6-1 antibody was tested in strain Que1 before it was known that Que1 was not *M. hapla*. Nevertheless the epitopes that the antibodies were made against were quite similar, especially in the C-terminus (Mh-HLH-6-2) as can be seen in Figure

3.2. The homologous epitope of *M. incognita* shares 9 of 14 amino acid residues with the *M. hapla* epitope for the Ab Mh-HLH-6-1, and 13 out of 14 residues with the Mh-HLH-6-2 epitope. It is therefore possible that the Mh-HLH-6-1 Ab would not work in *M. incognita* strain Que1, but the Mh-HLH-6-2 AB should still be able to function against HLH-6 in Que1.

Upon first review of the *Meloidogyne* IHC results there appeared to be a positive result where a glowing structure was identified in the glandular region. These glowing structures are noted with arrows in Figure 3.11F-H. The only other structure to light up non-specifically was the stomatal cavity at the anterior-most tip of the worm (arrowhead). The structure bound by the antibody in the posterior was not gland cells; however, upon close inspection it was found to be the excretory pore of the nematode. It shows variable placement from right next to the metacorpal pump to up to 20μm posterior to it. Ten nematodes were thus examined and all showed similar results.

3.5 cDNA - sequences and alignment

cDNA sequence data was produced for the genes *hlh-6, pha-4,* and *hlh-6-like* from *Meloidogyne* strain Que1. The antibodies had performed poorly, and because there was no EST data to support the predicted sequence of *hlh-6* it was cloned and sequenced along with the other genes of interest *pha-4* and *hlh-6-like.* As noted earlier, strain Que1 had initially been thought to be *M. hapla*. However in the attempt to align the generated sequence data from *hlh-6, pha-4,* and *hlh-6-like* to the *M. hapla* published genome, it was found that the *Meloidogyne* strain Que1 was in fact an isolate of *M. incognita,* and not *M. hapla* as had been initially assumed. This discovery was further supported with the

species specific primers as noted in section 3.2. The transcript sequences that were cloned are shown aligned to their closest homologous genomic sequence from *M. incognita* in Figure 3.12. An alignment of the Que1 sequences with all three *M. hapla*, *M. incognita*, and *M. floridensis*, is shown in Figure 6.1.

Some parts of the genes were difficult to clone, and some parts were not cloned and sequenced at all. It was only after sequencing that it was realized the difficulty was from the primers being developed from the wrong species of *Meloidogyne*. Each gene was broken down into subsections to make the PCR amplification, and the subsequent cloning and sequencing efforts simpler. Where the genes tended to be most highly conserved was the DBD (see Figure 3.2), and these regions were easier to amplify with PCR. The more divergent 5' and 3' regions were more difficult. After trying several different primers for the 5' end, splice leader (SL) primers were used as the forward primer in the reaction as seen in Figure 2.2. SLs are highly conserved genetic elements that act at the 5' end of nematode operons which help guide mRNA strands post-transcription (Guo et al., 2014). There are 4 known RKN SLs, and each was tested for the ability to act as a forward primer in a PCR amplification of the 5' ends of PHA-4, HLH-6, and HLH-6-LIKE from Que1 cDNA. SL1.1 appeared to work for PHA-4 and HLH-6-LIKE, and SL1.2 worked for HLH-6. The SLs initially appeared to work for each gene tested, based on gel results as seen in Figure 3.13.

In addition to the bands of correct size, other strong bands were noted; these were initially taken to be isoforms of the gene and for this reason were cloned and sequenced as well. This can clearly be seen in Figure 3.13 where there are multiple bands in some of the lanes. All three of these bands were extracted for cloning the 5' region of PHA-4.

This also occurred for the 5’ region of the HLH-6-LIKE cDNA, where two bands are noted, both of these were cloned.

However, because the template was not *M. hapla* as originally thought, low melting temperatures and nested PCRs were necessary to develop products as seen in Table 2.1. These non-stringent parameters in some cases led to the cloning of non-specific genes. Therefore what were initially taken to be isoforms were in fact other genes entirely. These non-specific genes were Minc15667 (no *C. elegans* homologue), Minc16854 (*Mi-pig-1),* and Minc16562 (*Mi-ubq-2-1*). Their complete sequences aligned to *M. incognita* can be seen in Figure 6.2. These accidental genes show that the SLs generally worked for the amplification of the 5’UTR. They showed that what we saw in the gel and thought were specific genes, and that isoforms of the genes was incorrect. The accidental genes did help prove the species was *M. incognita*, as the *M. incognita* genome provided the closest matches to these additional sequences, and the accidentally cloned sequences showed that predicted ATGs as shown on WormBase were frequently mis-predicted.

These non-stringent parameters also led to the SLs mis-priming within the targeted genes. Though the SL did bind in the 5’UTR of the HLH-6 cDNA, the SLs were not binding upstream of the ATG in PHA-4 or HLH-6-LIKE. See Figure 3.12 for alignment of sequenced Que1 cDNA to closest RKN matching gene and note location of primer annealing. Using these SL’s the 5’ end of the gene was cloned, and the ATG was found for HLH-6, but the ATGs of PHA-4 and HLH-6-LIKE were not found. The 3’ end of the gene, which also had low levels of cross species conservation, was unable to be sequenced to the stop codon because it could not be reliably amplified.

Paralogues of *M. incognita* genes were found in the process of piecing together the gene fragments from the sequence reads. Because of the complex hybrid origins of the *M. incognita* genome, many of its genes exist in 2 or 3 divergent copies (Abad et al., 2008; Cotton et al., 2014). Based on the published genome of *M. incognita, hlh-6* and *hlh-6-like* were both predicted to exist in two paralogous copies. The *pha-4* gene was not predicted at all in the *M. incognita* genome so it is difficult to know whether more than one paralogue was generated from our sequence data. For *hlh-6* and *hlh-6-like* the sequence reads were matched to one paralogue or the other based on the SNPs of the *M. incognita* predicted genes. Each separate region was amplified, cloned, and sequenced in 3 or 4 copies, and this resulted in both paralogues being represented in the reads for some regions. It was important when concatenating the gene to keep the correct paralogues together. The paralogues are very similar but SNP data was accurate enough to differentiate them. The presence of multiple paralogues of the genes was further proof that *M. incognita* was the species of strain Que1.

From these alignments it was clear that our strain Que-1 matched more closely with *M. incognita* than either of the other two published *Meloidogyne* genomes. The *M. incognita* homologues of HLH-6, Mi-HLH-6-1 and Mi-HLH-6-2, matched the Que-1-HLH-6 sequence at 96.02%, and 98.56%, compared to 95.06%, for both *M. floridensis* homologues (Mf-HLH-6-1, and Mf-HLH-6-2), and 83.04% for the lone *M. hapla* homologue (Mh-HLH-6). The HLH-6-LIKE sequence followed a similar pattern where one of the *M. incognita* homologues matched the highest at 96.90% identity, compared to 81.07% for the *M. hapla* homologue, and 83.95% for the *M. floridensis* homologue. The PHA-4 data was less conclusive as the *M. incognita* homologue did not show up in the

list of hits when Que1-PHA-4 was BLASTed against it. The cloned sequence of Que1-PHA-4 does however align more closely with the *M. floridensis* homologue at 97.09% than the *M. hapla* homologue at 73.37%. It would be predicted that if the Que1 strain is a tropical RKN like *M. incognita* or *M. floridensis*, and not a temperate RKN species like *M. hapla*, the Que1 cDNA would align most closely to *M. floridensis* in the absence of *M. incognita.*

3.6 Semi-quantitative Expression

A semi-quantitative expression assay, conducted to explore the roles of the HLH-6, HLH-6-LIKE, and PHA-4 transcription factors, revealed varying expression levels of the proteins through different life-stages of *Meloidogyne*. Gene transcript levels cannot be deduced from standard PCR methods as amplicon concentrations will plateau after a certain number of cycles and no longer show correlation between final concentration and starting copy number. In quantitative PCR (qPCR), initial transcript levels of RNA are estimated from comparing the PCR cycle number in which the gene begins amplifying exponentially. This cycle at which the PCR becomes exponential is considered proportional to its initial copy number, and can be described in relative fold increase to an internal control house-keeping gene. Semi-quantitative PCR can be used to estimate transcript levels of various genes without the use of a real-time PCR instrument. The concentration of specific amplicon DNA at a specific cycle number of the gene of interest is compared to a housekeeping gene while both are in exponential phase. This method was used to produce transcript levels of Mi-HLH-6, Mi-HLH-6-LIKE, and Mi-PHA-4, relative to the housekeeping gene Mi-PMP-3.

Semi-quantitative RT-PCR analysis was performed to determine their approximate expression levels throughout various life-cycles of *Meloidogyne* including the embryo, J2, and adult life stages. This method of analysis was chosen because of very lowly concentrated and difficult to obtain RNA samples. To obtain 100ng of RNA from J2s required the collection of at least 5000 J2 nematodes, which in turn required the individual collection of egg masses from roots and their subsequent hatching over the course of a week. To obtain similar RNA samples from adults required using a dissecting microscope and needles to excise mature females from the roots where they were completely covered by plant tissue. Methods to wash the egg masses off of roots with bleach and collect in sieves were attempted but this method produced very low yield of eggs in the final collection. Instead, egg masses were individually plucked off of heavily infected roots under a dissecting scope.

In addition to difficult to obtain samples, transcription factors can have low transcription levels, and HLH-6 in *C. elegans* is only produced in five cells in an organism of approximately 1000 cells. If HLH-6 is following the same pattern in *Meloidogyne* it will only be expressed in three cells. DNA-1000 Chips from Agilent, used on an Agilent Bio Analyzer are capable of quantitating DNA from a concentration as low as 0.1ng/µl, and these were thus used to quantitate amplicon DNA levels at specific cycle numbers of PCR.

Mh-PMP-3 was used as a house-keeping transcript as recommended by Hoogewijs *et al* (2008). pmp-3 normalized our data by providing a constitutively expressed sample to standardize to. Standardizing to PMP-3 minimized variation in sample preps, and loading volume.

For quality control of the RNA, the samples were run on an RNA 6000 pico chip on an Agilent Bio Analyzer following RNA 6000 Nano kit protocol (Agilent). The resulting electropherogram curve and gel images were analyzed for contaminants or degradation. RNA samples were shown to be pure and robust if they had strong peaks marking the 18S and 28S rRNA in the sample. See Figure 3.14 for a representative electropherogram and a gel image of all RNA samples tested.

Samples were consistently noted to be in exponential phase in cycles 34 and 36, thus after normalizing these data to PMP-3 data, cycles 34 and 36 were averaged and used as a semi-quantitative data point in the expression analysis. An example of the DNA concentrations used in the calculations is shown in a gel image produced by the Bio-analyzer in Figure 3.15. Three bio-replicates were performed for each gene, at each life-stage point, except for PHA-4 in embryos, which was only completed twice. This was the case despite repeated attempts, due to sample contamination and cDNA sample degradation. This method was adapted from Niu *et al* (2012).

Normalized and averaged DNA concentrations are summarized in Figure 3.16. The general trend of the expressed genes was from high expression levels in the embryo to lower expression levels in pre-J2s, and adult life stages. PHA-4 levels were highest in embryo, and lower but stable over the latter life-stages. However, only two bio-reps were conducted for PHA-4 in embryos, relative to three bio-reps for all other TFs at each life-stage. The first bio-rep for PHA-4 in embryos was noted to have contamination issues of cDNA in the water control. The experiment was repeated, but by that time the cDNA sample had degraded. Due to the lengthy culturing, and collection time for *Meloidogyne* samples, the proper re-analysis of PHA-4 in embryos could not be conducted within my

own research. The PHA-4 level in embryos is about 3.5 fold of PMP-3. In pre-J2s PHA-4 level is reduced to about 2 fold, and this level holds steady in adults as well.

HLH-6-LIKE levels were widely variable in the embryo stage, as can be seen from the large degree of standard deviation. This variation in the embryo stage is likely attributable to the embryos being of mixed stage. Embryos were collected from egg masses, which hold embryos from various levels of development from the two cell stage to J1s. Because of this variation it is impossible to say whether the expression levels remain the same in the pre-J2 stage or if they are reduced relative to embryo levels. However the results did indicate that HLH-6-LIKE was at its lowest levels in the adult tissue that was examined. HLH-6 levels were similar to HLH-6-LIKE in that they were comparable between the embryo stage, and the pre-J2 stage. In the adult stage the HLH-6 level decreased to its lowest level and was almost undetectable. Amongst the three TFs, HLH-6 had the lowest expression level across all life stages. HLH-6-LIKE and PHA-4 were comparable in the embryo and pre-J2 stages, but in the adults, PHA-4 was higher than HLH-6-LIKE.

Figures:

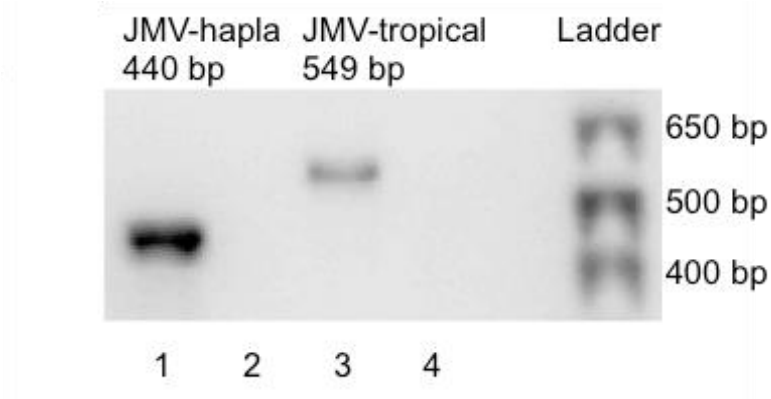


Figure 3.1 Genotyping gel of *Meloidogyne* strains Ont1 and Que1. Lane 1 and 2 use DNA template from Ont1, and lane 3 and 4 use DNA template from Que1.

```
Ce-HLH-6        MSISQNNFLTMFPVTYTFENGVYSTIINQNTIQTPIPNPIQHHIQNHIQTSIPNTNLLLE
Mh-HLH-6        ---------------------------MNNN--------PINDQ-------SVPLTIDILT
Mf-HLH-6-1      ---------------------------MNNT--------PINDQ-------SQPLTIDILT
Mi-HLH-6-1      ---------------------------MNNN--------PINDQ-------SQPLTIDILT
Que1-HLH-6      ---------------------------MNNN--------PINDQ-------SQPLTIDILT
                                           :*:.        **:.:       * * *  :*

Ce-HLH-6        NVQTDVQKLMVPLIDQQFHIPTSTPLQLAPIPTQIQ-------SQLQPQISQIPIHNQP-
Mh-HLH-6        N----EQ-QMLLERQQALCVAAAQAVNFANINMEASAFQRSSSVIVNPPIQEKGIKRKRG
Mf-HLH-6-1      N----QQQQMLLEHQQALCVAAAQTINFVNINLDASAFQRSSSVIVVPPSQEKGTKRKRK
Mi-HLH-6-1      N----EQQQMLLEHQQALCVAAAQTVNFANVNLEASAFQRSSSVIVVPPSQEKGTKRKRK
Que1-HLH-6      N----EQQQMLLERQQALCVAAAQTINFANINIDASAFQRSNSVIVVPPSQEKGTKRKRK
                *     *  *:   :* : : ::  :::. :  : .         : *  .:   :.:

Ce-HLH-6        QIQIQSQVQPQLPTQSQPKPSSKASLDTSSNAFKKYVNPFAPEATVPLPVELEDQYGPYS
Mh-HLH-6        DEIDQQEIQQKMPNMGCGM-----REPKEVVGVNRFISPFSEEAKIPLPHELNE-LSTHG
Mf-HLH-6-1      GDN-EQELQQKMPKMITPV-----KSPKEVLGVNRFISPFSEEAKIPLPHELNE-LSTHG
Mi-HLH-6-1      DDNGQQELQQKMPKMITPV-----KSPKEVLGVNRFISPFSEEAKIPLPHELNE-LSTHG
Que1-HLH-6      DDNGQQELQQKMPKMTTPV-----KSPKEVLGVNRFISPFSEEAKIPLPHELNE-LSTHG
                    :.::* ::*.             ..  ..::::.**: **.:*** **::  . :.

Ce-HLH-6        SSVWKRNERERCRVRNVNDGYERLRKHLPVHFDEKRISKVDTLRLAIRYIKHLDNLLRSE
Mh-HLH-6        TTVWKRNERERIRVREVNNGYTNLKSKLPLTENDRRLSKVDTLRLAIAYIHHLENLINEG
Mf-HLH-6-1      TTVWKRNERERIRVRQVNDGYTRLKNTLPLTENDRRLSKVDTLRLAIAYIHHLENLISEG
Mi-HLH-6-1      TTVWKRNERERIRVRQVNDGYTRLKNTLPLTENDRRLSKVDTLRLAIAYIHHLENLINEG
Que1-HLH-6      TTVWKRNERERIRVRQVNDGYTRLKNTLPLTENDRRLSKVDTLRLAIAYIHHLENLINEG
                ::********* ***:**:** .*:. **:  :::*:********** **:**:**: .

Ce-HLH-6        L-HQYNCKCFNGFQEESEGNILIDISTFNFNSSNNAM
Mh-HLH-6        VNHLIECQCFNYAMGECSSGENTPPPN----------
Mf-HLH-6-1      VNHLIECQCFNYVMGECSSAENTPPPN----------
Mi-HLH-6-1      VNHLIECQCFNYVMGECSSAENTPPPN----------
Que1-HLH-6      VNHLIECQCFNYVMGECSSA-----------------
                : *  :*:***    *...
```

Figure 3.2 Alignment of *C. elegans* HLH-6 with homologues in various RKN. DBD as predicted by Pfam is underlined, DBD as used in this report is shown in bold text. Epitopes to which antibodies were developed against, are bolded and italicized, these are always outside of the DBD. *C. elegans* HLH-6 is aligned using Clustal Omega against *M. hapla* HLH-6 (overall identity 34.47%, DBD identity 64.29%), *M. incognita* HLH-6 (overall identity 34.3%, DBD identity 67.86%), *M. floridensis* HLH-6 (overall identity 33.50%, DBD identity 67.86%), and Que1 HLH-6 (overall identity 36%, DBD identity 67.86%).

```
Ce-HLH-1          MNTETSTQSAPSDTYDTSIYYNSSPRVTANDITTLTSFAAPAPQVLDYANTQYDIYRNQP     60
Hb-HLH-1          -------------------MFQEIQRIAN--ILPIN-LLRTVSESIDFDGFHFP----KE     34
Hb-HLH-1edit      -------------------MFQEIQRIAN--ILPIN-LLRTVSESIDFDGFHFP----KE     34
                                     ::.  *::   *  :. :   . : :*:   ::     :

Ce-HLH-1          AYYLPSYAP-TAPTTFYSDFANFNVTRSQDFASVPAVANSSDVKPIIIKQEKSTPNATEL     119
Hb-HLH-1          TLIIPQISILMNDDSIFTDSNTFNPLRFLDS--DGNLKRYEEFIPFSIGK-------RQY     85
Hb-HLH-1edit      TLIIPQISILMNDDSIFTDSNTFNPLRFLDS--DGNLKRYEEFIPFSIGK-------RQY     85
                  :  :*. :      ::::*  .**  *  *      : . .:. *: * :        :

Ce-HLH-1          IIQSRVDSQHEDTTTSTAGGA-GVGGP-----RRTKLDRRKAATMRERRRLRKVNEAFEV     173
Hb-HLH-1          ILNS----GDSD--DGTMDAADPSNPPRRSKTDPILVDRRKAATMRERRRLR--------     131
Hb-HLH-1edit      ILNS----GDSD--DGTMDAADPSNPPRRSKTDPILVDRRKAATMRERRRLRKVNEAFEI     139
                  *::*     ..*   .*  .*     *         :***************

Ce-HLH-1          VKQRTCPNPNQRLPKVEILRSAIDYINNLERMLQQAGKMTKIMEQNQHLQMTQQINGAPP     233
Hb-HLH-1          --------------KVEILRSAIEYINKLEGMLQAEGKMTKIMAHNQQMTLHSQG----T     173
Hb-HLH-1edit      VKQRTCPNPNQRLEKVEILRSAIEYINKLEGMLQAEGKMTKIMAHNQQMTLHSQG----T     195
                              *********:***:** ***  ******* :**:: : .*

Ce-HLH-1          HDYVTSSHFASSSYNPENMFDDDDLTDSDDDRDHHKLGNAVDLRRRNSLDRLSRIVASIP     293
Hb-HLH-1          SDYLSSSTAQ-FPSSFDGGFDDDDVTDSDEPDS---PLPTVEGRKRTSLDRLSRIVANIA     229
Hb-HLH-1edit      SDYLSSSTAQ-FPSSFDGGFDDDDVTDSDEPDS---PLPTVEGRKRTSLDRLSRIVANIA     251
                   **::**       . :  *****:****:  .      :*: *:*.**********.*

Ce-HLH-1          NEEAMTDEQLLQPANDVIDGE---KKLEML     320
Hb-HLH-1          GEEGGVNG--VEHANDHVEGGDTEKKLVLL     257
Hb-HLH-1edit      GEEGGVNG--VEHANDHVEGGDTEKKLVLL     279
                   **. .:   :: *** ::*    *** :*
```

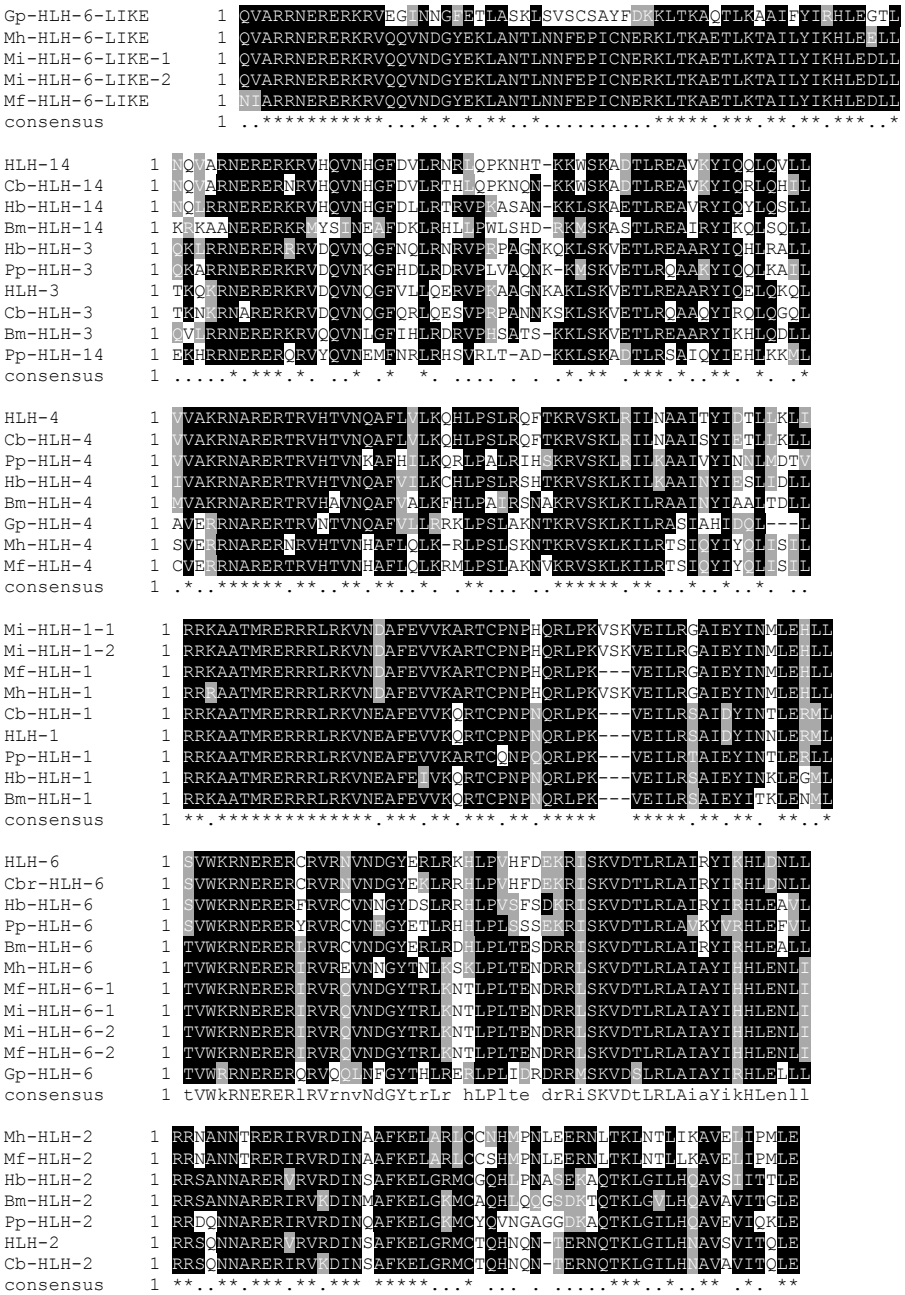
Figure 3.3 Clustal Omega protein alignment of Ce-HLH-1, Hb-HLH-1, and an edited version of Hb-HLH-1. Analysis of the original Hb-HLH-1 protein revealed missing data due to a mis-predicted splice site in the highly conserved DNA binding domain (DBD) region (bold amino acids). The genomic sequence was re-analyzed and the missing sequence was found and re-inserted into the coding sequence (white amino acids), producing the DBD that was used in other analyses in this report.

```
Mh-PHA-4-DBD           NFGPNKPPYSYISLISMAIQQSDRKMCTLSEIYNFIMEYFEYYKNH--RCSWQNSIRHSL   58
Mh-PHA-4-DBD-edit      NFGPNKPPYSYISLISMAIQQSDRKMCTLSEIYNFIMEYFEYYKNHNQRCSWQNSIRHSL   60
Que1-PHA-4-DBD         SYGPNKPPYSYISLISMAIQQSDRKMCTLSEIYNFIMAYFEYYKNHNQRCSWQNSIRHSL   60
                       .:********************************** ********  ************

Mh-PHA-4-DBD           SFNDCFVKVPRTPD-----RPGKGSFWTLHALCGDMFENGCFLRRQKRFKLKD 106
Mh-PHA-4-DBD-edit      SFNDCFVKVPRTPDRIKNCRPGKGSFWTLHALCGDMFENGCFLRRQKRFKLKD 113
Que1-PHA-4-DBD         SFNDCFVKVPRTPDRIKNCRPGKGSFWTLHALCGDMFENGCFLRRQKRFKLSK 113
                       **************     *******************************..
```

Figure 3.4 Clustal Omega alignment of original Mh-PHA-4 DNA binding domain (DBD) peptide sequence (top), aligned to the actual translation of the PHA-4 DBD from *M. incognita* strain Que1 (bottom). The sequenced cDNA was used to repredict the Mh-PHA-4 DBD peptide sequence (middle). Dashes in top sequence fall in exon junctions, and are examples of mis-predicted splice sites common in the published *Meloidogyne* genomes on WormBase-ParaSite.

A) bHLH proteins



B) forkhead proteins

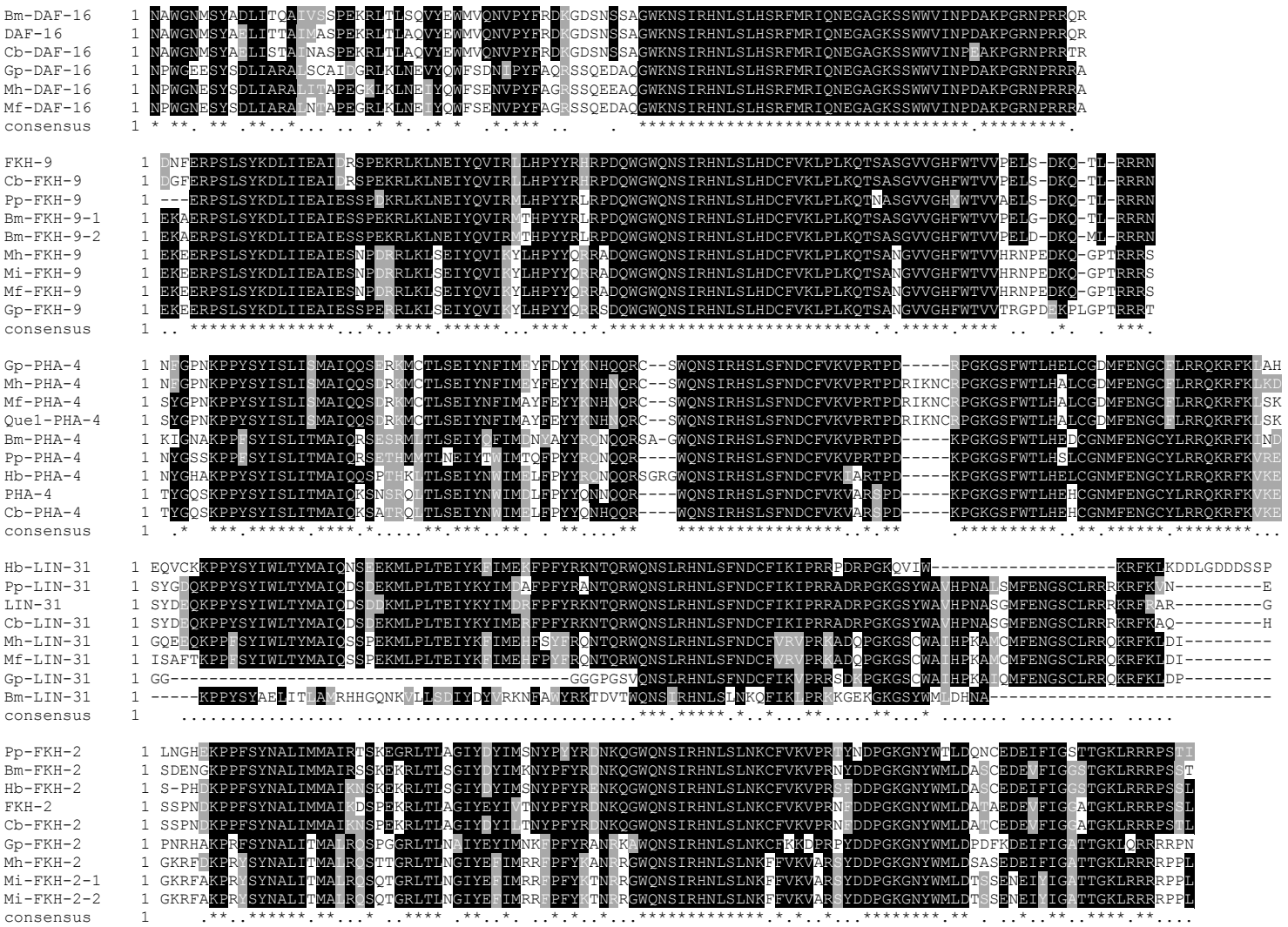
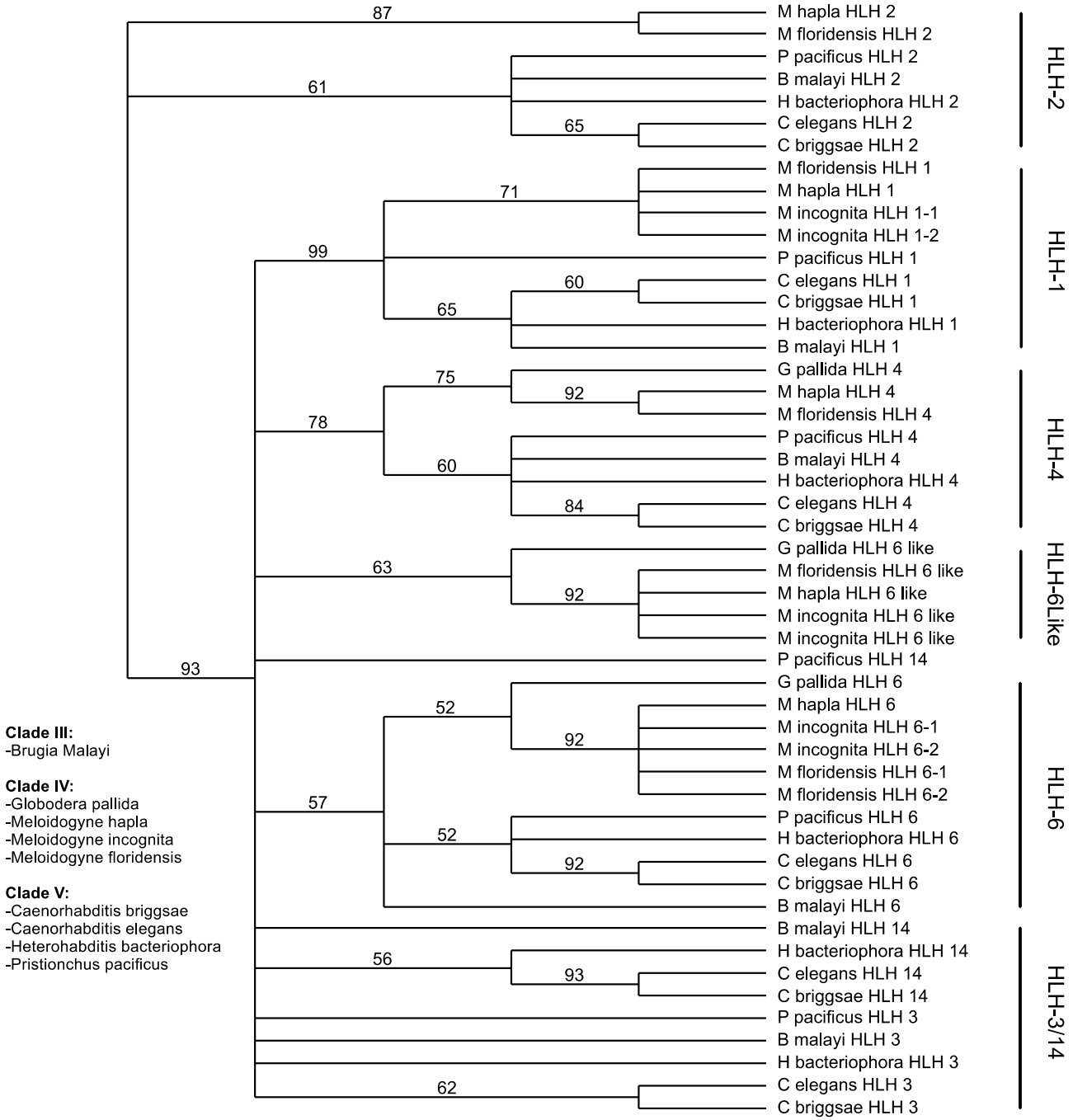


Figure 3.5 Complete T-Coffee alignment of DNA binding domains from A) bHLH transcription factors: HLH-6, HLH-6-LIKE, HLH-3, HLH-14, HLH-4, HLH-1, and HLH-2, and B) forkhead transcription factors: PHA-4, LIN-31, FKH-2, FKH-9, and DAF-16. Amino acids in black have the highest degree of identity, grey residues are less conserved, and white residues are non-conserved.

A) bHLH TFs – Maximum Parsimony



Clade III:
-Brugia Malayi

Clade IV:
-Globodera pallida
-Meloidogyne hapla
-Meloidogyne incognita
-Meloidogyne floridensis

Clade V:
-Caenorhabditis briggsae
-Caenorhabditis elegans
-Heterohabditis bacteriophora
-Pristionchus pacificus

B) Forkhead TFs – Maximum Parsimony

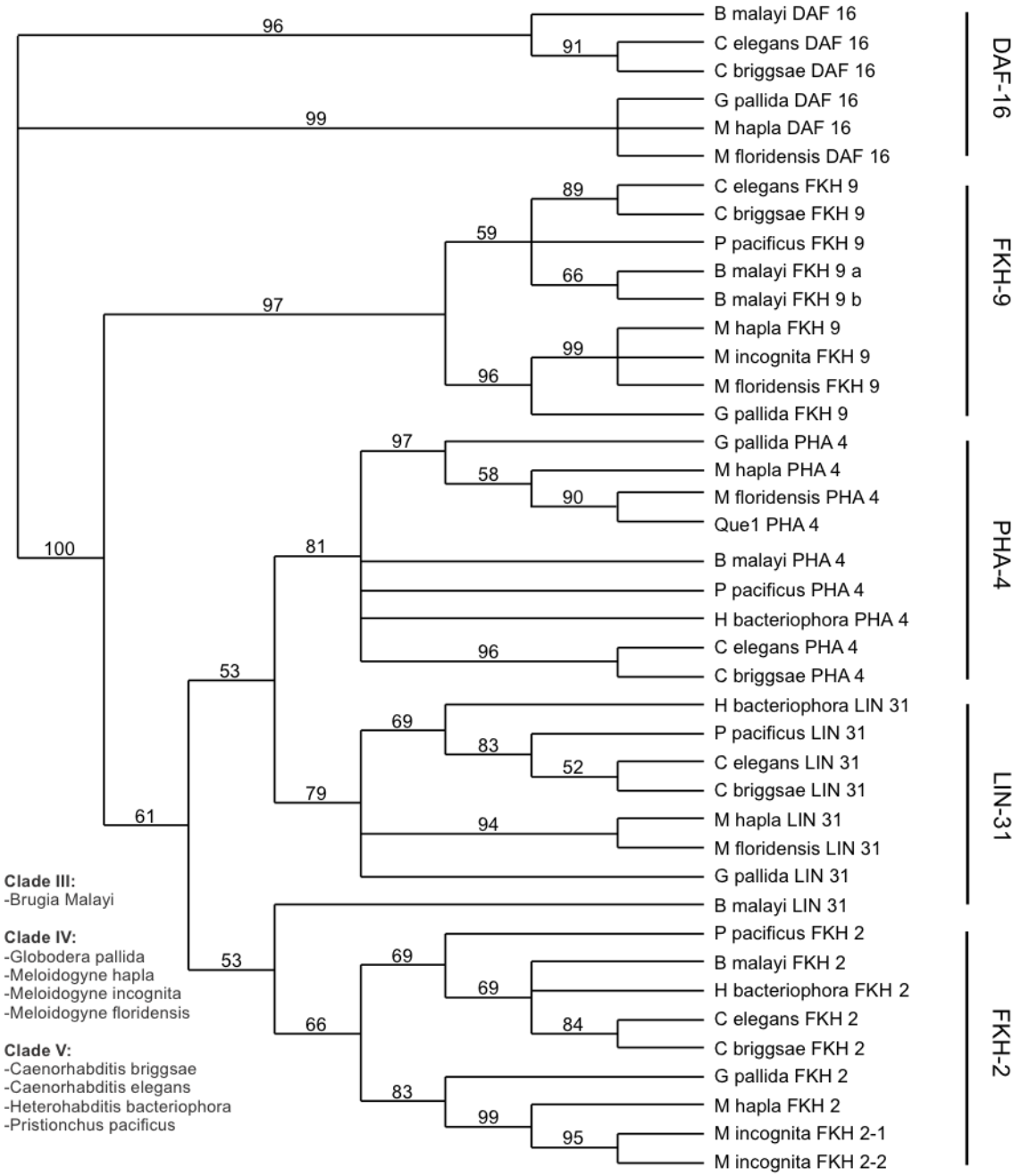
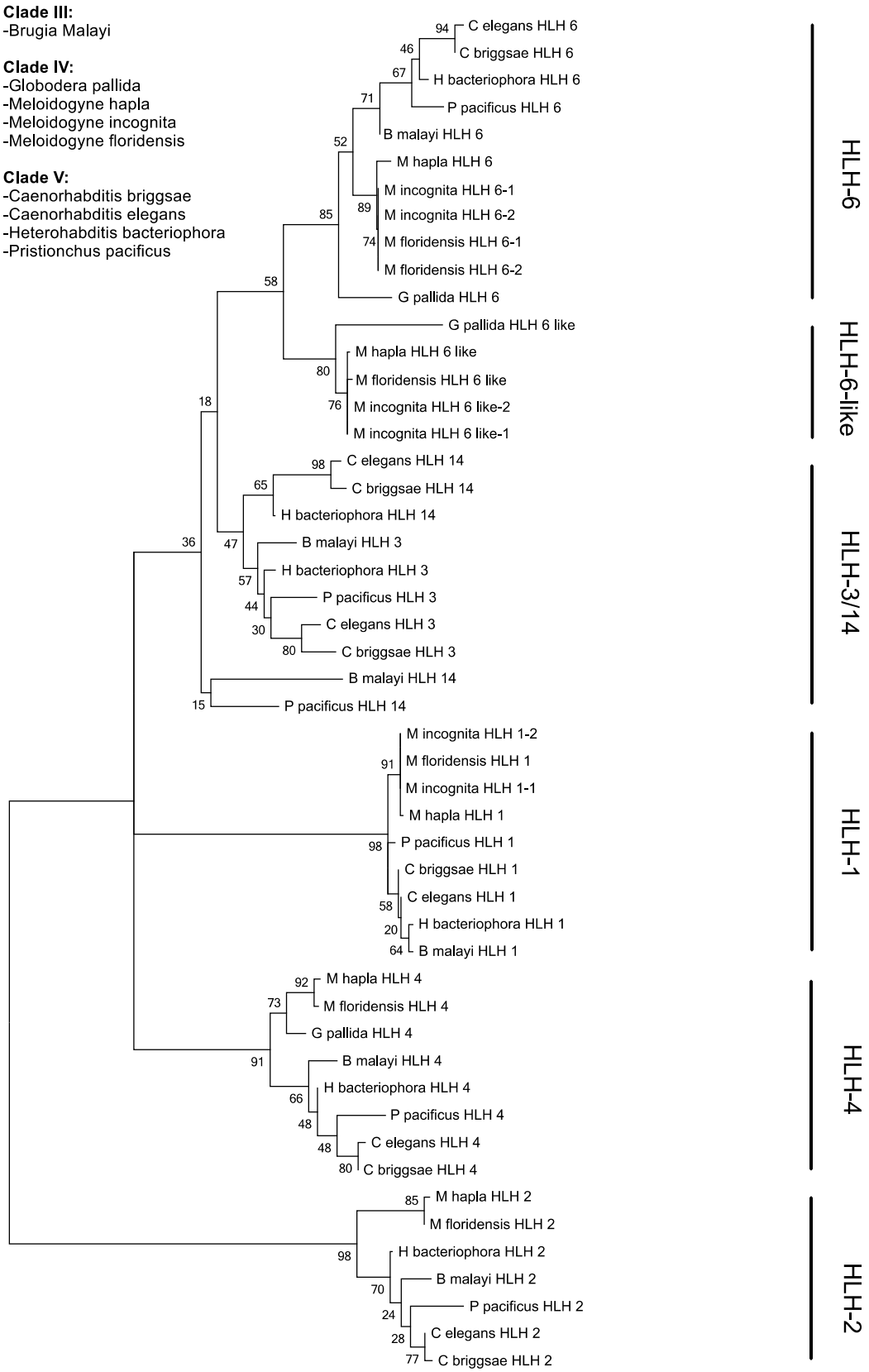


Figure 3.6 Molecular Phylogenetic analysis of DNA binding domains by maximum parsimony method. Nine nematodes across three different clades encompassing diverse lifestyles including plant, animal, and insect parasitism, and free dwelling bacteriovores were included. The species and their clades are listed in the legend. Numbers beside nodes indicate probability based on 1 million bootstraps. This analysis was conducted using PAUP4.0. Clades are named on the right hand side of the chart for the most likely *C. elegans* homologue. A) bHLH transcription factors of class II designation, except for the out-group class I bHLH protein HLH-2. B) Forkhead proteins are all FoxA-like TFs except the out-group FoxO protein DAF-16.

A) bHLH TFs – Maximum Likelihood



B) Forkhead TFs – Maximum Likelihood

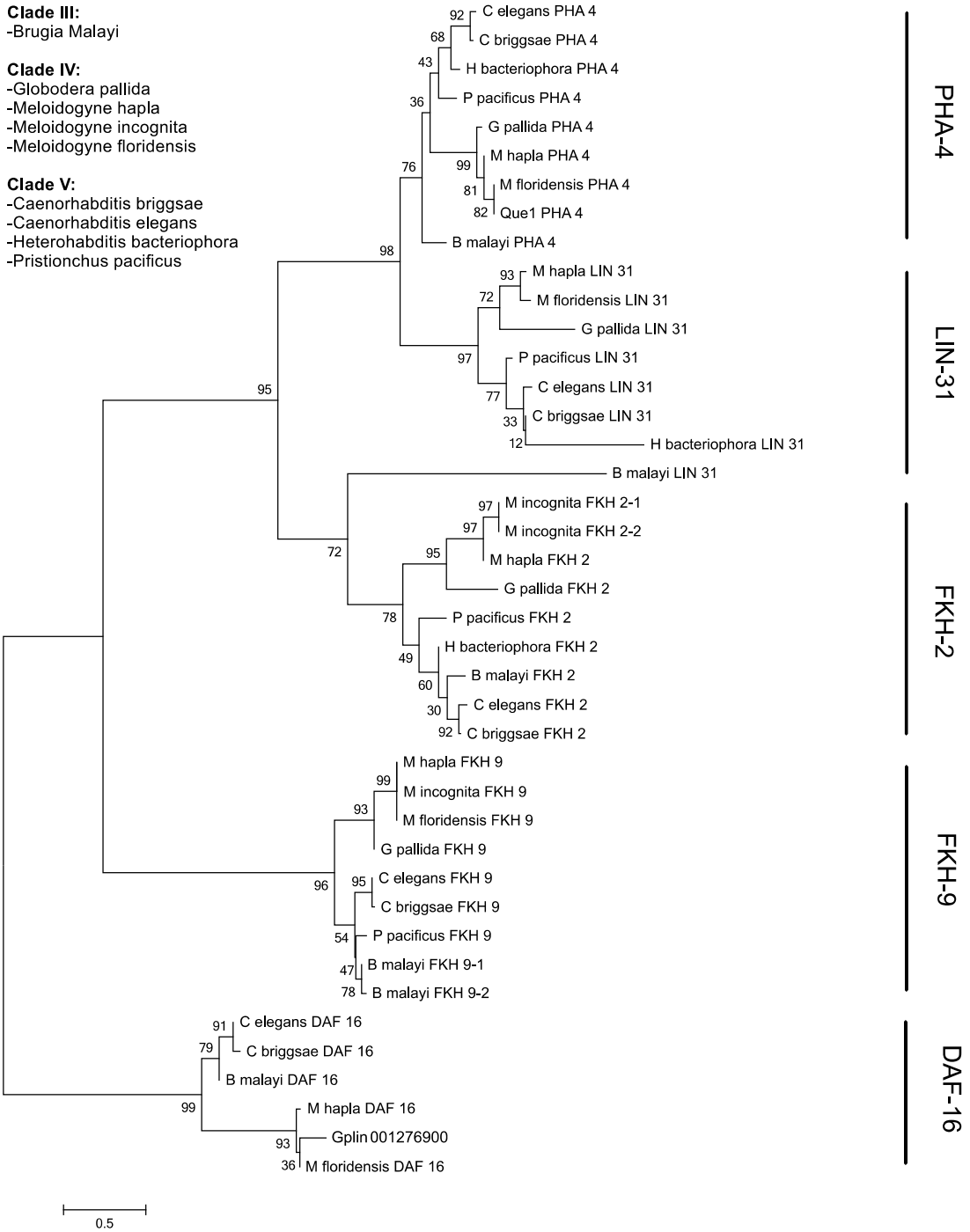


Figure 3.7 Molecular phylogenetic analysis of DNA binding domains by maximum likelihood method. Nine nematodes across three different clades encompassing diverse lifestyles including plant, animal, and insect parasitism, and free dwelling bacteriovores were included. The species and their clades are listed in the legend. Numbers beside nodes indicate probability based on 10000x bootstrap analysis. Branch lengths measure the number of substitutions per site. This analysis was conducted using MEGA6. Clade designations based on the most likely *C. elegans* homologue for each monophyletic group are shown on the right side of the chart. A) bHLH transcription factors of class II designation, except for the out-group class I bHLH protein HLH-2. B) Forkhead proteins are all forkheads, the out-group is FoxO protein DAF-16.

Table 3.1 PGM1 and PBS in proven or probable gland specific genes in A) *C. elegans* and B) PPN. PGM1-stringent fits the original sequence predicted from *phat* genes: CAnvTGhdYMAAY (Smit et al., 2008); PGM1-variable fits the adjusted sequence: CAnnTGnnYMAAY or CAnnTGnYMAAY (Ghai et al., 2012). PBS-variable is the canonical forkhead TF binding site: TRTTKRY; PBS-stringent is based off of a JASPAR Positional Weight Matrix: TRTTKRYHHD; PBS-embryo follows the sequence GAGAGAS (Zhong et al., 2010). Well studied glandular genes from PPN were analyzed for PGM1 and PBS elements within 1500bp of the gene's start codon. RKN homologues of gland genes were included for analysis whenever upstream data was unavailable for the original gene of study. SvG – Subventral Gland; DG – Dorsal Gland. (V=A, C or G, H=A, C or T, D=A, G or T, M=A or C, and Y=C or T, R=A or G, K=G or T, H=A, C, or T, S=C or G)

A)

gland specific gene	PBS-variable	PBS-stringent	PBS-embryo	PGM1	position	sequence	sub-gland specific	evidence
Ce-hlh-6	12	4	0	variable	-427	caggtgttaatta	all glands	Raharjo et al. (2007)
				variable	-347	catctgcttcagt		
B0507.1	8	3	0	stringent	-78	caggtgtgccaac		Smit et al. (2008)
phat-1	8	4	0	stringent	-69	gttgaaacatttg		
				stringent	-199	gtttgaacagatg		
phat-2	15	2	0	stringent	-143	gtttgtgcatctg		
phat-3	4	2	0	stringent	-53	cacctgttcaaat		
phat-4	1	0	1	stringent	-90	caggtgatcaaat		
phat-5	8	1	0	stringent	-213	cacctgttcaaat	g1A	
				stringent	-114	cagatgtgcaaat		
phat-6*	12	4	0	stringent	-1815	gtttgatcacctg		
dod-6	2	1	0	stringent	-111	caggtgaacaaat	all glands	

*phat-6 promoter region was extended beyond 1500bp to find PGM1

B)

Putative Gland Specific Genes	PBS-variable	PBS-stringent	PBS-embryo	PGM1	Position	Sequence	sub-gland specific	Evidence
Mh-hlh-6	7	4	1	stringent	-373	caattggaaaaac	?	n/a
Mf-hlh-6-1	5	1	1	stringent	-858	cagctgcacaaat		
Gp-hlh-6	7	1	2	stringent	-432	cagctgcacaaat		
Mh-hlh-6-like	21	8	0	no	n/a	n/a		
Mf-hlh-6-like	11	5	0	no	n/a	n/a		
Gp-hlh-6-like	5	1	2	no	n/a	n/a		
Mh-crt	12	2	0	stringent	-1253	caaatttgtaaac	SvG	Jaouannet et al. (2013)
				variable	-1021	caagtgatcaat		
Mi-crt	9	3	1	no	n/a	n/a		
Mh-8D05	11	4	0	no	n/a	n/a		Xue et al. (2013)
Mf-8D05	5	2	1	no	n/a	n/a		
Mh-eng-1	6	4	0	variable	-714	cagttgccaaac		Rosso et al. (1999)
Mi-eng-1	8	2	0	stringent	-1080	caaattgtggac		
Minc03866	8	2	3	no	n/a	n/a		Danchin et al. (2013)
Mha1_contig203.frz3.gene9	6	2	1	stringent	-781	atttaatcagatg		
Mh-D15	12	3	0	no	n/a	n/a	DG	Lin et al. (2012)
Mi-D15	10	3	1	variable	-939	caaatgctcaat		
				variable	-933	catttgttaaat		
Mh-eff-1	10	3	1	variable	-1006	caggtgatcaac		Jaouannet et al. (2012)
Mi-eff-1	12	6	1	no	n/a	n/a		

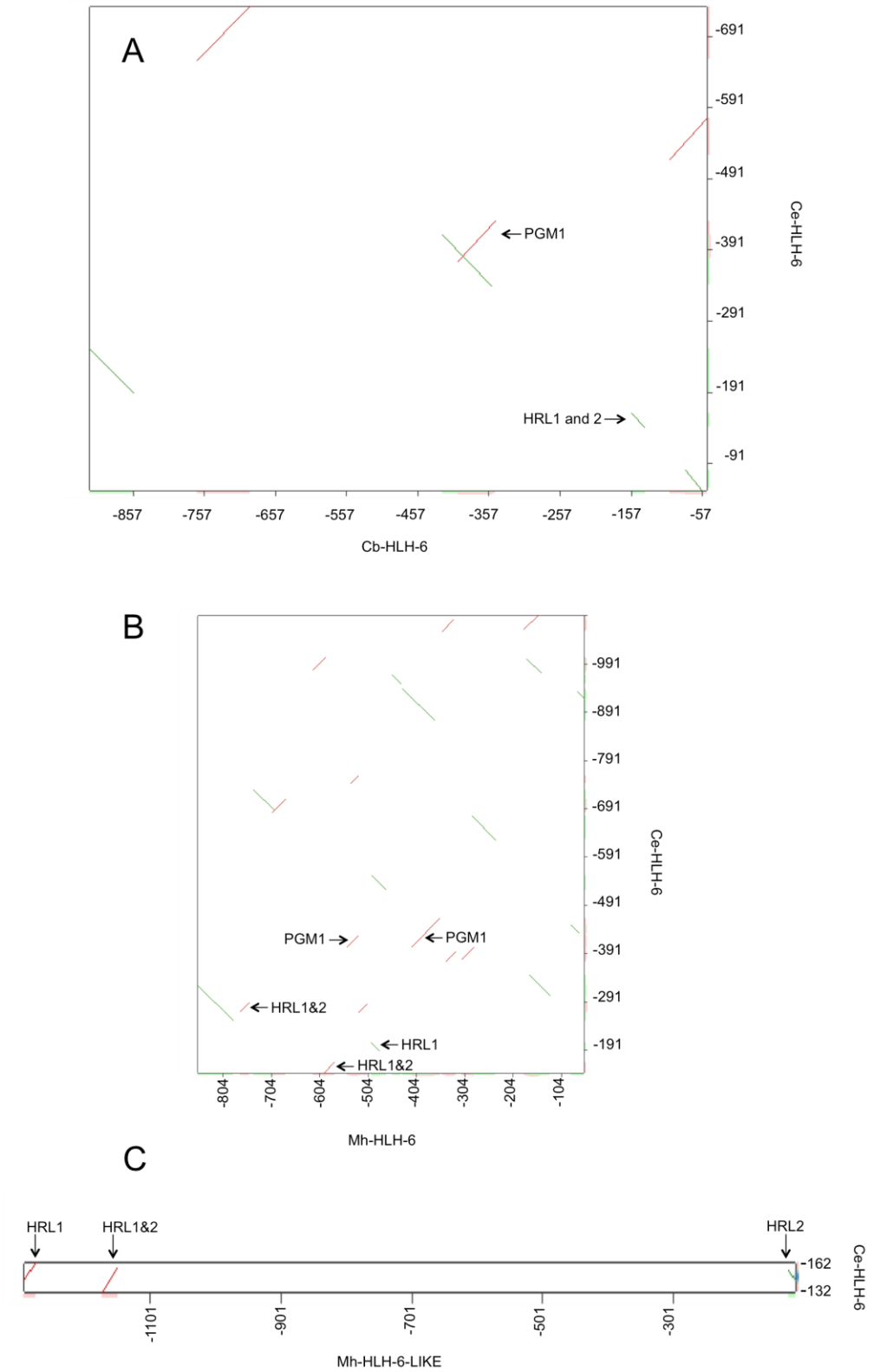


Figure 3.8 Dot plot analysis of *C. elegans hlh-6* promoter region relative to the homologous region in A) *C. briggsae hlh-6*, B) *M. hapla hlh-6*, and C) *M. hapla hlh-6-like*. Only the areas of synteny are shown. Units are displayed in number of base pairs away from the start codon. Elements shown to exist in *C. elegans* that had matching regions are pointed to and labeled with arrows.

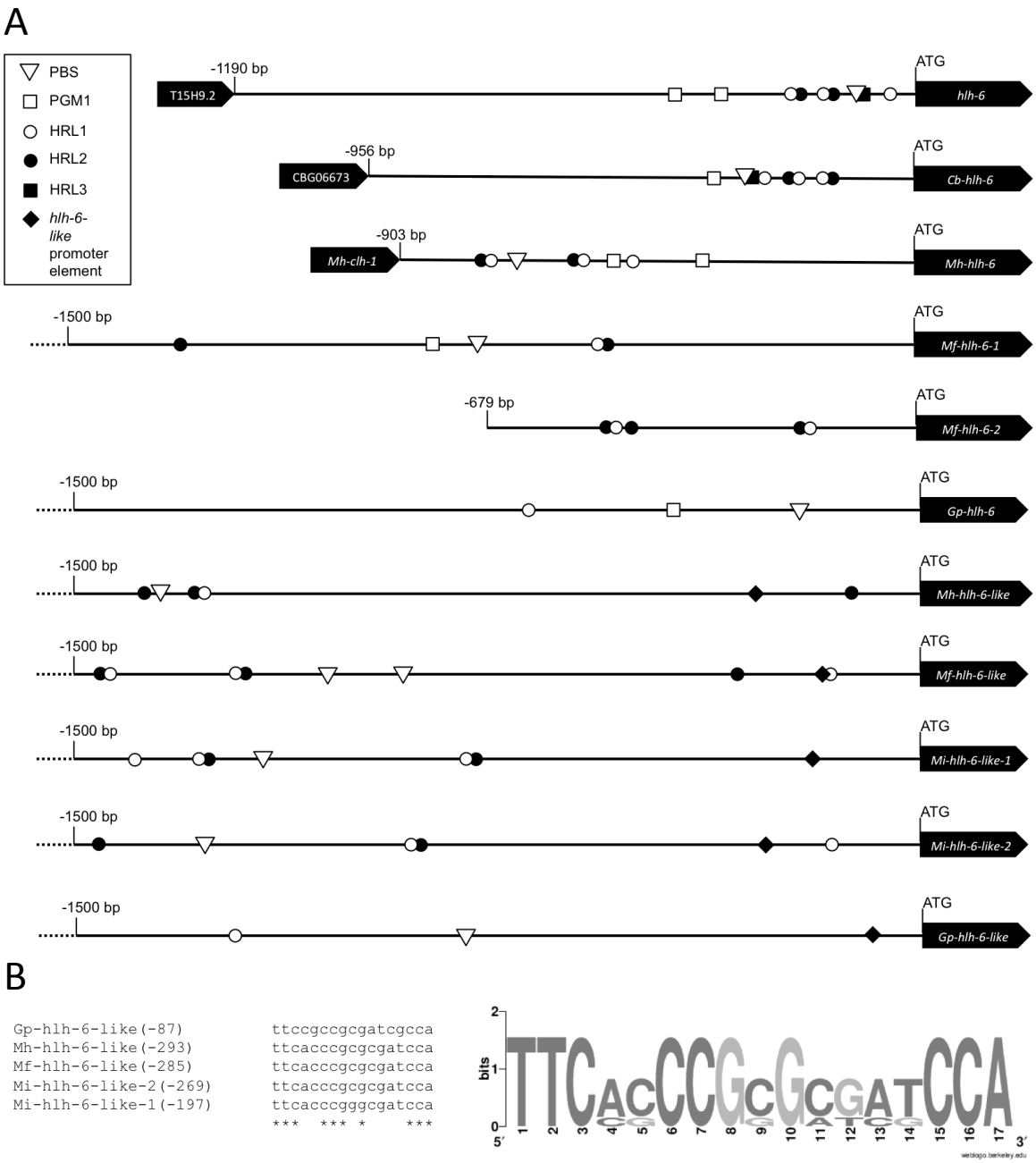


Figure 3.9 A) Promoters of *hlh-6* and *hlh-6-like* genes in *C. elegans*, *C. briggsae*, and the PPN *M. hapla*, *M. floridensis*, *M. incognita*, and *G. pallida.* All of the genes were searched 1500 bp upstream of their start codon, or until the next gene in the 5' direction. PBS (PHA-4 Binding Site) elements are shown in clear arrows, and were searched for using the Positional Weight Matrix from JASPAR (Mathelier et al., 2014); PGM1 (Pharyngeal Gland Motif 1) is shown in clear squares, and was found by searching for sequences matching the consensus CAnvTGnYMAAY, or CAnvTGnnYMAAY; HRL1 (*Hlh-6* Regulatory eLement) and HRL2 elements are shown in circles, and HRL3 is a dark square. HRLs were searched for using a dot plot analysis and hits were included that matched at least 4/7 or 5/8 for their respective *C. elegans* homologues. The sites in *C. elegans* have been previously verified experimentally (Ghai et al., 2012; I. Raharjo & Gaudet, 2007; Smit et al., 2008). The new *hlh-6-like* promoter element, represented here as a diamond, was predicted by the program Improbizer as the most likely regulatory region of these gene homologues. B) The sequence of the *hlh-6-like* promoter element is shown and relative position to the start codon given in brackets. PWM of the *hlh-6-like* predicted regulatory element is shown here as a WebLogo output (http://weblogo.berkeley.edu).

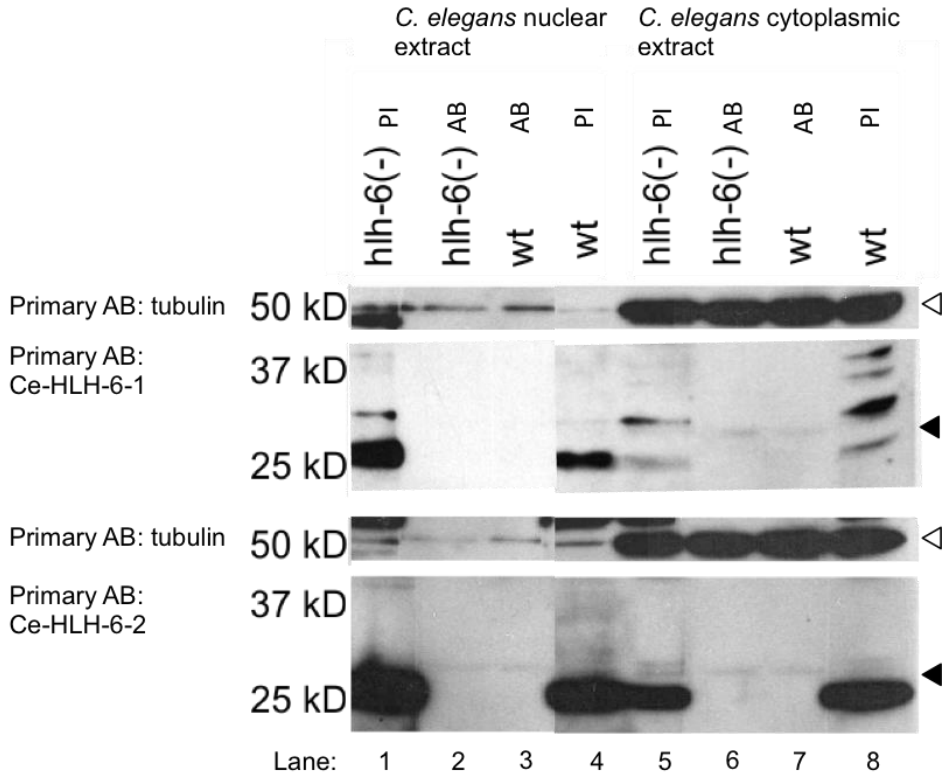
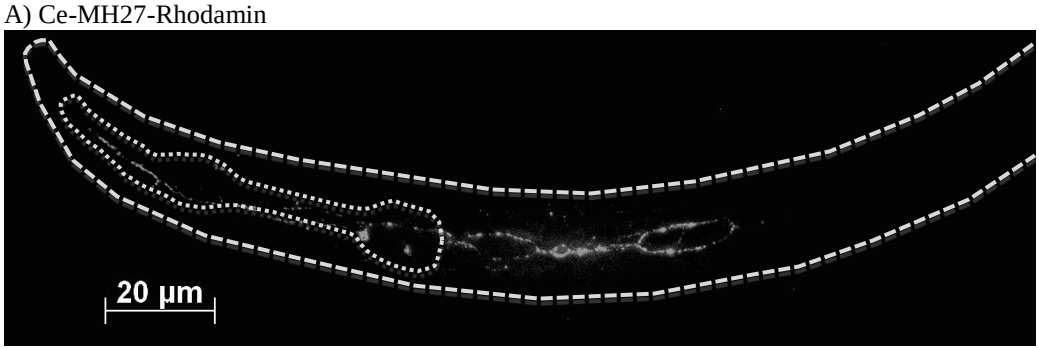
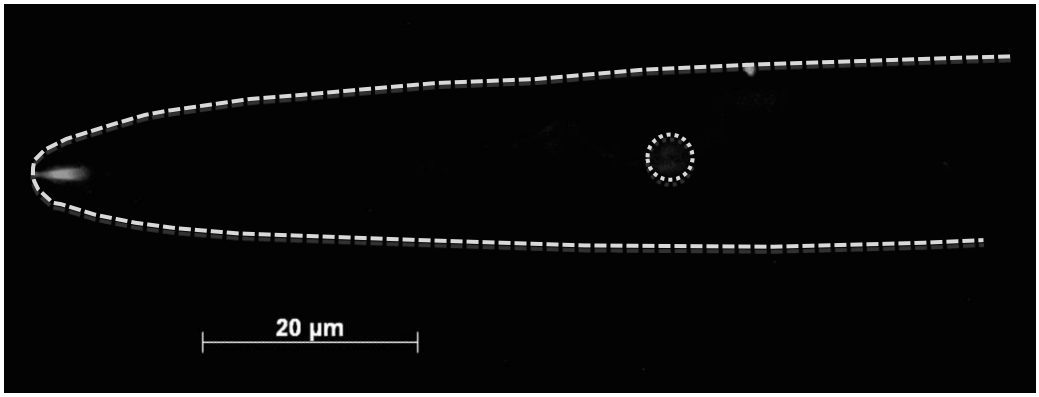


Figure 3.10 Western blot of antibodies developed against Ce-HLH-6. A nuclear and cytoplasmic fraction of *C. elegans* proteins were collected from a *hlh-6* null mutant strain (hlh-6(-)), and an N2 (wt) strain, and tested with Ce-HLH-6-1, Ce-HLH-6-2 primary antibodies (AB). Negative controls were run with Pre-immune (PI) serum in lanes 1, 4, 5, and 8. Blots were restained with anti-tubulin which is shown stained around the 50 kD mark (white arrows). HLH-6 should show up at 30.8 kD; solid arrows point to a band around 30 kD staining negative and positive control lanes.

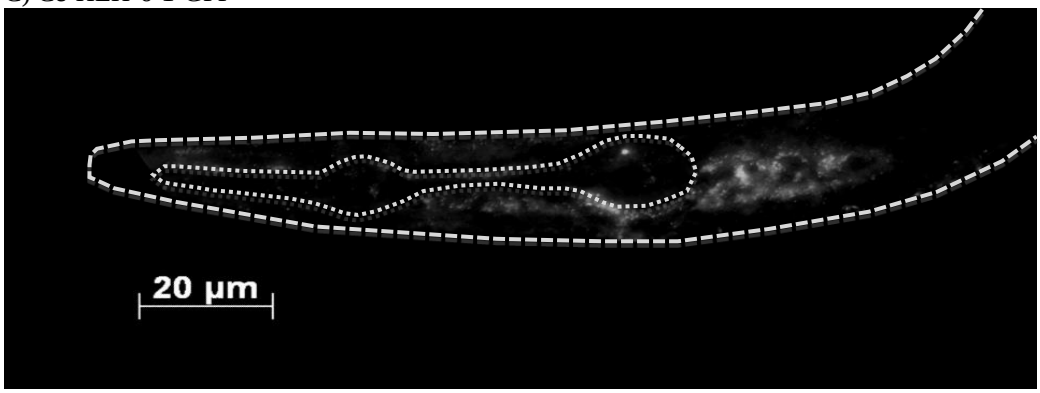
A) Ce-MH27-Rhodamin



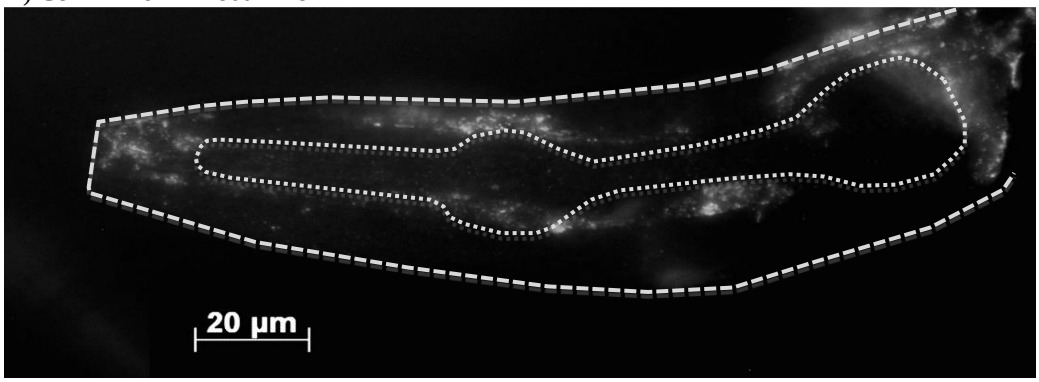
B) Que1-MH27-Rhodamine L2-0010



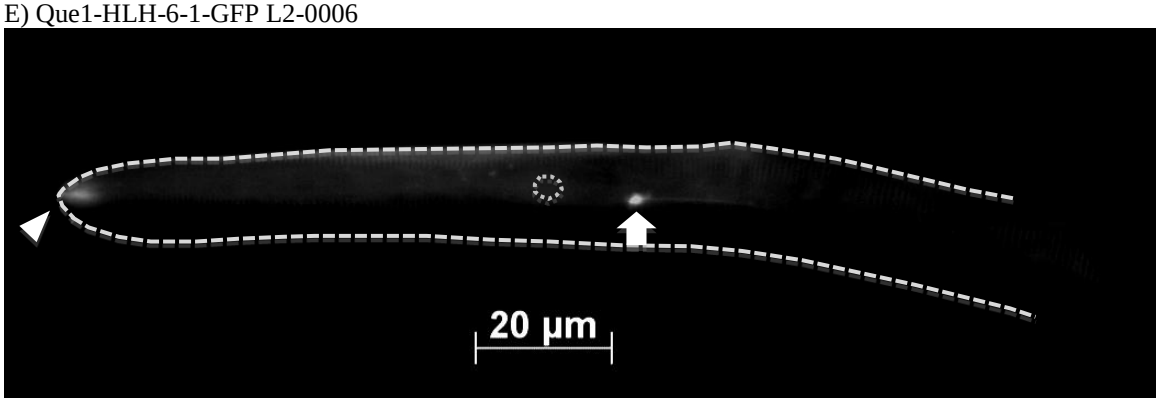
C) Ce-HLH-6-1-GFP



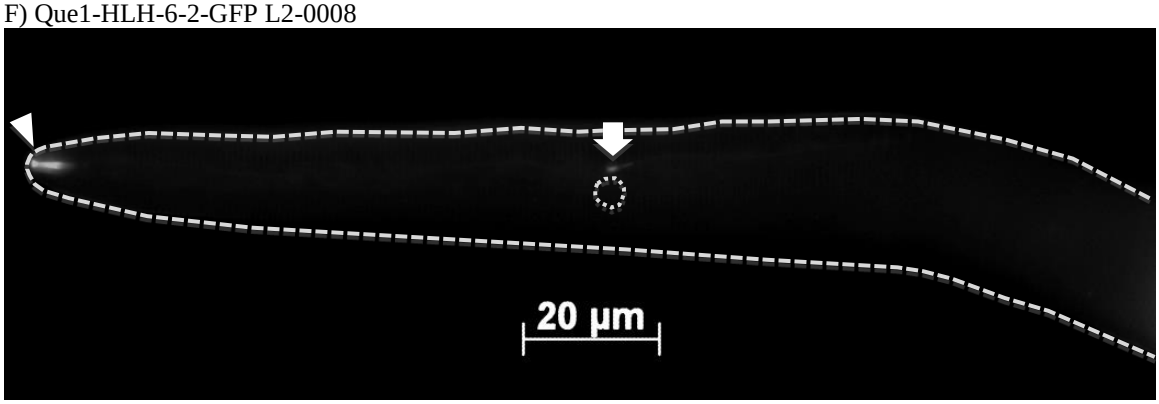
D) Ce-HLH-6-2-Rhodamine



E) Que1-HLH-6-1-GFP L2-0006



F) Que1-HLH-6-2-GFP L2-0008



G) Que1-HLH-6-2-GFP L2-0010

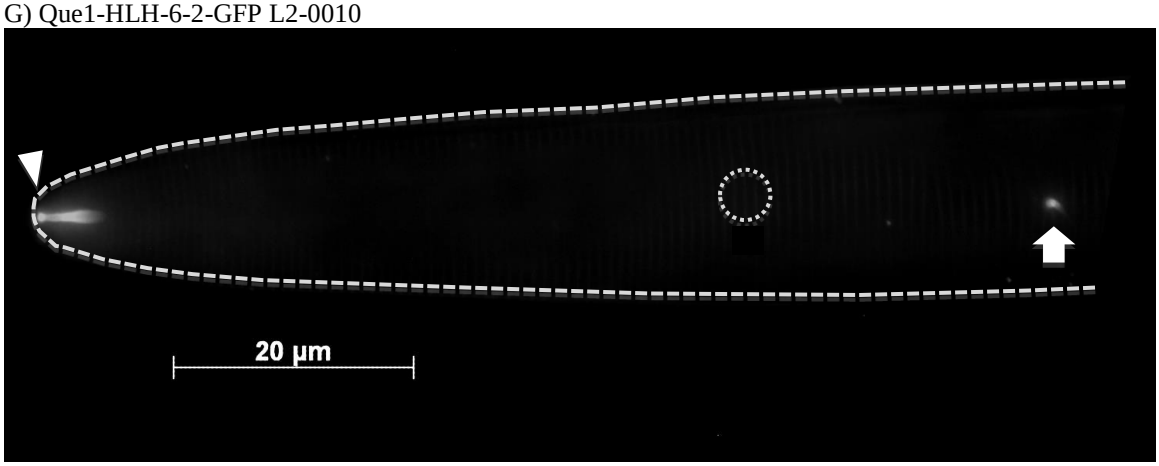


Figure 3.11 A) *C. elegans* L2s stained with antibody MH27 which binds adherence junctions in the epithelium. MH27 was used for creating positional context in the worm, and as a positive control for the IHC protocol. B) MH27 antibody did not show cross reactivity for binding adherens junctions in *M. incognita* strain Que1 J2s. C-D) antibody HLH-6-1 and HLH-6-2 did not binding specifically in the gland cells of the pharynx of *C. elegans* L2s. E) antibody Mh-HLH-6-1 not binding specifically; F-H) Mh-HLH-6-2 binding in stomatal cavity (arrowhead) and excretory pore (arrow). Small round dashes indicate pharynx in *C. elegans* and metacorpus pump in *M. incognita;* Long square dashes indicate the outside cuticle in both species. The gland cells are within the posterior bulb of the *C. elegans* pharynx. In *Meloidogyne* the glands are posterior to the metacorpus pump (circle). Anterior is to the left. All pictures at 40x magnification.

A) *hlh-6*

```
hlh6-Q1          GGTTTAATTACCCAATTTGAGTTTACAATGAATAATAATCCAATAAACGACCAATCACAA
Mi-HLH-6-2       ---------------------------ATGAATAATAATCCAATAAACGACCAATCACAA
                                            ********************************

hlh6-Q1          CCCTTAACTATAGATATTCTCACTAATGAACAACAACAAATGCTTCTAGAACGTCAACAA
Mi-HLH-6-2       CCCTTAACTATAGATATTCTCACTAATGAACAACAACAAATGCTTCTAGAACGTCAACAA
                 ************************************************************

hlh6-Q1          GCATTGTGTGTAGCTGCGGCGCAAACCATTAATTTTGCTAATATAAATATAGATGCTTCT
Mi-HLH-6-2       GCATTGTGTGTAGCTGCGGCGCAAACCATTAATTTTGCTAATATAAATATAGATGCTTCT
                 ************************************************************

hlh6-Q1          GCTTTTCAAAGAAGCAACAGCGTAATAGTAGTTCCACCTAGCCAAGAAAAAGGGACTAAA
Mi-HLH-6-2       GCTTTTCAAAGAAGCAACAGCGTAATAGTAGTTCCACCTAGCCAAGAAAAAGGGACTAAA
                 ************************************************************

hlh6-Q1          AGAAAAAGAAAAGATGATAATGGGCAACAAGAATTACAGCAAAAAATGCCTAAAATGACA
Mi-HLH-6-2       AGAAAAAGAAAAGATGATAATGGGCAACAAGAATTACAGCAAAAAATGCCTAAAATGACA
                 ************************************************************

hlh6-Q1          ACACCTGTTAAAAGCCCAAAAGAGGTTTTAGGTGTAAATCGTTTTATTAGTCCCTTCTCT
Mi-HLH-6-2       ACACCTGTTAAAAGCCCAAAAGAGGTTTTAGGTGTAAATCGTTTTA----TCCCTTCTCT
                 *********************************************    **********

hlh6-Q1          GAAGAGGCTAAAATACCGTTGCCGCATGAATTGAATGAATTGAGCACACACGGAACTACC
Mi-HLH-6-2       GAAGAGGCTAAAATACCGTTGCCGCATGAATTGAATGAATTTAGCACACACGGAACTACC
                 ****************************************** ******************

hlh6-Q1          GTCTGGAAGCGTAATGAACGAGAAAGGATTCGCGTCAGACAAGTTAATGATGGATATACG
Mi-HLH-6-2       GTCTGGAAGCGTAATGAACGAGAAAGGATTCGCGTCAGACAAGTTAATGATGGATATACG
                 ************************************************************

hlh6-Q1          AGGTTGAAGAACACGCTGCCTCTGACTGAAAATGATCGCCGACTTAGCAAAGTTGATACG
Mi-HLH-6-2       AGGTTGAAGAACACGTTGCCTTTGACTGAAAATGATCGCCGACTTAGCAAAGTTGATACT
                 *************** ***** ************************************

hlh6-Q1          CTTCGCCTAGCTATTGCATATATTCACCATTTGGAAAACTTAATTAATGAAGGGGTTAAT
Mi-HLH-6-2       CTTCGCCTAGCTATTGCGTATATTCACCATTTGGAAAATTTAATTAATGAAGGGGTTAAT
                 ***************** ******************* **********************

hlh6-Q1          CATCTAATCGAGTGCCAATGTTTTAATTATGTAATGGGAGAATGTAGCAGTGCTG-----
Mi-HLH-6-2       CACCTAATCGAGTGCCAATGTTTTAATTATGTAATGGGAGAGTGTAGCAGCGCTGAAAAT
                 ** ************************************ ******** ****

hlh6-Q1          ------------------
Mi-HLH-6-2       ACACCGCCGCCAAATTAA
```

B) *hlh-6-like*

```
Que1-hlh-6-like-1      ------------------------------------------------------------
Mi-hlh-6-like-1        ATGCGTGAATGTTCTTCTTCTTCTTCCTCTTCCCCATCTTGGCCTTCTAACCAACAA


Que1-hlh-6-like-1      ------------------------------------------------------GGTTTA
Mi-hlh-6-like-1        AATACTTCCTCATATTCTTCATCATTTTCTTCCTCTCCTTCAATTAATTCTTCCAATTTA
                                                                              ****

Que1-hlh-6-like-1      ATTACCCAAGTTTAAGACAATTATCAATCAACTTCTTTAATTAAAAATTCCAACAATTTT
Mi-hlh-6-like-1        ATTTATTTTCCTTTAGACAATTATCAATCAACTTCTTTAATTAAAAATTCCAACAATTTT
                       ***        ** ********************************************

Que1-hlh-6-like-1      TATTTAAATAATATAAAAACAAATTCAAAACAACAAAAATTAGATAAAAATGGAAAAGAA
Mi-hlh-6-like-1        TATTTAAATAATATAAAAACAAATTCAAAACAACAAAAATTAAATAAAAATGCAAAAGAG
                       ****************************************** ********* ******
```

```
Que1-hlh-6-like-1       AATCAAAAAAATAAGATTACTATATCAAAAAGAAAAAGAAAAAATCCAAATTTAATTATT
Mi-hlh-6-like-1         AATCAAAAAAATAAGATTACTATATCAAAAAGAAAAAGAAAAAATCCAAATTTAATTATT
                        ************************************************************

Que1-hlh-6-like-1       GGAATTAAATTTGGGGGTAAAAAGCCCCATCAAGTAGCTAGAAGAAATGAAAGGGAACGT
Mi-hlh-6-like-1         GGAATTAAATTTGGGGGTAAAAAGCCCCATCAAGTAGCTAGAAGAAATGAAAGGGAACGT
                        ************************************************************

Que1-hlh-6-like-1       AAAAGAGTTCAGCAAGTAAATGATGGATATGAAAAATTGGCAAACACTTTAAATAATTTC
Mi-hlh-6-like-1         AAAAGAGTTCAGCAAGTAAATGATGGATATGAAAAATTGGCAAACACTTTAAATAATTTC
                        ************************************************************

Que1-hlh-6-like-1       GAGCCAATTTGCAATGAAAGAAAATTAACAaaagcagaaacattaaaaacagccaTTTTA
Mi-hlh-6-like-1         GAGCCAATTTGCAATGAAAGAAAATTAACAAAAGCAGAAACATTAAAAACAGCCATTTTA
                        ************************************************************

Que1-hlh-6-like-1       TATATTAAACACCCGGAGGATTTATTAAAACAACAACCTTTGGAAAAACAAAATAAAATT
Mi-hlh-6-like-1         TATATTAAACATCTCGAAGATTTATTAAAACAACAACCTTTGGAAAAACAAAATAAAATT
                        *********** *  ** ****************************************

Que1-hlh-6-like-1       AAAAATGAAAATAATTCTGAAAATTCAACAAATTTTTGTCAGAATTTAAATATTTCCGAA
Mi-hlh-6-like-1         AAAAATGAAAATAATTCTGAAAATTCAACAAATTTTTGTCAGAATTTAAATTTTTCCGAA
                        ************************************************** ********

Que1-hlh-6-like-1       GAAAATCAAAAACAATTTTTAACCAAGAATTCAAATACTCCAAACTTTATTTATTCACCA
Mi-hlh-6-like-1         GAAAATCACAAACAATTTTTAACCAAAAATTCAAATACTCCAAACTTTATTTATTCACCA
                        ******** ***************** **********************************

Que1-hlh-6-like-1       CCACCACCAACAATAATTAATACAAATTATAATAATGTAAATAATAATAATATTTATTTA
Mi-hlh-6-like-1         CCACCACCACCAACAATTAATACAAATTATAATAATATAAATAATTATTACACTTAA---
                        ********* *** ********************** ******** ** * * ***

Que1-hlh-6-like-1       AATAATAATAATAA
Mi-hlh-6-like-1         --------------
```

C) *pha-4*

```
Que1-pha-4      ------------------------------------------------------------
Mf-pha-4        ATGCAACAACAATTTATTTCTTCTCTCGATATTTCCTCTACTACAAATCCTTTAATAGAA

Que1-pha-4      ------------------------------------------------------------
Mf-pha-4        ATGCCTTCAAAAACAACAACAACAGCACCAACAGATTATTTAAATGATCAATGGAACATT

Que1-pha-4      ------------------------------------------------------------
Mf-pha-4        TACAACAATTCTTCTTCTTCCCCAACAACAAATTTAAATATTTCTTTAATTGGCCAAACA

Que1-pha-4      ------------------------------------------------------------
Mf-pha-4        CCCTCTTCTTCTGTTTATCCTCCTTCCCAACCTCCACCACCTCCTCCACCAATTTATCCC

Que1-pha-4      ------------------------------------------------------------
Mf-pha-4        CCTTCACAATATACTTCTCTAAATATTTATTCAAACAATTTATATATCCCCTCTTCTTCC

Que1-pha-4      --------------GGTTTAATTACCCAAGTTTAAGCAACAACAATTTATTTCTTCTCTC
Mf-pha-4        CCTTCCTCCTCCTCTTCTTCAAATCCCNNNNNNNNNNTCCAACAATTTATTTCTTCTCTC
                              ** *   ***           **********************

Que1-pha-4      GATATTTCCTCTACTACAAATCCTTTAATAGAAATGCCTTCAAAAACAACAACAACAGCA
Mf-pha-4        GACATTTCCTCTACTACAAATCCTTTAATAGAAATGCCTTCAAAAACAA---CAACAGCA
                ** *******************************************   ********

Que1-pha-4      CCAACAGATTATTTAAATGATCAATGGAATATTTATAACAACAACAATTCTTCCCCTTCC
Mf-pha-4        CCAACAGATTATTTAAATGATCAATGGAATATTTATAACAACAACAATTCTTCCCCTTCC
                ************************************************************
```

```
Que1-pha-4      CCAACAACAAATTTAAATATTTCCTTAATTGGCCAAACACCCTCTTCTTCTGTTTATCCT
Mf-pha-4        CCAACAACAAATTTAAATATTTCCTTAATTGGCCAAACACCCTCTTCTTCTGTTTATCCT
                ************************************************************

Que1-pha-4      CCTTCCCAACC---------TCCACCTCCAATTTATCCTCCCCCTTCACAATATACTTCT
Mf-pha-4        CCTTCCCAACCTCCACCACCTCCACCTCCAATTTATCCTCCCCCTTCACAATATACTTCT
                ***********         ****************************************

Que1-pha-4      CTAAATATTTATTCAAACAATTTATATATCCCTTCTTCTTCCCCTTCCTCTTCTTCAAAT
Mf-pha-4        CTAAATATTTATTCAAACAATTTATATATCCCTTCTTCTTCCCCTTCCTCTTCTTCAAAT
                ************************************************************

Que1-pha-4      CCACAATTTAGTCAATATTTACAGCAACAATTAATTGATAATAATTCTTTAAATTTATCT
Mf-pha-4        CCACAATTTAGTCAATATTTACAGCAACAATTAATTGATAATAATTCTTTAAATTTATCT
                ************************************************************

Que1-pha-4      TCATATCCTTCAGCTTTTCAATCAACAATTAATTCAGTAATTAATAGTAGAGGGGAAGAA
Mf-pha-4        TCATATCCTTCAGCTTTTCAATCAACAATTAATTCAGTAATTAATAGTAGAGGGGAAGAA
                ************************************************************

Que1-pha-4      GTTGAAATAAATAAACAATTAAAAAATTGTAGAAATAAGACATCATCTACTAATGGAGAA
Mf-pha-4        GTTGAAATAAATAAACAATTAAAAAATTGTAGAAATAAGACATCATCTACTAATGGAGAA
                ************************************************************

Que1-pha-4      TCTACAAATACAACAAATCAAAGTCCACCAATACACAATATGGAGGAATTAACTGTCCAC
Mf-pha-4        TCTACAAATACAACAAATCAAAGTCCACCAATACACAATATGGAGGAATTAACTGTTCAC
                ************************************************************ ***

Que1-pha-4      GAAATGCAAAAAATTAAGAATCAAGGAAGTTATGGTCCAAATAAGCCTCCATATTCTTAT
Mf-pha-4        GAAATGCAAAAAATTAAAAATCAAGGAAGTTATGGTCCAAATAAGCCTCCATATTCTTAT
                ***************** ******************************************

Que1-pha-4      ATTTCTTTAATTTCTATGGCTATACAACAAAGTGATCGAAAAATGTGTACATTGAGTGAA
Mf-pha-4        ATTTCTTTAATTTCTATGGCTATACAACAAAGTGATCGAAAAATGTGTACATTGAGTGAA
                ************************************************************

Que1-pha-4      ATATATAATTTTATTATGGCATATTTTGAATATTACAAAAATCATAATCAACGTTGTTCC
Mf-pha-4        ATATATAATTTTATTATGGCATATTTTGAATATTACAAAAATCATAATCAACGTTGTTCC
                ************************************************************

Que1-pha-4      TGGCAAAATTCAATTAGGCATTCTTTATCTTTTAATGATTGTTTTGTTAAAGTTCCTCGA
Mf-pha-4        TGGCAAAATTCAATTAGGCATTCTTTATCTTTTAATGATTGTTTTGTTAAAGTTCCTCGA
                ************************************************************

Que1-pha-4      ACCCCCGACAGAATAAAAAATTGCAGACCGGGCAAAGGTTCTTTTTGGACTTTACATGCT
Mf-pha-4        ACCCCCGACAGAATAAAAAATTGCAGACCGGGCAAAGGTTCTTTTTGGACTTTACATGCT
                ************************************************************

Que1-pha-4      TTGTGTGGAGATATGTTCGAAAATGGATGTTTTTTGAGAAGACAAAAACGTTTTAAACTT
Mf-pha-4        TTGTGTGGAGATATGTTCGAAAATGGATGTTTTTTGAGAAGACAAAAACGTTTTAAACTT
                ************************************************************

Que1-pha-4      AGCAAAACTGAGAAACCAGAAAGGCAAAGAAGAAATAATAAATTAAATTTGGGGGGAGAA
Mf-pha-4        AGCAAAACTGAGAAACCAGAAAGGCAAAGAAGAAATAATAAATTAAATTTGGGGGGAGAA
                ************************************************************

Que1-pha-4      AAGAAGGCTAACAACAATAATTTAAATAAACAGAAAAAACAACAAAAAAAATTTAATTCT
Mf-pha-4        AGGAAGGCTAATAACAACAATTTAAATAAACAGAAAAAACAACAAAAAAATTTTAATTCT
                * ********* ***** ******************************** *********

Que1-pha-4      TTAATTTTAACTGAACAAAAATTAAAAAAAGA----------------------------
Mf-pha-4        TTAATTTTGACTGAACAAAAATTAAAAAAAGAAATATCAGACTCTTCCACAATTGATTTA
                ******** ***********************

Que1-pha-4      ------------------------------------------------------------
Mf-pha-4        AAACCTTTAATTCAAACAAATATATTAGAACAACACCCCCAACACCCCCACTCCCACCAC


Que1-pha-4      ------------------------------------------------------------
Mf-pha-4        CAANNNNNNNNNNCTGAGAAACCAGAAAGGCAAAGAAGAAATAATAAATTAAATTTGGGG
```

```
Que1-pha-4      ------------------------------------------------------------
Mf-pha-4        GGAGAAAAGAAGGCTAACAACAACAATTTAAATAGTAAACAGAAAAAACAACAAAAAAAT


Que1-pha-4      ----------------------------------------AATATCAGACTCTTCCACA
Mf-pha-4        TTTAATTCTTTAATTTTGACTGAACAAAAGTTAAAAAAGGAAATATCAGATTCTTCTACA
                                                        ********* ***** ***

Que1-pha-4      ATTGATTTAAAACCTTTAATTCAAACAAATATATTAGAACACCCCCACCCCCATCACCAC
Mf-pha-4        ATTGATTTAAAACCTTTAATTCAAACAAATATATTAGAACACCCCCACTCCCAACACCAC
                ********************************************** **** ******

Que1-pha-4      CAACAATTAATTGGAGAAGATGGAAATAATAATAATAATTTAATATTAGATATTATTGGT
Mf-pha-4        CAACAATTAATTGGAGAAGATGGAAATAATAATAATAATTTAATATTAGATATTATTGGT
                ************************************************************

Que1-pha-4      AATGAAGAAGAAGAAAAATATCAAAAAATAAATAATTTTAATTTAATTACTAATAGAAGA
Mf-pha-4        AATGAAGAAGAAGGAAAATATCAAAAAATAAATAATTTTAATTTAATTACTAATAGAAGA
                ************* ********************************************

Que1-pha-4      AGACAAGAAGAAAATACAAATAATTTACAACAACAAATTGAAATATTTAATAATAATTTT
Mf-pha-4        AGACAAGAAGAAAATACAAATAATTTACAGCAACAAATTGAAATATTTAATAATAATTTT
                ***************************** ******************************

Que1-pha-4      GAATTAGCTAATTTAACACAAATTCA---ATGTCCCTCAAGTTCCTCTGAAATTTTACAA
Mf-pha-4        GAATTAGCTAATTTAACACAAATTCAACAATGCCCCTCAAGTTCTTCTGAAATTTTACAA
                ************************   *** *********** ***************

Que1-pha-4      GGAAATAATTTGTGTGGATCTCCTTCTGTTATTTCTTCTATTGGAGGAGGAGGGGGAGGA
Mf-pha-4        GGAAATAATTTGTGTGGATCTCCTTCTGTTATTTCTTCTATTGGAGGAGGAGAGCAACAA
                ************************************************* *  *  *

Que1-pha-4      CAACAATTATTATCTTCCTCTTCCTCAATTTATTGTCATCAATTCCCAAATTCCTCAGCA
Mf-pha-4        TTATTATCTTCCTCTTCTTCTTCCTCAATTTATTGTCATCAATTCCCAAATTCCTCAGCA
                  *  **  *  ***** ****************************************

Que1-pha-4      ACAGCTATAATCCATCCAACAACAATAACTACACAACAACAACAACAACAATTTCCCCAA
Mf-pha-4        ACAGCTATAATCCATCCAACAACAATAACTACACAACAACAACAACAA---TTTCCCCAA
                ********************************************   *********

Que1-pha-4      CAATTTTATTCTTCTTCAAATAATTTAAATTGTGCTGCTGCTGCAAGTCCTCAACAGTTA
Mf-pha-4        CAATTTTATTCTTCTTCAGCAAG------------------------TCCTCAACAATTA
                ******************   *                        ********* ***

Que1-pha-4      AT------TAATTCTCAACAACAATTATTTAACCCCCAACCTTTAGTTAATATTCATTTA
Mf-pha-4        ATTAATTCTCAACAACAACAACAATTATTTAACCCCCAACCTTTAGTTAATATTCATTTA
                **      * *    *******************************************

Que1-pha-4      AATATTAATCAACCCCTTGTTGACTATCAAATGTTA-------------------------
Mf-pha-4        AATATTAATCAACACCTTGTTGACTATCAAANNNNNNNNNNNAATGATGTATAGTGGTGGT
                ************* *****************

Que1-pha-4      ------------------------------------------------------------
Mf-pha-4        AATATATCTGGGGGAGGAGGGGGGCAAACAACACAGACAATTTATAATGAAAAGTAA
```

Figure 3.12 cDNA sequences of Que1-hlh-6, Que1-hlh-6-like and Que1-pha-4 aligned to their closest RKN matches. The SL primer, which was supposed to bind upstream of the ATG is underlined. In Que1-hlh-6-like and Que1-pha-4 the SL primer annealed within the cDNA sequence, rather than upstream of the ATG as would have been predicted. The ATGs are shaded.

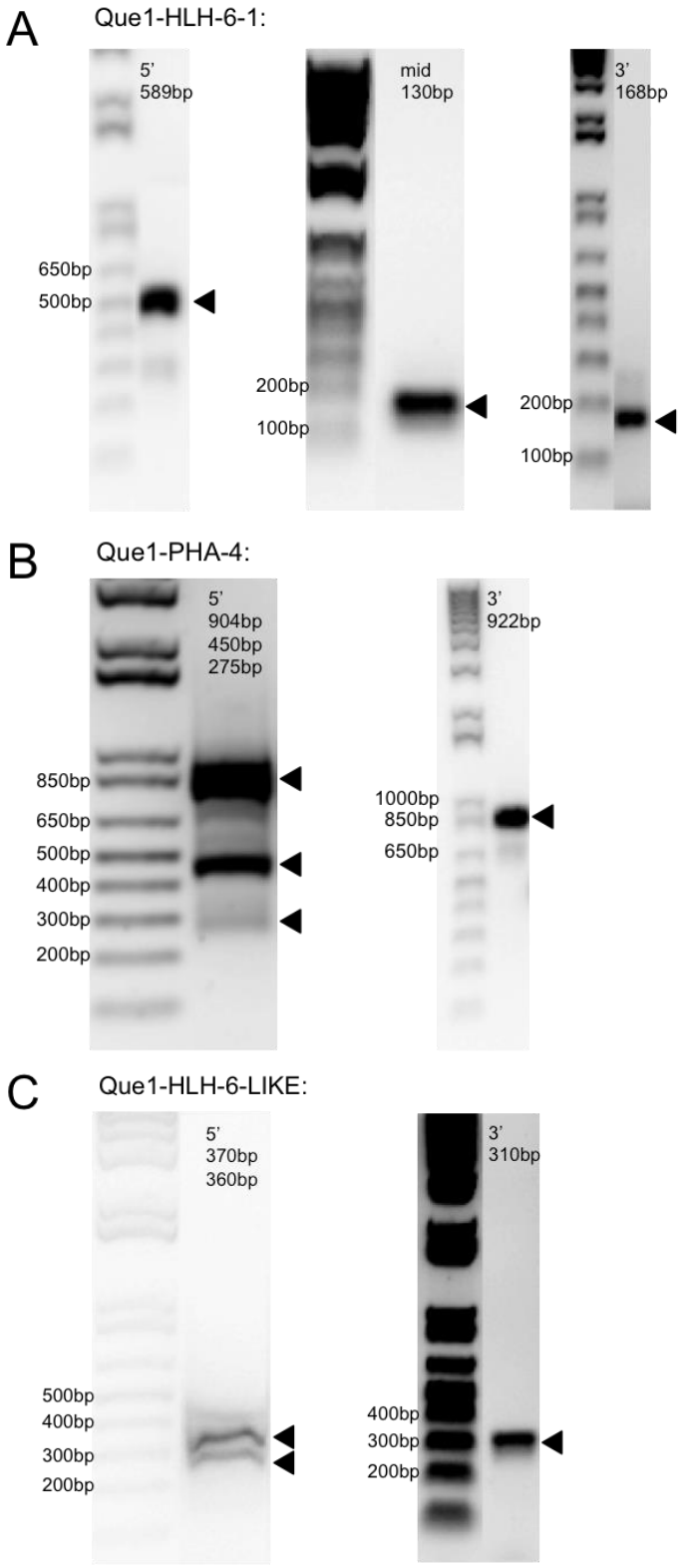


Figure 3.13 Agarose gel electrophoresis image of fragments that were cloned. These bands were excised from gels and used in cloning and sequencing experiments. cDNA of A) HLH-6 B) PHA-4 C) HLH-6-LIKE from *Meloidogyne* strain Que1 are represented in each lane. Multiple bands in the 5' lane were thought to be isoforms of the same gene.

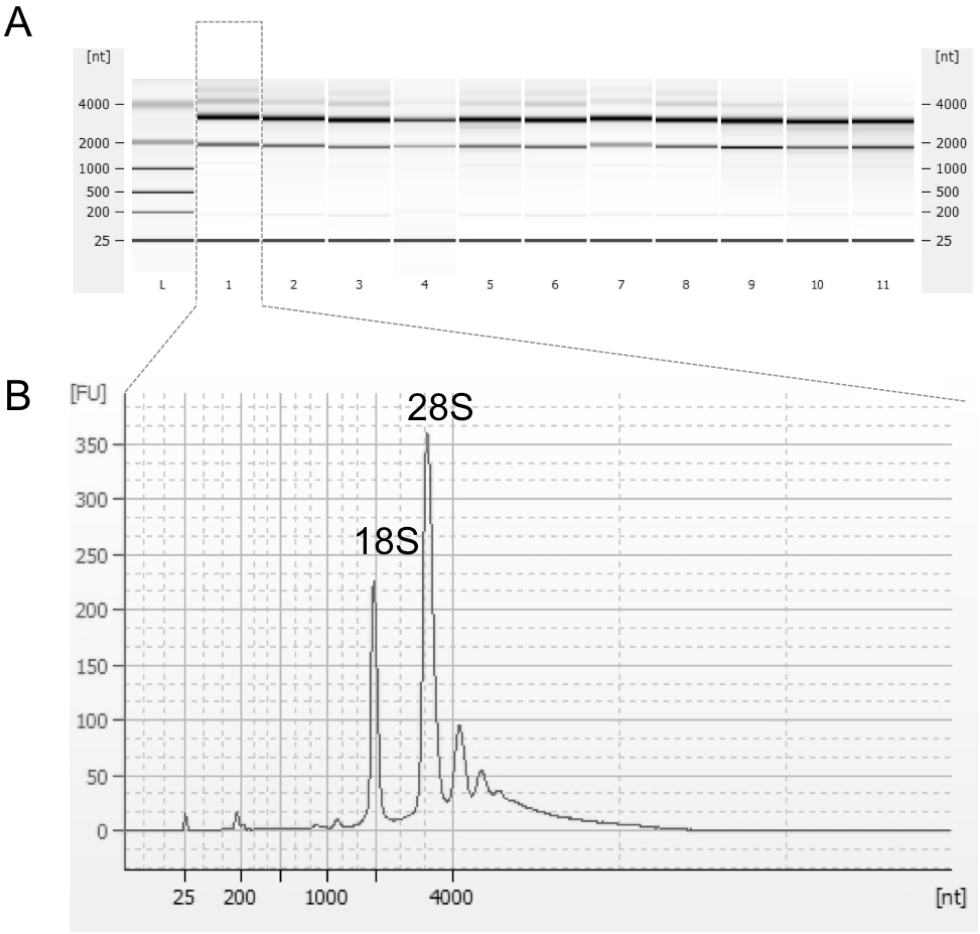


Figure 3.14 RNA 6000 Chip results. A) gel image showing strong bands of the 18S and 28S rRNA amongst the RNA samples tested (lanes 1-11) for later use in cDNA synthesis. B) Lane 1 (embryo RNA) of previous gel image shown as an electropherogram – note the strong 18S and 28S rRNA peaks.

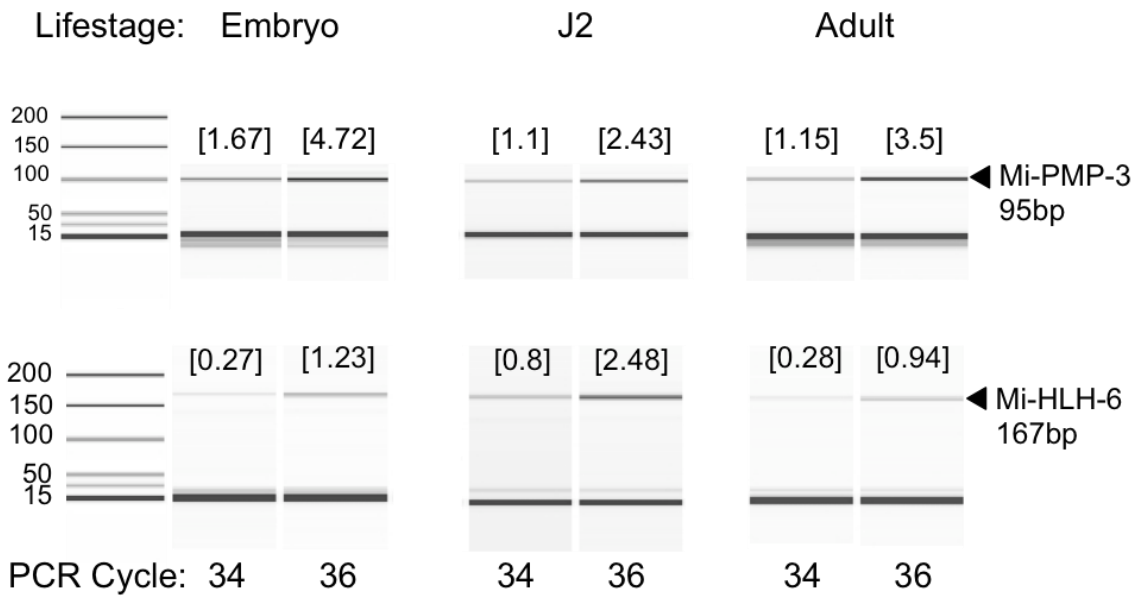


Figure 3.15 Bands representing the transcripts of Mi-PMP-3 and Mi-HLH-6 are seen to increase exponentially from PCR cycle 34 to 36. Concentrations are given above each band in [ng/µL]. Mi-HLH-6 transcript levels were normalized to Mi-PMP-3, the mean from cycles 34 and 36 was calculated and the standard error from three bio replicates was used to estimate transcript levels of the analyzed genes across three life-stages.

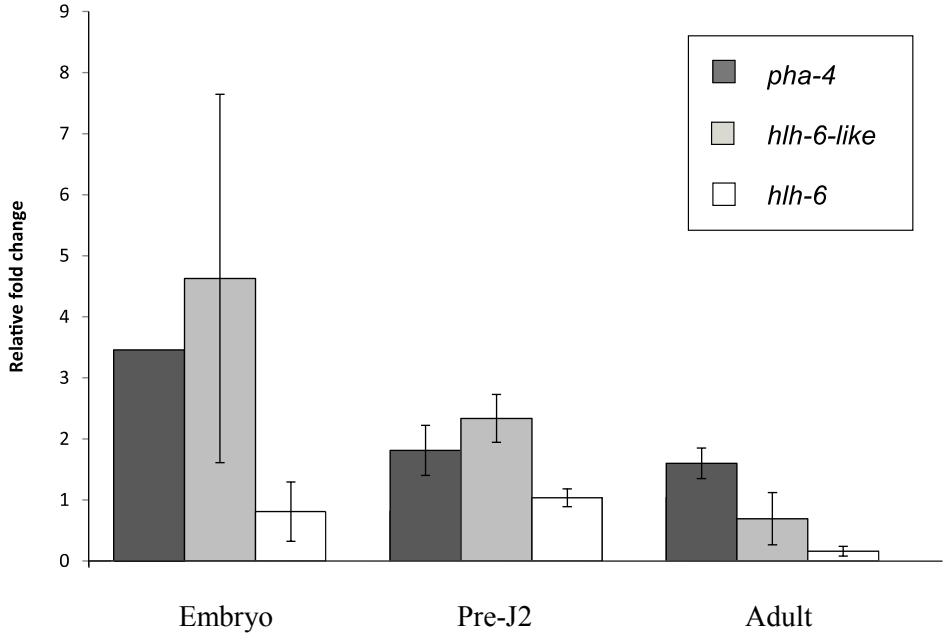


Fig 3.16 Semi-quantitative expression results of three *Meloidogyne* TFs through three lifecycle stages of *Meloidogyne* strain Que1. Expression level was normalized to the house-keeping gene *pmp-3.* Error bars show standard deviation from three bio-replicates; data for *pha-4* in embryos was only collected twice and therefore does not have error bars.

4. Discussion

Based on my findings I have found proof of PHA-4 and HLH-6 in plant parasitic nematodes (PPN). The active regions of the PPN homologues, including specific promoter elements and their DNA binding domains, appear to have been maintained within the genetic sequence since PPN diverged from their free-dwelling ancestors. Closely matching homologues of PHA-4 and HLH-6 were discovered in PPN, and were found to be expressed throughout their lifecycle. One surprising finding was the existence of a PPN-specific HLH-6-LIKE protein that may have evolved to aid in the expanded role of gland cells in their parasitic interactions.

4.1 Homologues of HLH-6 and PHA-4 RKN

I have identified the homologues of the transcription factors HLH-6 and PHA-4 in *M. incognita*, *M. hapla,* and *M. floridensis*. These are the three species of root knot nematode (RKN) that have had their genome sequenced and published. The alignments of the transcription factor DNA binding domains (DBDs) show strong conservation between nematode clades III, IV, and V, as noted in Figure 3.5, and displayed in relative evolutionary terms in phylogenetic trees in Figure 3.6 and 3.7. All of the sequences for the phylogenetic comparisons were found in the genomic databases on Wormbase, except for *Mi-pha-4*. This sequence was deduced from cDNA from our own strain of *M. incognita*, Que1, thus proving that PHA-4 exists and is transcribed in *M. incognita*.

The cDNA data provided from Que1 also highlighted the fact that the published genome of *M. incognita* has some gaps and errors. Some of the challenges in sequencing the *M. incognita* genome are due to the lack of a genetic map, and *M. incognita's*

complex hybrid origins (Lunt et al., 2014). *M. incognita* is mitotically parthenogenetic, and therefore it cannot be crossed. Males occasionally develop, but the nuclei of the sperm degenerate before fertilization can occur (Castagnone-Sereno, 2006). Because of this, no linkage map could be generated for *M. incognita*, as was developed for *M. hapla* before it was sequenced (Opperman et al., 2008). As such the *M. incognita* genome has less reliability. As well, the *M. incognita* genome is hypo-triploid, the product of probably two inter-species hybridization events (Castagnone-Sereno, 2006; Lunt et al., 2014). This unique background has led to diverged alleles for many genes, and the appearance of two or three paralogues for many genes (Abad et al., 2008). These paralogues were noted in the search for *Meloidogyne hlh-6* and *pha-4* homologues. The divergent copies can be seen in phylogenetic trees in Figures 3.4, and 3.5 for the *M. incognita* genes *hlh-1, hlh-6, hlh-6-like,* and *fkh-2.* Notably missing from the *M. incognita* genome were any homologues of *hlh-3, hlh-14, hlh-2, hlh-4, lin-31, daf-16,* and *pha-4.* As mentioned earlier *Mi-pha-4* was found by directly cloning it from our own strain of *M. incognita.* The lack of *hlh-3* and *hlh-14* homologues may be true because these genes are also missing in the other PPN surveyed. However *hlh-2, hlh-4, lin-31, daf-16,* and *pha-4* are present in the genomes of the other PPN surveyed and their absence in *M. incognita* are more likely representative of errors and gaps in the genomic sequence.

The existence of HLH-6 and PHA-4 homologues in RKN has been previously predicted (Smit et al., 2008). Indeed their absence would have been quite surprising, as these transcription factors are evolutionarily conserved across a wide range of metazoa. *hlh-6* homologues involved in gland development and function include *Sgn1* in

mammals, and *Ash2* in jellyfish (Seipel et al., 2004; Yoshida et al., 2001). In *Drosophila* there is a similar class II bHLH protein called *sage,* that functions specifically in salivary gland development and function, though it is not the closest *hlh-6* homologue in flies (Abrams et al., 2006).

Homologues of *pha-4* also show conserved function in other organisms. In *Drosophila, fkh-1* functions with the aforementioned *sage* to block gland cell apoptosis and thus allow for the organ to form (Abrams et al., 2006). In mammals, the *pha-4* homologues, *FoxA1* and *FoxA2* (formerly *HNF3A*(*α* and *ß*)) function cooperatively in the modeling of the foregut, and also in other key endoderm structures such as the liver, pancreas, lungs, and prostate (Friedman & Kaestner, 2006).

To accommodate larger and more complex biological design, transcription factor families have grown and diversified. For instance, whereas *C. elegans* has 42 basic helix-loop-helix (bHLH) proteins (Reece- Hoyes et al., 2005), the human genome encodes more than 100 (Simionato et al., 2007). More complex regulation has been proposed to lead to greater diversity of form and function (Grove et al., 2009).

HLH-6 and PHA-4 exist amongst the metazoa and have similar roles that vary in specific function from organism to organism. As the organ gains in complexity, as we see in fly gland formation, Fkh-1 becomes less of a direct "master regulator" of all gland specific genes as we see in *C. elegans*; rather it functions at the top of a hierarchical system, whereby it indirectly controls gland specific genes, through the aid of intermediaries such as the transcription factor CrebA (Abrams et al., 2006). For the HLH-6 homologue in mammals, Sgn1, we see that it also has evolved a slightly different role in the gland cells. Sgn1 negatively regulates other bHLHs rather than functioning as a

transcription factor initiating transcription of gland specific genes as we see in *C. elegans* (Yoshida et al., 2001). The question pursued in my own research was whether HLH-6 and PHA-4 have similar, reduced, or expanded roles in RKN gland cells relative to the *C. elegans* model.

A key observation from the phylogenetic analysis was that PPN have a group of HLH-6-LIKE transcription factors that consistently form a clade in the maximum likelihood (ML) and maximum parsimony (MP) trees as seen in Figures 3.6 and 3.7. This clade topologically falls between the HLH-6 proteins, and HLH-3 or HLH-14 proteins. Therefore this PPN specific group of proteins is likely neuron specific like HLH-3 and HLH-14 are, or it is a new HLH-6 protein, possibly having evolved to aid in the new complexities of the gland cell function (Krause et al., 1997).

Meloidogyne gland cells appear much more complex than *C. elegans'*. Although *Meloidogyne* only have three gland cells to the *C. elegans* five, these cells have many more secreted proteins, and play diverse roles throughout the life of the organism. In *C. elegans* the glands generally have uniform secretions, with some exceptions such as PHAT-5 only being produced in the g1As (Smit et al., 2008). In *Meloidogyne* the SvGs (sub-ventral glands) function earlier in the life of the nematode, during the J2 stage primarily, and decrease in size and secretions as the nematode begins to feed. At this stage the DG (dorsal gland) grows in size and produces the majority of the secretions used to create and maintain giant cells (Williamson & Gleason, 2003). The complexity is also apparent in the type of proteins being secreted by the glands. While *C. elegans* produces proteins to primarily aid in lubrication of the pharynx and molting (Smit et al., 2008), *Meloidogyne* glands produce hundreds of various proteins contributing to roles in

cell wall degradation, plant immune system suppression, and giant cell formation and maintenance (Bellafiore et al., 2008). Thus it is most likely that new transcription factors have evolved to aid in the functional regulation of glands in RKN. These genes would have developed *via* horizontal gene transfer (HGT), gene duplication, or have been co-opted from redundant functions.

If *hlh-6-like* is a copied and diverged form of *hlh-6,* and not a homologue of the ACSL (Achaete-scute like) genes *hlh-3/14,* then the *hlh-3/14* genes may have been lost in PPN. HLH-3/14 are proneural transcription factors that function with HLH-2 to establish neuronal identity in the *C. elegans* embryo (Thellmann et al., 2003). In flies the ACSL complex is required for external sense organ development (Frank et al., 2003). As seen in the phylogenetic trees in Figure 3.4 and 3.5 *hlh-3/14* genes appear in the free-dwelling and animal parasitic nematodes, but in none of the PPN species tested. These cladistically consistent results suggest that the *hlh-3/14* absence in PPN is not due to a simple genomic sequencing error, unlike the previous oversight of the presence of *pha-4* in *M. incognita*.

It is not entirely surprising for PPN to have lost neuron-specific genes. PPN spend the majority of their lives inside the relative safety of plant root tissue, and the J3, J4, and adult life-stages in the female are entirely sedentary (Taylor & Sasser, 1978). Only the J2 stage is active in the open environment. A notable feature of *Meloidogyne* genomes is their overall reduction in size. For example *C. briggsae* and *C. elegans* have approximate genome sizes of 104 Mbp and 100 Mbp respectively, while *M. incognita's* genome is 86 Mbp, and *M. hapla's* genome is only 56 Mbp (Abad et al., 2008; Guo et al., 2014; Stein et al., 2003). One of the most reduced gene families in *Meloidogyne* is the

GPCRs (G-protein coupled receptor). The GPCRs are neuro-related receptor genes, and have been reduced from 1101 genes in *C. elegans,* to 108 in *M. incognita,* and 147 in *M. hapla* (Abad et al., 2008; Opperman et al., 2008). Thus my own research, and the current body of literature on bHLH genes suggest that PPN may have lost the proneural genes HLH-3/14, and gained HLH-6-LIKE to function in the expanded roles of gland development and function.

4.2 Promoter homology exists in RKN gland specific genes

Some gland specific genes in *Meloidogyne* have PGM1 (Pharyngeal Gland Motif-1) in their upstream promoter regions. In *C. elegans* PGM1 is the binding site of HLH-6, and if HLH-6 is functioning the same way in *Meloidogyne,* it will also bind this site to activate transcription of downstream targets. In *C. elegans,* HLH-6 functions in all of the gland cells, but in more complex organisms such as mammals, the HLH-6 homologue functions in only a subset of foregut gland cells (Smit et al., 2008; Yoshida et al., 2001). Because RKN appear to have more complex gland cells than *C. elegans*, it is possible HLH-6 may be functioning in only a subset of the glands. Therefore *Meloidogyne* genes expressed sub-specifically to the SvGs or DG were looked at for an HLH-6 PGM1 in their upstream regions.

As seen in Table 3.1 six of the twelve *Meloidogyne* sub-gland specific genes in both the SvG, and in DG have copies of PGM1 in their upstream promoter regions. This was not an inclusive search of the hundreds of secretome genes, but does provide evidence that HLH-6 is functioning in at least a sub-set of the gland specific genes in all three of the esophageal gland cells. Since the different glands are active at different times

throughout the life-cycle, this provides some evidence that HLH-6 is also active throughout much of the life-cycle, and especially during times of migration and feeding, when secretions from gland cells are higher than non-feeding or non-migratory times.

All RKN gland specific genes analyzed also have PHA-4 binding sites (PBS) in their upstream regions as seen in Table 3.1. This finding fits with the model seen in *C. elegans* whereby PHA-4 acts as a pioneer TF opening up heterochromatin to other TFs in pharynx specific genes; this process occurs in *C. elegans* gland cells where PHA-4 works in conjunction with HLH-6 to enable most gland specific genes (Cirillo et al., 2002; Smit et al., 2008). Interestingly, in *C. elegans*, PHA-4 has an embryo specific binding site (GAGAGAG/C) through which it can activate embryo specific genes (Zhong et al., 2010). This binding site occurs at the lowest frequency of genes in all species surveyed for PBS elements, and is often entirely absent from the promoter region (Table 3.1). This concurs with the pattern that the genes surveyed are secretome genes, and likely not transcribed until post-embryonic life stages.

Not all gland specific genes are under direct regulation by HLH-6. In *C. elegans* HLH-6 appears to be a dominant transcription factor promoting gland specificity in a gene; however there are some gland specific genes which are not under control of HLH-6 such as *Y8A9A.2* and *nas-12* (Ghai et al., 2012). This is likely also the case in *Meloidogyne*. As seen in Table 3.1 some of the homologous copies of gland specific genes we looked at in the various *Meloidogyne* species do not contain a PGM1 in their promoter region. It should be noted though that the PGM1 could have a slightly different appearance in *Meloidogyne* from *C. elegans*, or could be placed further away from the gene (greater than 1500 bp) and could therefore have been missed in the search. One gene

in particular, *8D05*, had no PGM1 in the *M. hapla* or the *M. floridensis* homologues that were analyzed.

Some of the secretome genes were almost certainly introduced into the genome of RKN via horizontal gene transfer (HGT) (Abad et al., 2008; Opperman et al., 2008). HGT has been commonly noted amongst prokaryotes, but with the increase in genomic sequence data, it has increasingly been observed in eukaryotes. HGT is less common in eukaryotes, as the added barrier of the nuclear membrane impedes foreign DNA. However HGT events do occur in eukaryotes, and occur at elevated rates when two organisms are in consistent close contact like those between symbiotic organisms, and even more so when one organism is feeding upon another, as seen in host-parasite interactions (Soucy et al., 2015). For example, plant-like genes like beta-1,4-endoglucanase and rhizobacteria-like genes such as polygalacturonase appear to function in the secretome of RKN (Bellafiore et al., 2008). It is interesting to see that these genes, which were acquired via HGT, have possibly been placed under the transcriptional control of an anciently conserved transcription factor like HLH-6.

Information from the HLH-6-like promoter region suggests that this gene is not actually functioning like HLH-6. *hlh-6* homologues in both *C. elegans* and RKN have a PGM1 in their upstream regions. This feature is suggestive of autoregulation as is common amongst transcription factors. For example the *C. elegans* transcription factors *ceh-22*, and *hlh-1*, specific to pharyngeal muscle and body-wall muscle cells respectively, both autoregulate (Lei et al., 2010; Vilimas et al., 2004). The PPN *hlh-6-like* genes however do not have PGM1 elements in their upstream regions as seen in figure 3.9. Thus the RKN promoter region of HLH-6-LIKE is not homologous to HLH-6 and

suggests that if it is autoregulating, it is autoregulating through a different enhancer than the PGM1 seen upstream of HLH-6.

A new element was uncovered in the promoter regions of *hlh-6-like*, and is noted as the diamond shape in Figure 3.9. Although 1500 bps were used to find this element, the element appears no further than 300 bp away from the start codon of the downstream gene. This is comparable to the *hlh-6* promoter where most of the promoter elements are near the start codon, falling no more than 500 bp away. The finding of a strong consensus promoter element amongst these genes, one not shared with *hlh-6* genes, and different from PGM1, is a suggestion of alternative function to the HLH-6-specific role. However, this new sequence did not match any elements on the JASPAR or Transfac promoter element databases, and as such will need to be explored further to deduce functionality.

4.3 Antibody analysis

To explore whether HLH-6 was gland specific in RKN, antibodies were developed to test its locations *in vivo*. However, as described in the results, these antibodies were not shown to react specifically to HLH-6. We began the studies thinking we were working with a wild strain of *M. hapla* that would have minor differences to the sequenced strain VW9. Thus it was thought that primers and antibodies would not work as efficiently as on strain VW9. As research progressed it became apparent that the strain we were working with was in fact a different *Meloidogyne* species, *M. incognita*, and that this may have been one reason for the poor antibody function in immunohistochemistry (IHC) experiments. Indeed the epitope for the antibody Mh-HLH-6-1 was not completely conserved from *M. hapla* to *M. incognita*, and only nine out of fourteen residues were

conserved. However, in the antibody Mh-HLH-6-2, thirteen out of fourteen amino acids were identical between the targeted *M. hapla* epitope, and that in *M. incognita*. Thus reasons for this antibody not working, may have been due to a relatively low abundance of HLH-6 in pre-parasitic J2s, or the fact that HLH-6 functions at the top of a battery of genes, and the genes downstream of it are expressed at higher levels than HLH-6 itself. Through IHC, the antibodies were not shown to locate specifically to gland cells.

Two structures that did routinely show antibody binding in *Meloidogyne* were the stomata and the excretory pore as seen in Figure 3.11. This may have been due to non-specific binding, or simply a gathering of the antibody in external orifices. Initially the stain on the excretory pore was thought to be a positive for the gland cell, but upon closer inspection the profile of the stain was consistently a single canal on the outside of the nematode, rather than one to three stained nuclei of the gland cells. Contributing to the confusion was that the gland cells and the excretory pore are relatively near to each other in the overall body plan of *Meloidogyne* (Taylor & Sasser, 1978).

As noted earlier, PGM1s are present in the promoters of gland specific genes, suggesting the functioning of HLH-6 in the transcription of these effector molecules. Thus from this information we may predict that HLH-6 will be functioning during migration and feeding times. Unfortunately the expression levels of HLH-6 were not tested during parasitic-J2, or J3 life-cycle stages. These are stages of migration, and initial feeding and may tell us more about the function of HLH-6. These stages are the most difficult to stage and to collect, but could be looked at in future work.

4.3 Esophageal Gland TFs are expressed throughout RKN lifecycle

The *Meloidogyne* antibodies were primarily tested on IHC in J2s, and all results appeared negative. In order to deduce whether this was because the protein levels were too low to detect, the transcript levels of HLH-6 were tested across the life-cycle of *Meloidogyne*, in addition to several other important transcription factors, including PHA-4 and the new PPN gene HLH-6-LIKE. While the semi-quantitative method as noted in Figures 3.14 - 3.16 is not as accurate as qPCR in mRNA quantification, it does enable a quick view of the relative levels of the genes of interest, and provide evidence that the genes are being transcribed roughly in a fashion similar to the pattern seen in *C. elegans.* Namely that PHA-4 levels are generally higher than HLH-6 levels, and that the genes are immediately important early in development of the organism, but maintain transcription throughout the life of the nematode. PHA-4 levels are higher, as we would expect them to be if they are expressed in all of the cells of the pharynx (~100 cells), compared to HLH-6 which is only expressed in the gland cells (three cells in *Meloidogyne*, and five cells in *C. elegans*) (Mango, 2009).

The expression levels also informed us that HLH-6 was expressed in the J2 life-stage, in fact this may have been its highest level of expression of the stages tested. Therefore the choice of life-stage for IHC experiments was not the reason for that experiment's failure. The antibodies failed due inadequate sequence data on HLH-6, and *M. incognita* rather than *M. hapla* being used as the target species.

5. Conclusions and Future Directions

5.1 Summary

Meloidogyne strains were established in the Kormish laboratory for the purpose of exploring the genetics of their gland cells in comparison with the genetic model organism *C. elegans.* Culturing techniques, and methods to speciate the strains were developed for *Meloidogyne* to pursue this research. Through this work it was uncovered that *Meloidogyne* strains Que1 and Que2 were *Meloidogyne incognita,* and strain Ont1 was *M. hapla*.

The foregut developmental transcription factors *pha-4* and *hlh-6* were uncovered in the published PPN genomes of *M. incognita, M. floridensis, M. hapla,* and *G. pallida*. To create context and to understand the evolution of these transcription factors, the homologues were also identified in free-dwelling species *C. briggsae, P. pacificus, H. bacteriophora,* and the animal parasite *B. malayi.* The uncovered putative homologues were proven to be the most likely homologues using phylogenetic analyses. In addition to verifying *hlh-6* and *pha-4* in PPN, these analyses revealed a new PPN-specific gene *hlh-6-like,* termed thusly for its similarities to *hlh-6*. The homologue search also revealed that, due to lacking cDNA evidence, predicted gene splice sites were not accurate. Cloned and sequenced cDNA of *hlh-6, hlh-6-like,* and *pha-4* from *M. incognita* strain Que1 was used to prove these sequence inaccuracies were real, and revealed the true sequence of these genes.

Antibodies against HLH-6 were developed for use in *C. elegans* and in *M. hapla*, and unfortunately both were shown to be ineffective. Therefore in order to study these genes further, a bioinformatics analysis was used to deduce functionality of HLH-6, and

HLH-6-LIKE in PPN. The bioinformatics work explored the promoter regions of the *Meloidogyne* and *Globodera* genes *hlh-6* and *hlh-6-like,* as well as the promoters of their putative targets, revealing important conserved promoter elements in PPN *hlh-6* including PGM1, HRL1, HRL2, and PBS. Some of these same elements were found in the *hlh-6-like* promoter but were notably missing the PGM1 element. By comparing all of the *hlh-6-like* genes from the various PPN species examined, a specific conserved region was found that may be a *cis*-regulatory region for *hlh-6-like.* This sequence was not found in *hlh-6-like's* strongest *C. elegans* homologues *hlh-6, hlh-3, or hlh-14* suggesting a possible unique function for this new gene. Importantly the HLH-6 binding site was found upstream of some *Meloidogyne* secretome genes, strongly suggesting the maintained role of *hlh-6* in controlling glandular secretions in *Meloidogyne* as in *C. elegans.*

A semi-quantitative expression assay revealed varying levels of PHA-4, HLH-6, and HLH-6-LIKE throughout the *Meloidogyne* life-cycle. These expression levels show the genes are active in each life-cycle stage of *Meloidogyne,* though they appear least important in the final adult stage.

5.2 Proposed Model

Collectively these results indicate a model of gene function for *hlh-6* and *pha-4* similar to that seen in *C. elegans*. In *C. elegans, pha-4* binds PBS upstream of *hlh-6,* initiating its transcription early in embryo development (Raharjo & Gaudet, 2007). PBS elements were also found upstream of *Meloidogyne hlh-6,* and both *hlh-6* and *pha-4* are transcriptionally active in *Meloidogyne* embryos. HLH-6 is active in *C. elegans* larvae

and adults, binding the *cis*-regulatory region PGM1, with a likely heterodimerization partner HLH-2, upstream of gland-secreted mucin-like PHAT proteins (Grove et al., 2009; Smit et al., 2008). The PHATs aid in lubricating the pharynx for passage of food (Smit et al., 2008). Similarly in *Meloidogyne,* HLH-6 was expressed in L2 and adult life stages, suggesting the continued importance throughout the life-span of *Meloidogyne* as a gene necessary for gland function. Additionally the *hlh-2* homologue was found in the genomes of *Meloidogyne* species, and may therefore also be the bHLH class I binding partner of *Meloidogyne* HLH-6 as is seen in *C. elegans*. PGM1 elements were found upstream of about half of the secretome genes analyzed, meaning that *Meloidogyne* HLH-6 could be activating transcription of secreted gland proteins as is seen in *C. elegans*.

A major difference between *C. elegans* and *Meloidogyne* gland cells is the larger number of gland secreted proteins and their increased complexity in *Meloidogyne* (Mitchum et al., 2013). These secretome proteins are primarily what allow *Meloidogyne* to be such an effective plant parasite. While HLH-6 can be knocked down in *C. elegans* and the nematode can still survive if fed a specific diet, knocking down even one of the secretome genes can have disastrous effects on *Meloidogyne* survival (Damiani et al., 2012; Jaouannet, Magliano, et al., 2012; Smit et al., 2008; Xue et al., 2013). A major difference seen between the plant parasitic nematodes and the other species surveyed in the phylogenetic analysis was the presence of the new gene *hlh-6-like. hlh-6-like* does not have an obvious homologue in *C. elegans* and as such, its role in *Meloidogyne* is unknown. Given the increased complexity of PPN gland cells, and their increased

importance, *hlh-6-like* may have developed to aid in the control of the increased battery of gland secreted products seen in PPN.

5.3 Future

One limitation of working with PPN is our inability to create transgenic strains *via* injection of genetic constructs, as has been commonly done with *C. elegans* for some time (Mello et al., 1991). However, *C. elegans* itself can be used to test models of gene pathways in *Meloidogyne,* such as those proposed within this report. To test whether *hlh-6* plays the same role in *Meloidogyne* as *C. elegans*, the gene could be injected into *C. elegans* to test for functional similarities. The upstream promoter region of *M. incognita hlh-6* has already been cloned in our lab, and it can now be placed upstream of a fluorescent protein tag, and expressed in *C. elegans*, and thus tested for tissue location. Given the sequence similarities and conservation of promoter elements already noted, it is quite possible that the transcriptional reporter would show specificity to *C. elegans* gland cells as has been seen with the *C. elegans hlh-6* promoter (Smit et al., 2008). Next, the entire *Mi-hlh-6* gene could be injected into the *C. elegans* strain GD251, a null mutant for *hlh-6*. If a mini-gene construct, consisting of the *Mi-hlh-6* upstream region and cDNA, could rescue strain GD251 from its feeding defects, this would strengthen the notion of *hlh-6* having similar roles in the free living nematodes and PPN. Such experiments using *C. elegans* as a heterologous system to study PPN gene functionality have effectively been used before (Britton & Murray, 2006; Qin et al., 1998).

In addition to these experiments, an RNAi study could be pursued to test the importance of *hlh-6* and *hlh-6-like* in *Meloidogyne,* and whether or not they have similar

functions. While transgenic animals are not possible in PPN, the RNAi pathway, initially discovered in *C. elegans,* is viable in PPN (Fire et al., 1998; Urwin et al., 2002). The RNAi pathway has been used to test functionality of various parasite-specific genes, including those of the secretome (Quentin et al., 2013). *hlh-6* and *hlh-6-like* could be knocked down by soaking *Meloidogyne* in dsRNA of *hlh-6* and separately of *hlh-6-like*. The transcript levels of various secretome genes in addition to *hlh-6* and *hlh-6-like* could then be tested to see whether knocking down *hlh-6* or *hlh-6-like* also reduces the transcript levels of secretome genes. By testing *hlh-6-like* in parallel to *hlh-6,* the importance of *hlh-6-like* within the gland cells could be deduced. In addition to observing transcript levels of secretome genes, these knockdowns could then be tested for reduced viability, by testing the ability of the knockdown animals to infect roots. Knocking down even one secrotome gene has been noted to have dramatic effects on the parasitic abilities of RKN (Quentin et al., 2013). If *hlh-6* or *hlh-6-like* were important in regulating secretome transcription dynamics I would predict the infection efficacy of nematodes lacking their transcription to be highly reduced.

Understanding gland development and function in *Meloidogyne* is becoming increasingly important as researchers work to find new control strategies for this agricultural pest. As indicated within this report, *hlh-6* is a strong candidate for an important regulator of *Meloidogyne* gland development and function. Further research on its functions will provide researchers with new targets for RNAi, and thus enable new strains of plants capable of resisting this troublesome parasite.

References

Abad, P., Gouzy, J., Aury, J.-M., Castagnone-Sereno, P., Danchin, E. G. J., Deleury, E., … Wincker, P. (2008). Genome sequence of the metazoan plant-parasitic nematode Meloidogyne incognita. *Nature Biotechnology*, *26*(8), 909–915. http://doi.org/10.1038/nbt.1482

Abrams, E. W., Mihoulides, W. K., & Andrew, D. J. (2006). Fork head and Sage maintain a uniform and patent salivary gland lumen through regulation of two downstream target genes, PH4alphaSG1 and PH4alphaSG2. *Development (Cambridge, England)*, *133*(18), 3517–3527. http://doi.org/10.1242/dev.02525

Agriculture and Agri-Food Canada. September 18, 2012. Integrated Management of Potato Cyst Nematodes.

Altun, Z.F. and Hall, D.H. 2009. Alimentary System, Pharynx. In *WormAtlas* doi:10.3908/wormatlas.1.3 Edited for the web by Laura A. Herndon. Last revision: July 21, 2012.

Ao, W., Gaudet, J., Kent, W. J., Muttumu, S., & Mango, S. E. (2004). Environmentally induced foregut remodeling by PHA-4/FoxA and DAF-12/NHR. *Science*, *305*(5691), 1743–1746. http://doi.org/10.1126/science.1102216

Artimo, P., Jonnalagedda, M., Arnold, K., Baratin, D., Csardi, G., De Castro, E., … Stockinger, H. (2012). ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Research*, *40*(W1), 1–7. http://doi.org/10.1093/nar/gks400

Bai, X., Adams, B. J., Ciche, T. a., Clifton, S., Gaugler, R., Kim, K. S., … Grewal, P. S. (2013). A Lover and a Fighter: The Genome Sequence of an Entomopathogenic Nematode Heterorhabditis bacteriophora. *PLoS ONE*, *8*(7), 1–13. http://doi.org/10.1371/journal.pone.0069618

Bellafiore, S., Shen, Z., Rosso, M.-N., Abad, P., Shih, P., & Briggs, S. P. (2008). Direct identification of the Meloidogyne incognita secretome reveals proteins with host cell reprogramming potential. *PLoS Pathogens*, *4*(10), e1000192. http://doi.org/10.1371/journal.ppat.1000192

Bird, D. M., & Opperman, C. H. (2009). The secret(ion) life of worms. *Genome Biology*, *10*(1), 205. http://doi.org/10.1186/gb-2009-10-1-205

Blaxter, M. L., De Ley, P., Garey, J. R., Liu, L. X., Scheldeman, P., Vierstraete, A., … Thomas, W. K. (1998). A molecular evolutionary framework for the phylum Nematoda. *Nature*, *392*(6671), 71–75. http://doi.org/10.1038/32160

Bossinger, O., Klebes, A., Segbert, C., Theres, C., & Knust, E. (2001). Zonula Adherens Formation in Caenorhabditis elegans Requires dlg-1 , the Homologue of the Drosophila Gene discs large, *42*, 29–42.

Brenner, S. (1974). The Genetics of Caenorhabditis elegans. *Methods*, *77*(1), 71–94. http://doi.org/10.1111/j.1749-6632.1999.tb07894.x

Brito, J. A., Dickson, D. W., Kaur, R., Vau, S., & Stanley, J. D. (2015). The Peach Root-Knot Nematode : Meloidogyne floridensis , and its Potential Impact for the Peach Industry in Florida 1, (224).

Britton, C., & Murray, L. (2006). Using Caenorhabditis elegans for functional analysis of genes of parasitic nematodes. *International Journal for Parasitology*, *36*(6), 651–

659. http://doi.org/10.1016/j.ijpara.2006.02.010
Buermans, H. P. J., & den Dunnen, J. T. (2014). Next generation sequencing technology: Advances and applications. *Biochimica et Biophysica Acta*, *1842*(10), In Press. http://doi.org/10.1016/j.bbadis.2014.06.015
Burns, A. R., Luciani, G. M., Musso, G., Bagg, R., Yeo, M., Zhang, Y., … Roy, P. J. (2015). Caenorhabditis elegans is a useful model for anthelmintic discovery. *Nature Communications*, *6*(May), 7485. http://doi.org/10.1038/ncomms8485
Castagnone-Sereno, P. (2006). Genetic variability and adaptive evolution in parthenogenetic root-knot nematodes. *Heredity*, *96*(4), 282–289. http://doi.org/10.1038/sj.hdy.6800794
Chitwood DJ. 2003. Research on plant-parasitic nematode biology conducted by the United States Department of Agriculture–Agricultural Research Service. *Pest Management Science 59*: 748–753.
Cirillo, L. A., Lin, F. R., Cuesta, I., Friedman, D., Jarnik, M., & Zaret, K. S. (2002). Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. *Molecular Cell*, *9*(2), 279–289. http://doi.org/10.1016/S1097-2765(02)00459-8
Coordinators, N. R. (2016). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res*, *41*(Database issue), D8–D20. http://doi.org/10.1093/nar/gks1189
Cotton, J. a, Lilley, C. J., Jones, L. M., Kikuchi, T., Reid, A. J., Thorpe, P., … Urwin, P. E. (2014). The genome and life-stage specific transcriptomes of Globodera pallida elucidate key aspects of plant parasitism by a cyst nematode. *Genome Biology*, *15*(3), R43. http://doi.org/10.1186/gb-2014-15-3-r43
Damiani, I., Baldacci-Cresp, F., Hopkins, J., Andrio, E., Balzergue, S., Lecomte, P., … Hérouart, D. (2012). Plant genes involved in harbouring symbiotic rhizobia or pathogenic nematodes. *New Phytologist*, *194*(2), 511–522. http://doi.org/10.1111/j.1469-8137.2011.04046.x
Danchin, E. G. J., Arguel, M.-J., Campan-Fournier, A., Perfus-Barbeoch, L., Magliano, M., Rosso, M.-N., … Abad, P. (2013). Identification of Novel Target Genes for Safer and More Specific Control of Root-Knot Nematodes from a Pan-Genome Mining. *PLoS Pathog*, *9*(10), e1003745. http://doi.org/10.1371/journal.ppat.1003745
Dieterich, C., Clifton, S. W., Schuster, L. N., Chinwalla, A., Delehaunty, K., Dinkelacker, I., … Sommer, R. J. (2008). The Pristionchus pacificus genome provides a unique perspective on nematode lifestyle and parasitism. *Nature Genetics*, *40*(10), 1193–1198. http://doi.org/10.1038/ng.227
Eisenback, J. D. (1989). Detailed Morphology and Anatomy of Second-Stage Juveniles, Males, and Females of the Genus Meloidogyne (Root-Knot Nematodes). In *Journal of Chemical Information and Modeling* (Vol. 53, p. 160). http://doi.org/10.1017/CBO9781107415324.004
Finn, R. D., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., … Bateman, A. (2015). The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Research*, *44*(December 2015), gkv1344. http://doi.org/10.1093/nar/gkv1344
Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E., & Mello, C. C. (1998). Potent and specific genetic interference by double-stranded RNA in Caenorhabditis

elegans. *Nature*, *391*(6669), 806–811. http://doi.org/10.1038/35888 [doi]
Fosu-Nyarko, J., & Jones, M. G. K. (2015). *Application of Biotechnology for Nematode Control in Crop Plants*. *Plant Nematode Interactions A view on compatible interrelationships*. Elsevier Ltd. http://doi.org/10.1016/bs.abr.2014.12.012
Francis, R., Waterston, R. H. (1991). Muscle Cell Attachment in Caenorhabditis elegans. *Cell Biology*, *114*(3), 465–479.
Frank, C. A., Baum, P. D., & Garriga, G. (2003). HLH-14 is a C. elegans achaete-scute protein that promotes neurogenesis through asymmetric cell division. *Development (Cambridge, England)*, *130*, 6507–6518. http://doi.org/10.1242/dev.00894
Friedman, J. R., & Kaestner, K. H. (2006). The Foxa family of transcription factors in development and metabolism. *Cellular and Molecular Life Sciences*, *63*(19-20), 2317–2328. http://doi.org/10.1007/s00018-006-6095-6
Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., & Bairoch, A. (2005). Protein Identification and Analysis Tools on the ExPASy Server. *The Proteomics Protocols Handbook*, 571–607. http://doi.org/10.1385/1-59259-890-0:571
Gaudet, J., & Mango, S. E. (2002). Regulation of organogenesis by the Caenorhabditis elegans FoxA protein PHA-4. *Science (New York, N.Y.)*, *295*(February), 821–825. http://doi.org/10.1126/science.1065175
Gengyo-Ando, K., & Mitani, S. (2000). Characterization of mutations induced by ethyl methanesulfonate, UV, and trimethylpsoralen in the nematode Caenorhabditis elegans. *Biochemical and Biophysical Research Communications*, *269*(1), 64–69. http://doi.org/10.1006/bbrc.2000.2260
Ghai, V., & Gaudet, J. (2008). The CSL transcription factor LAG-1 directly represses hlh-6 expression in C. elegans. *Developmental Biology*, *322*(2), 334–44. http://doi.org/10.1016/j.ydbio.2008.07.018
Ghai, V., Smit, R. B., & Gaudet, J. (2012). Transcriptional regulation of HLH-6-independent and subtype-specific genes expressed in the Caenorhabditis elegans pharyngeal glands. *Mechanisms of Development*, *129*(9-12), 284–97. http://doi.org/10.1016/j.mod.2012.06.005
Ghedin, E., Wang, S., Spiro, D., Caler, E., Zhao, Q., Crabtree, J., … Scott, A. L. (2007). Draft Genome of the Filarial Nematode Parasite Brugia malayi. *Science*, *317*, 1756–1760.
Goggin, F. L., Jia, L., Shah, G., Hebert, S., Williamson, V. M., & Ullman, D. E. (2006). Heterologous expression of the Mi-1.2 gene from tomato confers resistance against nematodes but not aphids in eggplant. *Molecular Plant-Microbe Interactions : MPMI*, *19*(4), 383–388. http://doi.org/10.1094/MPMI-19-0383
Goverse, A., & Smant, G. (2014). The Activation and Suppression of Plant Innate Immunity by Parasitic Nematodes. *Annual Review of Phytopathology*, *52*, 243–265. http://doi.org/10.1146/annurev-phyto-102313-050118
Green, M.R.; Sambrook, J. (2012). *Analysis of Proteins by Immunoblotting.* Molecular Cloning: A Laboratory Manual (Fourth Edition), 1616-1623.
Grove, C. A., De Masi, F., Barrasa, M. I., Newburger, D. E., Alkema, M. J., Bulyk, M. L., & Walhout, A. J. M. (2009). A Multiparameter Network Reveals Extensive Divergence between C. elegans bHLH Transcription Factors. *Cell*, *138*(2), 314–327. http://doi.org/10.1016/j.cell.2009.04.058

Guiliano, D. B., & Blaxter, M. L. (2006). Operon conservation and the evolution of trans-splicing in the phylum Nematoda. *PLoS Genetics*, *2*(11), e198. http://doi.org/10.1371/journal.pgen.0020198

Guo, Y., Bird, D. M., & Nielsen, D. M. (2014). Improved structural annotation of protein-coding genes in the Meloidogyne hapla genome using RNA-Seq. *Worm*, *3*(May), e29158. http://doi.org/10.4161/worm.29158

Haegeman, A., Jones, J. T., & Danchin, E. G. J. (2011). Horizontal gene transfer in nematodes: a catalyst for plant parasitism? *Molecular Plant-Microbe Interactions : MPMI*, *24*(8), 879–887. http://doi.org/10.1094/MPMI-03-11-0055

Haegeman, A., Mantelin, S., Jones, J. T., & Gheysen, G. (2012). Functional roles of effectors of plant-parasitic nematodes. *Gene*, *492*(1), 19–31. http://doi.org/10.1016/j.gene.2011.10.040

Hall, B. G. (2011). *Phylogenetic Trees Made Easy* (4th ed.). Sunderland, Massachusetts: Sinauer.

Hall, B. G. (2013). Building Phylogenetic Trees from Molecular Data with MEGA. *Molecular Biology and Evolution*, *30*(5), 1229–1235. http://doi.org/10.1093/molbev/mst012

Hewezi, T., & Baum, T. (2012). Manipulation of Plant Cells by Cyst and Root-Knot Nematode Effectors. *Molecular Plant-Microbe Interactions*, *26*(1), 120718074931008. http://doi.org/10.1094/MPMI-05-12-0106-FI

Hillis DM, Bull JJ. 1993. An empirical-test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Syst Biol 42* (2): 182–92.

Hoogewijs, D., Houthoofd, K., Matthijssens, F., Vandesompele, J., & Vanfleteren, J. R. (2008). Selection and validation of a set of reliable reference genes for quantitative sod gene expression analysis in C. elegans. *BMC Molecular Biology*, *9*(1), 9. http://doi.org/10.1186/1471-2199-9-9

Hope, I. A., Mounsey, A., Bauer, P., & Aslam, S. (2003). The forkhead gene family of Caenorhabditis elegans. *Gene*, *304*, 43–55. http://doi.org/10.1016/S0378-1119(02)01175-7

Horner, M. A., Quintin, S., Domeier, M. E., Kimble, J., Labouesse, M., & Mango, S. E. (1998). pha-4, an HNF-3 homolog, specifies pharyngeal organ identity in Caenorhabditis elegans. *Genes & Development*, *12*(13), 1947–52. http://doi.org/10.1101/gad.12.13.1947

Howard, R. M., & Sundaram, M. V. (2002). C. elegans EOR-1/PLZF and EOR-2 positively regulate Ras and Wnt signaling and function redundantly with LIN-25 and the SUR-2 Mediator component. *Genes & Development*, *11*, 1815–1827. http://doi.org/10.1101/gad.998402.other

Huang, G., Allen, R., Davis, E. L., Baum, T. J., & Hussey, R. S. (2006). Engineering broad root-knot resistance in transgenic plants by RNAi silencing of a conserved and essential root-knot nematode parasitism gene. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(39), 14302–14306. http://doi.org/10.1073/pnas.0604698103

Huang, G., Gao, B., Maier, T., Allen, R., Davis, E. L., Baum, T. J., & Hussey, R. S. (2003). A profile of putative parasitism genes expressed in the esophageal gland cells of the root-knot nematode Meloidogyne incognita. *Molecular Plant-Microbe Interactions : MPMI*, *16*(5), 376–381. http://doi.org/10.1094/MPMI.2003.16.5.376

Jaouannet, M., Magliano, M., Arguel, M.-J., Gourgues, M., Evangelisti, E., Abad, P., & Rosso, M.-N. (2012). The root-knot nematode calreticulin Mi-CRT is a key effector in plant defense suppression. *Molecular Plant-Microbe Interactions*, *26*(1), 120802061400003. http://doi.org/10.1094/MPMI-05-12-0130-R

Jaouannet, M., Perfus-Barbeoch, L., Deleury, E., Magliano, M., Engler, G., Vieira, P., … Rosso, M. N. (2012). A root-knot nematode-secreted protein is injected into giant cells and targeted to the nuclei. *New Phytologist*, *194*(2008), 924–931. http://doi.org/10.1111/j.1469-8137.2012.04164.x

Jones, S. (2004). An overview of the basic helix-loop-helix proteins. *Genome Biology*, *5*(6), 226. http://doi.org/10.1186/gb-2004-5-6-226

Kalb, J. M., Lau, K. K., Goszczynski, B., Fukushige, T., Moons, D., Okkema, P. G., & McGhee, J. D. (1998). pha-4 is Ce-fkh-1, a fork head/HNF-3alpha,beta,gamma homolog that functions in organogenesis of the C. elegans pharynx. *Development (Cambridge, England)*, *125*, 2171–2180.

Kolaczkowski, B., & Thornton, J. W. (2004). Performance of maximum parsimony and likelihood phylogenetics when evolution is heterogeneous. *Nature*, *431*(7011), 980–984. http://doi.org/10.1038/nature02917

Koressaar, T., & Remm, M. (2007). Enhancements and modifications of primer design program Primer3. *Bioinformatics*, *23*(10), 1289–1291. http://doi.org/10.1093/bioinformatics/btm091

Kormish, J. D., Gaudet, J., & McGhee, J. D. (2010). Development of the C. elegans digestive tract. *Current Opinion in Genetics & Development*, *20*(4), 346–354. http://doi.org/10.1016/j.gde.2010.04.012

Krause, M. W., Fire, A., Harrison, S. W., Priess, J. and Weintraub, H. (1990). CeMyoD accumulation defines the bodywall muscle cell fate during C. elegans embryogenesis. *Cell 63*, 907-922.

Krause, M., Park, M., Zhang, J. M., Yuan, J., Harfe, B., Xu, S. Q., … Fire, a. (1997). A C. elegans E/Daughterless bHLH protein marks neuronal but not striated muscle development. *Development (Cambridge, England)*, *124*(11), 2179–2189.

Kwon, E.-S., Narasimhan, S. D., Yen, K., & Tissenbaum, H. A. (2010). A new DAF-16 isoform regulates longevity. *Nature*, *466*(7305), 498–502. http://doi.org/10.1038/nature09184

Le, S. Q., & Gascuel, O. (2008). An Improved General Amino Acid Replacement Matrix. *Molecular Biology and Evolution*, *25*(7), 1307–1320. http://doi.org/10.1093/molbev/msn067

Lei, H., Fukushige, T., Niu, W., Sarov, M., Reinke, V., & Krause, M. (2010). A Widespread Distribution of Genomic CeMyoD Binding Sites Revealed and Cross Validated by ChIP-Chip and ChIP-Seq Techniques. *PLoS ONE*, *5*(12), e15898. http://doi.org/10.1371/journal.pone.0015898

Li, J., Todd, T. C., Lee, J., & Trick, H. N. (2011). Biotechnological application of functional genomics towards plant-parasitic nematode control. *Plant Biotechnology Journal*, *9*(9), 936–944. http://doi.org/10.1111/j.1467-7652.2011.00601.x

Lin, B., Zhuo, K., Wu, P., Cui, R., Zhang, L.-H., & Liao, J. (2012). A novel effector protein MJ-NULG1a targeted to giant cell nuclei plays a role in Meloidogyne javanica parasitism. *Molecular Plant-Microbe Interactions*, *26*(1), 120703085016009. http://doi.org/10.1094/MPMI-05-12-0114-FI

Lipman, S. F. A. W. G. W. M. E. W. M. D. J. (1990). {B}asic local alignment search tool. *J. Mol. Biol.*, *215*, 403–410.

Liu, Q. L., Thomas, V. P., & Williamson, V. M. (2007). Meiotic Parthenogenesis in a Root-Knot Nematode Results in Rapid Genomic Homozygosity. *Genetics*, *176*(3), 1483–1490. http://doi.org/10.1534/genetics.107.071134

Lowry, O. H., Rosebrough, N. J., Farr, L., & Randall, R. (1951). Protein measurement with the folin phenol reagent. *J. Biol. Chem.*, *193*(1), 265–275.

Lunt, D. H., Kumar, S., Koutsovoulos, G., & Blaxter, M. L. (2014). The complex hybrid origins of the root knot nematodes revealed through comparative genomics. *PeerJ*, *2*, e356. http://doi.org/10.7717/peerj.356

Mango, S. E. (2009). The molecular basis of organ formation: insights from the C. elegans foregut. *Annu Rev Cell Dev Biol*, *25*, 597–628. http://doi.org/10.1146/annurev.cellbio.24.110707.175411

Mango, S. E., Lambie, E. J., & Kimble, J. (1994). The pha-4 gene is required to generate the pharyngeal primordium of Caenorhabditis elegans. *Development (Cambridge, England)*, *120*(10), 3019–3031.

Massari, M. E., & Murre, C. (2000). MINIREVIEW Helix-Loop-Helix Proteins : Regulators of Transcription in Eucaryotic Organisms. *Molecular and Cellular Biology*, *20*(2), 429–440. http://doi.org/10.1128/MCB.20.2.429-440.2000.Updated

Mathelier, A., Zhao, X., Zhang, A. W., Parcy, F., Worsley-Hunt, R., Arenillas, D. J., … Wasserman, W. W. (2014). JASPAR 2014: an extensively expanded and updated open-access database of transcription factor binding profiles. *Nucleic Acids Research*, *42*(D1), D142–D147. http://doi.org/10.1093/nar/gkt997

McCarter, J. P. (2008). Molecular Approaches Toward Resistance to Plant-Parasitic Nematodes. *Plant Cell Monographs*, 239–267. http://doi.org/10.1007/7089

Mello, C. C., Kramer, J. M., Stinchcomb, D., & Ambros, V. (1991). Efficient gene transfer in C. elegans: extrachomosomal maintenance and integration of transforming sequences. *EMBO Journal*, *10*(1), 3959 – 3970.

Mitchum, M. G., Hussey, R. S., Baum, T. J., Wang, X., Elling, A. A., Wubben, M., & Davis, E. L. (2013). Nematode effector proteins: An emerging paradigm of parasitism. *New Phytologist*, *199*(4), 879–894. http://doi.org/10.1111/nph.12323

Molin, L., Mounsey, A., Aslam, S., Bauer, P., Young, J., James, M., … Hope, I. A. (2000). Evolutionary conservation of redundancy between a diverged pair of forkhead transcription factor homologues. *Development (Cambridge, England)*, *127*(22), 4825–4835.

Okkema, P. G., Ha, E., Haun, C., Chen, W., & Fire, a. (1997). The Caenorhabditis elegans NK-2 homeobox gene ceh-22 activates pharyngeal muscle gene expression in combination with pha-1 and is required for normal pharyngeal development. *Development (Cambridge, England)*, *124*(20), 3965–3973.

Opperman, C. H., Bird, D. M., Williamson, V. M., Rokhsar, D. S., Burke, M., Cohn, J., … Windham, E. (2008). Sequence and genetic map of Meloidogyne hapla: A compact nematode genome for plant parasitism. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(39), 14802–14807. http://doi.org/10.1073/pnas.0805946105

Qin, L., Smant, G., Stokkermans, J., Bakker, J., Schots, A., & Helder, J. (1998). Cloning of a trans-spliced glyceraldehyde-3-phosphate-dehydrogenase gene from the potato

cyst nematode Globodera rostochiensis and expression of its putative promoter region in Caenorhabditis elegans. *Molecular and Biochemical Parasitology*, *96*(1-2), 59–67. http://doi.org/10.1016/S0166-6851(98)00108-X

Quentin, M., Abad, P., & Favery, B. (2013). Plant parasitic nematode effectors target host defense and nuclear functions to establish feeding cells. *Frontiers in Plant Science*, *4*(March), 53. http://doi.org/10.3389/fpls.2013.00053

Raharjo, I., & Gaudet, J. (2007). Gland-specific expression of *C. elegans hlh-6* requires the combinatorial action of three distinct promoter elements. *Developmental Biology*, *302*(1), 295–308. http://doi.org/10.1016/j.ydbio.2006.09.036

Raharjo, W. H., Logan, B. C., Wen, S., Kalb, J. M., & Gaudet, J. (2010). In vitro and in vivo characterization of Caenorhabditis elegans PHA-4/FoxA response elements. *Developmental Dynamics : An Official Publication of the American Association of Anatomists*, *239*(8), 2219–32. http://doi.org/10.1002/dvdy.22359

Rosso, M. N., Favery, B., Piotte, C., Arthaud, L., De Boer, J. M., Hussey, R. S., … Abad, P. (1999). Isolation of a cDNA encoding a beta-1,4-endoglucanase in the root-knot nematode Meloidogyne incognita and expression analysis during plant parasitism. *Molecular Plant-Microbe Interactions : MPMI*, *12*(7), 585–591. http://doi.org/10.1094/MPMI.1999.12.7.585

Schneider, C. A., Rasband, W. S., & Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nature Methods*, *9*(7), 671–675. Retrieved from http://dx.doi.org/10.1038/nmeth.2089

Seipel, K., Yanze, N., & Schmid, V. (2004). Developmental and evolutionary aspects of the basic helix–loop–helix transcription factors Atonal-like 1 and Achaete-scute homolog 2 in the jellyfish. *Developmental Biology*, *269*(2), 331–345. http://doi.org/10.1016/j.ydbio.2004.01.035

Smit, R. B., Schnabel, R., & Gaudet, J. (2008). The HLH-6 transcription factor regulates C. elegans pharyngeal gland development and function. *PLoS Genetics*, *4*(10), e1000222. http://doi.org/10.1371/journal.pgen.1000222

Sommer, J. R. (2006). Pristionchus pacificus. *WormBook: The Online Review of C. Elegans Biology*, 1–8. http://doi.org/10.1895/wormbook.1.1

Soucy, S. M., Huang, J., & Gogarten, J. P. (2015). Horizontal gene transfer: building the web of life. *Nature Reviews Genetics*, *16*(8), 472–482. http://doi.org/10.1038/nrg3962

Stein, L. D., Bao, Z., Blasiar, D., Blumenthal, T., Brent, M. R., Chen, N., … Waterston, R. H. (2003). The genome sequence of Caenorhabditis briggsae: A platform for comparative genomics. *PLoS Biology*, *1*(2). http://doi.org/10.1371/journal.pbio.0000045

Sulston, J.E., Schierrenberg, E., White, J.G., Thomson, J.N., 1983. The embryonic cell lineage of the nematode Caenorhabditis elegans. *Dev. Biol. 100*, 64–119

Tamura, K., Stecher, G., Peterson, D., Filipski, A., & Kumar, S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Molecular Biology and Evolution*, *30*(12), 2725–2729. http://doi.org/10.1093/molbev/mst197

Taylor, A. L., & Sasser, J. N. (1978). *Biology, Identification and Control of Root-Knot Nematodes*. Department of Plant Pathology North Carolina State University.

Thellmann, M., Hatzold, J., & Conradt, B. (2003). The Snail-like CES-1 protein of C. elegans can block the expression of the BH3-only cell-death activator gene egl-1 by

antagonizing the function of bHLH proteins. *Development (Cambridge, England)*, *130*(17), 4057–4071. http://doi.org/10.1242/dev.00597

Tenuta, M., Tenuta, A. (2015). Slowing the spread of Soybean Cyst Nematode. *Crops and Soils. 48*(5), 7.

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., & Rozen, S. G. (2012). Primer3-new capabilities and interfaces. *Nucleic Acids Research*, *40*(15), 1–12. http://doi.org/10.1093/nar/gks596

Urwin, P. E., Lilley, C. J., & Atkinson, H. J. (2002). Ingestion of double-stranded RNA by preparasitic juvenile cyst nematodes leads to RNA interference. *Mol Plant Microbe Interact*, *15*(8), 747–752. http://doi.org/10.1094/MPMI.2002.15.8.747

Vilimas, T., Abraham, A., & Okkema, P. G. (2004). An early pharyngeal muscle enhancer from the Caenorhabditis elegans ceh-22 gene is targeted by the Forkhead factor PHA-4. *Developmental Biology*, *266*(2), 388–398. http://doi.org/10.1016/j.ydbio.2003.10.015

Williams, B. D., Schrank, B., Huynh, C., Shownkeen, R., & Waterston, R. H. (1992). A genetic mapping system in Caenorhabditis elegans based on polymorphic sequence-tagged sites. *Genetics*, *131*(3), 609–624.

WILLIAMSON, V., & KUMAR, A. (2006). Nematode resistance in plants: the battle underground. *Trends in Genetics*, *22*(7), 396–403. http://doi.org/10.1016/j.tig.2006.05.003

Williamson, V. M., & Gleason, C. A. (2003). Plant–nematode interactions. *Current Opinion in Plant Biology*, *6*(4), 327–333. http://doi.org/10.1016/S1369-5266(03)00059-1

Wishart, J., Phillips, M. S., & Blok, V. C. (2002). Ribosomal intergenic spacer: A polymerase chain reaction diagnostic for Meloidogyne chitwoodi, M-fallax, and M-hapla. *Phytopathology*, *92*(8), 884–892. http://doi.org/Unsp P-2002-0610-02r

Xue, B., Hamamouch, N., Li, C., Huang, G., Hussey, R. S., Baum, T. J., & Davis, E. L. (2013). The 8D05 parasitism gene of Meloidogyne incognita is required for successful infection of host roots. *Phytopathology*, *103*(2), 175–81. http://doi.org/10.1094/PHYTO-07-12-0173-R

Yoshida, S., Ohbo, K., Takakura, A., Takebayashi, H., Okada, T., Abe, K., & Nabeshima, Y. (2001). Sgn1, a basic helix-loop-helix transcription factor delineates the salivary gland duct cell lineage in mice. *Developmental Biology*, *240*(2), 517–530. http://doi.org/10.1006/dbio.2001.0473

Zanin, E., Dumont, J., Gassmann, R., Cheeseman, I., Maddox, P., Bahmanyar, S., . . . Desai, A. (2011). *Affinity Purification of Protein Complexes in C. elegans* (2nd ed., Vol. 106, Methods in Cell Biology). Waltham, MA: Elsevier. doi:10.1016/B978-0-12-544172-8.00011-6

Zhong, M., Niu, W., Lu, Z. J., Sarov, M., Murray, J. I., Janette, J., … Snyder, M. (2010). Genome-Wide Identification of Binding Sites Defines Distinct Functions for Caenorhabditis elegans PHA-4/FOXA in Development and Environmental Response. *PLoS Genetics*, *6*(2), e1000848. http://doi.org/10.1371/journal.pgen.1000848

Zorn, A. M., & Wells, J. M. (2010). Vetebrate endoderm development. *Annual Review of Cell and Developmental Biology*, 221–251. http://doi.org/10.1146/annurev.cellbio.042308.113344.Vertebrate

Appendix

Table 6.1 Homologous Gene Names and WormBase Gene IDs in order as they appear in phylogenies and promoter analysis.
*splice sites re-predicted

Homologous Gene Name	Worm Base Gene ID
hlh-6	*hlh-6*
Cb-hlh-6	*Cbr-hlh-6*
Mh-hlh-6	*MhA1_Contig384.frz3.gene12*
Mi-hlh-6-1	*Minc00130*
Mi-hlh-6-2	*Minc07019*
Mf-hlh-6-1	*maker-nMf.1.1.scaf00640-augustus-gene-0.13*
Mf-hlh-6-2	*maker-nMf.1.1.scaf11064-augustus-gene-0.3**
Gp-hlh-6	*GPLIN_001235500*
Hb-hlh-6	*Hba_15973**
Bm-hlh-6	*Bm5805*
Pp-hlh-6	*Ppa_Contig33:1,320,300..1,322,299**
hlh-1	*hlh-1*
Cb-hlh-1	*Cbr-hlh-1*
Mh-hlh-1	*MhA1_contig1774.frz3.gene3**
Mi-hlh-1-1	*Minc10641*
Mi-hlh-1-2	*Minc05405*
Mf-hlh-1	*augustus_masked-nMf.1.1.scaf12118-processed-gene-0.0*
Hb-hlh-1	*Hba_04209**
Bm-hlh-1	*Bm1750*
Pp-hlh-1	*Ppa08246*
hlh-2	*hlh-2*
Cb-hlh-2	*Cbr-HLH-2*
Mh-hlh-2	*MhA1_contig1203.frz3.gene2*
Mf-hlh-2	*maker-nMf.1.1.scaf08683-snap-gene-0.4*
Hb-hlh-2	*Hba_05393*
Bm-hlh-2	*Bm5103*
Pp-hlh-2	*Ppa24788**
hlh-3	*hlh-3*
Cb-hlh-3	*Cbr-HLH-3*
Hb-hlh-3	*Hba_02087*
Bm-hlh-3	*Bm10445*
Pp-hlh-3	*Ppa22802**

hlh-4	*hlh-4*
Cb-hlh-4	*Cbr-HLH-4*
Mh-hlh-4	*MhA1_Contig880.frz3.gene1**
Mf-hlh-4	*augustus_masked-nMf.1.1.scaf00213-processed-gene-0.1*
Gp-hlh-4	*GPLIN_000401900*
Hb-hlh-4	*Hba_08861*
Bm-hlh-4	*Bm5590*
Pp-hlh-4	*Ppa05241*
hlh-14	*HLH-14*
Cb-hlh-14	*Cbr-HLH-14*
Hb-hlh-14	*Hba_02090*
Bm-hlh-14	*Bm2289*
Pp-hlh-14	*Ppa28351*
Mh-hlh-6-like	*MhA1_Contig809.frz3.gene10*
Mi-hlh-6-like-1	*Minc17316*
Mi-hlh-6-like-2	*Minc14681*
Mf-hlh-6-like	*augustus_masked-nMf.1.1.scaf04284-processed-gene-0.0*
Gp-hlh-6-like	*Gpal_scaffold_120:2945-3351**
pha-4	*pha-4*
Cb-pha-4	*Cbr-PHA-4*
Mh-pha-4	*MhA1_Contig1988.frz3.gene1*
Mi-pha-4	n/a
Gp-pha-4	*Gplin_000566600*
Hb-pha-4	*Hba_19194*
Bm-pha-4	*Bm3905*
Pp-pha-4	*Ppa14055*
Mf-pha-4	*maker-nMf.1.1.scaf00075-augustus-gene-0.15-mRNA-1**
lin-31	*LIN-31*
Cb-lin-31	*Cbr-LIN-31*
Mh-lin-31	*MhA1_Contig168.frz3.gene9**
Gp-lin-31	*Gplin_001431700**
Hb-lin-31	*Hba_14851**
Bm-lin-31	*Bm-scaffold1191:-660:1444:1**
Pp-lin-31	*Ppa10081**
Mf-lin-31	*maker-nMf.1.1.scaf03071-snap-gene-0.6*
fkh-2	*fkh-2*
Cb-fkh-2	*Cbr-FKH-2*

Mh-fkh-2	*MhA1_Contig83.frz3.gene6**
Mi-fkh-2-1	*Minc01693*
Mi-fkh-2-2	*Minc04531*
Gp-fkh-2	*Gplin_001134200*
Hb-fkh-2	*Hba_4886**
Bm-fkh-2	*Bm5802b*
Pp-fkh-2	*Ppa18101*
fkh-9	*fkh-9*
Cb-fkh-9	*Cbr-FKH-9*
Mh-fkh-9	*MhA1_Contig2150.frz3.gene5*
Mi-fkh-9	*Minc17526*
Gp-fkh-9	*Gplin_000643600*
Bm-fkh-9-1	*Bm4783*
Bm-fkh-9-2	*Bm8986*
Pp-fkh-9	*Ppa29963**
Mf-fkh-9	*augustus_masked-nMf.1.1.scaf03615-processed-gene-0.0*
daf-16	*daf-16*
Cb-daf-16	*Cbr-DAF-16*
Mh-daf-16	*MhA1_Contig353.frz3.fgene4*
Gp-daf-16	*Gplin_001276900*
Bm-daf-16	*Bm5392*
Mf-daf-16	*snap_masked-nMf.1.1.scaf25582-processed-gene-0.1*
Mi-ubq-2-1	*Minc16562*
Mi-pig-1	*Minc16584*
Minc15667	*Minc15667*
Mh-crt	*MhA1_Contig309.30*
Mi-crt	*Minc06693*
Mh-8D05	*MhA1_Contig2374.frz3.gene4*
Mf-8D05	*snap_masked-nMf.1.1.scaf08123-processed-gene-0.5*
Mh-eng-1	*MhA1_Contig344.frz3.gene3*
Mi-eng-1	*Minc13221*
Minc03866	*Minc03866*
Mha1_contig203.frz3.gene9	*Mha1_contig203.frz3.gene9*
Mh-D15	*MhA1_Contig1354.frz3.gene2*
Mi-D15	*Minc09973*
Mh-eff-1	*MhA1_Contig138.frz3.gene15*
Mi-eff-1	*Minc17998*

A)

```
Mh-HLH-6          ATGAACAATAATCCAATAAATGACCAATCAGTCCCTTTAACAATAGATATTCTCACCAAT
Mf-HLH-6-2        ATGAATAATACTCCAATAAATGACCAATCACAACCCTTAACTATAGATATCCTCACCAAT
Mf-HLH-6-1        ATGAATAATACTCCAATAAATGACCAATCACAACCCTTAACTATAGATATCCTCACCAAT
Mi-HLH-6-1        ATGAATAATAATCCAATAAACGACCAATCACAACCCTTAACTATAGATATTCTCACTAAT
Mi-HLH-6-2        ATGAATAATAATCCAATAAACGACCAATCACAACCCTTAACTATAGATATTCTCACTAAT
Que-1-HLH-6       ATGAATAATAATCCAATAAACGACCAATCACAACCCTTAACTATAGATATTCTCACTAAT
                  ***** **** ********* *********   ** ***** ******** ***** ***

Mh-HLH-6          GA---ACAACAAATGCTTCTAGAACGTCAACAAGCATTATGTGTAGCTGCTGCTCAGGCT
Mf-HLH-6-2        CAACAACAACAAATGCTTCTAGAACATCAACAAGCATTGTGTGCTGCTGCGGCGCAAAGT
Mf-HLH-6-1        CAACAACAACAAATGCTTCTAGAACATCAACAAGCATTGTGTGCTGCTGCGGCGCAAAGT
Mi-HLH-6-1        GAGCAACAACAAATGCTTCTAGAACATCAACAAGCATTGTGTGTAGCTGCGGCGCAAACC
Mi-HLH-6-2        GAACAACAACAAATGCTTCTAGAACGTCAACAAGCATTGTGTGTAGCTGCGGCGCAAACC
Que-1-HLH-6       GAACAACAACAAATGCTTCTAGAACGTCAACAAGCATTGTGTGTAGCTGCGGCGCAAACC
                   *   ******************** ************ ****  ***** ** **

Mh-HLH-6          GTCAATTTTGCTAATATAAATATGGAAGCTTCTGCCTTTCAAAGAAGCAGCAGTGTAATA
Mf-HLH-6-2        GTTAATTTTGCTAATATAAATTTAGATGCTTCTGCTTTTGAAAGAAGCAGCAGCGTAATA
Mf-HLH-6-1        GTTAATTTTGCTAATATAAATTTAGATGCTTCTGCTTTTGAAAGAAGCAGCAGCGTAATA
Mi-HLH-6-1        GTTAATTTTGCTAATGTAAATTTAGAAGCTTCTGCTTTTCAAAGAAGCAGCAGCGTAATA
Mi-HLH-6-2        ATTAATTTTGCTAATATAAATATAGATGCTTCTGCTTTTCAAAGAAGCAACAGCGTAATA
Que-1-HLH-6       ATTAATTTTGCTAATATAAATATAGATGCTTCTGCTTTTCAAAGAAGCAACAGCGTAATA
                   * ************ ***** * ** ******** *** ********* *** ******

Mh-HLH-6          GTTAATCCACCAATTCAAGAAAAAGGAATTAAAAGGAAAAGAGGAGATGAAATCGACCAA
Mf-HLH-6-2        GTTGTTCCACCTAGCCAAGAAAAAGGGACTAAAAGGAAAAGAAAAGATGATAATGGGCAA
Mf-HLH-6-1        GTTGTTCCACCTAGCCAAGAAAAAGGGACTAAAAGGAAAAGAAAAGATGATAATGGGCAA
Mi-HLH-6-1        GTTGTTCCACCTAGCCAAGAAAAAGGGACTAAAAGGAAAAGAAAAGATGATAATGGGCAA
Mi-HLH-6-2        GTAGTTCCACCTAGCCAAGAAAAAGGGACTAAAAGAAAAAGAAAAGATGATAATGGGCAA
Que-1-HLH-6       GTAGTTCCACCTAGCCAAGAAAAAGGGACTAAAAGAAAAAGAAAAGATGATAATGGGCAA
                  **   ****** *  *********** * ****** ******  ****** *  *  ***

Mh-HLH-6          CAAGAAATACAACAAAAAATGCCTAATATGGGTTGCGGGATGCGGGAACCAAAAGAAGTT
Mf-HLH-6-2        CAAGAATTACAGCAAAAAATGCCCAAAATGATAACACCAGTAAAAAGCCCAAAAGAGGTG
Mf-HLH-6-1        CAAGAATTACAGCAAAAAATGCCCAAAATGATAACACCAGTAAAAAGCCCAAAAGAGGTG
Mi-HLH-6-1        CAAGAATTACAGCAAAAAATGCCCAAAATGATAACACCAGTAAAAAGCCCAAAAGAGGTG
Mi-HLH-6-2        CAAGAATTACAGCAAAAAATGCCTAAAATGACAACACCTGTTAAAAGCCCAAAAGAGGTT
Que-1-HLH-6       CAAGAATTACAGCAAAAAATGCCTAAAATGACAACACCTGTTAAAAGCCCAAAAGAGGTT
                  ****** **** *********** ** ***          *       ******** **

Mh-HLH-6          GTAGGTGTAAATCGTTTTATTAGTCCTTTTTCGGAAGAGGCTAAAATACCGTTGCCACAC
Mf-HLH-6-2        TTAGGCGTAAATCGTTTTATTAGTCCCTTCTCAGAAGAGGCTAAAATACCGTTGCCGCAT
Mf-HLH-6-1        TTAGGCGTAAATCGTTTTATTAGTCCCTTCTCAGAAGAGGCTAAAATACCGTTGCCGCAT
Mi-HLH-6-1        TTAGGTGTAAATCGTTTTATTAGTCCCTTCTCGGAAGAAGCTAAGATACCGTTACCGCAT
Mi-HLH-6-2        TTAGGTGTAAATCGTTTTA----TCCCTTCTCTGAAGAGGCTAAAATACCGTTGCCGCAT
Que-1-HLH-6       TTAGGTGTAAATCGTTTTATTAGTCCCTTCTCTGAAGAGGCTAAAATACCGTTGCCGCAT
                   **** *************    *** ** ** ***** ***** ******** ** **

Mh-HLH-6          GAATTGAATGAGTTAAGCACACATGGTACAACCGTTTGGAAACGCAATGAAAGGGAAAGA
Mf-HLH-6-2        GAATTGAATGAATTGAGCACACACGGAACTACAGTTTGGAAGCGTAATGAACGAGAACGG
Mf-HLH-6-1        GAATTGAATGAATTGAGCACACACGGAACTACAGTTTGGAAGCGTAATGAACGAGAACGG
Mi-HLH-6-1        GAATTGAATGAATTAAGCACACACGGAACAACCGTTTGGAAGCGTAATGAACGAGAACGG
Mi-HLH-6-2        GAATTGAATGAATTTAGCACACACGGAACTACCGTCTGGAAGCGTAATGAACGAGAAAGG
Que-1-HLH-6       GAATTGAATGAATTGAGCACACACGGAACTACCGTCTGGAAGCGTAATGAACGAGAAAGG
                  *********** ** ******** ** ** ** ** ***** ** ****** * *** *

Mh-HLH-6          ATTAGAGTTCGAGAAGTTAACAATGGATATACGAATTTGAAAAGCAAGTTGCCTTTGACT
Mf-HLH-6-2        ATTCGAGTTAGACAAGTTAATGATGGATATACGAGGTTGAAGAACACTCTACCTTTGACT
Mf-HLH-6-1        ATTCGAGTTAGACAAGTTAATGATGGATATACGAGGTTGAAGAACACTCTACCTTTGACT
Mi-HLH-6-1        ATTCGAGTTAGACAAGTTAATGATGGATATACGAGATTGAAGAACACGCTGCCTCTGACT
Mi-HLH-6-2        ATTCGCGTCAGACAAGTTAATGATGGATATACGAGGTTGAAGAACACGTTGCCTTTGACT
Que-1-HLH-6       ATTCGCGTCAGACAAGTTAATGATGGATATACGAGGTTGAAGAACACGCTGCCTCTGACT
                  *** * **  ** *******  ************  ***** * **   * *** *****
```

```
Mh-HLH-6          GAAAACGATCGACGTCTTAGCAAAGTTGATACTCTTCGCCTAGCTATTGCATATATTCAT
Mf-HLH-6-2        GAAAATGATCGCCGACTTAGCAAAGTTGATACGCTTCGCCTAGCTATTGCATATATTCAC
Mf-HLH-6-1        GAAAATGATCGCCGACTTAGCAAAGTTGATACGCTTCGCCTAGCTATTGCATATATTCAC
Mi-HLH-6-1        GAAAATGATCGCCGACTTAGCAAAGTTGATACGCTTCGCCTAGCTATTGCATATATTCAC
Mi-HLH-6-2        GAAAATGATCGCCGACTTAGCAAAGTTGATACTCTTCGCCTAGCTATTGCGTATATTCAC
Que-1-HLH-6       GAAAATGATCGCCGACTTAGCAAAGTTGATACGCTTCGCCTAGCTATTGCATATATTCAC
                  ***** ***** ** ***************** ***************** ********

Mh-HLH-6          CATTTGGAAAATTTAATTAATGAAGGAGTAAATCATTTGATTGAGTGCCAATGTTTTAAT
Mf-HLH-6-2        CATTTGGAAAACTTAATTAATGAAGGGGTTAATCATCTAATCGAGTGCCAATGTTTTAAT
Mf-HLH-6-1        CATTTGGAAAACTTAATTAATGAAGGGGTTAATCATCTAATCGAGTGCCAATGTTTTAAT
Mi-HLH-6-1        CATTTGGAAAACTTAATTAATGAAGGGGTTAATCATCTAATCGAGTGCCAATGTTTTAAT
Mi-HLH-6-2        CATTTGGAAAATTTAATTAATGAAGGGGTTAATCACCTAATCGAGTGCCAATGTTTTAAT
Que-1-HLH-6       CATTTGGAAAACTTAATTAATGAAGGGGTTAATCATCTAATCGAGTGCCAATGTTTTAAT
                  *********** ************** ** *****  * ** ******************

Mh-HLH-6          TATGCGATGGGGGAATGCAGCAGTGGAGAAAATACACCACCACCAAATTAA
Mf-HLH-6-2        TATGTAATGGGAGAATGTAGCAGTGCTGAAAATACGCCTCCACCAAATTAA
Mf-HLH-6-1        TATGTAATGGGAGAATGTAGCAGTGCTGAAAATACGCCTCCACCAAATTAA
Mi-HLH-6-1        TATGTAATGGGAGAATGTAGCAGTGCTGAAAATACGCCTCCACCAAATTAA
Mi-HLH-6-2        TATGTAATGGGAGAGTGTAGCAGCGCTGAAAATACACCGCCGCCAAATTAA
Que-1-HLH-6       TATGTAATGGGAGAATGTAGCAGTGCTG-----------------------
                  ****  ***** ** ** ***** *  *
```

RKN Homologues:	Mi-HLH-6-2	Mi-HLH-6-1	Mf-HLH-6-1	Mf-HLH-6-2	Mh-HLH-6
%ID with: Que-1-hlh-6	98.56	96.02	95.06	95.06	83.04

B)

```
Mh-PHA-4          ------------------------------------------------------------
Mf-PHA-4          ATGCAACAACAATTTATTTCTTCTCTCGATATTTCCTCTACTACAAATCCTTTAATAGAA
Que1-PHA-4        ------------------------------------------------------------

Mh-PHA-4          ------------------------------------------------------------
Mf-PHA-4          ATGCCTTCAAAAACAACAACAACAGCACCAACAGATTATTTAAATGATCAATGGAACATT
Que1-PHA-4        ------------------------------------------------------------

Mh-PHA-4          ------------------------------------------------------------
Mf-PHA-4          TACAACAATTCTTCTTCTTCCCCAACAACAAATTTAAATATTTCTTTAATTGGCCAAACA
Que1-PHA-4        ------------------------------------------------------------

Mh-PHA-4          ------------------------------------------------------------
Mf-PHA-4          CCCTCTTCTTCTGTTTATCCTCCTTCCCAACCTCCACCACCTCCTCCACCAATTTATCCC
Que1-PHA-4        ------------------------------------------------------------

Mh-PHA-4          ----------ATGA-------------TATCATGGGAAGTTGTAGAAGAAGTAAGACAAC
Mf-PHA-4          CCTTCACAATATACTTCTCTAAATATTTATTCAAACAATTTATATATCCCCTCTTCTTCC
Que1-PHA-4        ------------------------------------------------------------

Mh-PHA-4          AAATACAACA-ACAACAAGAAAAGCAACAACAACAACAGCAACAATTTATTTCTTCTCTT
Mf-PHA-4          CCTTCCTCCTCCTCTTCTTCAAATCCCNNNNNNNNNNTCCAACAATTTATTTCTTCTCTC
Que1-PHA-4        -----------------------------------CAACAACAATTTATTTCTTCTCTC
                                                     ********************

Mh-PHA-4          GATATTTCTTCAACAAATTCTTTAACAGGAGGAATGTCTTCAACAACAACAACAACAGAT
Mf-PHA-4          GACATTTCCTCTACTACAAATCCTTTAATAGAAATGCCTTCAAAAACAA---CAACAGCA
Que1-PHA-4        GATATTTCCTCTACTACAAATCCTTTAATAGAAATGCCTTCAAAAACAACAACAACAGCA
                  ** ***** ** ** *    *     *  ** **** ****** *****   ******
```

```
Mh-PHA-4        TATCTAAATGACCAACAACAATGGAATAATATTTATCAAAATAATAATTC------TTCA
Mf-PHA-4        CCAACAGATTATTTAAATGATCAATGGAATATTTATAACAACAACAATTCTTCCCCTTCC
Que1-PHA-4      CCAACAGATTATTTAAATGATCAATGGAATATTTATAACAACAACAATTCTTCCCCTTCC
                     * ** *   * *  *       ********* * ** ** *****      ***

Mh-PHA-4        CCAAATTCTTCTTCCACAATTTCATTAATTAACCAAAACCCTTCATTAACTAATTCT---
Mf-PHA-4        CCAACAACAAATTTAAATATTTCCTTAATTGGCCAAACACCCTCTTCTTCTGTTTATCCT
Que1-PHA-4      CCAACAACAAATTTAAATATTTCCTTAATTGGCCAAACACCCTCTTCTTCTGTTTATCCT
                ****   *   **  *  ***** ******  *****  ** ** *   **  ** *

Mh-PHA-4        ----------------------TTATATTCTTCCCCAATTTATCCTCCACAATATGCTTCT
Mf-PHA-4        CCTTCCCAACCTCCACCACCTCCACCTCCAATTTATCCTCCCCCTTCACAATATACTTCT
Que1-PHA-4      CCTTCCCAACC---------TCCACCTCCAATTTATCCTCCCCCTTCACAATATACTTCT
                                  *  * *         *   *** ******** *****

Mh-PHA-4        TTAAATGGTTTTCAAAATAATCTTTATATTCCTTCTTCTTCCCCTTCTTCTTCTTCAAAT
Mf-PHA-4        CTAAATATTTATTCAAACAATTTATATATCCCTTCTTCTTCCCCTTCCTCTTCTTCAAAT
Que1-PHA-4      CTAAATATTTATTCAAACAATTTATATATCCCTTCTTCTTCCCCTTCCTCTTCTTCAAAT
                 *****  ** *  *** *** * ***** ***************** ************

Mh-PHA-4        CCACAAATTTGTCAATATATACAACAACAACAATTATTAGAACAAAATAATTCTTTAAAT
Mf-PHA-4        CCACAATTTAGTCAATATTTACAGCAACAATTAATTGATAA---TAATTCTTTAAATTTA
Que1-PHA-4      CCACAATTTAGTCAATATTTACAGCAACAATTAATTGATAA---TAATTCTTTAAATTTA
                ****** ** ******** **** ******  * *     *    ***  **

Mh-PHA-4        TTATCATATACTTCTGCTTTTCAATCAGCAATTAATTCTGCTACAAATAATTTAAATTCT
Mf-PHA-4        TCTTCATATCCTTCAGCTTTTCAATCAACAATTAATTCAGTAATTAATAGT---------
Que1-PHA-4      TCTTCATATCCTTCAGCTTTTCAATCAACAATTAATTCAGTAATTAATAGT---------
                *  ****** **** ************ ********** *  *  **** *

Mh-PHA-4        TCAATTAGAGGGGGAGAAGAACAAAAATATATTGAAATAAATAAACAATTAAAAAATAGT
Mf-PHA-4        ------AGAGGGGAAGA------------AGTTGAAATAAATAAACAATTAAAAAATTGT
Que1-PHA-4      ------AGAGGGGAAGA------------AGTTGAAATAAATAAACAATTAAAAAATTGT
                      ******* ***             ************************** **

Mh-PHA-4        AAAAATCAGTCAACAAATAA------TCAAATTGGAGATTCTAATAATAATCAAAGTCCA
Mf-PHA-4        AGAAATAAGACATCATCTACTAATGGAGAATCTACAAATACAACAAATCAAAGTCCACCA
Que1-PHA-4      AGAAATAAGACATCATCTACTAATGGAGAATCTACAAATACAACAAATCAAAGTCCACCA
                * **** ** ** **  **         **  *  * ** * *  *** *       ***

Mh-PHA-4        ACAATAAATACTATGGAATTAACTGTTCATGAAATGCAAAAAATTAAAAATCAAGGAAAT
Mf-PHA-4        ATACACAATATGGAGGAATTAACTGTTCACGAAATGCAAAAAATTAAAAATCAAGGAAGT
Que1-PHA-4      ATACACAATATGGAGGAATTAACTGTCCACGAAATGCAAAAAATTAAGAATCAAGGAAGT
                * *   ****    ************ ** ***************** ********** *

Mh-PHA-4        TTTGGTCCAAATAAACCTCCTTATTCTTATATTTCTTTAATTTCTATGGCTATACAACAA
Mf-PHA-4        TATGGTCCAAATAAGCCTCCATATTCTTATATTTCTTTAATTTCTATGGCTATACAACAA
Que1-PHA-4      TATGGTCCAAATAAGCCTCCATATTCTTATATTTCTTTAATTTCTATGGCTATACAACAA
                * ************ ***** ****************************************

Mh-PHA-4        AGTGACCGAAAAATGTGTACATTGAGTGAGATTTATAATTTTATAATGGAATACTTTGAA
Mf-PHA-4        AGTGATCGAAAAATGTGTACATTGAGTGAAATATATAATTTTATTATGGCATATTTTGAA
Que1-PHA-4      AGTGATCGAAAAATGTGTACATTGAGTGAAATATATAATTTTATTATGGCATATTTTGAA
                ***** ********************** ** *********** **** *** ******

Mh-PHA-4        TATTACAAAAATCATAATCAACGTTGTTCTTGGCAAAATTCAATAAGACATTCTTTATCT
Mf-PHA-4        TATTACAAAAATCATAATCAACGTTGTTCCTGGCAAAATTCAATTAGGCATTCTTTATCT
Que1-PHA-4      TATTACAAAAATCATAATCAACGTTGTTCCTGGCAAAATTCAATTAGGCATTCTTTATCT
                **************************** ************** ** ************

Mh-PHA-4        TTTAATGATTGTTTTGTTAAAGTACCTCGAACACCTGATAGAATAAAAAATTGCAGACCA
Mf-PHA-4        TTTAATGATTGTTTTGTTAAAGTTCCTCGAACCCCCGACAGAATAAAAAATTGCAGACCG
Que1-PHA-4      TTTAATGATTGTTTTGTTAAAGTTCCTCGAACCCCCGACAGAATAAAAAATTGCAGACCG
                ********************** ******** ** ** ********************

Mh-PHA-4        GGCAAAGGTTCTTTTTGGACATTACATGCATTGTGTGGAGATATGTTCGAAAATGGATGT
Mf-PHA-4        GGCAAAGGTTCTTTTTGGACTTTACATGCTTTGTGTGGAGATATGTTCGAAAATGGATGT
Que1-PHA-4      GGCAAAGGTTCTTTTTGGACTTTACATGCTTTGTGTGGAGATATGTTCGAAAATGGATGT
                ******************* ******** ******************************
```

```
Mh-PHA-4        TTTTTGAGAAGACAAAAACGTTTTAAATTAAAAGATAAAGAAAAACCAGAAAGGCAAAGA
Mf-PHA-4        TTTTTGAGAAGACAAAAACGTTTTAAACTTAGCAAAACTGAGAAACCAGAAAGGCAAAGA
Que1-PHA-4      TTTTTGAGAAGACAAAAACGTTTTAAACTTAGCAAAACTGAGAAACCAGAAAGGCAAAGA
                ************************** * *   * *  ** ******************

Mh-PHA-4        AGAACTAAAAATATTTCGAATAAATTAACTTT------AACAGAAAGAAAATTACAAAAA
Mf-PHA-4        AGAAATAATAAATTAAATTTGGGGGGAGAAAGGAAGGCTAATAACAACAATTTAAATAAA
Que1-PHA-4      AGAAATAATAAATTAAATTTGGGGGGAGAAAAGAAGGCTAACAACAATAATTTAAATAAA
                **** *** **  *           *            *   * *  ** *** * ***

Mh-PHA-4        CAACAAAAACAAATGATTAATTCTGCTTCTTCTATTTTATTAGAACAAAAACCAAAAAGG
Mf-PHA-4        CAGAAAAAACAACAAAAAAATTTTAATTCTTTAATTTTGACTGAACAAAAATTAAAAAAA
Que1-PHA-4      CAGAAAAAACAACAAAAAAAATTTAATTCTTTAATTTTAACTGAACAAAAATTAAAAAAA
                **  ********   *  ** * *  *****  *****    *********  *****

Mh-PHA-4        GAAATAGAATCTTCAATAGATTTAAAATTAGAACCTTTAATTCAAACAGATATAGGACTA
Mf-PHA-4        GAAATATCAGACTCTTCCACAATTGATTTAAAACCTTTAATTCAAACAAATATATTAGAA
Que1-PHA-4      GA------------------------------------------------------------
                **

Mh-PHA-4        CGACTACTACAATTAATAA-------------------------CAGAAGAAATAAATA
Mf-PHA-4        CAACACCCCCAACACCCCCACTCCCACCACCAANNNNNNNNNNCTGAGAAACCAGAAAGG
Que1-PHA-4      ------------------------------------------------------------


Mh-PHA-4        ATAATAATAATTTAATAGATATTATTGG-GGATGGAAATAATTCTTCCAATA--------
Mf-PHA-4        CAAAGAAGAAATAATAAATTAAATTTGGGGGGAGAAAAGAAGGCTAACAACAACAATTTA
Que1-PHA-4      ------------------------------------------------------------


Mh-PHA-4        -------AAGAAAAACATTTAAATAATAATTTAATTCAATTAATAAGGCCATCAGCAGAA
Mf-PHA-4        AATAGTAAACAGAAAAAACAACAAAAAAATTTTAATTCTTTAATTTTGA---CTGAACAA
Que1-PHA-4      ------------------------------------------------------------


Mh-PHA-4        AATTTAAATAATAATTTA-ATTTTACCTTCACAATTTAATAATAATTTTGAATT------
Mf-PHA-4        AAGTTAAAAAAGGAAATATCAGATTCTTCTACAATTGATTTAAAACCTTTAATTCAAACA
Que1-PHA-4      --------------AATATCAGACTCTTCCACAATTGATTTAAAACCTTTAATTCAAACA
                              **       * *  ****** * * * **  ** ****

Mh-PHA-4        -ATCTCATTTAACAACACAATGCCCTTCAAGTTCTTCTGGAATTTTACAAGGAAATTCTA
Mf-PHA-4        AATATATTAGAACACCCCCACTCCCAACACCACCAACAATTAATTGGAGAAGATGGAAAT
Que1-PHA-4      AATATATTAGAACACCCCCACCCCCATCACCACCAACAATTAATTGGAGAAGATGGAAAT
                 ** *  *  **** * * *  ***  **    *  *    * **    * **

Mh-PHA-4        ATTTATGTGGATCTCCTGTTA----TTTCTTCTATTGGAGGGGGACAAATGTTATCATCA
Mf-PHA-4        AATAATAATAATTTAATATTAGATATTATTGGTAATGAAGAAGAAGGAAAATATCAAAAA
Que1-PHA-4      AATAATAATAATTTAATATTAGATATTATTGGTAATGAAGAAGAAGAAAAATATCAAAAA
                * * **    ** *  * ***    **  *  ** ** **  * *  **  *    *  *

Mh-PHA-4        TTATCTTCTGCTTATTGTCATCAATTTCCAAATTCAACAACAGCTATAGTCCATCCAACA
Mf-PHA-4        ATAAA------TAATTTTAATTTAATTACTAATAGAAGAAGACAAGAAGAAAATACAAA-
Que1-PHA-4      ATAAA------TAATTTTAATTTAATTACTAATAGAAGAAGACAAGAAGAAAATACAAA-
                 **        * *** * **  * ** * ***  ** ** *     **   ** ***

Mh-PHA-4        ATAACACAACAATTTCCCCATCAATTTTATACTTC---AAATAATTCTGATTTAGATAAT
Mf-PHA-4        --------TAATTTACAGCAACAAATTGAAATATTTAATAATAATTTTGAATTAGCTAAT
Que1-PHA-4      --------TAATTTACAACAACAAATTGAAATATTTAATAATAATTTTGAATTAGCTAAT
                         * ** *  ** *** ** * *  *     ******* *** **** ****

Mh-PHA-4        TGTTCAAGTCCTCAATCAACAAATAATAATTTAAATTCACAACAATTAACATGCCTTAAT
Mf-PHA-4        TTAACACAAATTCAACAATGCCCCTCAAG---TTCTTCTGAAATTTTACAAGGAAATAAT
Que1-PHA-4      TTAACACAAATTCA---ATGTCCCTCAAG---TTCCTCTGAAATTTTACAAGGAAATAAT
                *   **     ***   *         *        **  **   ***  * *   ****

Mh-PHA-4        TCTTCTGGACCTTTTTTAAATATTAATTTAAATATA-AATCAAAATTTAATTGATTATCA
Mf-PHA-4        TTGTGTGGATCTCCTTCTGTTATTTCTTCTATTGGAGGAGGAGAGCAACAATTATTATCT
Que1-PHA-4      TTGTGTGGATCTCCTTCTGTTATTTCTTCTATTGGAGGAGGAGGGGGAGGACAACAATTA
                *  * **** **  **    ****  **  * *  *  *  *           *  **
```

```
Mh-PHA-4        ACCAATAATTCCTCCTAATTTTCCACCAGAATATTCAATGATTTATG-------------
Mf-PHA-4        TCCTCTTCTTCTTCCTCAATTTATTGTCATCAATTCCCAAATTCCTCAGCAACAGCTATA
Que1-PHA-4      TTATCTTCCTCTTCCTCAATTTATTGTCATCAATTCCCAAATTCCTCAGCAACAGCTATA
                     *   ** **** * ***          ****    ***  *

Mh-PHA-4        -GAGGACCAGCACAGATTAATACAATTTATAATGGAAAGTTAGTTTTTTTTAAATATTTT
Mf-PHA-4        ATCCATCCAACAACAATAACTACACAACAACAAC---AACAA---TTTCCCCAACAATTT
Que1-PHA-4      ATCCATCCAACAACAATAACTACACAACAACAAC---AACAACAATTTCCCCAACAATTT
                      *** **   ** * ****    *  *     *   *   ***    ** * ***

Mh-PHA-4        TGTTTAATGTC--------------------------------TTTAAATAT---TTTTT
Mf-PHA-4        TATTCTTCTTCAGCAAG-----------------------TCCTCAACAATTAATTAAT
Que1-PHA-4      TATTCTTCTTCAAATAATTTAAATTGTGCTGCTGCTGCAAGTCCTCAACAGTTAAT----
                * **     **                                  * **   *   *

Mh-PHA-4        TTTAA-------------------------------------------------------
Mf-PHA-4        TCTCAACAACAACAACAATTATTTAACCCCCAACCTTTAGTTAATATTCATTTAAATATT
Que1-PHA-4      --TAATTCTCAACAACAATTATTTAACCCCCAACCTTTAGTTAATATTCATTTAAATATT
                  * *

Mh-PHA-4        -------------------------------------------------------------
Mf-PHA-4        AATCAACACCTTGTTGACTATCAAANNNNNNNNNNAATGATGTATAGTGGTGGTAATATA
Que1-PHA-4      AATCAACCCCTTGTTGACTATCAAATGTTA-----------------------------


Mh-PHA-4        ----------------------------------------------------
Mf-PHA-4        TCTGGGGGAGGAGGGGGGCAAACAACACAGACAATTTATAATGAAAAGTAA
Que1-PHA-4      ----------------------------------------------------
```

RKN Homologues:	Mf-PHA-4	Mh-PHA-4
%ID with: Que1-PHA-4	97.09%	73.37%

C)

```
Mh-HLH-6-LIKE       ------------------------------------------------------------
Mf-HLH-6-LIKE       ---------ATGCGTGAATGTTCTTCTTCTTCTTCTTCTTCCCCATCTTGGCCTTCTAAC
Mi-HLH-6-LIKE-2     ATGCGTGAATGTTCTTCTTCTTCTTCTTCTTCTTCTTCCCCATCTTGGCCTTCTAAC
Que1-HLH-6-LIKE     ------------------------------------------------------------
Mi-HLH-6-LIKE-1     ------ATGCGTGAATGTTCTTCTTCTTCTTCCTCTTCCCCATCTTGGCCTTCTAAC


Mh-HLH-6-LIKE       ------------------------------------------------------------
Mf-HLH-6-LIKE       CAACAAAATAATTCTTCATATTCCTCATCATTTTCTTCATCTCCTTCAATTAATTCTTCA
Mi-HLH-6-LIKE-2     CAACAAAATAATTCTTCATATTCTTCATCATTTTCTTCATCTCCTTCAATTAATTCTTCA
Que1-HLH-6-LIKE     ------------------------------------------------------------
Mi-HLH-6-LIKE-1     CAACAAAATACTTCCTCATATTCTTCATCATTTTCTTCCTCTCCTTCAATTAATTCTTCC


Mh-HLH-6-LIKE       ------------------------------------------------------------
Mf-HLH-6-LIKE       AATTTAATTTATTTTCCTCTAAACAATTGTCAATCAACTTCTTTAATTAAAGATTCCAAC
Mi-HLH-6-LIKE-2     AATTTAATTTATTTTCCTCTAAACAATTGTCAATCAACTTCTTTAATTAAAGATTCCAAC
Que1-HLH-6-LIKE     ----------------------ACAATTATCAATCAACTTCTTTAATTAAAAATTCCAAC
Mi-HLH-6-LIKE-1     AATTTAATTTATTTTCCTTTAGACAATTATCAATCAACTTCTTTAATTAAAAATTCCAAC


Mh-HLH-6-LIKE       ------------------------------------------------------------
Mf-HLH-6-LIKE       AATTTTTATTTAAATATAAAAACAAAT---TCAAAACAACAAAAATTAAATAAAAATGGA
Mi-HLH-6-LIKE-2     AATTTTTATTTAAATATAA---AAACAAATTCAAAACAACAAAAATTAAATAAAAATGGA
Que1-HLH-6-LIKE     AATTTTTATTTAAATAATATAAAAACAAATTCAAAACAACAAAAATTAGATAAAAATGGA
Mi-HLH-6-LIKE-1     AATTTTTATTTAAATAATATAAAAACAAATTCAAAACAACAAAAATTAAATAAAAATGCA


Mh-HLH-6-LIKE       ---------------------------------------------ATGTTTGAATGTTCTTCT
Mf-HLH-6-LIKE       AAAGAAAATCAAAAAAATAATATATTTATAAAAATGCGTGAATGTTCTTCTTCTTCCTCT
Mi-HLH-6-LIKE-2     AAAGAAAATCAAAAAAA---TAATATTATATCAAAAAGAAAAAGAAAA------------
Que1-HLH-6-LIKE     AAAGAAAATCAAAAAAATAAGATTACTATATCAAAAAGAAAAAGAAAA------------
Mi-HLH-6-LIKE-1     AAAGAGAATCAAAAAAATAAGATTACTATATCAAAAAGAAAAAGAAAA------------
```

```
Mh-HLH-6-LIKE       TCTTCATCTTGGCTTTCTAATCAACAAAATAATTCTTCATATTCTTCATCATTTTCCCCT
Mf-HLH-6-LIKE       TCCCCATCTTGGCCTTCTAACCAACAAAATACTTCTTCTTCATATTCCTCATCATTTTCT
Mi-HLH-6-LIKE-2     ------------------------------------------------------------
Que1-HLH-6-LIKE     ------------------------------------------------------------
Mi-HLH-6-LIKE-1     ------------------------------------------------------------

Mh-HLH-6-LIKE       ATTAATTCTTCTAATTTATATCAATCAATATCTTTAAATAATAATTCTAACAATTTTTAT
Mf-HLH-6-LIKE       TCCTCTCCCTC------A------------------ATTAATTCTTCAAATTTAATTTAT
Mi-HLH-6-LIKE-2     ------------------------------------------------------------
Que1-HLH-6-LIKE     ------------------------------------------------------------
Mi-HLH-6-LIKE-1     ------------------------------------------------------------

Mh-HLH-6-LIKE       TTAAATAATACAAAAATAAATTTAAATTCAAAACAACAAAAATTAAATAAAAATGGAAAA
Mf-HLH-6-LIKE       TT--------------------CCCTTTAAACAATT-----GCCAATCAACTTCTTCAA
Mi-HLH-6-LIKE-2     ------------------------------------------------------------
Que1-HLH-6-LIKE     ------------------------------------------------------------
Mi-HLH-6-LIKE-1     ------------------------------------------------------------

Mh-HLH-6-LIKE       GAAAAACAAAAAAATATTTTATCAAAAAGAAAAAGAAATAATAAAAATAATTTAAATGTT
Mf-HLH-6-LIKE       TTAAAA-ATTCCAACAATTTTTATTTAAATATAAAAACAAATTCAAATCAACAAAAATTA
Mi-HLH-6-LIKE-2     ----------------------------------------AATCCAAATTTAATT---ATT
Que1-HLH-6-LIKE     ----------------------------------------AATCCAAATTTAATT---ATT
Mi-HLH-6-LIKE-1     ----------------------------------------AATCCAAATTTAATT---ATT
                                                            ***  ****         *

Mh-HLH-6-LIKE       GGAATTAAATTTGGAGGTAAAAAGCCACATCAAGTTGCTAGAAGAAATGAAAGAGAACGT
Mf-HLH-6-LIKE       AATAAAAATGGAAAAGAAAATCAAAAAAATAATATAGCTAGAAGAAATGAAAGGGAACGT
Mi-HLH-6-LIKE-2     GGAATTAAATTTGGGGGTAAAAAGCCCCATCAAGTAGCTAGAAGAAATGAAAGGGAACGT
Que1-HLH-6-LIKE     GGAATTAAATTTGGGGGTAAAAAGCCCCATCAAGTAGCTAGAAGAAATGAAAGGGAACGT
Mi-HLH-6-LIKE-1     GGAATTAAATTTGGGGGTAAAAAGCCCCATCAAGTAGCTAGAAGAAATGAAAGGGAACGT
                       *  **       *  **  *     ** *  * ***************** ******

Mh-HLH-6-LIKE       AAAAGAGTTCAACAAGTAAATGATGGATATGAAAAATTAGCAAATACTTTAAATAATTTC
Mf-HLH-6-LIKE       AAAAGAGTTCAGCAAGTAAATGATGGATATGAAAAATTGGCAAACACTTTAAATAATTTC
Mi-HLH-6-LIKE-2     AAAAGAGTTCAGCAAGTAAATGATGGATATGAAAAATTGGCAAACACTTTAAATAATTTC
Que1-HLH-6-LIKE     AAAAGAGTTCAGCAAGTAAATGATGGATATGAAAAATTGGCAAACACTTTAAATAATTTC
Mi-HLH-6-LIKE-1     AAAAGAGTTCAGCAAGTAAATGATGGATATGAAAAATTGGCAAACACTTTAAATAATTTC
                    *********** ************************* ***** ***************

Mh-HLH-6-LIKE       GAACCTATTTGCAATGAAAGAAAATTAACAAAAGCAGAAACATTAAAAACAGCAATTTTA
Mf-HLH-6-LIKE       GAGCCAATTTGCAATGAAAGAAAATTAACAAAAGCAGAAACATTAAAAACGGCCATTTTA
Mi-HLH-6-LIKE-2     GAGCCAATTTGCAATGAAAGAAAATTAACAAAAGCAGAAACATTAAAAACAGCCATTTTA
Que1-HLH-6-LIKE     GAGCCAATTTGCAATGAAAGAAAATTAACAaaagcagaaacattaaaaacagccaTTTTA
Mi-HLH-6-LIKE-1     GAGCCAATTTGCAATGAAAGAAAATTAACAAAAGCAGAAACATTAAAAACAGCCATTTTA
                    ** ** ****************************************** ** ******

Mh-HLH-6-LIKE       TATATTAAACATTTAGAAGAATTATTAAAACAACAACAACCTTTAAAAATAAATTTAGAA
Mf-HLH-6-LIKE       TATATTAAACATCTCGAAGATTTATTAAAACAGCAACCTTTGGA---------------A
Mi-HLH-6-LIKE-2     TATATTAAACATCTCGAAGATTTATTAAAACAGCAACCTTTAGA---------------A
Que1-HLH-6-LIKE     TATATTAAACACCCGGAGGATTTATTAAAACAACAACCTTTGGA---------------A
Mi-HLH-6-LIKE-1     TATATTAAACATCTCGAAGATTTATTAAAACAACAACCTTTGGA---------------A
                    ***********    ** ** *********** ****                     *

Mh-HLH-6-LIKE       AAACAAAATAAAATTAAAAATGAAAATAATATTGAAAATTCAAATAATTTTTATCAGAAT
Mf-HLH-6-LIKE       AAACAAAATAAAATTAAAAATGAAAATAATTCTGAAAATTCAACAAATTTTTGTCAGAAT
Mi-HLH-6-LIKE-2     AAACAAAATAAAATTAAAAATGAAAATAATTCCGAACATTCAACAAATTTTTGTCAGAAT
Que1-HLH-6-LIKE     AAACAAAATAAAATTAAAAATGAAAATAATTCTGAAAATTCAACAAATTTTTGTCAGAAT
Mi-HLH-6-LIKE-1     AAACAAAATAAAATTAAAAATGAAAATAATTCTGAAAATTCAACAAATTTTTGTCAGAAT
                    ****************************   *** ******  ******* *******

Mh-HLH-6-LIKE       TTAAATTCTTCTCAAGAAAATCAAAAAGAACAGCAATTCTTAACAAAACATTCACCAACT
Mf-HLH-6-LIKE       AATTTTGTAATTTAA---------------------------------------------
Mi-HLH-6-LIKE-2     TTAAATTTTCCCAAGAAAATCAAAAACAATTTTTAACCAAAAATTCAAATACTCCAAAC
Que1-HLH-6-LIKE     TTAAATATTTCCGAAGAAAATCAAAAACAATTTTTAACCAAGAATTCAAATACTCCAAAC
Mi-HLH-6-LIKE-1     TTAAATTTTTCCGAAGAAAATCACAAACAATTTTTAACCAAAAATTCAAATACTCCAAAC
                         *       **
```

```
Mh-HLH-6-LIKE        AATTTTATTTAT---------TCACAATCAATAAATACAAATTATAATAATACAAATAAT
Mf-HLH-6-LIKE        ------------------------------------------------------------
Mi-HLH-6-LIKE-2      TTTATTTGTTCACCACCACCACCAACAATTAA---TACAAATTATAATAATTTAAATAAT
Que1-HLH-6-LIKE      TTTATTTATTCACCACCACCACCAACAATAATTAATACAAATTATAATAATGTAAATAAT
Mi-HLH-6-LIKE-1      TTTATTTATTCACCACCACCACCACCAACAATTAATACAAATTATAATAATATAAATAAT


Mh-HLH-6-LIKE        AATAATATTTATTTAAATAATAATTCATTTCAAAATTGTTTATTTCCTTCTAAACAAAAA
Mf-HLH-6-LIKE        ------------------------------------------------------------
Mi-HLH-6-LIKE-2      AATAATATTTATTTAAATAATAATTCATTTCAAAATTGTTCATTTCCTTATATAAAACAA
Que1-HLH-6-LIKE      AATAATATTTATTTAAATAATAATAATAA------------------------------
Mi-HLH-6-LIKE-1      TATTACACTTAA-----------------------------------------------


Mh-HLH-6-LIKE        TAA---
Mf-HLH-6-LIKE        ------
Mi-HLH-6-LIKE-2      AAATAA
Que1-HLH-6-LIKE      ------
Mi-HLH-6-LIKE-1      ------
```

RKN Homologues	Mf-HLH-6-LIKE	Mi-HLH-6-LIKE-2	Mh-HLH-6-LIKE	Mi-HLH-6-LIKE-1
% ID with: Que1-HLH-6-LIKE	83.95	95.25	81.07	96.90

Appendix Figure 6.1 – Predicted cDNA of *Meloidogyne* genes A) HLH-6 B) PHA-4, and C) HLH-6-like after alignment to Que1 cDNA; aligned with Clustal Omega. Sequenced cDNA from *Meloidogyne* strain Que1 is not the complete gene, it includes as much as could be cloned at the time, but does include entire DBD region. ATGs are highlighted. %identity as determined by Clustal Omegas is shown below each alignment set.

A)

93.92% identity

```
Que1-15667      GGTTTAATTACCCAATTTGAGCAGAAATGTTTATTGAACCTACAATTATTGATAAATCCT
Minc15667       ----------------------------ATGTTTATTGAACCTACAATTATTGATAAATCCT
                                             *********************************

Que1-15667      TTCCGACAGAACATATACATTGTGTCAATTGTATTAAATTAAATTGTAAACAAGGTGATC
Minc15667       TTCCAACAGAACATGTACATTGTGTCATTTGTATTAAATTAAATTGTAAACAAGAAGATT
                **** ********* ************ *************************  ***

Que1-15667      TATGTCCACTTATTCATTGTAACTTCTGTGGTGTTCGTTTACATAAATGCAAATTGGAAG
Minc15667       TATGTCCGCTTATTCATTGTAACTTCTGTGGCGTTCGTTTACATAAATGCAAATTGGGAG
                ******* ********************** ************************ **

Que1-15667      ATCATATTGAACACATTTGTGGAAAGGCCCCCTGCACTTGCCCGAACACAATTTATGGCT
Minc15667       ATCATATTGAACACATTTGTGGAAAGGCCCCCTGCAGTTGTCCGAATACAATTTATGGCT
                ************************************ *** ***** **************

Que1-15667      GCAATGTTCGACTTCGTCGACAACTTATCGCTGAACATATTGCTTTCTGTTGTGTTTCTT
Minc15667       GCAATGTTCGACTTCGTCGACAACTTATTGCTGAACATATTGCTTTCTGTTGTGCTTCTT
                *************************** ************************ *****

Que1-15667      TAGTTTTTTGTCCATTTGTTCGCAATCGACAGTTTTACACTGCAAAAGCTAAAAAATTGT
Minc15667       TAGTTTTTTGTCCATTTGTTCGCAATCGACAGTTTTACTCTGCAAAAGCTAAAAAATTGT
                ************************************ ********************

Que1-15667      TAAAAAGAATTGCCAGAACTGGAGAAAAAATATTTCCCGAAAATGAGAAAACAATGGAAG
Minc15667       TAAAAAGAATTGCCAGAACTGGAGAAAAAATATTTCCTGAAAATGAGAAAACAATGGAAG
                ************************************ ********************

Que1-15667      AATTAAATGAACTTGAAGTTCCAGACATCGCAATTGCTTTAATTGATCAGAAAGTTCTAA
Minc15667       AATTAAATGAACTTGAAGTTCCAGACATTGCAATTGCTTTAATTGATCAGAAAATTCTAA
                *************************** ************************ ******
```

```
Que1-15667      CAAAATTCTTTCAAATTCCTCGTAAAAAACGTTCTACTCTTCGCC-TA--GCTATTGCA-
Minc15667       CAAAATTCTTCCAAATTCCACGTAAAAAGCGTTCTACTCTTCGCCCTACCGCAATTGCAA
                ********** ******** ******** ***************** **  *********

Que1-15667      ------------------------------------------------------------
Minc15667       TGGATTTCATCAAACAAAATTCTTTAGTAATAAATAAATATAAAAATCGGCAAAATGAAG

Que1-15667      ------------------------------------------------------------
Minc15667       TAGTAAATGAAGCTCCATTGGAAGACAAAAATTTTTACAAAAGAAAATTTAGTTTAAATG

Que1-15667      ------------------------------------------------------------
Minc15667       AACCAGATTCTGATTATTTTGATAGCAGTGATGAAGAAAAACAAGAAATTGAAAACAAAA

Que1-15667      ------------------------------------------------------------
Minc15667       ACAAAAAAATTCGTGAGCCTTTTGCTGGTTGTCGTTTATGCAAGTTAGATCCAGGAAGTC

Que1-15667      ------------------------------------------------------------
Minc15667       AACATTTACATCAATTGGGTTCAATGAAGGAAAACAAAGAAAATGGGGATTTTGAACAAA

Que1-15667      ------------------------------------------------------------
Minc15667       AGAAGAAAAATATTTTAAAATTATTTGATAATAATTCAATTTTGAGTTTTTTACCTGAAT

Que1-15667      ------------------------------------------------------------
Minc15667       TTTACCGGAAGAGAAGCTTGCTTGTCACTCCAAATATTCCATTTCTTAGCATTGTTTTGG

Que1-15667      ------------------------------------------------------------
Minc15667       GTGAAGAATATCTCAGCGTAAACTACCCAATTTTTGGTTGTTCTGCTTGCTATATCCCCC

Que1-15667      ------------------------------------------------------------
Minc15667       GTAGCCAATTTAACTCGCACTATGCTTCACACAATATTCTGGATGAATCTACGCCTTATA

Que1-15667      ------------------------------------------------------------
Minc15667       ACTTTATTGCAAGATGTCCGAACTGGCAGAGGGGATATAATTCTTTCTTGTCTGCGTGTG

Que1-15667      ------------------------------------------------------------
Minc15667       TATATGAACCGATGCCTAATTTGCCTATTTTGATGCGCGTGCCAACAGCCTCTTCCTCGG

Que1-15667      ------------------------------------------------------------
Minc15667       GTCTCTTGCCGAAATGCCGAAATGCCTTCACATTCCTTGCCGAAATGCCTTCACATTTTT

Que1-15667      ------------------------------------------------------------
Minc15667       ATTTGCTCATTGAGAGATGGATGGATGATGCCTCTCTTCGCATATTATCTGTTACTTGTC

Que1-15667      ------------------------------------------------------------
Minc15667       GTCGAATGTATCAATTATTACCTAAATTATTACCCCATCGTTTATGTGTTGAAATAAAAT

Que1-15667      ------------------------------------------------------------
Minc15667       GGAAAAAGGAACGAGTTGTTGGAGATGGATTTGATGGCAATAGATTTGTTGAAGACGGAT

Que1-15667      ------------------------------------------------------------
Minc15667       TTTACGAGAGTTTCTCAGCTGTCCAAGGCTCAATACCTTCACTAATTGTTCAACCTCAAG

Que1-15667      ------------------------------------------------------------
Minc15667       GACCAATAATTGATCATTTACAAAATTGTGACTATCGAGACGTTTATGACTTGGGAATTG
```

```
Que1-15667      --------------------------------------------------
Minc15667       AACCGATTGAATTGAAAAAAGTACACGGATCATTTGTCTCGGCCAATTAA
```

B)

98.37% ID

```
Que1-pig-1      GGTTTAATTACCCAAGTTTAAGACTAATTTTTTGGTAGATTTAATATGCCAAAAGGTGAA
Minc16584       ------------------------------------------ATGCCAAAAGGTGAA
                                                           ***************

Que1-pig-1      AACATGGATTGTGGTGCTAATCTCATCTGTTCTAATGACATTCCTTCAATACAAACAGAA
Minc16584       AACATGGATTGTGGTGCTAATCTCATCTGTTCTAATGACATTCCTTCAATACAAACAGAA
                ************************************************************

Que1-pig-1      TCTAGACAACGCCAAATACTTTCTAGTAAGAGCAATAATTATTATGAATGTGAAAATTAT
Minc16584       TCTAGACAACGCCAAATACTTTCTAATAAGAGCAATAATTATTATGAATGTGAAAATTAT
                ************************.***********************************

Que1-pig-1      TCTTCCGATCAGAATTGTATGGAAAGAATGTACCTCTTTTTGGATGAATTAGGTTCTGGT
Minc16584       TCTTCCGATCAGAATTGTATGGAAAGAATGTACCTCTTTTTGGATGAATTAGGTTCTGGT
                ************************************************************

Que1-pig-1      GGTTTCGGGAAAGTTAAATTGGCAAAGCATATCCTCACGGGGGACCAAGTTGCGATTAAA
Minc16584       GGTTTCGGGAAAGTTAAATTGGCAAAGCATATCCTCACGGGGGACCAAGTTGCGATTAAA
                ************************************************************

Que1-pig-1      ATTATTGATAAGAAGAGCATTCCTAATGATTTACCGCGAGTATTCAGCGAGATGGAAGCC
Minc16584       ATTATTGATAAGAAGAGCATTCCTAATGATTTACCGCGAGTATTCAGCGAGATGGAAGCC
                ************************************************************

Que1-pig-1      TTAAAGCTGTTGGCGCACCAAAATATTTGTCGTCTATTTCAGTTCCGCGAAACGGAAGAC
Minc16584       TTAAAGCTGTTGGCGCACCAAAATATTTGTCGTCTATTTCAGTTCCGCGAAACGGAAGAC
                ************************************************************

Que1-pig-1      AGGTTTTATATTGTTATGGAGTATTGCAATGGTGGAGATATGTTCGAAAATGGA------
Minc16584       AGGTTTTATATTGTTATGGAGTATTGCAATGGTGGAGAGATGTTCGATTATATTGTTCGC
                ************************************ ********::**. :

Que1-pig-1      ------------------------------------------------------------
Minc16584       AAGGAGCGGCTTGGAGAATCTGAGGCACGCCATTTTTTTAGACAGAGTGGCTTACATTCA

Que1-pig-1      ------------------------------------------------------------
Minc16584       TTCAATGGGTTTGTTGTGATTTTAATGTTTGGTTTAAAAAGTTTAGGTTTTGCTCATCGT

Que1-pig-1      ------------------------------------------------------------
Minc16584       GATTTAAAACCTGAGAATTTGTTGTTGACACAAGAACTCCAATTAAAGGTGATTGATTTC

Que1-pig-1      ------------------------------------------------------------
Minc16584       GGGCTTTGTGCAAGACCTTTAAATGGTTTGACTAGGCCACTTGAAACTTGTTGTGGTTCC

Que1-pig-1      ------------------------------------------------------------
Minc16584       CCTGCCTATGCTGCGCCCGAACTTATTCAGAATCAATCCTATTTTGGGAACGAAGCCGAC

Que1-pig-1      ------------------------------------------------------------
Minc16584       ATTTGGTCAATGGGTGTTTTGTTGTATGCATTGCTTTGTGATGACACAAATATGCCTCGA

Que1-pig-1      ------------------------------------------------------------
Minc16584       TTGTACAAACAAATTACGCTGGGACAATATTTCGAACCTGATTTCTTGTCACAATCTTGC

Que1-pig-1      ------------------------------------------------------------
Minc16584       AAAGATTTATTACGTTCAATGCTAACTGTTGACCCAAAGAAACGTGCAACTATTCAACAA
```

```
Que1-pig-1      ------------------------------------------------------------
Minc16584       ATACTTGTCCATCCATGGCTTAATTCCAATTATGAACAACAAATTAAATGGCAAAGTATT

Que1-pig-1      ------------------------------------------------------------
Minc16584       TATAATCGAGATATTGTGGATGAAGAAATTATTAGAGAATTAGTTTGCTTTTTTAAAATA

Que1-pig-1      ------------------------------------------------------------
Minc16584       AATACCGAGAATATGATTGCTCGAATAAAACAGTGGCGCTTTGACTATATTTGTGCAACC

Que1-pig-1      ------------------------------------------------------------
Minc16584       TATCTTATTCTTTTACAAAGAAAGGAACAAAAAAAGAATTTTGCACTCCCTTTCTATGAT

Que1-pig-1      ------------------------------------------------------------
Minc16584       AAGAACTTCAACTGCAATTTTCCTTGCTCGCCAACTATTCACTCTTCATTGGAAAATGAT

Que1-pig-1      ------------------------------------------------------------
Minc16584       TTGGATTTATCTGGCTTTGAAGATATTATTAAATTCTCTAATCCCGATAACGCAAATGAA

Que1-pig-1      ------------------------------------------------------------
Minc16584       ATTATGACAGAGTCCTTGGCTCTACTTTCTCCACCTGTCTTCACAAATACAAAACAAATG

Que1-pig-1      ------------------------------------------------------------
Minc16584       ATCTTACCTCCAATTCCTTCCACATTCTCTCCGTATGAAGGTGGTAAACAAGGCCAAAGA

Que1-pig-1      ------------------------------------------------------------
Minc16584       CTTGACGCTAATTCGATTACACCTCGTAAACCACAGCAACGTAAACCTTCTTATGAGGCA

Que1-pig-1      ------------------------------------------------------------
Minc16584       CCCACACATCATAGACCTGCACCTTCGAACAAAGGTGGTTGTAGTGTTTATACAACACCT

Que1-pig-1      ------------------------------------------------------------
Minc16584       AGAAGAGGCACTGTTGAATTACCTGGACCAATTGTAGGAGGATTTGCTTATCGCCGTCCG

Que1-pig-1      ------------------------------------------------------------
Minc16584       AATTCTGTGGATAGACCACCTTCCTTGAGATTAGAGAAGAACGATCCTGCTGATGAAAAT

Que1-pig-1      ------------------------------------------------------------
Minc16584       TCTCCACCAAGTCTTCATTTTAACAACAATACCCAAAAGACTGAAAGTTGTGAAAGAAGA

Que1-pig-1      ------------------------------------------------------------
Minc16584       TCGCGTATTGATATGTGCCGAGCTTATAACAATTGTAATAACAATATTAAACAGTTACAA

Que1-pig-1      ------------------------------------------------------------
Minc16584       AATAACAAATCACCGCATCGGTTCCGCCAACGCGTGTTTTCGTCATTAGAACGAAAAGCT

Que1-pig-1      ------------------------------------------------------------
Minc16584       GATAAAATGATTAGTTTGCTAACTCCAAAGAAGATTAAATCAGAAGCTCCAACTAAACTT

Que1-pig-1      ------------------------------------------------------------
Minc16584       AAATGTACAGAGCCAATGGCTAACGTTTCGGTTACATCGTCTAGTGATCCAGTAAAAGTG

Que1-pig-1      ------------------------------------------------------------
Minc16584       CGTGAGGAATTGTGGAAAGTACTTTGCAATTTGGGAATGAATGCTACACAGAATGGATGG
```

```
Que1-pig-1      ------------------------------------------------------------
Minc16584       AAAGTTAGTGGATGTCGAAAGAATTCAAATAACCCGAAAAAAGATGCCACTGTCGAACTT

Que1-pig-1      ------------------------------------------------------------
Minc16584       GAAGTTGTTTCTATTGAGGGTTTGCAATATGTTGGTGTGAAACGCAGAAGAGTTCATGGT

Que1-pig-1      ------------------------------------------------------------
Minc16584       GAACCATTCACCTATAAAAGAATTTGCGAAGAATTTGTAGCATTTGCCACAGCAGACTCA

Que1-pig-1      ------------------------------------------------------------
Minc16584       AAATCAATGGAAAATGATGACGGATTGTATGTTGGCCAACTTGCTCGTATCATGTTGAAC

Que1-pig-1      --------------------------------------------
Minc16584       ATCCTTTCATACGTTGCCTGTGTTTTACAAGTAAGTAGAGCATGA
```

C)

95.48% ID

```
Que-1-ubq-2-2       GGTTTAATTACCCAAGTTTAAGCTCATTTAAACAATCTACGGA-AGATTAGCATTAAATA
Minc16562a          ---------------CAAAATGCTCATTTAAACAATCTACGGAAGATCAAGCATTAAATA
                                   * **********************      ***********

Que-1-ubq-2-2       TGCAGATTTTTGTAAAGACTCTCACCGGAAAAACTATTACTCTCGAGGTTGAGGCTTCTG
Minc16562a          TGCAGATTTTTGTAAAGACTCTCACCGGAAAAACTATTACTCTCGAGGTTGAGGCTTCTG
                    ************************************************************

Que-1-ubq-2-2       ATACCATTGAGAATGTTAAGGCAAAAATTCAAGATAAAGAGGGTATCCCGCCTGATCAAC
Minc16562a          ATACCATTGAGAATGTTAAGGCAAAAATTCAAGATAAAGAGGGTATCCCGCCTGATCAAC
                    ************************************************************

Que-1-ubq-2-2       AGCGTTTGATCTTTGCTGGCAAGCAACTTGAAGATGGACGAACCTTGGCTGATTATAACA
Minc16562a          AGCGTTTGATCTTTGCTGGTAAGCAACTTGAAGATGGACGAACCTTGGCTGATTATAACA
                    ****************** ****************************************

Que-1-ubq-2-2       TCCAAAAGGAGTCTACACTTCACTTAGTTTTACGTCTTCGTGGTGGAAAGGTTCACGGTT
Minc16562a          TCCAAAAGGAGTCTACACTTCACTTAGTTTTACGTCTTCGTGGTGGAAAGGTTCACGGTT
                    ************************************************************

Que-1-ubq-2-2       CATTGGCTCGTGCTGGAAAGGTTCGTGCTCAAACTCCTAAGGTCGAAAAGCAGAAACATT
Minc16562a          CATTGGCTCGTGCTGGAAAGGTTCGTGCTCAAACTCCTAAGGTCGAAAAGCAGGAACATA
                    ************************************************** *****

Que-1-ubq-2-2       AAAAACAGCA----------------------------------------------------
Minc16562a          AGAAAAAGAAGCGCGGCCGTGCTTTCCGTCGCATTCAATATAACCGTCGCTTCACCAATG
                    * *** ** *

Que-1-ubq-2-2       ------------------------------------------------------------
Minc16562a          TTGCTACTTCTGGGGCGGGACGCCGTCGTGGCCCTAACTCCAACGCTGCATAAGAGAATG

Que-1-ubq-2-2       ------------------------------------------------------------
Minc16562a          GTCGTATCTTGATGAATGTATGGTGATATAATCAATTTAATACATTCGACTTTATGAAGT

Que-1-ubq-2-2       --------------------------------------------
Minc16562a          TTTCTGTTATTCAAGATAAATCTTTTTGTTGATATATTTGTT
```

Figure 6.2 – Non-specific genes cloned from *Meloidogyne* strain Que1 using a Splice Leader as the forward primer. These genes were accidentally cloned and sequenced but do provide genetic information on the strain and species. The genes are aligned to their closest available homologue, all of which belong to *M. incognita*. *M. incognita* gene (*C. elegans* homologue): A) *Minc15667* (no *C. elegans* homologue), B) *Minc16584* (*pig-1*), C) *Minc16562* (*ubq-2*). Primers are underlined, ATGs are shaded.