Comparative genomics of nematodes: *Caenorhabditis elegans* as a tool to study the *Haemonchus contortus* genome

Gary Ian Saunders

UNIVERSITY
of
GLASGOW

For the degree of

DOCTOR OF PHILOSOPHY

Author's declaration

The work presented in this thesis is entirely my own and carried out with the help of those people mentioned in the acknowledgements. It has not been previously submitted to any university for the award of a degree. The following publications include work contained in this thesis –

Geldhof, P., Visser, A., Clark, D., Saunders, G., Britton, C., Gilleard, J., Berriman, M., & Knox, D. 2007, "RNA interference in parasitic helminths: current situation, potential pitfalls and future prospects", *Parasitology*, vol. 134, no. Pt 5, pp. 609-619.

Redman, E., Grillo, V., Saunders, G., Packard, E., Jackson, F., Berriman, M., & Gilleard, J. S. 2008, "Genetics of mating and sex determination in the parasitic nematode Haemonchus contortus", *Genetics*, vol. 180, no. 4, pp. 1877-1887.

Gary I. Saunders

August, 2009

Acknowledgements

I am extremely grateful for the unfaltering support and encouragement of my supervisors Professor J.S Gilleard and Dr C. Britton. Sincere thanks to Ms Gill McCormack and Ms. Roz Laing for their help with the generation of transgenic *C. elegans* lines and *H. contortus* RT-PCR data, respectively. Thanks also to Dr W. Weir for helpful advice with my numerous computer issues. A massive thanks to all members of the department, past and present, for their help, advice, humour and patience - it really has been a great place to work.

I would also like to thank Dr R.N Beech for introducing me to the world of bioinformatics, and Dr M. Berriman and his team for helpful advice when required.

Finally, I would like to thank all friends and family who have supported me, and my acknowledgments to the University of Glasgow for the financial support enabling this PhD project.

Abstract

The genome of the Trichostrongylid nematode parasite of small ruminants *Haemonchus contortus* is being sequenced at the Pathogen Sequencing Unit of the Wellcome Trust Sanger Institute, Cambridge, UK. Currently, in excess of 800 Mb of genomic sequence is available for this on-going project (http://www.sanger.ac.uk/Projects/H_contortus/). Once available, the fully sequenced and assembled genome of *H. contortus* will be an extremely valuable resource for both novel drug discovery and biological research into this important pathogen.

H. contortus resides in the same Clade (Clade V) of the phylum Nematoda as the free-living model organism *Caenorhabditis elegans*. Therefore, it is ideally placed to extrapolate the wealth of genomic and biological data available for *C. elegans*. The extent to which such data can be applied to parasitic nematode research was a major focus of this project. I have concentrated on two well documented and important gene classes: those comprising the β-tubulin gene family and those of the RNA-interference (RNAi) pathway.

Control methods for *H. contortus* are becoming increasingly restricted due to the rise in resistance to current anthelmintic drugs. Benzimidazoles (BZ) are a class of anthelmintic to which there is widespread resistance. Mutations and deletions in both of the β-tubulin genes previously identified from *H. contortus*, *isotypes-1* and *2*, have been shown to correlate with BZ resistance. I have identified an additional two β-tubulin loci within the *H. contortus* genome, which now gives a total of four genes for this family. Using *C. elegans* as a surrogate expression system together with antibody immunolocalisation in *H. contortus* I have investigated the expression pattern of three of these *H. contortus* β-tubulin genes and encoded proteins, and compared these with those of the *C. elegans* β-tubulin gene family. In addition, I have characterised the phylogenetic relationships of all available Trichostrongylid β-tubulin polypeptide sequences. This has allowed the determination of the evolution of this gene family, and possible association of *isotypes-1* and *2* with BZ resistance, across these nematodes.

RNAi is a well established technique used in *C. elegans* to silence gene expression using double stranded RNA (dsRNA). However, RNAi is far less effective and repeatable in parasitic nematode species. Using gene searching techniques I have examined whether genes required for RNAi in *C. elegans* are present and conserved in the *H. contortus* genome. Although I identified putative homologues of Dicer (*dcr-1*) and several other RNAi genes, no sequence homologous to the *C. elegans rde-4* gene could be found. This gene is essential for the generation of small inhibitory RNAs (siRNAs) in the *C. elegans* RNAi pathway. Furthermore, no *H. contortus* genomic sequence encoding a homologue of Ce-SID-2 was identified. SID-2 is essential for dsRNA uptake from the environment, and sequence differences between *C. elegans* and *C. briggsae* SID-2 are responsible for the lack of environmental RNAi in the latter. I have also searched the available genomic sequence databases of *Pristionchus pacificus* and *Brugia malayi* for RNAi pathway genes and concluded that components of the RNAi pathway may not be conserved across the phylum Nematoda, although full genome sequences will be required to confirm these findings.

Abbreviations and Symbols

°C	Degrees Celsius
A	Adenine
Ab	Antibody
Amp	Ampicillin
β	Beta
BAC	Bacterial Artificial Chromosome
BLAST	Basic Local Alignment Search Tool
bp	Base pair
BSA	Bovine Serum Albumin
BZ	Benzimidazole
C	Cystosine
cDNA	Complementary DNA
Contigs	Contiguous sequences
C-terminus	Carboxyl terminus of a protein
DNA	Deoxyribonucleic acid
dNTPs	Deoxyribonucleic triphosphates
ds	Double stranded
EDTA	Ethylenediamine tetra acetic acid
e.g.	Example given
EST	Expressed Sequence Tag
Ex	Exsheathed
FACS	Fluorescence activated sorting
FITC	Fluorescein isothiocyanate
g	Gram
G	Guanine
Gb	Gigabase (one billion bases)
Hsp90	eat shock protein 90
IFA	indirect antibody fluorescent
IVR	Ivermectin
Kb	kilobase (one thousand bases)
l	Litre(s)
LEV	Levamisole
Mb	Mega (million) base pairs
ml	Millilitre(s)
min	Minute(s)
mm	Millimeters
mM	Millimolar
mRNA	Messenger RNA
mtDNA	Mitochondrial DNA
MW	Molecular weight
Mya	Million years ago
PCR	polymerase chain reaction
PCR-RFLP	PCR - restriction fragment length polymorphism
rDNA	ribosomal DNA
RNA	Ribonucleic acid0
RNAi	RNA interference
rRNA	ribosomal RNA
RT	Room Temperature

RT-PCR	Reverse Transcription PCR
s	Second(s)
SDS	Sodium Dodecyl Sulphate
spp	Species
T	Thymine
Taq	*Thermus aquaticus* polymerase
USA	United States of America
µg	Microgrammes
µl	Microlitres
µM	Micromolars

Table of contents

List of tables

List of figures

Chapter 1: General introduction

1.1 The parasitic helminth *Haemonchus contortus*

1.1.1 Host range and life cycle

H. contortus worms are 2-3 cm at the adult stage, and are most readily identified by the white ovaries winding spirally around the blood-filled intestine which results in a 'barbers pole' appearance of fresh specimens. In both sexes there are cervical papillae and a tiny lancet within the buccal capsule. Males and females are able to be distinguished as the male has an asymmetrical dorsal lobe and barbed spicules, whereas the females have a vulval flap. *H. contortus* are known to parasitize predominantly sheep and goats. However, infection of cattle, deer, camel and llama has been reported (Taylor *et al.*, 2007).

The life cycle of *H. contortus* is direct (Figure 1.1; Urquhart *et al.*, 1996). Female worms are prolific egg layers. Eggs hatch on the pasture and develop through stages L1 to L3 in as little as five days, although this can be delayed for weeks or even months in cool conditions. L3 larvae are ingested and exsheath in the abomasum where they molt twice to adults. It is just before the final molt of the life cycle that the worms develop the piercing lancet. At this final molt and during adult life stages the worms move freely to the mucosa and feed on host blood. The prepatent period of *H. contortus* infection is 2-3 weeks in sheep and four weeks in cattle.

1.1.2 Veterinary importance

Each feeding *H. contortus* worm is able to remove 0.05 ml of blood per day, and burdens of up to 50,000 worms can occur in one host. Such hyper-acute haemonchosis can result in sudden death of apparently healthy sheep due to severe haemorrhagic gastritis. Typical infections of ~5,000 worms result in host anaemia approximately two weeks post infection. This is characterised by a progressive and steady fall in packed red blood cell volume. Continual loss of iron and protein results in marrow exhaustion and the haematocrit falls even further prior to death.

The parasite is optimally adapted to warm climates and is predominantly a disease of small ruminants in the tropical and sub-tropical regions of the globe. However, *H. contortus* is able to adapt to cooler environments and has been reported as far north as the Arctic Circle (Lindqvist *et al*., 2001).

Control strategy varies depending on global positioning due to variations on the duration and number of periods of heavy, prolonged rainfall in conjunction with high temperatures. At such times anthelmintic treatments are advised at 2-4 weekly intervals for the majority of the year. However, such high levels of anthelmintic dosing has led to extremely high selection pressures on *H. contortus* populations resulting in wide-spread resistance to all classes of anthelmintic drug (Bartley *et al*., 2003; Sargison *et al*., 2001; SAC report, 2000 and Jackson *et al*., 1992).

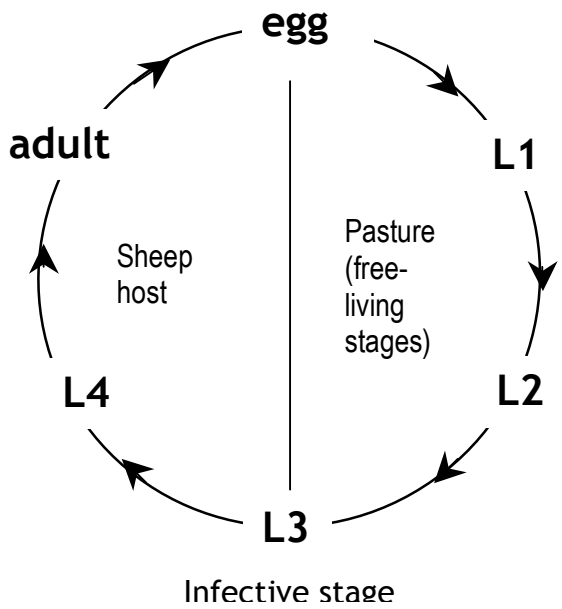


Figure 1.1 *H. contortus* life cycle
Eggs are passed out in faeces which hatch on the pasture and worms develop through stages L1 to L3. Ingestion of L3 infective *H. contortus* larvae allows the parasite to enter the parasitic stages of the life cycle.

(Adapted from Couthier, 2004)

1.2 Anthelmintic resistance

Resistance in parasitic nematodes to all three classes of anthelmintic drugs (Benzimidazoles (BZ), Levamisole (LEV) and Macrocyclic Lactones (ML)) is widespread, and continually rising. The resistance problem is most severe in trichostrongylid nematodes of sheep and goats (Table 1.1; Gilleard *et al*., 2006). However, the situation is progressively worsening in nematodes infecting cattle and, worryingly, evidence is accumulating that resistance is beginning to emerge in human infective parasitic nematode populations (Prichard, 2007). Extensive research is on-going to uncover the mechanisms underpinning such resistances. This section shall focus primarily on the current understanding of the genetic mechanisms of anthelmintic resistance in parasitic nematodes.

Table 1.1: World-wide prevalence of anthelmintic resistance as indicated by the percentage of sheep farms tested

Country	Benzimidazoles	Levamisoles	Macrocyclic Lactones
Australia	90 %	80 %	60 %
South Africa	79 %	73 %	73 %
Argentina	40 %	22 %	6 %
Brazil	90 %	84 %	13 %
Paraguay	73 %	68 %	73 %
Uruguay	80 %	71 %	1 %
Scotland	64 %	8 %	2 %

(Bartley *et al*., 2003; Sargison *et al*., 2001; SAC report, 2000 and Jackson *et al*., 1992)

1.2.1 Benzimidazole resistance

Resistance in parasitic nematodes to BZ is the most extensively investigated of all the three anthelmintic classes (Von Samson-Himmelstjerna, 2007). Consequently, the genetic mechanism underpinning this phenotype is best understood for this drug class. BZ drugs are known to exert a lethal effect by

interfering with tubulin polymerisation which in turn leads to a disruption in microtubule formation and ultimately the loss of cellular integrity in the nematode gut (Kohler, 2001), through inhibition of secretory vesicle transport (Jasmer *et al.*, 2000). Specific, known mutations within β-tubulin polypeptide sequences are able to confer a BZ resistance phenotype to fungi (Fujimura *et al.*, 1992; Jung *et al.*, 1992). In addition, the mapping of 28 different mutations which conferred BZ resistance to the free-living genetic model organism *Caenorhabditis elegans* all mapped to one locus of the genome, that of the β-tubulin *ben-1* gene (Driscoll *et al.*, 1989). This data provided the suggestion that similar mutations in β-tubulin genes may be responsible for conferring BZ resistance in parasitic nematode species. Roos *et al.* (1990) were able to show a reduced (in terms of allele numbers) restriction fragment length polymorphism (RFLP) pattern for the *isotype-1* locus within *H. contortus* populations resistant to BZ drugs, when compared to that from a BZ susceptible populations. Furthermore, similar results were later published using the parasitic nematode *Trichostrongylus colubriformis* (Grant and Mascord, 1996). From sequence analysis of this data, the only polymorphism that consistently differed between the susceptible and resistant populations was a substitution at P200 of the Isotype-1 polypeptide from a phenylalanine to tyrosine residue (Kwa *et al.*, 1993). This data provided strong evidence linking this specific mutation to the BZ resistant phenotype in parasitic nematodes.

Elegant use of *C. elegans* as a heterologous system has unequivocally highlighted the importance of the single amino acid substitution at P200 of the protein encoded by the *H. contortus* β-tubulin *isotype-1* gene in conferring BZ resistance in transgenic *C. elegans* worms (Kwa *et al.*, 1995). Transgenic expression of *H. contortus isotype-1* alleles encoding a phenylalanine residue at position 200 of the polypeptide increased the susceptibility of BZ resistant *C. elegans* worms to the drug, whereas alleles which encoded a tyrosine at this residue had no effect on the worms. This unequivocally demonstrated the significance of this point mutation to the BZ resistance phenotype. Due to the relative simplicity of the mechanism that leads to BZ resistance, PCR techniques have been developed to detect the frequency of resistant alleles in *H. contortus* populations (Von Samson-Himmelstjerna *et al.*, 2007).

However, the BZ resistance phenotype has also been linked with different mutations leading to single amino acid substitutions within the *H. contortus* ISO-1 protein (Von Samson-Himmelstjerna, 2007). In addition, the deletion of a second β-tubulin locus, *H. contortus* β-tubulin *isotype-2*, has also been correlated with a heightened resistance to BZ drugs (Kwa *et al.*, 1993; Beech *et al.*, 1994). Furthermore, recent investigation highlighted a significant increase in the frequency of one allele of a P-glycoprotein membrane transport encoding locus in BZ resistant strains of *H. contortus* when compared with susceptible (Blackhall *et al.*, 2008).

Therefore, although the parasitic nematode BZ resistance mechanism has been well studied, it appears that our understanding is still incomplete. Although found to be responsible for conferring BZ resistance in fungi and in *C. elegans*, parasitic nematodes are the only species for which mutations in more than one β-tubulin locus have been implicated in the resistance mechanism (Von Samson-Himmelstjerna, 2007).

1.2.2 Levamisole and Macrocyclic lactone resistance

Our understanding of the mechanisms underpinning both LEV and ML resistance in parasitic nematodes lags far behind that of BZ. Polymorphisms in a number of different genes have been linked with resistant phenotypes but, importantly, the functional significance has not been investigated via either directly *in vivo* or, *in vitro* techniques or by the use of heterologous expression systems. In the case of LEV resistance, mutations in cholinergic receptors have been linked with the phenotype (Neveu *et al.*, 2007). For ML the mechanism appears more complex, and has so far been hypothesised to potentially involve a glutamate-gated chloride, a γ-aminobutyric acid channel, a P-glycoprotein and a putative amino-acid gated anion channel subunit HG1 (Sangster *et al.*, 1999; Blackhall *et al.*, 1998; Blackhall *et al.*, 2003; Prichard, 2001). Interestingly, mutations within the *H. contortus* β-tubulin *isotype-1* gene have also been linked with ML resistance (Eng *et al.*, 2006). Again, it is important to stress that the functional significance of such mutations has not been reported. However the findings of this project may further progress understanding of these resistance mechanisms.

1.3 Phylum Nematoda

1.3.1 Overview

Nematodes are the most abundant genus on earth. The phylum Nematoda contains a diverse range of free-living and parasitic species which inhabit a huge variety of ecological niches. In the late 1990s Mark Blaxter and colleagues revolutionised the traditional view of the phylum with the construction of a phylogenetic tree based on 18S ribosomal RNA sequence (Blaxter *et al*., 1998). This tree is vastly different from those based on morphological characterisations of nematode organisms, and although the classification of organisms remains an issue surrounded by much debate, the Blaxter classification is the most widely accepted (Figures 1.2 and 3.1; Gilleard, 2005). It is this classification of the phylum Nematoda that will be considered when discussing all evolutionary aspects of comparative genomic investigations in this thesis. This tree disperses plant and animal parasitic species along with free-living nematodes within four of the five distinct Clades (Clade III contains only animal parasites). This phylogeny therefore suggests that parasitism has emerged independently a number of times over evolutionary time scales.

1.3.2 Classification of Haemonchus contortus

H. contortus branches within Clade V of nematode phylogeny within the order Strongylida and the superfamily Trichostrongyloidae (Figures 1.2 and 3.1; Blaxter *et al*., 1998). Trichostrongylid nematodes are common and exceptionally pathogenic in grazing ruminants, and other domesticated animal species can also act as hosts. The abomasum and small intestine are the location of parasitic stages for the majority of the trichostrongylid species.

1.3.3 Classification of Caenorhabditis elegans

The genetic model organism *C. elegans* also branches in Clade V of the phylum Nematoda, within the order Rhabditida and superfamily Rhabditoidea (Figures 1.2 and 3.1; Blaxter *et al*., 1998). As both *H. contortus* and *C. elegans* reside within the same Clade of nematode phylogeny it can be reasonably hypothesised

that they are more similar to one another at the molecular level than when comparing *C. elegans* with other species outwith this Clade. This makes *H. contortus* well placed to exploit the wealth of data available for this model organism. The extent to which *C. elegans* can be exploited as a tool to investigate *H. contortus* is a question of major importance to parasitologists and evolutionary biologists alike.

1.4 Genomic resources for the phylum Nematoda

1.4.1 Free-living species

The free-living Clade V nematode species *C. elegans* was the first eukaryotic, multi-cellular organism to have its genome fully sequenced (*C. elegans* sequencing consortium, 1998). *C. elegans* has a genome size of ~100 Mb, predicted to encode 19,000 proteins. There are now hundreds of completed genome projects for a variety of eukaryotic species. However, as a whole, the phylum Nematoda currently lacks sufficient genomic resources required to investigate this highly diverse range of organisms.

Since the completion of the *C. elegans* genome project, this species has been the mainstay of nematode investigation. *C. elegans* is the organism of choice for many research groups worldwide investigating such diverse topics as aging, apoptosis, development, cell signalling and drug metabolism to name a few. A major branch of nematode genomic research has focussed on the sequencing of the genomes of other species for the *Caenorhabditis* genus (Sternberg *et al.*, 2003). Since the initiation of this project, the complete genome sequence of *C. briggsae* has been released (Stein *et al.*, 2003), and currently the genome sequencing projects of C. *remanei*, *C. japonica* and *C. brenneri* are each at 9, 6 and 9 X coverage, respectively (Sternberg *et al.*, 2003; http://genomeold.wustl.edu/genome_group.cgi?GROUP=6).

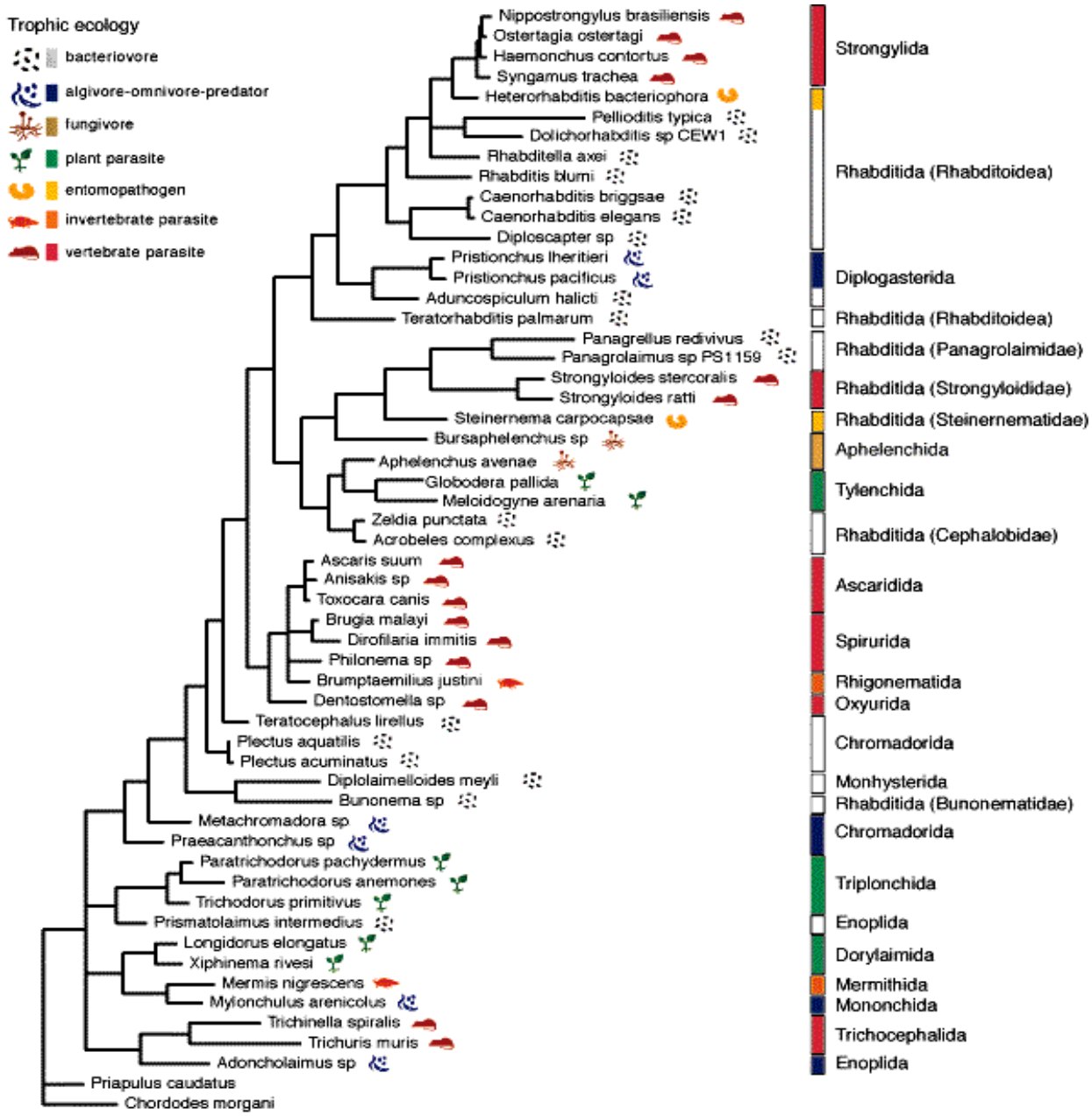


Figure 1.2 Phylogenetic classification of 53 species of nematode based on 18S RNA sequence

This Maximum Parsimony phylogenetic tree is based on pair-wise alignments of 18S RNA sequence from 53 nematode species. Branch lengths are proportional to the number of nucleotide changes inferred.

(Blaxter *et al.*, 1998)

As stated, *C. elegans* and *C. briggsae* are the only two free-living nematode species with a completed genome sequence available (*C. elegans* sequencing consortium, 1998; Stein *et al.*, 2003). Investigations of the conservation of biological processes between these species has relied heavily on comparative genomic investigations, as it can be reasonably hypothesised that if gene families have been conserved across species, the biological processes that the gene products are involved in are also likely to be conserved. Based on such analysis of 18,808 *C. elegans* and 19,507 *C. briggsae* genes, 12,155 show a one-to-one orthologous relationship. In addition, when the total number of proteins from each species were analysed for specific domains, very similar percentages of each dataset were given identical gene ontology annotations (Stein, *et al.*, 2003). This is indicative of a high level of conservation of biological processes between these species.

Importantly, the median percentage identity between protein products of *C. briggsae* and *C. elegans* 1:1 orthologous genes is 80 %. This is very similar to the 78.5 % median identity of mouse/human orthologues and is indicative that cross genus orthologue identification is possible for nematode comparative genomics. Further to this, the *C. elegans* and *C. briggsae* genes are relatively similar in terms of structure and overall length (Stein *et al.*, 2003).

In addition to orthologue conservation and gene structure, the extent to which genome organisation is conserved across the *Caenorhabditis* genus has important relevance to nematode comparative genomic investigations. There are almost 5,000 blocks of syntenic sequence, which cover 84.6 and 80.8 % of the *C. elegans* and *C. briggsae* genomes respectively. Rearrangement events of each genome are predicted to have been common, with an estimated 11,461 genomic breakpoint events during the *C. elegans* and *C. briggsae* divergence. However, interestingly, when the positions of only the 1:1 orthologous genes are considered, 95 % remain on the same autosome in the genome of the respective species (Hillier *et al.*, 2007). This is indicative of a bias toward intrachromosomal rearrangement events, as apposed to interchromosomal rearrangements, something which contrasts with similar investigation of mammalian genome conservation (Mouse Genome Sequencing Consortium, 2002).

Around 15 % of all *C. elegans* genes are arranged into transpliced operon structures. This gives a total of roughly 1,000 operons for this species. It is postulated that 96 % of these operons are conserved in the *C. briggsae* genome (Hillier *et al.*, 2007). The extent to which these structures are conserved in other species of the *Caenorhabditis* genus remains unreported, as the genome projects of the other three *Caenorhabditis* species are not at a stage which would permit such investigations. Interestingly, analysis of the current sequence of the most advanced of these projects, C. *remanei*, show that it is virtually impossible to align any common intergenic or intronic sequences of the three nematode species *C. elegans*, *C. briggsae* or C. *remanei* to one another (Hillier *et al.*, 2007). Again this is in stark contrast with the mammalian comparative genomics story. This informs us of two crucial things:

1. The genomes are extremely divergent for those assigned to a single genus
2. When compared with mammals, much less of the nematode genome is devoted to conserved regulatory regions

It has been argued that the full utility of nematode genomics has not been exploited due to the lack of a sister species to *C. elegans* (Thomas, 2008). Identification and complete genome sequencing of such a species would allow better identification and understanding of regulatory regions within the *C. elegans* genome. Examples of the advantages of such a genomic resource have been illustrated in yeast, insect and mammalian comparative genomic studies (Clark *et al.*, 2007; Borneman *et al.*, 2007; Hahn *et al.*, 2007; Gibbs *et al.*, 2007) However, although the search has started for a species sufficiently closely related to *C. elegans*, an example of this remains elusive (Thomas, 2008).

1.4.2 Parasitic species

Despite the caveat outlined above, comparative genomic investigations across the phylum Nematoda is an intriguing prospect. Such investigations will help provide information on the evolutionary basis behind the biology of parasitism and on the diverse range of organisms that are found in each of the five distinct Clades (Blaxter *et al.*, 1998). Comparative genomic analysis also has the potential to identify parasite specific genes which can potentially be new targets for control methods. Furthermore, comparative genomic analysis can provide an

added element of investigation on the genetic basis of drug resistance. However, genomic resources for parasitic nematode species lag far behind the majority of other pathogens (Gilleard, 2005). Comparative genomic investigations for parasitic species have therefore largely been performed on a 1:1 basis in comparisons with the available resources for the *Caenorhabditis* genus. However, there is currently a significant effort being focussed on improving the impoverished state of genomic resources for parasitic nematodes (http://www.sanger.ac.uk/Projects/Helminths/; http://genomeold.wustl.edu/genome_group.cgi?GROUP=6). With the advent of next generation sequencing technologies, I can confidently predict a vast improvement of parasitic nematode genomic data in the near future.

1.4.2.1 Clade V

Due to the close phylogenetic relationship with *C. elegans*, it is parasitic nematode species within Clade V of the phylum Nematoda that will benefit most from the wealth of data available for the free-living model organism, and for which comparative analysis with *C. elegans* will have most relevance. There is only one report of the size of the haploid genome of *H. contortus*, where it was predicted to be ~52 Mb using flow cytometry (LeRoy *et al.*, 2003). However this estimate is thought to be relatively conservative based on the assembly statistics of current *H. contortus* genomic sequence datasets (M. Berriman, Wellcome Trust Sanger Institute, personal communication).

There is currently in excess of 800 Mb genomic sequence available for *H. contortus* (http://www.sanger.ac.uk/cgi-bin/blast/submitblast/h_contortus). This is therefore the furthest advanced genome sequencing project of all Clade V parasitic nematodes. Currently the majority of this information resides within the core shotgun sequence reads due to difficulties in assembly. Chapter 3 presents detailed findings and discusses issues associated with the *H. contortus* genome sequencing project.

P. pacificus also resides within Clade V of nematode phylogeny and is the only species within this Clade with a complete genome sequence that does not belong to the *Caenorhabditis* genus. *P. pacificus* is a necromenic nematode of beetles (Dieterich *et al.*, 2008). Necromenic species associate themselves with a host,

wait for its death and survive by feeding on the microbes that develop on the carcass of the host. Necromenic species are hypothesised to provide a link between free-living and parasitic life forms. The genome of *P. pacificus* is 169 Mb and is therefore the largest nematode genome sequence available, and is predicted to contain 23,500 protein-encoding genes. The expansion in the encoded proteome in comparison to species of the *Caenorhabditis* genus is predicted to be due to the complexity of niches that *P. pacificus* inhabits throughout its life cycle. This genome sequence will therefore provide important insights into the evolution of parasitism specifically within Clade V of the phylum Nematoda.

1.4.2.2 Clade IV

Plant parasitic nematodes provide the best examples of parasitic nematode species with available genomic resources (Rosso *et al.*, 2009). There are now complete genome sequences for *Meloidogyne incognita*, *Meloidogyne hapla*, *Heterodera glycines* and the project of *Globodera pallida* is nearing completion (Abad *et al.*, 2008; Opperman *et al.*, 2008; http://genamics.com/cgi-bin/genamics/genomes/genomesearch.cgi?field=ID&query=802; http://www.sanger.ac.uk/Projects/Helminths/). All these species reside within Clade IV of the phylum Nematoda. These genome sequences will allow specific investigations into the evolution of plant parasitism and the identification of biological pathways that are both specific to this class of species, and those which are conserved in other Clades. However, as these species branch outwith Clade V of nematode phylogeny, they are not ideally positioned to exploit the wealth of data, or for functional genomic investigations, using the free-living nematode species *C. elegans*.

1.4.2.3 Clade III

There is only one nematode with completed genome sequence from Clade III of the phylum Nematoda. The genome sequence of the human filarial parasite *Brugia malayi* was the first parasitic nematode sequence reported (Ghedin *et al.*, 2007). The completed nuclear genome size is predicted to be 95 Mb and contain 14,500 to 17,800 protein-encoding genes. This gene number is far lower than the current figures of 19,762 and 19,507 like-wise genes from the genomes

of *C. elegans* and *C. briggsae*, respectively. Approximately 50 % of all *B. malayi* genes share 1:1 ortholgous relationships with *Caenorhabditis* genes. However, the average pair wise identity of the protein products of these genes is only 48 % (Ghedin *et al.*, 2007). Similar to *C. elegans*, an estimated 15 % of *B. malayi* genes reside within operons. However, most *B. malayi* operons represent combinations of genes that are not contained within *C. elegans* operons. These investigations have led to the postulation that *Brugia* and *Caenorhabditis* last shared a common ancestor ~350 million years ago (Mya) (Scott and Ghedin, 2008). It can therefore be reasonably predicted that the use of *C. elegans* will be less useful in the investigation of Clade III nematodes when compared with Clade V species. However, the *B. malayi* genome sequence will provide significant insights into the core elements common to all nematode species.

1.4.2.4 Clades I and II

There are no completed genome sequences from any species belonging to Clades I or II of the phylum Nematoda. The genome project of the Clade I parasitic nematode of humans, pigs, and rats *Trichinella spiralis* is at an advanced stage, nearing completion (Mitreva and Jasmer, 2008). This genome sequence will provide an interesting insight into the biology of this organism and others that branch within this Clade of the phylogeny. In addition, this genome sequence will help identify biological features conserved across all five Clades of the phylum Nematoda. An interesting potential application of the genome sequence of this organism is the investigation of parasitism, at differing intervals across the phylum Nematoda. The phylogenetic classification of these species has allowed the hypothesis that parasitism has evolved several times independently within each of the five distinct Clades. Potentially, comparisons of the genome data for Clade I parasites with those of *B. malayi*, *P. pacificus* and *H. contortus* can reveal genomic perspectives on this evolution (Mitreva and Jasmer, 2008).

1.5 Investigation of parasitic nematode gene function and regulation

The lack of *in vitro* culture methods, and/or availability of surrogate host laboratory animals for most species, makes the investigation of parasitic

nematode gene function and regulation extremely difficult. This section will review current *in vivo* methodologies available for investigation of parasitic nematode genes.

1.5.1 Transgenic studies in parasitic nematodes

There have been recent advances in the development of transgenic techniques applicable to investigate parasitic nematode genes. The majority of species for which such tools are now available belong to Clade IV, and branch within the Strongyloidoidea superfamily which have a free-living cycle, amenable to genetic manipulation. *Parastrongyloides trichosuri* was the first parasitic nematode species for which heritable DNA transformation was demonstrated (Grant *et al.*, 2006). Since this report, reliable methods for gene transfer of GFP-reporter constructs injected into the gonad of the human parasitic nematode *Strongyloides stercoralis* have also been demonstrated (Lok and Massey, 2002; Li *et al.*, 2006). However, these two parasitic species are the only two from all Clades of the phylum Nematoda for which methods of heritable DNA transformation have been reported.

The only other example of parasitic nematode species expressing DNA and RNA constructs are for the filarial species *Litomosoides sigmodontis* and *Brugia malayi*, and the intestinal parasite *Ascaris suum* (Jackstadt *et al.*, 1999; Davis *et al.*, 1999; Higazi, 2002). However, in each of these cases, only transient expression of the construct was reported. Therefore, transgenesis methodologies have been reported for only a select few parasitic nematode species. In addition, all of these species branch within Clades III or IV of the phylum Nematoda. There have been no reports of such techniques for any Clade I, II or V parasitic nematode species. Given the complexity of the life-cycle for most of these parasitic nematodes, and the lack of *in vitro* culture systems, it is reasonable to speculate that no methods of stable transgenesis will be forthcoming in the foreseeable future (Gilleard, 2005).

1.5.2 RNA interference (RNAi) in parasitic nematodes

RNAi is a reverse genetic technique to down-regulate the expression of genes by gene-specific double-stranded RNA (dsRNA). The use of dsRNA to silence genes

was first carried out in *C. elegans*, and has since been applicable to the investigation of gene function for a number of species (Fire *et al.*, 1998; Ullu *et al.*, 2004; Hannon, 2002). The mechanism underpinning this process is best understood in *C. elegans* through the identification and characterisation of RNAi deficient mutant worms (Figure 4.1; Grishok, 2005). Briefly, the dsRNA is cleaved into small inhibitory RNAs (siRNAs) which are the molecules responsible for gene silencing. The generation of siRNAs activates the RNA induced silencing complex (RISC) which is responsible for targeting homologous mRNA transcripts. Once bound, the RISC complex signals these mRNA sequences for degradation, resulting in expression knock-down or total silencing of the gene.

Since its discovery, RNAi has been applied to investigate almost every *C. elegans* gene (Kamath, R.S and Ahringer, J., 2003; www.wormbase.org). Interestingly, RNAi is more applicable to *C. elegans* than to any other species from the *Caenorhabditis* genus (Winston *et al.*, 2007). It is only within this species that the dsRNA can be delivered to the worms by either micro-injection of the dsRNA into the adult worm, soaking any life cycle stage of the worm in dsRNA or by feeding worms on bacteria which express dsRNA (Strange, 2006). In all three examples of delivery, an RNAi systemic effect is achieved in *C. elegans*. This simply means that the gene knock down effect is not only seen in the tissues directly accessible to the dsRNA, but is also spread throughout the worm. In addition to this, the *C. elegans* RNAi effect is heritable (Johnson *et al.*, 2005). It has been shown that the major reason for *C. briggsae* not showing systemic RNAi is that this species lacks the *sid-2* gene (Winston *et al.*, 2007). This gene encodes a transmembrane protein responsible for the uptake of dsRNA from the environment. Despite the lack of environmental RNAi, injection of dsRNA is a viable technique to reliably investigate gene functionally induced in species other than *C. elegans* from the *Caenorhabditis* genus (Winston *et al.*, 2007).

Although readily applicable to the *Caenorhabditis* genus, RNAi as a method to investigate virtually all other classes of nematode species has been sparsely reported and appears to be extremely variable. Even within other non-parasitic nematode species there has only been one successful report of RNAi to investigate a *P. pacificus* gene (Pires da Silva, 2006). In 2002 Hussein and colleagues reported the first example of RNAi as a method of gene silencing in a parasitic nematode species (Hussein *et al.*, 2002). Knock-down of the

acetylcholinesterase genes of the rat parasite *Nippostrongylus brasiliensis* during *in vitro* culture was reported, although no phenotype was observed. However, this example typifies the majority of similar RNAi investigations with parasitic nematodes, being difficult to repeat using the same methodologies reported in the initial publication (G. Ball, Imperial College, personal communication). Chapter 4 discusses the variability of RNAi in parasitic nematode species in greater detail, and all published reports of RNAi investigations using parasitic nematode species are detailed in Table 4.1.

1.6 The use of *Caenorhabditis elegans* as a surrogate system to investigate parasitic nematode genes

As discussed above there are currently no reliable methods of *in vivo* gene analysis for most parasitic nematodes. However, it is now generally accepted that, due to the close phylogenetic relationship, *C. elegans* can be exploited as a surrogate expression system in particular to investigate Clade V nematode genes (Gilleard, 2005; Britton and Murray, 2006). The use of this genetic model organism to investigate species that branch outwith Clade V is not as routine (Gilleard, 2005).

1.6.1 Investigation of parasite gene function

The first report of a parasitic nematode gene being functionally expressed in *C. elegans* was that of the *H. contortus* β-tubulin *isotype-1* gene (Kwa *et al.*, 1995). Injection of the complete parasitic genomic *isotype-1* locus amplified and cloned from a *H. contortus* BZ sensitive population was able to confer sensitivity to this class of anthelmintic drug in previously resistant *C. elegans* worms. In addition, injection of the same genomic locus when amplified from BZ resistant *H. contortus* worms, and injected into *C. elegans*, had no effect on BZ sensitivity, and instead the worms remained resistant to this class of drug. This work was able to link this parasite gene with the ability to modulate susceptibility to the BZ class of anthelmintic. Moreover, this work was also able to demonstrate the functional significance of a single point mutation in BZ resistance. As the genomic locus had been injected into *C. elegans*, this work was able to

demonstrate that the *H. contortus* gene was correctly spliced and apparently expressed in the appropriate tissues in order to be functional in *C. elegans*.

A second example of using *C. elegans* as a tool to investigate Clade V parasitic nematode gene function, again, comes from the species *H. contortus*. In this case, expression of the *H. contortus cathepsin-L-1* (*cpl-1*) was shown to rescue the embryonic lethal phenotype of *C. elegans cpl-1* RNAi or genetic mutants (Britton and Murray, 2002). This report provides evidence that the parasitic gene is able to replace the function of the mutant native gene, which in turn suggests that they are functionally very similar.

Another example is the investigation of the *H. contortus elt-2* gene (Couthier *et al.*, 2004). This gene encodes a transcription factor homologous to the *C. elegans elt-2* gene which is known to be highly active during endodermal differentiation and has a major role in control of gut specific gene expression (Fukushige *et al.*, 1998). Expression of this parasite gene under the control of a heat inducible promoter was able to change the fate of *C. elegans* blastomeres in a fashion similar to over-expression of the endogenous *C. elegans elt-2* (Couthier *et al.*, 2004). This is strongly suggestive that these genes are functionally very similar.

Each of these examples has used different methods to investigate Clade V parasitic nematode gene function, but all have used *C. elegans* as a heterologous system. The diverse range of genetic techniques that are applicable to investigate *C. elegans* biology can also be applied to transgenic *C. elegans* strains which express parasitic nematode genes. This is perhaps of most importance when investigating genes which have an essential function, or those which may result in phenotypic changes when over-expressed or expressed in the wrong tissues (Gilleard, 2005). Published reports of the success, or lack of success, of future similar investigations will allow conclusions to be drawn on the degree to which such investigations are useful. However, with the lack of reliable methods to examine gene function directly in parasitic species, using *C. elegans* as a tool to investigate the function of parasitic nematode genes is one of the most promising applications of parasitic nematode research.

1.6.2 Investigation of parasite gene regulation and expression

Investigation of *C. elegans* gene expression pattern is now routine and there are a number of elegant expression vectors available to examine activity of regulatory regions (Fire, 1998). Once injected and incorporated into a transgenic line these constructs are inherited in a non-Mendelian fashion as extra-chromosomal arrays and the expression of the reporter gene can be examined via relatively basic live worm microscopy. Such investigations are quick, cheap and simple and have revolutionised *C. elegans* gene expression investigations, to the extent that the majority of *C. elegans* genes have now been investigated in this manner (www.wormbase.org; http://elegans.bcgsc.ca/perl/eprofile/index).

There is little published data available on parasitic nematode gene regulation, however a number of studies have used *C. elegans* as a heterologous system to examine parasite promoter activity (Table 1.2). Recent investigations into DNase-1 hyper-sensitive regions of both *C. elegans* and *Homo sapiens* complete genome sequences suggest that most regulatory sequences reside within 2 kb, both upstream and downstream, and in the first introns of gene structures (Shi *et al.*, 2009; Bernat *et al.*, 2006; Sabo *et al.*, 2006). However, it is important to remember that *C. elegans* regulatory elements have been identified as far as 6 kb upstream of the ATG initiator site (Grandien and Sommer, 2001). From this data it would be advisable that parasitic DNA constructs are generated containing as much 5' flanking region as possible in addition to the complete gene structure and 2 kb of 3' sequence. A general rule has been observed with *H. contortus* genes, where most tend to be larger, and contain more intronic sequence than their *C. elegans* counterparts (J.S Gilleard, University of Calgary, and C. Britton, University of Glasgow, personal communication). In addition, the upper insertion size limit for expression vectors is 6-7 kb (A.D. Winter, University of Glasgow, personal communication). Therefore, it is not a viable option to routinely clone complete parasite gene structures plus 2 kb 5' and 3' flanking sequence into the available expression vectors to generate reporter gene expression patterns.

A potential route to overcoming such difficulties is the use of PCR-fusion based reporter gene expression constructs (Hobert, 2002). These constructs are

generated by fusing together two PCR products, one of which contains the regulatory region of the gene of interest (transcriptional or translational) and the other contains the reporter gene fragment of one of the Fire vectors. The size of these constructs is not limited by plasmid insertion upper size limits as there are no cloning steps involved. In addition, the use of such constructs has, so far, generated comparable results with the patterns obtained using cloning vectors (Hobert, 2002; Gilleard, 2005).

Due to the relative ease with which such experiments can be carried out, there are a number of studies in which parasite DNA reporter constructs have been generated and putative expression patterns determined in transgenic *C. elegans* of genes (Table 1.2). However, although an expression pattern has been obtained for the majority of these investigations, how this pattern relates to the true gene expression in the native host is an important consideration and a question that should be investigated as thoroughly as possible. From comparisons of parasite reporter gene expression patterns with the expression pattern of the homologous *C. elegans* gene, or immunolocalisation in the parasitic species, a general rule is starting to emerge that, despite little sequence conservation of regulatory regions between putative *C. elegans* and parasitic nematode homologous genes, there is extensive conservation of spatial expression but little conservation of temporal expression (Britton *et al.*, 1999; Gilleard, 2005).

1.7 *Haemonchus contortus* β-tubulin nomenclature

Different nomenclature has been used in reference to the two *H. contortus* β-tubulin loci that have previously been identified (Geary *et al.*, 1992; Kwa *et al.*, 1993; Beech *et al.*, 1994; Geldhof *et al.*, 2005). For clarity, and in order to avoid confusion, the nomenclature that will be used to refer to these previously identified genes, and additional *H. contortus* β-tubulin genes identified by work detailed in this thesis, are shown in Table 1.3.

1.8 Project aims

The overall aim of this project was to use comparative genomics along with functional analysis in *C. elegans* to help annotate the *H. contortus* genome and

study gene function in this parasite. Focus has primarily been on the identification and characterisation of the *H. contortus* β-tubulin gene family as well as genes involved in the RNAi mechanism. Such analysis also allowed general investigation of the current *H. contortus* genomic sequence databases, analysing the information present, and identifying possible reasons underlying the poor assembly statistics for this genome sequencing project.

Table 1.2: Investigation of parasitic nematode gene regulation and expression using *C. elegans* as a heterologous system (Adapted from Gilleard, 2004)

Parasite species	Clade	Gene	Reporter fusion	5' flanking (kb)	Expression pattern	Reference
Brugia malayi	III	*alt-1* (abundant larval transcript)	Transcriptional and translational within 3rd exon	1.9	Gut	Gomez-Escobar *et al*. 2002
Brugia malayi	III	*alt-2* (abundant larval transcript	Transcriptional	2.5	Gut	Gomez-Escobar *et al*. 2002
Brugia malayi	III	*Bm-phy-1* (proly1-4 hydroxylase)	Translational within 1st exon	2	Hypodermis	Winter *et al*. 2003
Onchocerca volvulus	III	*Ov*-GST1A (glutathione-S-transferase)	Translational within 3rd exon	2.6	Hypodermis	Krause *et al*. 2001
Globodera rostochiensis	IV	*Gpd* (glyceraldehydes-3-P-dehydrogenase)	Translational within 1st exon	0.8	Body wall muscle	Qin *et al*. 1998
Globodera pallida	IV	*Gp*-ace-2 (AChE)	Transcriptional	2.1	Neuronal cells and head ganglia	Costa *et al*. 2009
Parastrongyloides trichosuri	IV	*Pt-hsp-1*	Transcriptional	3	Gut	Newton-Howes *et al*. 2006
Haemonchus contortus	*V*	*β-tubulin isotype-1*	Genomic fragment	<2	Phenotypic rescue	Kwa *et al*. 1995
Haemonchus contortus	*V*	*AC-1* (cysteine protease)	Transcriptional	2.3	Gut	Britton *et al*. 1999

Haemonchus contortus	*V*	*Pep-1* (pepsinogen)	Transcriptional	2	Gut	Britton *et al*. 1999
Haemonchus contortus	*V*	*Hc-elt-2* (Transcription Factor)	Translational within 1st exon	3.3	Gut	Couthier *et al*. 2004
Teladorsagia circumcincta	*V*	*Colost-1* (cuticle collagen)	Transcriptional	1.3	Hypodermis	Britton *et al*. 1999

Table 1.3 *H. contortus* β-tubulin nomenclature

Genbank accession number		Previous nomenclature used	Nomenclature to be used in thesis		
gDNA	cDNA		Gene	Protein	3-letter
X67489	M76493	*Hc-isotype-1* *Hc-tub-8-9* *Hc-gru-1* *Hc-ben-1*	*Hc-isotype-1*	Hc-Isotype-1	Hc-ISO-1
N/A	M76491	*Hc-isotype-2* *Hc-tub-12-16*	*Hc-isotype-2*	Hc-Isotype-2	Hc-ISO-2
N/A	N/A	N/A	*Hc-isotype-3*	Hc-Isotype-3	Hc-ISO-3
N/A	N/A	N/A	*Hc-isotype-4*	Hc-Isotype-4	Hc-ISO-4

(Geary *et al.*, 1992; Kwa *et al.*, 1993; Beech *et al.*, 1994; Geldhof *et al.*, 2005)

Chapter 2: Materials and Methods

2.1 *Caenorhabditis elegans* Methods

2.1.1 Culture and maintenance of Caenorhabditis elegans

All *C. elegans* strains were obtained from the *Caenorhabditis* Genetics Centre (CGC) (University of Minnesota, Minnesota, USA) or from Shohei Mitani, National Bioresource Project for *C. elegans* (Tokyo Women's Medical College, Tokyo, Japan). Worms were grown and maintained as described previously (Brenner, 1974) on NGM-Agar inoculated with *E.coli* strain OP50 as a food source. All strains of *C. elegans* were maintained at 20^{o}C.

2.1.2 Freezing of Caenorhabditis elegans for prolonged storage

Worms were washed from plates that were just clearing bacteria, using 1 ml of 1 X M9 buffer, and added to 15 ml Falcon tubes (Starstedt). 1 ml of 1 X freezing solution was added to the sample and tubes thoroughly mixed by inverting 4-6 times. 500 µl of solution was transferred to a 1.5 ml screw top microfuge tube and stored in an insulted container at -80^{o}C or in liquid nitrogen. Cultures could be recovered by thawing tubes and pouring the contents onto NGM-agar plates seeded with OP50 *E.coli* bacteria and incubating at 20^{o}C.

2.1.3 Transformation of Caenorhabditis elegans

2.1.3.1 *Caenorhabditis elegans* microinjection procedure

C. elegans transformation by microinjection was performed as described by Mello and Fire (1995). DNA was injected into the cytoplasm of the syncytical, mitotically active gonad of well fed, healthy adult hermaphrodites. Double stranded DNA promoter-GFP reporter constructs were injected as neat PCR product at a final concentration of 2-28 ng/µl (Hobert, 2002). I would like to thank Gill M^{c}Cormack, University of Glasgow, for the generation and maintenance of C. elegans transgenic lines generated for this project.

2.1.3.2 Selection of transformed *Caenorhabditis elegans* lines

Oocyte nuclei in the adult gonad of *C. elegans* all share a common core of cytoplasm. As they mature, plasma membranes form to separate individual nuclei along with a portion of the core cytoplasm. DNA that is injected into this syncitium can also become incorporated into the developing oocyte. Due to regions of homology between co-injected DNA sequences, recombination events can occur in a fraction (1-10%) of the F1 progeny which results in the formation of long tandem arrays. Large arrays can become heritable as extrachromosomal elements and once assembled these elements are no longer targets for further inter-array recombination. The arrays are then transmitted at a frequency of between 5-95% and the lines can be maintained indefinitely by the selection of the phenotype conferred by the selectable marker DNA. Only one F2 line was selected from any one positive F1 plate of transformants so that generated lines were the result of different array formations. Generally, at least three lines were examined for each set of injections in order to ensure that a result found was not due to the particular arrangement of DNA within a particular array in a single line.

2.1.4 Microscopy of live Caenorhabditis elegans

C. elegans worms transformed with promoter-reporter constructs were transferred from NGM-OP50 plates to 2% agarose pads on microscope slides. A small volume of 0.1% sodium azide in 1 X PBS buffer was added to the pad to anesthetise the worms and to prevent drying of samples. A coverslip was added to seal the samples which were viewed immediately using an Axioskop 2 plus microscope with an AxioCam Mrm camera. Nomarski bright field and GFP fluorescence images were processed using Axiovision software.

2.2 *Haemonchus contortus* methods

2.2.1 Haemonchus contortus strain used

The MHco3 (ISE) strain of *H. contortus* was used for all work in this thesis. This is an inbred strain of SE, an outbred population from Edinburgh, UK (Roos *et al*,

2004). This strain remains susceptible to all three major classes of anthelmintic in current general use. This is also the strain that is being used by the Sanger Institute for the *H.contortus* genome project.

2.2.2 Haemonchus contortus maintenance and culturing

H. contortus adult worms were a kind gift from Dr. Frank Jackson (Moredun Research Institute, Edinburgh, UK). *H. contortus* L3 larvae were obtained by coproculture in distilled water of faeces from infected sheep (Seaton, 1989) and were a gift of Prof. Dave Knox (Moredun Research Institute, Edinburgh, UK). *H. contortus* L3 were stored in distilled water at 4^{o}C and used within three months.

2.2.3 Exsheathing Haemonchus contortus L3 for immunolocalisation

H. contortus L3 stage larvae were exsheathed in 0.25% sodium hypochlorite (Sigma)/1 X PBS for 5 minutes at room temperature. Exsheathed L3s were washed three times using 1 X PBS to remove all traces of sodium hypochlorite.

2.3 General Molecular Biology Techniques

2.3.1 Genomic DNA isolation

Genomic DNA was extracted from *C. elegans* strains grown on NGM plates seeded with OP50 bacteria and from *H. contortus* MHco3 (ISE) adult material (donated by Dr Frank Jackson and Prof Dave Knox, Moredun Research Institute, Edinburgh). The same method of genomic DNA isolation was used for both species. Worms were washed using 1 x M9 solution and then concentrated by centrifugation. Six volumes of 1 x worm lysis buffer was added to 200-500 µl of worm pellet. Worm suspensions were disrupted in a glass hand held homogeniser then incubated at 65^{o}C for 4 hours. Debris was removed by centrifugation and DNA purified by repeated phenol:chloroform and chloroform extractions. DNA was treated with RNAse A (DNAse free) (Sigma) at a final concentration of 100 µg/ml for 1 hour at 37^{o}C then extracted with phenol:chloroform, chloroform

alone and precipitated in 100% ethanol at -20°C. 100-200 ng of genomic DNA was used in 20 µl PCR reactions.

2.3.2 Preparation of individual worm lysates

All individual *H. contortus* worms were identified in accordance to the Manual of Veterinary Parasitology Laboratory Techniques using male tail and spicule morphology (Maff, 1986) for separation of male and female *H. contortus* adults. Healthy hermaphrodite *C. elegans* adults were picked from NGM plates seeded with OP50 bacteria. Individual adult worms or *H. contortus* exsheathed L3 larvae were picked into single 0.2 µl tubes containing 50 µl or 20 µl of 1 X single worm lysis buffer, respectively. All lysates were placed at -80°C for a minimum of 10 minutes before incubation at 60°C for 98 minutes followed by 20 minutes at 94°C to denature the proteinase K. Individual worm lysates were stored at -80°C. One µl of a 1:50 dilution neat single adult lysate or a 1:10 dilution of single L3 lysate was used as PCR template. Dilutions of several aliquots of lysate buffer, made in parallel, were used in PCR as negative controls.

2.3.3 Total RNA isolation

RNA isolation from *H. contortus* worms at L3 and adult life cycle stages was carried out by the Trizol method (Gibco BRL Life Technologies) (Burdine and Stern, 1996). Briefly, 4 x volume of Trizol was added to 400-600 µl of concentrated worm pellet. Suspensions were vortexed to solubilise and disrupt worms then left at room temperature for 30 minutes. Insoluble material was removed by centrifugation at 4°C at 14,000 rpm. Supernatants were transferred to fresh 1.5 ml eppendorf tubes and 500 µl of isopropanol added to each sample. Solutions were frozen at -80°C for 1-8 hours for precipitation of RNA. Following defrosting at room temperature, samples were centrifuged at 4°C at 14,000 rpm to pellet the RNA. RNA was washed with 75% ethanol (25% DEPC-treated H_2O), spun at 7,500 rpm for 5 minutes and resuspended in 87.5 µl of DEPC-treated H_2O. To purify RNA and remove any contaminating DNA, samples were subjected to 2 X DNase Digestion (Invitrogen) followed by RNeasy Mini Kit RNA Cleanup (Invitrogen). Solutions and protocols used for both procedures are as described in the manufacturers' instructions accompanying the kits. A GeneQuant *pro*

spectrophotometer (Amersham Biosciences) was used to estimate the concentration of the RNA samples.

2.3.4 Synthesis of first strand cDNA

First strand cDNA was synthesised using an RT-PCR kit (Cloned AVM First Strand cDNA Synthesis Kit, Invitrogen) according to the manufacturer's instructions. Five µg of total RNA was used per reaction. To obtain the 5' end of genes of interest, first strand cDNA was synthesised using the random-hexamer primers provided with the kit. 0.5 µl of cDNA was used as template per 20 µl PCR reaction.

2.3.5 Polymerase chain reaction

Genomic DNA regions were PCR amplified using a Go*Taq* (Promega) and *Pfu* Turbo (Stratagene) DNA polymerase mix at a 5:1 ratio. Go*Taq* DNA polymerase was selected for its high yield benefits; *Pfu* Turbo DNA polymerase was used due to its proof-reading capacity. The final concentration of polymerase mix in the 20 µl PCR reaction was 0.075 U/µl. dNTPs (dATP, dCTP, dGTP and dTTP) were used at a final concentration of 1 mM each and were supplied by Promega. All oligonucleotides used for PCR were synthesised by either Invitrogen or MWG and used at 0.5 µM each. PCR reactions were performed using a *Flexi*gene thermocylcer (Techne). Specific details of PCR programmes and primer sequences will be presented in the appropriate chapter or Table.

2.3.6 Agarose gel electrophoresis

For visualising PCR products, 1 to 3% (w/v) agarose gels were used. Gels were prepared by dissolving powdered agarose (Sigma-Aldrich) in 1 x TAE solution. Ethidium bromide was added to a final concentration of 1 µg/ml. Gels were cast and run in 1 x TAE using horizontal gel tanks (Anachem Bioscience). Two lanes of 1 kb DNA ladder (Invitrogen) were included per row of wells to allow estimation of DNA product size and concentration. DNA samples and markers were diluted in loading buffer and gels were electrophoresed at 90-120 V to allow separation of amplicons. Agarose gels were viewed under UV exposure using a FluorChem 5500 transilluminator (Alpha Innotech).

2.3.7 TA cloning of PCR products

PCR products were ligated into the pCR-4 vector (Invitrogen) and chemically transformed into TOP10 *E. coli* competent cells (Invitrogen) according to the manufacturer's instructions. 50 and 100 µl of each transformation was plated out onto separate LB plates containing 100 µg/ml ampicillin and incubated at 37°C overnight.

2.3.8 Identification of bacterial transformants

To identify bacterial colonies containing PCR insert DNA, selected colonies were picked and touched on to a LB + ampicillin (100 µg/ml) master plate. The plate was allowed to grow overnight in a 37°C incubator. A toothpick was used to transfer overnight colonies to PCR reaction tubes. Overnight colonies were then used as template for PCR reactions without any additional purification. PCR primers were complementary to the vector backbone and provided by the manufacturer.

2.3.9 Plasmid DNA preparation

Transformed bacterial colonies containing the PCR insert were grown overnight in 7 mls of L-broth + Ampicillin (100 µg/ml) at 37°C in an orbital shaker. Mini plasmid purification kits (QIAGEN) were used to prepare the plasmid DNA from the bacterial cultures using solutions and protocols supplied by the manufacturer. The protocol takes advantage of the selective renaturation of plasmid DNA at high salt concentrations. Eluted DNA was examined by agarose gel electrophoresis to estimate concentration.

2.3.10 Sequencing of pCR-4 plasmid inserts

For each plasmid to be sequenced, DNA preparations from two representative samples each containing 700 – 1000 ng were dried overnight at 37°C in 1.5 ml screw-top eppendorf tubes. Each tube was then sent to MWG to be sequenced via its Value Read service using primers M13 uni (-21) and M13 rev (-29). Both primers were provided by the manufacturer and are complementary to the

vector backbone. In cases where large inserts were to be sequenced, sequences of primers used are in Appendix 2.

2.3.11 Generation of Ce-ben-1 and Ce-mec-7 promoter-GFP reporter gene fusions

All promoter-reporter fusions were generated by the fusion PCR method described by Hobert, 2002. Promoter regions for both *C. elegans* loci were identified from cosmid DNA sequences, available from Wormbase (www.wormbase.org). For *Ce-ben-1*, a 1.971 kb region directly 5' of the ATG translational start site was PCR amplified from N2 gDNA, cloned into the pCR-4 vector (Invitrogen) and sequenced in triplicate (MWG). Expression from the *Ce-mec-7* promoter has previously been reported (Hamelin *et al.*, 1992). This study showed that 850 bp of sequence directly 5' of the ATG start codon was sufficient for tissue-specific expression. I therefore amplified the same 850 bp region for the *Ce-mec-7* promoter-GFP reporter construct. Gene specific reverse primers contained a 23 bp sequence identical to the 5' end of the GFP sequence to be used in my fusion constructs to allow overlap of this region.

Plasmid DNA of these cloned promoter fragments was used as template in a second round PCR containing the same forward promoter primer as used in the first round PCR amplification, and a reverse primer designed to the 3' end of the GFP gene contained in Fire vector pPD95.67, present in the second round PCR (kindly provided by A. Fire, Stanford School of Medicine, USA). Although this vector contains a nuclear localisation signal (NLS) non-localisation of the GFP signal has been reported, hypothesised as being due to the protein diffusing out of the nucleus as a result of its relatively small size (Roberts, *et al.*, 2003; http://www.addgene.org/labs/Fire/Andrew/Vec95.pdf). Sequence for all primers used to generate first and second round products can be found in Appendix 2.

DNA constructs were co-injected with the selectable marker plasmid pRF-4 (Mello *et al*, 1991) (a gift from A. Fire, Carnegie Institute of Washington, Baltimore). Plasmid pRF-4 contains a semi-dominant allele of the collagen gene *rol-6* (*su1006*). This allele confers a right-hand roller phenotype which allowed

identification of transformed progeny using low power light microscopy. The pRF-4 plasmid DNA was injected at a final concentration of 50-100 ng/µl. Injection mixes were typically made up to a final DNA concentration of 120-200 ng/µl by the addition of pBluescript SKM then centrifuged at 13,000 g for 5 minutes to pellet debris and the supernatant used for microinjection. Injection needles, 1.2 mm O.D X 0.69 mm I.D borosilicate glass capillaries with standard wall and inner filament (Clark Electromedical Instruments), were pulled on a computer controlled electrode puller model 773 (Campden Instruments). Nematodes for injection were mounted on 2% agarose pads on large coverslips that had been baked for 2-12 hours at 80^{o}C. *C. elegans* worms were added to agarose pads under mineral oil (Sigma) and injected using an Axiovert-100 inverted microscope (Zeiss) equipped with a flat, free-sliding glide stage with centred rotation and DIC/Nomarski optics. DNA was injected into the hermaphrodite gonad using a micromanipulator guided needle and pressurised nitrogen gas. Recovery buffer was added to injected worms on pads, after which they were transferred to fresh NGM OP50 plates in a pool of recovery buffer. Transformed progeny from injected hermaphrodites (F1 generation) were selected clonally or in small groups of 3-7, and the next generation (F2) surveyed for stable, transmitting transformants.

2.3.12 Generation of Haemonchus contortus β-tubulin promoter-GFP promoter fusion constructs

H. contortus promoter-reporter constructs were generated by the PCR fusion method described above. Due to the incomplete nature of the *H. contortus* genome project, regions directly 5' of the ATG translational start codon were identified from shotgun sequence reads overlapping with coding sequence of the *H. contortus isotype-1* and *isotpye-2* genes. Four sequence reads from the 23 which show significant alignment with the *Hc-iso-1*, cover the ATG start codon. Scaffolding analysis allowed us to deduce that the *H. contortus* genomic shotgun sequence read haem-1857d17.p1k lay 5' of read haem-1857d17.q1k in relation to the Hc-*isotype-1* initiating codon. Primers were designed to amplify the gap between these genomic shotgun sequence reads and a 2.4 kb fragment was obtained, cloned into pCR-4 (Invitrogen) and sequenced in triplicate (MWG) (primer sequences, Appendix 2). This sequence overlapped with both reads for >

100 bp and showed 100 % similarity over these regions. This sequence allowed us to design primers to amplify 1.971 kb of 5' flanking region of the *H. contortus* β-tubulin *isotype-1* locus (primer sequences in Appendix 2). This product was separately TA cloned into vector pCR-4 and sequenced in triplicate for three individual clones. Once this fragment had been obtained it was fused in a second round PCR with the GFP coding region of Fire vector pPD95.67 to create a promoter-reporter construct for the *H. contortus isotype-1* locus (Hobert, 2002).

Of the 25 genomic shotgun sequences aligning to *Hc-iso-2*, five cover the ATG start codon. Through the use of bioinformatic scaffolding, I were able to hypothesise that sequence read haem-1049c22.p1K lay 5' of read haem-1049c22.q1k relative to the *Hc-iso-2* initiating codon. Primers were designed to amplify the gap between these sequence reads, and a 400 bp fragment was amplified (primer sequences in Appendix 2). The product obtained was TA cloned into vector pCR-4 and sequenced in triplicate for three individual clones (MWG). The resulting sequence overlapped with both sequence reads for > 80 bp at 100 % similarity. Primers were designed to amplify the 802 bp contiguous region 5' of the *isotype-2* translational start site in order to construct a promoter-reporter fusion (primer sequences Appendix 2). This product was separately TA cloned into vector pCR-4 and sequenced in triplicate for three individual clones. Once this fragment had been obtained it was fused with the GFP coding region of Fire vector pPD95.67 to create a promoter-reporter construct for the *H. contortus isotype-2* locus (Hobert, O., 2002; Chapter 2).

The *H. contortus* β-tubulin *isotype-3* gene resides in the centre of a fully sequenced and finished contiguous 408, 911 bp stretch of sequence from the X-chromosome (Chapter 5; Redman *et al.*, 2008). PCR primers were designed to amplify 2.056 kb of the 5' region directly upstream of the initiating codon of the *H. contortus* β-tubulin *isotype-3* gene (primer sequences Appendix 2; Figure 6.1).

2.4 Antibody localisation methods

2.4.1 Preparation of freeze-cracked nematode specimens for immunolocalisation

Nematode embryos, larval and adult stages for immunolocalisation experiments were prepared according to the method described for *C. elegans* embryos by Miller and Shakes (Miller *et al*. 1995). Briefly, slides for immunolocalisation assay were prepared by adding a drop of poly-L-lysine (Sigma). Poly-L-lysine is an adhesive solution used for adhering tissue sections to glass slides. *C. elegans* specimens for immunolocalisation assay were grown on NGM plates seeded with OP50 bacteria. When at the desired life cycle stage, worms were transferred to approximately 5 µl 1 X M9 buffer on a non seeded NGM plate and left for 5 minutes to allow removal of bacteria from the cuticle. To obtain embryos, adult *C. elegans* were added to 50 µl of 1 x M9 buffer in a clear watch-glass and laterally sliced. Free embryos were then mouth pippetted to a non seeded NGM plate. For *H. contortus*, only exsheathed L3 larvae stages, which were readily available, were used in immunolocalisation. For all nematode stages, specimens were then placed into a drop of 1 X M9 buffer on a poly-L-lysine coated slide and a coverslip added. Slides were immediately stored on a metal block at -80^{o}C for 15 minutes. Fracturing of the worms was carried out using a scalpel blade to pop off the coverslip. Slides were immediately fixed in 100% methanol at -20^{o}C for 10 minutes and then in 100% acetone at -20^{o}C for 10 minutes. Slides were then left to air dry at room temperature. Slides could be stored at -20^{o}C or used immediately for immunolocalisation experiments.

2.4.2 Freeze-crack immunolocalisation of nematode specimens

Freeze cracked nematode samples on glass slides were re-hydrated with 150 µl 1 X PBST for 5 minutes then blocked with 150 µl 1 X PBST + 1% dry milk for 20 minutes. 150 µl of primary antibody mix was added at concentrations shown in table 6.3. Slides were then incubated at 4^{o}C overnight in a humid chamber to keep the slides in the dark and prevent dehydration. Slides were then washed 3 times for 10 minutes with 150 µl 1 X PBST to remove all traces of unbound primary antibody. 150 µl of secondary antibody (Alexa Flour 448 goat anti-rabbit

IgG, Invitrogen) was added to the slides at 1:200 dilution in 1 X PBST + 1% soluble milk and slides left for 1 hour at room temperature in a humid chamber. Slides were washed 3 times for 10 minutes using 150 µl 1 X PBST. Samples were mounted in 10 µl mounting media and coverslips gently placed on top. Slides were either viewed immediately or stored at 4°C in a humid chamber and viewed within 24 hours. Nomarski bright field and fluorescence images were processed using Axiovision software.

2.4.3 Ruvkun fixed nematode specimen indirect fluorescence antibody test

Larval and adult stage nematode specimens were fixed, permeabilised and stained as per Finney, *et al.*, 1990. Worms were harvested from NGM plates by washing with 1 ml of M9 buffer and collected in a 1.5 ml Eppendorf tube. Tubes were spun at 7, 000 rpm for 2 minutes in a micro-centrifuge, washed three times in M9 before being resuspended in 100-150 µl M9 and chilled on wet ice. Ice cold 2X Ruvkun buffer and 40 % formaldehyde were added to a final concentration of 1X and 2%, respectively. Eppendorf tubes were vigorously mixed and frozen on dry ice. Frozen pellets were freeze-thawed three times and incubated on ice with agitation for 1 hour. Worms were then washed twice in Tris-Triton buffer and resuspended in Tris-Triton buffer, 1 % β-mercaptoethanol overnight at 37°C with mild agitation to reduce the disulfide linkages and permeabilise the highly cross-linked nematode cuticle. To complete the reduction reaction, the worms were washed in ten volumes of 1X BO_3 buffer, 0.01 % Triton and incubated for 15 minutes at room temperature with agitation. Worms were then briefly washed again in 1X BO_3 buffer, 0.01 % Triton and once for 15 minutes in AbB buffer. The worms were then stored in AbA buffer at 4°C or used for the subsequent staining steps.

Worms were stained by incubating 25 µl of fixed worms in the appropriate dilution of primary antibody (Table 6.3) in AbA buffer overnight at 4 °C with mild agitation. The worms were then washed for three hours on a rocker at room temperature in several changes of AbB buffer and then rinsed once in AbA buffer prior incubation in secondary antibody. Secondary antibody (Alexa Flour 448 goat anti-rabbit IgG, Invitrogen) was used at 1:200 dilution in AbA. Worms were left

at 4 °C overnight with gentle agitation. Following a final wash for three hours in AbB at room temperature, the worms were mounted in 10 µl mounting media for observation.

2.4.4 Generation of anti-peptide antibodies directed against nematode β-tubulin proteins

To allow localisation of *C. elegans* and *H. contortus* β-tubulin proteins, anti-peptide antibodies were generated. β-tubulin peptide synthesis and rabbit immunisations were carried out by CovalAb, Cambridge, UK. In each case the peptide sequence used for the immunisation protocol was located at the hyper variable C-terminal region of the β-tubulin protein (Table 2.1). All peptide sequences were used as queries for BLAST searching of the relevant genome database to ensure that the best effort was made to select unique sequences. Two rabbits were immunised with the selected peptide for each project. Booster immunisations were performed at days 0, 21, 42, 74, 91 and 105. Serum samples were taken from each rabbit at 0, 53, 74 and 116 days post primary immunisation with terminal bleeds performed at 130 days. Enzyme-linked immunosorbent assay (ELISA) of each serum sample at each test bleed was performed by CovalAb, allowing the immunoreactivity of the antisera to be monitored. IgG antibody (2ml) was immunoaffinity purified by CovalAb from terminal bleed serum, pooled from both rabbits in each project. Final documentation contained ELISA titre information for the four purified IgG antibody preparations (so that the correct antibody dilution could be used for the desired procedure).

Table 2.1: *C. elegans* and *H. contortus* β-tubulin C-terminal peptide sequence selected for antibody generation

Species	Gene Name	C-terminal Protein Sequence	Name of purified antiserum
C.elegans	*ben-1*	EDGELDGTDGDAE	C-EDGELDGTDGDAE-coOH
H.contortus	*isotype-1*	DLDAEGGEEAYPE	C-DLDAEGGEEAYPE-coNH2
H.contortus	*isotype-2*	EMEGAVENDTYAEE	C-EMEGAVENDTYAEE-coOH
H.contortus	*isotype-3*	DEEPAETFEAE	C-DEEPAETFEAE-coOH

2.4.5 Pre-absorption of BEN-1 primary antibody against Caenorhabditis elegans mutant strain tm234 acetone powders

A *ben-1* mutant strain of *C. elegans*, allele *tm234*, was obtained from the Tokyo Women's Medical University as a gift of Dr Shohei Mitani. This strain contains a 2136 bp deletion covering the first three exons of the *Ce-ben-1* gene and is anticipated to be a null allele. Mutants are homozygous viable (www.wormbase.org). Worms were grown on NGM plates to obtain enough for ~5 ml of tightly packed worms. Three volumes of 0.1M NaCl was added to the worms and they were lysed by sonication. The homogenate was transferred to a 100 ml glass graduated cylinder and five volumes of acetone added. Worms were mixed by inverting and allowed to sediment for ~30 minutes to allow precipitation of proteins. Supernatant was removed. The cylinder was filled with acetone in order to remove any remaining lipids. Precipitate/acetone suspension was poured through 3 MM Whatmann paper to collect and material was dried overnight at 37^{o}C. Material was crushed into a fine powder using a mortar and pestle. Fine powder was added to a 15 ml falcon tube and 15 ml PBS added (for 0.9g of dry powder). The solution was vortexed and distributed into 1.5 ml eppendorf tubes. Each tube was microfuged for ten minutes, supernatant discarded and stored at -20^{o}C. For pre-absorption, 1/3 volume packed moist acetone powder was mixed with primary antibody solution mix for 30 minutes with gentle agitation, spun for ten minutes in a microfuge and supernatant was transferred to a new tube. Pre-absorbed antibodies were used at the same concentrations as non-preabsorbed (Table 6.3)

2.5 General Bioinformatic Analysis Techniques

2.5.1 Sequencing analysis software and computer analysis of DNA and protein sequences

DNA sequence data was analysed, assembled and compared using the Vector NTI (Invitrogen) applications Contig Express and Align X, and the Basic Local Alignment Search Tool (BLAST) (Altschul *et al.*, 1990) from the National Centre

for Biotechnology Information (NCBI) (http://ncbi.nlm.nih.gov/). Vector NTI (Invitrogen) and its associated software was used throughout this project for the routine viewing and annotation of DNA and protein sequences. Data on sequenced *C. elegans* and *C. briggsae* genes was accessed using the Wormbase (http://wormbase.org/) website.

2.5.2 BLAST searching of nematode sequence databases

A local BLAST server was used to search the EST, contig and shotgun sequence reads from the *H .contortus* genome sequencing project, being carried out by the Pathogen Sequencing Unit at the Wellcome Trust Sanger Institute (http://www.sanger.ac.uk/Projects/H_contortus/) The local server (http://130.209.234.35/blast/blast.html) contains all available *H. contortus* sequence datasets that have been released to date from the ongoing genome project. This server was also used to search the genomic and protein sequence datasets of *C.elegans* (release WS150). The Wormbase BLAST web page (http://www.wormbase.org/db/searches/blast_blat) was used for searching of the *C. briggsae* genomic and protein sequence databases (release WS195). BLAST searching of the *P. pacificus* genome assembly version 'P.pac California assembly 2' was performed using the online BLAST web page available on the *P. pacificus* genome website (http://www.pristionchus.org/cgi-bin/blast_iframe.pl). The genomic shotgun sequence reads for the genome sequencing project of the filarial parasitic nematode *B. malayi* were BLAST searched using the server available at The Institute of Genome Research, Washington, USA (http://blast.jcvi.org/er-blast/index.cgi?project=bma1). This server was also used to search the polypeptide dataset of *B. malayi* (release BMA1_pep).

2.5.3 Drosophila melanogaster gene information and sequence homology searching of sequence databases

The *D. melanogaster* genome website Flybase was used to obtain all information on *D. melanogaster* genes (http://flybase.org/). This web page also has an on-line BLAST server that was used to search the *D. melanogaster* protein dataset for putative homologues to *C. elegans* proteins and/or genes (http://flybase.org/blast/).

2.5.4 Hierarchical scaffolding of Haemonchus contortus shotgun sequence reads

The shotgun approach to genome sequencing entails sequencing both ends of sheared genomic DNA fragments that are typically 3-8 kb in length. Current sequencing technologies are limited to generating sequence reads of 600-1200 nucleotides and therefore the central region of each sheared fragment usually remains unsequenced. The paired relationship of these end-sequences has been used as a basis for hierarchical scaffolding and software has been developed to facilitate the assembly of these structures (Pop *et al*., 2004). A Perl Script was written to construct such scaffolds specifically using *H. contortus* shotgun genomic sequence reads (R.N. Beech, McGill University, personal communication). To summarise, a single shotgun read is used as a seed to generate a sequence scaffold for a particular locus. This initial sequence read is used as a query in a BLAST search of a database of all *H. contortus* shotgun genomic reads, in order to identify aligning and overlapping sequences. The name of each overlapping target sequence is used to identify its associated read pair and this paired-sequence is used as a query for further BLAST searching of the shotgun read database. Thus, scaffold building continues in an iterative manner until the process terminates when no further matching sequence reads can be found or a predetermined BLAST query sequence limit is reached. The tolerances used in scaffold construction may be adjusted by modifying BLAST parameters in the Perl Script such as minimum percentage of identity (i.e. quality of match) and minimum alignment length (i.e. length of overlap).

The following sections detail methods specific for certain chapters of this thesis.

2.6 Chapter 3 Materials and Methods

2.6.1 Searching of the Haemonchus contortus genome sequence databases for evidence of contaminating information

2.6.1.1 Bacterial contamination

The *H. contortus* genomic sequence database hc_allreads_112607.fasta is available at http://ubiquitin.mine.nu/blast/db/. This database contains 1,126,320 shotgun genomic sequence reads that have been generated from the *H. contortus* genome sequencing project. To identify sequences arising from bacterial contamination, this dataset was searched for sequences which showed a high homology to the completed 5,153,435 bp genome of *E. coli* strain H10407 (http://www.sanger.ac.uk/Projects/E_coli_H10407/) using stand alone BLAST (Altschul *et al.*, 1990), software that is freely available from the NCBI web site (http://www.ncbi.nlm.nih.gov/). The Microsoft Excel application of Windows version 2003 was used to filter these results so that each query sequence was represented once and that only hits of >90% identity for stretches of >50 bp contributed to the final result.

2.6.1.2 Host contamination

The same *H. contortus* genomic sequence database as above was selected and 296,710 of the reads were selected at random and used as query sequences to identify regions of similarity when compared to the 2.73 Gb of *Bos taurus* (*B. taurus*) assembled genome sequence (release bos.Tau4 available at ftp://hgdownload.cse.ucsc.edu/goldenPath/bosTau4/bigZips/) using stand alone BLAST (Altschul *et al.*, 1990), software that is freely available from the NCBI web site (http://www.ncbi.nlm.nih.gov/). The Microsoft Excel application of Windows version 2003 was used to filter these results so that each query sequence was represented once and that only hits of >90% identity for stretches of >100 bp contributed to the final result.

2.6.2 Searching for single cluster EST sequences within Haemonchus contortus genomic sequence databases

To investigate the representation of single cluster *H. contortus* MHco3 (ISE) EST sequences within current *H. contortus* genomic sequence datasets, a database containing 5,906 single cluster EST sequences was used as query sequence in searching the *H. contortus* genomic sequence databases hc_allreads_112607.fasta and combined_worms_contigs_121107.fasta using stand alone BLAST (Altschul *et al.*, 1990) software. The combined_worms_contigs_121107.fasta database contains homology based contiguous sequences assembled by the Pathogen Sequencing Unit of the Wellcome Trust Sanger Institute, Cambridge, UK using the information within the hc_allreads_112607.fasta database. Both databases are available at http://130.209.234.35/blast/blast.html. A cut-off value of e=3e-08 was used to identify significant alignments. This was the lowest value of any sequence read used to calculate the coverage of 19 single cluster EST sequences. The Microsoft Excel application of Windows 2003 was used to calculate the percentage of single cluster EST sequences with at least one significant hit within each of *H. contortus* genomic sequence databases.

In order to calculate the coverage of 19 randomly selected *H. contortus* single cluster EST sequences, each was used as query sequence for stand alone BLAST searching of the hc_allreads_112607.fasta database (http://130.209.234.35/blast/blast.html). Only alignments of >50bp showing >92% identity were considered as real hits. Regions where >10 *H. contortus* shotgun genomic sequence reads aligned to one region of EST sequence were identified as simplistic sequence and not included in any further analysis. All *H. contortus* shotgun genomic sequence reads that passed the searching criteria were assembled using the Contig Express application of Vector NTI (Invitrogen). All assembled contigs and/or single sequence reads were then realigned to the original identifying single cluster EST using BLAST2 (http://130.209.234.35/blast/wblast2.html). In cases where more than one contig or single sequence read aligned to one region of EST sequence, the contig or single sequence read which showed the greatest percentage identity to the

EST sequence was selected to represent that region. The percentage coverage of each single cluster EST sequence was subsequently calculated.

2.6.3 Allelic sequences within Haemonchus contortus genomic sequence datasets

The *H. contortus* genomic sequence database hc_sw_contigs200808.fasta contains 87,735 sequences. This homology based contiguous sequence database was generated using the 803,014 shotgun genomic sequence reads contained within database single_worm_unassembledReads070308.fasta. All sequences were generated and assembled by the Pathogen Sequencing Unit, Sanger Institute, UK. Both sequence databases can be found at http://130.209.234.35/blast/blast.html. I have used stand alone BLAST (Altschul *et al.*, 1990) software to investigate these 87,735 sequences for evidence of the presence of allelic sequences. The Microsoft Excel application of Windows version 2003 was used to filter these results so that each query sequence was represented once and that only hits of >90% identity for stretches of >300 bp contributed to the final result.

2.7 Chapter 5 Materials and Methods

2.7.1 β-tubulin sequences within Haemonchus contortus EST datasets

There are two searchable *H. contortus* EST sequence datasets publicly available which between them contain in excess of 30,000 sequences, consisting of more than 14 Mb of sequence data (http://xyala.cap.ed.ac.uk/services/blastserver/; http://www.sanger.ac.uk/cgi-bin/blast/submitblast/h_contortus). The identification of β-tubulin sequences within this combined dataset was aided by the fact that the β-tubulin polypeptide sequences for all species, for which β-tubulin sequence data exists, show a very high level of identity and similarity (Njue and Prichard, 2003). Each of the six *C. elegans* β-tubulin polypeptides was used as the query in tBLASTx searching of the *H. contortus* complete EST sequence dataset (http://130.209.234.35/blast/blast.html). All

identified sequences were used in reciprocal BLASTx searching of the *C. elegans* Wormpep dataset (release WS150; http://130.209.234.35/blast/blast.html) to ensure that each was more related to a *C. elegans* β-tubulin polypeptide than it was to any other protein. All sequences which passed this criterion were aligned to available fully sequenced *H. contortus* β-tubulin cDNA sequences for characterisation (Geary *et al.*, 1992; AlignX, Invitrogen). Prior to this project, full length cDNA sequence data was already available for the two previously characterised *H. contortus* β-tubulin genes, *isotype-1* and *isotype-2* (Accession numbers: M76493 and M76491; Geary *et al.*, 1992). In addition the full genomic locus for the *H. contortus isotype-1* gene was also available (Accession number: X67489.1).

2.7.2 Sequencing of the Haemonchus contortus β-tubulin isotype-2 genomic locus

Primer sequences were designed to be specific for the *H. contortus* β-tubulin *isotype-2* genomic locus using highly conserved regions of the two fully sequenced EST sequences that are available for this gene (Geary *et al.*, 1992). Sequences for the primers used to amplify this genomic locus are detailed in Appendix 2. PCR amplification of this region was carried out using pooled MHco3 (ISE) L3 genomic DNA as template and the resulting product cloned into the pCR-4 TOPO vector (Invitrogen) and sequenced in triplicate for three individual plasmid clones (MWG). A consensus sequence was generated using the ContigExpress application of VectorNTI (Invitrogen) which was used to represent the *H. contortus* β-tubulin *isotype-2* genomic locus for this work (Appendix 2).

2.7.3 Identification and sequencing of a Haemonchus contortus β-tubulin locus containing Bacterial Artificial Chromosome

A BLAST searchable *H. contortus* BAC-end genomic sequence database has been generated by the Sanger Institute, potentially enabling researchers to identify BAC sequences which contain their loci of interest (http://www.sanger.ac.uk/cgi-bin/blast/submitblast/h_contortus). The 16 Mb of sequence in this database represents 10,414 BACs. The complete sequence

length of a typical BAC is >100 kb therefore this database represents only a fraction of each BAC sequence. tBLASTn searching of this database using the *H. contortus* β-tubulin Isotype-1 protein sequence identified a BAC-end (haem_ends15g16.q1kT7) containing β-tubulin sequence (John Gilleard, University of Calgary, personal communication). This BAC was selected for complete sequencing and finishing (Pathogen Sequencing Unit, Sanger Institute, UK). In using this completed BAC sequence to identify other BAC-end sequences, in conjunction with a BAC-fingerprinting approach, a 'BAC walk' was generated consisting of five overlapping sequenced and finished BAC sequences. The total length of this sequence is 408,911 bp and it represents the largest contiguous sequence fragment that has been generated thus far from the *H. contortus* genome sequencing project.

2.7.4 Identification of Haemonchus contortus loci homologous to Caenorhabditis elegans protein sequences

The contiguous BAC sequence described above was split into 50 kb sections and used in BLASTx searching of the *C. elegans* Wormpep dataset (release WS150; http://130.209.234.35/blast/blast.html) to identify *H. contortus* loci homologous to *C. elegans* protein sequences. In order to avoid ambiguous or paralogous relationships, only BLAST alignments with *E*-value of $<1e^{-08}$ with the next best alignment *E*-value being >0.01 were considered. Regions which passed the searching criteria but were putatively homologous to a *C. elegans* protein that was a member of a protein family were discounted as it was not possible to clearly distinguish these relationships based solely on BLASTx analysis (Redman et *al.*, 2008).

2.7.5 Identification and characterisation of β-tubulin-containing sequences within Haemonchus contortus shotgun genomic sequence reads

Each of the six *C. elegans* β-tubulin protein sequences were also used as query in tBLASTn searching of the *H. contortus* shotgun sequence reads database which contains in excess of 800 Mb of sequence data (http://130.209.234.35/blast/blast.html). Only reads which gave a *C. elegans* β-

tubulin polypeptide as a top hit in reciprocal BLASTx searching of *C. elegans* Wormpep (release WS150, http://130.209.234.35/blast/blast.html) were considered as real hits.

All reads that passed the searching criteria were aligned to each of the fully sequenced and available genomic loci for *H. contortus* β-tubulin *isotypes-1*, *2* and *3* using BLAST2 (http://130.209.234.35/blast/wblast2.html). Outputs resulting in full length high scoring pair-wise alignments showing >90 % ID to only one of the three loci were designated as representing this *H. contortus* genomic locus.

2.7.6 PCR amplification, cloning and sequencing of cDNA sequences transcribed from the Haemonchus contortus β-tubulin isotype-3 and -4 loci

PCR primers were designed to the flanking regions of the *H. contortus* β-tubulin *isotype-3* locus, based on a gene model that was manually curated onto a contiguous 408,911bp stretch of *H. contortus* genomic DNA (primer sequences Appendix 2).

Alignments of many eukaryotic β-tubulin polypeptides with the nine *H. contortus* shotgun genomic sequence reads that represent the *H. contortus* β-tubulin *isotype-4* locus provides evidence that these sequences are insufficient to complete the protein encoding region for this locus. Based on this analysis, sequence information is missing for the initial ~600 bp of this transcribed region of this locus. Primers were designed to amplify the available transcribed region of this β-tubulin encoding locus (primer sequences Appendix 2).

PCR products obtained for the *H. contortus* β-tubulin *isotypes-3* and *-4* were cloned into vector pCR-4 (Invitrogen) and sequenced in triplicate for three plasmid clones (MWG). The sequences were assembled, respective to which of the two loci each represented, and a consensus was deduced for each (ContigExpress, Invitrogen; Appendix 2). The *isotype-3* consensus contained sequence flanking both the initiation and stop codons indicating that the

complete protein coding region for this locus had been obtained. Sequence data flanking the Hc-ISO-4 stop codon was generated, allowing determination of the C-terminal region for this β-tubulin protein.

2.7.7 Identification of β-tubulin loci from Wormbase datasets and genomic sequence databases for four species from the Caenorhabditis genus, and Pristionchus pacificus

Protein sequences for the six *C. elegans* β-tubulin family members were used in BLAST (tBLASTn and/or BLASTp) homology searching of the available sequence databases for the Clade V nematode species *C. briggsae*, *C. remanei*, *C. japonica*, *C. brenneri* and *P. pacificus* to identify β-tubulin sequences (http://www.wormbase.org/db/searches/blast_blat). Only sequences which contained enough data to putatively encode for a complete protein in addition to giving a *C. elegans* β-tubulin sequence as the top hit in reciprocal BLAST (BLASTx or BLASTp) searching of *C. elegans* Wormpep (release WS150, http://130.209.234.35/blast/blast.html), were considered as real hits. Wormbase contains manually curated datasets for all five of these species. In cases where such information was available, the protein encoded by the manually curated gene was taken to represent the locus. For instances where such data was not available, β-tubulin polypeptide sequences were deduced from gene models that were constructed based on high-scoring pairwise alignments to *C. elegans* β-tubulin proteins and conserved eukaryotic exon/intron boundary sequence information. This allowed putative homologous relationships to be inferred based on reciprocal BLASTp alignments to the *C. elegans* Wormpep dataset (release WS150, http://130.209.234.35/blast/blast.html).

2.7.8 Phylogenetic analysis

Sequences were aligned and phylogenetic trees constructed using MEGA version 3.1 (Kumar *et al.*, 2004). The statistical significance of all trees was tested by bootstrap analysis using 500 replicates.

2.7.9 Identification of nematode full-length β-tubulin protein sequences from the Uniprot database

All available Uniprot sequence databases were searched for β-tubulin sequences of species from the phylum Nematoda (http://www.uniprot.org/). Only sequences which had supporting transcript data were selected. Sequences from the free living nematode species *C. elegans*, *C. briggsae*, *C. japonica*, C. *remanei* and *C. brenneri* as well as those from *P. pacificus* and *H. contortus* were ignored as the β-tubulin gene families of these species were investigated separately. Using this criterion a total of 75 sequences were identified. Each of these polypeptide sequences contained both the β-tubulin auto-regulatory and signature motifs (Figure 5.9; Cleveland, 1988; Hesse *et al.*, 1987). Sequences which showed >99% pair-wise identity to another from the same species were identified as allelic (AlignX, Invitrogen).

2.8 Chapter 6 Materials and Methods

2.8.1 β-tubulin sequences within the Caenorhabditis elegans EST dataset

There are a total of 367, 044 *C. elegans* ESTs available for homology BLAST searching through the Wormbase webpage (release WS198; http://www.wormbase.org/db/searches/blast_blat). Each of the six *C. elegans* β-tubulin polypeptides were used as query in tBLASTx searching of this complete EST sequence dataset. This method identified a total of 481 *C. elegans* β-tubulin ESTs, which equates to 0.13 % of the entire database. All identified EST sequences contained regions encoding for the signature sequence pattern which distinguishes α, β and γ tubulin family members 'GGGTGSG' at residues 140-146 and/or the characteristic β-tubulin auto-regulation recognition element 'MREI' at codon positions 1-4 of the translated protein, in cases where EST sequence was available that encoded for either of these regions (Hesse *et al.*, 1987; Cleveland, D.W., 1988).

2.8.2 β-tubulin gene tags within the Caenorhabditis elegans SAGE datasets

SAGE datasets have been generated specific to each of the six *C. elegans* life-cycle stages (http://www.sagenet.org/; http://elegans.bcgsc.ca/home/sage.html). Identification of tags representing transcripts for each of the six *C. elegans* β-tubulin genes within each dataset was possible by simply selecting the required database to search and using the Wormbase accession number for each of the genes as the query term in turn (http://www.wormbase.org/). When searching each database, the output of each was normalised for 100, 000 tags. This option allows comparisons of the number of SAGE tags found for each gene at different life-cycle stages, irrespective of differences in the total number of tags in each database. A total of 451 *C. elegans* β-tubulin SAGE tags were identified across all six databases searched. The same number of SAGE tags was selected to represent each of the six *C. elegans* β-tubulin genes. Each tag selected was re-aligned to the *C. elegans* N2 genome (release WS150) to ensure that each was specific to one locus. Secondly, each tag used was aligned to each of the six *C. elegans* β-tubulin cDNA sequences (www.wormbase.org) to further ensure that each was specific to only one β-tubulin gene.

Chapter 3: The *Haemonchus contortus* genome sequencing project

3.1 Introduction

3.1.1 General introduction

The phylum Nematoda is split in to five distinct Clades (Blaxter, 1998; Figure 3.1), all of which harbour both parasitic and non-parasitic species. To date the vast majority of genomic sequence information stems from Clade V. This Clade incorporates the free living genetic model organism *C. elegans* along with *C. briggsae* and *P. pacificus*, all of which have fully sequenced genomic sequences available (*C. elegans* sequencing consortium, 1998; Stein *et al.*, 2003 and Dieterich *et al.*, 2008). The genus *Caenorhabditis* is the furthest advanced in terms of genomic sequence information available across the whole Nematoda phylum (Thomas, 2008). Clade V also includes the parasitic nematode *H. contortus* whose genome is currently being sequenced at the Pathogen Sequencing Unit of the Sanger Institute, Cambridge, UK.

Parasitic nematode research has still to fully enter the genomic era. The major limiting effect on this transition is the low number of fully sequenced genome sequences that are available for this class of species. To date there are only four parasitic nematode species with a genome sequencing project that is complete, or nearing completion; these are *B. malayi*, *M. incognita*, *M. hapla* and *P. pacificus* (Ghedin *et al.*, 2007; Abad *et al.*, 2008; Opperman *et al.*, 2008 and Zarlenga, *et al.*, 2008). Many parasitic nematode genome sequencing projects are planned to commence in the near future in an attempt to radically improve the impoverished state of the available parasitic nematode genomic sequence data. The majority of these projects will take place at the Pathogen Sequencing Unit of the Sanger Institute, UK (http://www.sanger.ac.uk/Projects/Helminths/) and at the Washington University Genome Sequencing Centre, USA (http://genome.wustl.edu/genome_group.cgi?GROUP=8).

The genome sequencing projects of parasitic nematode species are complicated. The majority of this complexity stems from the high levels of genetic variation that is present within nematode parasite populations (Gilleard and Beech, 2007). Within this class of species, the Trichostrongylid nematodes, of which *H. contortus* is an example, are predicted to be the most diverse (Blouin, 1998). Genetic variation is directly linked to population size and mutation rate (Anderson *et al.*, 1998). Both of these are huge in parasitic nematodes. Adult parasite burdens of *H. contortus* consist of many thousands of worms, with each female adult laying as many as 4,000 eggs per day (Coyne and Smith, 1992). In addition, due to the long branch lengths associated with many nematode phylogenetic studies the mutation rate is suggested to be incredibly high within this class of species (Anderson *et al.*, 1998; Blouin *et al.*, 1998 and 2000). Recent evidence has also been provided that *H. contortus* mating is polyandrous which again leads to an increase of genetic variation within populations (Redman *et al.*, 2008). The work contained within this chapter and thesis looks to address the extent to which genetic variation has inhibited the progress of the *H. contortus* genome project.

3.1.2 Current status of the Haemonchus contortus genome sequencing project

The *H. contortus* genome sequencing project was initiated in 2004 by the Pathogen Sequencing Unit of the Sanger Institute, UK (http://www.sanger.ac.uk/Projects/H_contortus/). The overall aim of the project is to produce a high quality reference genome for this species. Importantly, it is the MHco3 (ISE) strain of *H. contortus* that was selected for this project. This was the most inbred strain available when the project first commenced, however recent analysis has suggested that this strain is not as inbred as first thought (Roos *et al.*, 2003; Redman *et al.*, 2008).

There have been three main approaches to the *H. contortus* genome sequencing project. Originally, a BAC clone-by-clone sequencing approach was adopted as a way of generating large contiguous stretches of genomic DNA sequence. However, this approach was found to be problematic when uncertainty was found in deciding if BAC sequences were from the same locus of the genome.

The two other approaches have employed the Sanger capillary sequencing method. This technique generates paired-end reads of, on average, 600-800 bp in length. Originally this technique was used to generate 373,348, 198 bp of genomic DNA sequence from template recovered from a starting population of roughly fifteen mixed sex *H. contortus* worms. More recently, a total of 556,465,542 bp of *H. contortus* genomic DNA sequence has been generated from template obtained from a single male *H. contortus* worm. In total there is now more than 900 Mb of genomic sequence data available for this project. The only published estimate for the size of the *H. contortus* haploid genome size is ~53 Mb (Leroy *et al.*, 2003). Using this estimate, there is currently in excess of 10X coverage. The core sequence data, and subsequent homology based contiguous sequence databases assembled from each of these approaches are available via the Sanger ftp site for downloading, or on the BLAST server for homology based searching (ftp://ftp.sanger.ac.uk/pub/pathogens/Haemonchus/contortus/genome/; Table 3.1).

3.1.3 Contamination of genome sequencing project datasets

Many genome sequencing project datasets have problems with contaminating sequence data. An example of such a project is that of the filarial nematode parasite *B. malayi*. The majority of filarial nematode parasites, including *B. malayi*, harbour the obligate alpha-proteobacterial endosymbiont *Wolbachia* (McLaren *et al.*, 1975) in 100% of individuals within populations (Taylor and Hoerauf, 1999). As a result of this obligate relationship, as much as 15% of sequence data generated for the *B. malayi* genome sequencing project had to be discarded (E. Ghedin, The University of Pittsburgh, personal communication). *Wolbachia* is known not to reside within non-filarial nematodes (Bordenstein *et al.*, 2003) and as a result I can be almost certain that the genomic sequence datasets from the genome sequencing project of the trichostrongylid parasitic nematode *H. contortus* do not contain contaminating *Wolbachia* sequence. It is possible however that the *H. contortus* datasets do contain contaminating bacterial sequence data from other species.

The pH of the sheep abomasum rises during *H. contortus* infection (Simcock *et al.*, 2005). This rise in pH results in an increased number of viable bacteria present in parasitized sheep (Urquhart *et al.*, 1966 and Simcock *et al.*, 1999). Bacterial contamination of *H. contortus* populations that have been used for experimental procedures has previously been reported (Simcock *et al.*, 2005). I have used the assembled genome of the bacterial *E. coli* strain H10407 (http://www.sanger.ac.uk/Projects/E_coli_H10407/) to investigate the *H. contortus* shotgun genomic sequence reads database hc_all_reads_112607.fasta for any evidence of contaminating bacterial sequence.

Another possible source of contamination of genomic sequence datasets for parasitic species is with host DNA. Parasitic nematode specimens that are used to generate sequence databases are de-contaminated as a precautionary step; however, this does not always prevent contamination with host DNA. *B. malayi* is, again, an example of such a project where host sequence information has been a serious issue (T. Unnasch, The University of Alabama at Birmingham, personal communication). As *B. malayi* passes through all four of its life cycle stages within one of two host species it is not difficult to envisage the reasons behind host contamination being an issue with this genome sequencing project (http://www3.niaid.nih.gov/news/newsreleases/2007/brugiaImage.htm). In the case of *H. contortus*, all genomic sequence datasets that have been generated thus far have used worms at the adult stage of the life cycle. At this stage *H. contortus* worms reside within the abomasum of their ovine host (http://www.goatbiology.com/animations/haem.html). I have used the wealth of genome sequence data from the bovine genome sequencing project (http://www.ncbi.nlm.nih.gov/sites/entrez?db=genomeprj&cmd=Retrieve&dopt=Overview&list_uids=10708) to investigate the possibility of ovine genome sequence contamination within the *H. contortus* genome sequencing project datasets. The bovine and ovine genome sequences show 96% and 92% homology within coding and non-coding regions, respectively (J. McEwan, AgResearch, personal communication). I have chosen to use the *Bos taurus* (*B. taurus*) genomic sequence dataset, release bos.Tau4 (ftp://hgdownload.cse.ucsc.edu/goldenPath/bosTau4/bigZips/) due to the far

greater amount of data that is available for this project in comparison with the ovine genome sequencing project at this time.

3.1.4 Single cluster EST sequence representation within Haemonchus contortus genomic sequence databases

The size of the *H. contortus* haploid genome has been estimated by flow cytometry at ~53 Mb (Leroy *et al.*, 2003). The flow cytometry method of haploid genome size estimation is widely used however it can be variable, leading to incorrect estimates. Examples of this can be found in the same publication in which the *H. contortus* haploid genome was estimated at ~53 Mb. The *B. malayi* and *M.incognita* haploid genome lengths were estimated at 81 and 47 Mb respectfully (Leroy *et al.*, 2003). The draft haploid length of the *B. malayi* genome is 90 Mb (Ghedin *et al.*, 2007) and the completed haploid genome length of *M.incognita* is 86 Mb (Abad *et al.*, 2008). It is therefore important that flow cytometry estimates of haploid genome length are treated as crude estimates.

There are a number of *H. contortus* MHco3 (ISE) EST sequence databases publicly available. The databases differ by the life cycle stage analysed and/or what type of tissue was used to generate the sequence database. I have used the information in all *H. contortus* MHco3 (ISE) EST databases (available at http://www.nematodes.org/nematodeESTs/nembase.html) to evaluate the representation and coverage of EST sequences within the genomic sequences databases of the *H. contortus* genome sequencing project.

3.1.5 Allelic sequences within Haemonchus contortus genomic sequence datasets

There are high levels of natural genetic variability between allelic sequences within parasitic nematode populations, with the Trichostrongylid nematode species the most variable of this class (Blouin, 1998). The most recent genomic sequence datasets released from the *H. contortus* genome sequencing project (http://www.sanger.ac.uk/Projects/H_contortus/) have been generated using only a single worm as starting material. This strategy of single worm genome sequencing was employed for the *H. contortus* genome sequencing project due

to poor assembly statistics using 547,477 shotgun genomic sequence reads generated from a starting population of fifteen worms (Table 3.2). As *H. contortus* is a diploid organism, DNA sequence data generated from a single *H. contortus* male worm should consist of a maximum of two copies per autosome locus and the X-chromosome should only be single copy. A widespread method used to evaluate homology based contiguous sequence databases is to use the N50. The N50 is reported in kb and is calculated by first ordering all sequences by size and then adding the lengths of each (starting with the longest). The N50 is taken as the length of last sequence added that results in the summed length exceeding 50% of the total length of the sequence database. The N50 for this assembled sequence database is 1.456 kb, using 373.3 Mb of genomic sequence. The emphasis on the poor assembly of this sequence information can be seen when it is considered that in using only 400,000 *C. briggsae* shotgun genomic sequence reads, a contig database with an N50 of 2.3 kb was assembled (http://korflab.ucdavis.edu/Datasets/genome_completeness/). Natural genetic variation between *H. contortus* sequence reads from the same genetic locus is thought to be the main reason for the poor assembly of this sequence information.

Recently, a database was assembled at the Pathogen Sequencing Unit of the Sanger Institute using 803,014 shotgun genomic sequence reads generated from a single *H. contortus* worm. This database was assembled using the same software and parameters used to assemble the sequence data generated from fifteen *H. contortus* worms. Comparing the assembly statistics of both databases, it is noticeable that the N50 of the single worm contig database is larger than that of the fifteen worm dataset (Table 3.2). However, an N50 of 2.032 kb is still considered to be extremely low for the amount of sequence information available (M. Berriman, Wellcome Trust Sanger Institute, personal communication).

A potential reason for the poor assembly statistics associated with all *H. contortus* genomic sequence databases is that the level of genetic variation within even a single worm is far too high to confidently build large stretches of contiguous sequence. If the two alleles of each locus present in a single *H. contortus* worm are sufficiently different that they cannot assemble together to

form a contiguous sequence, it can be hypothesised that they would assemble independently.

3.1.6 Aims of this chapter

There are four basic possible reasons for the poor assembly statistics currently associated with the *H. contortus* genome sequencing project:

1. Genome is far greater in size than the current estimate

2. Natural genetic variation is preventing allelic sequences from assembling

3. There are contaminating sequences from other genomes in current databases

4. AT richness, repetitive sequences, etc.

This chapter shall aim to address points 1-3, evaluating the impact that each is exerting to influence the poor assembly of the ~850 Mb of *H. contortus* genomic sequence that is currently available (ftp://ftp.sanger.ac.uk/pub/pathogens/Haemonchus/contortus/genome/).

3.2 Results

3.2.1 Searching of the Haemonchus contortus genome sequence databases for evidence of contamination

3.2.1.1 Bacterial contamination

Searching of 1,126,320 *H. contortus* shotgun genomic sequence reads using the 5,153,435 bp genome of *E. coli* strain H10407 resulted in 43 significant BLAST alignments. All 43 *H. contortus* genomic sequence reads aligned to 9 different regions of the bacterial genome. 43 reads are equivalent to 0.0000381% total of the *H. contortus* genomic shotgun sequence read database hc_allreads_112607.fasta. In using each of the 9 target regions of the *E. coli* genome that were identified by my BLAST searching as query sequences against the NCBI nucleotide collection database (http://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Nucleotides&PROGRAM=blastn&MEGABLAST=on&BLAST_PROGRAMS=megaBlast&PAGE_TYPE=BlastSearch&SHOW_DEFAULTS=on&LINK_LOC=blasthome), none of the regions encoded for known genes or transposable elements or were from regions of simplistic sequence. The average length of the 43 *H. contortus* genomic shotgun sequence reads identified is 364 bp. This compares with an average of 668 bp for the whole database. As only 0.0000381% of the all *H. contortus* genomic shotgun reads were identified as showing significant homology to the bacterial genome of *E. coli* strain H10407, this indicates that there is no significant bacterial contamination of the *H. contortus* genomic shotgun sequence reads database hc_allreads_112607.fasta.

3.2.1.2 Host contamination

Of the 297,710 *H. contortus* shotgun genomic sequence reads selected at random to be used as query sequences to search the 2.73 Gb of *B. taurus* assembled genomic sequence (release bos.tau4), 382 showed alignments of >100bp at >90% identity. All 382 *H. contortus* genomic sequence reads aligned to only 4 different loci of the *B. taurus* genome. Details of the alignments, and what each *B. taurus* genomic region is comprised of are shown in Table 3.3. Two of the four *B. taurus*

loci have been annotated as repetitive sequence. Although searches were performed using the automated simplistic sequence filter, it is not uncommon for repetitive regions to remain within query sequences. The other two *B. taurus* genomic loci that showed significant homology to the *H. contortus* shotgun genomic sequence reads are both ribosomal RNA genes. These genes are present in all eukaryotic ribosomal DNA sequences at high levels of sequence identity (Długosz E. and Wiśniewski M., 2006). As a result it is not unsurprising, and perhaps reassuring, that these sequences were identified. As these were the only four *B. taurus* loci identified I have no evidence to suggest that the *H. contortus* genomic shotgun sequence reads database hc_allreads_112607.fasta is contaminated with any host DNA sequence information.

3.2.2 Using single cluster Haemonchus contortus MHco3 (ISE) EST DATA to investigate Haemonchus contortus genome coverage

As there is only one published estimate of the size of the *H. contortus* genome, it remains unclear to what extent current genomic sequence databases represent this sequence (Leroy *et al.*, 2003). It can be reasonably assumed that, if the current databases represent a significant coverage of the genome, then the majority of available sequenced, single cluster *H. contortus* MHco3 (ISE) EST sequences should be contained within such databases. The aim of this work was to firstly investigate the representation of 5,906 single cluster *H. contortus* MHco3 (ISE) EST sequences within the current *H. contortus* genomic sequence databases; and then to take a small random sample of these sequences and subject these to thorough, detailed analysis in order to investigate the coverage of each within the *H. contortus* genomic sequence reads database.

Significant alignments were found for 5,657 of the 5,906 (95.8%) *H. contortus* single cluster EST sequences used to search 1,126,320 *H. contortus* shotgun genomic sequence reads (Table 3.4). In using the same 5,906 *H. contortus* single cluster EST sequences to search the *H. contortus* genomic sequence database combined_worms_contigs_121107.fasta, 5,269 (89.2%) sequences showed a significant alignment (Table 3.4). This database contains homology based

contiguous sequences assembled using 780,947 of the 1,126,320 (69.3%) shotgun genomic sequence reads within database hc_allreads_112607.fasta.

Each of the 19 *H. contortus* single cluster EST sequences subjected to detailed coverage analysis were used as query sequence in searching the hc_allreads_112607.fasta database, and each resulted in at least one significant alignment. Details of each of the 19 EST sequences and respective database hits can be found in Table 3.5. The average number of genomic shotgun sequence reads aligning to each EST was 8.5. This result, based on these 19 randomly selected EST sequences, suggests that >80% of the transcribed sequence from within the *H. contortus* genome sequence is present in the shotgun genomic sequence database hc_allreads_112607.fasta.

3.2.3 Allelic sequences within Haemonchus contortus genomic sequence datasets

A single male *H. contortus* is known to be diploid (Redman *et al.*, 2008). Therefore, in a shotgun sequence read database generated from only single worm data, I would expect there to be, at a maximum, two allelic copies of each autosomal locus, and that the X-chromosome should be clonal. I have investigated the extent to which the allelic copies of each autosomal locus are able to assemble into one contiguous sequence by searching for the presence of allelic copies of the same locus within a homology based contiguous sequence database, generated using only shotgun sequence read data from a single *H. contortus* male worm.

40,426 of the 87,735 *H. contortus* genomic sequences within the single worm contiguous sequence database hc_sw_contigs200808.fasta show a significant alignment to another sequence in the database. 40,426 sequences equates to 46.1% of the database. Not all loci across both allelic copies of the genome present in a single *H. contortus* will be sufficiently variable to assemble into distinct contiguous sequences. In addition, this sequence database does not contain enough data to sufficiently cover the *H. contortus* haploid genome twice. Therefore it is not possible that the database contains two sequences for each locus. In spite of these caveats, with 46.1% of the sequences within the

database significantly aligning to another within the same database I can confidently hypothesise that these sequences are allelic. The total length of sequence within database hc_sw_contigs200808.fasta is 166.5 Mb. The length of all sequences that do not find a significant hit to another sequence within the database plus half of the total length of all sequences which do find another match within this database give a total of 128.3 Mb. Therefore 128.3 Mb represents the estimated length of unique sequence within this sequence database.

3.3 General Discussion

The results in this chapter support the hypothesis that the major underlying reason for the *H. contortus* shotgun genomic sequence read data failing to assemble into large contiguous sequences is that high levels of genetic variation between allelic sequences exist both between and within *H. contortus* worms.

No evidence of sequence contamination was found within the *H. contortus* genomic sequence databases. Although 43 shotgun genomic sequence reads were identified to be highly similar to bacterial sequence, it is likely that these sequences are artefacts of high throughput automated 'clipping' of shotgun genomic sequence read databases. This technique is employed to rid sequence databases of any bacterial sequence. However, on such a large scale, it is almost inevitable that some sequence will be missed. All *H. contortus* sequences identified as being highly similar to bovine genomic sequence aligned to either repetitive regions or ribosomal RNA gene loci of the bovine genome. Nematode parasite genomes are known to be rich in terms of simplistic sequence (i.e. repetitive sequence, transposable elements, etc.) (Ghedin *et al.*, 2007; Abad, P., *et al.*, 2008; Opperman, C.H., *et al.*, 2008 and Zarlenga, D.S., *et al.*, 2008) and the ribosomal genes identified reside within the ribosomal DNA sequence of all eukaryotic species, at very high levels of sequence identity (Długosz and Wiśniewski, 2006).

Significant alignments were found within the *H. contortus* shotgun genomic sequence reads for 95.8% of all available *H. contortus* single cluster EST sequences. The average representation of a small subset of these EST sequences was found to be more than 80%. This suggests that the sequence data for the vast majority of all transcribed loci within the *H. contortus* genome is present within the available genomic sequence databases. As a result, it can be concluded from this work that it is possible to use the available sequence databases to confidently piece together gene pathways and families from within the *H. contortus* genome. This makes the current *H. contortus* genomic sequence datasets useful tools for researchers, as at least partial sequence would be expected to be found for every locus from the genome. However, at the current time, doing such is laborious as small fragments have to be pieced

together. This is mainly due to problems with the automated assembly of the *H. contortus* sequence data, arising from the high level of sequence variation present within *H. contortus* populations (Blouin, 1998).

A difference of only 6.6% was found for the *H. contortus* single cluster EST representation percentages between the *H. contortus* shotgun genomic sequence reads database and a sequence database containing the homology based assembly of these sequences. There are 345,373 sequences within the *H. contortus* shotgun genomic sequence reads that were not incorporated into the assembly database. As an additional ~30% of shotgun genomic sequence read data results in an increase of only 6.6% single cluster EST representation, it may be suggested that the assembled sequence database from the *H. contortus* genome sequencing project is starting to plateau in terms of new transcribed genomic loci sequence information. A potential caveat with this analysis is that it compares unassembled and assembled genomic sequence databases. In order to fully investigate this suggestion it would be important to repeat this analysis once the next batch of shotgun genomic sequence read information has been released, using genomic sequence databases which contain similar types of data.

With large population sizes and high mutation rates, a high level of natural genetic variation is expected between alleles of parasitic nematode genome sequences (Anderson *et al.*, 1998; Gilleard and Beech, 2007). Of this class, the trichostrongylid parasitic nematodes are predicted to be most variable (Blouin, 1998). It is hypothesised that the difficulties in assembling the shotgun genomic sequence reads, generated from a starting population of fifteen worms, for the *H. contortus* genome sequencing project are primarily due to this genetic variation (J.S. Gilleard, University of Calgary, and M. Berriman, Wellcome Trust Sanger Institute, personal communication). The latest *H. contortus* shotgun genomic sequence read databases have been generated from a single worm in order to minimise the effect that this variation has on the assembly of the *H. contortus* genome sequence (http://www.sanger.ac.uk/Projects/H_contortus/). The N50 of the homology based assembly of the single worm shotgun genomic sequence data is larger than that of the multi-worm (Table 3.2). However, this increase is not as much as expected for the amount of genomic sequence data that the shotgun genomic sequence read databases contain (M. Berriman, Wellcome Trust Sanger Institute, personal communication). The results

presented in this chapter suggest that the genetic variation within a single *H. contortus* worm is sufficient for allelic sequences to assemble independently. 46.1% of the 87,735 assembled sequences using shotgun genomic sequence reads generated from a single *H. contortus* worm have another significantly homologous sequence within the same database. The total length of all sequence data within this database is insufficient to sequence the complete haploid genome of *H. contortus* at each locus twice. In addition to this, not all loci within the *H. contortus* genome will be sufficiently variable to assemble independently. It is therefore my conclusion that the 46.1% of sequences identified represents allelic sequences that are sufficiently different to assemble as independent sequences. This work has allowed us to estimate the length of unique sequence contained within this database at 128.3 Mb. The sequenced haploid genome length of the genetic model organism *C. elegans* is 97 Mb (*C. elegans* sequencing consortium, 1998). Both *C. elegans* and *H. contortus* reside within Clade V of nematode phylogeny (Blaxter, 1998; Figure 3.1). There are often more, and larger, introns within *H. contortus* gene models compared with that of the *C. elegans* homologue (Redmond *et al.*, 2001; Liu *et al.*, 2004; Britton and Murray, 2002 and Chapter 6). It is therefore unsurprising that my estimation of unique genomic sequence assembled from shotgun genomic sequence reads from a single *H. contortus* worm is larger than the haploid genome estimated by flow cytometry (Leroy *et al.*, 2003) or that of *C. elegans*.

The conclusion that the 46.1% of sequences within the assembled sequence database are allelic, using only shotgun genomic sequence data generated from a single *H. contortus* worm, is based solely on sequence homology. In this analysis no attempt is made to investigate these sequences for gene duplication events, gene families or for the presence of simplistic sequence. The fact that the general structure of the *H. contortus* genome is poorly understood, at this time, means that the conclusions drawn from the data presented in this chapter must be considered preliminary. A possible avenue to further test the conclusions made would be to investigate the sequences identified as being allelic for regions of coding sequence. The close phylogenetic relationship and wealth of data available for the free living nematode *C. elegans* could be exploited in this regard. There are many instances, and many more cases predicted to be found, where the proteins encoded by putative homologous

genes from the genomes of *H. contortus* and *C. elegans* have very similar sequence (data not shown). If the 46.1% of sequences predicted to be allelic are so, it could be predicted that in many cases both sequences would contain regions that are highly homologous to the same *C. elegans* protein. Such analysis would aid in our understanding of the genetic diversity between allelic sequences within a single *H. contortus* worm. This basic understanding is of critical importance for the development and production of a fully sequenced haploid genome sequence for this organism.

Overall, the data presented in this chapter suggests that both polymorphism and genome size are contributing to the problems seen when assembling *H. contortus* genomic sequence data. From these results it is apparent that, at ~53 Mb, the *H. contortus* haploid genome is underestimated (Leroy *et al.*, 2003). However, with the advent of next generation sequencing technologies, it is not as difficult, or as expensive, to address this situation as it may have been in the past. What is more difficult to address is the issue of polymorphism. Recent analysis suggests that the *H. contortus* MHco3 (ISE) strain is not as inbred as originally thought (Redman *et al.*, 2008). It is an important consideration that the template for future *H. contortus* genomic sequencing projects may have to be changed for a more inbred strain. Furthermore, examples of such strains are now available in comparison to when this genome sequencing project originated. In order to achieve large contiguous stretches of *H. contortus* genomic DNA data, it may prove fruitful to select one of these strains as template for the next generation sequencing technologies to act upon.

3.4 Tables and Figures

Table 3.1: Sizes of genomic sequence databases generated thus far for the *H. contortus* genome sequencing project.

Sequencing strategy	Database name	Total number of sequences	Total size of database (bp)
Shotgun multi-worm capillary sequencing	HAEM_031105.reads	547,477	373,348,198
Shotgun single-worm capillary sequencing	single_worm.fasta	803,014	556,465,542
BAC sequence database	Hc_bac_contigs.fasta	22	2,244,527

Table 3.2: Assembly statistics for the *H. contortus* multi- and single worm contiguous sequence databases.

Database	Number of worms	Number of reads	Number of assembled reads	Number of contigs	Total assembled length (Mb)	N50 (kb)
Multi-worm	15	547,477	305,748	58,172	93	1.456
Single worm	1	803,014	498,165	87,735	167	2.032

These assembly statistics were generated and provided by the Pathogen Sequencing Unit of that Sanger Institute, Cambridge, UK (M. Berriman, Wellcome Trust Sanger Institute, personal communication).

Table 3.3: Details of the 382 *H. contortus* genomic shotgun sequence reads aligning to the *B. taurus* genome.

Locus of *B. taurus* genome		Number of aligning *H. contortus* reads	Details of *B. taurus* genomic locus
Chromosome	Region (bp)		
4	16197-16509	90	28s rRNA gene
14	13238132 – 16238300	4	Repetitive region
21	520442265 – 52043089	69	Repetitive region
27	7419065 – 7419338	219	18s rRNA gene

Of 296,710 *H. contortus* shotgun genomic sequence reads used to search the 2.73 Gb of *B. taurus* assembled genomic sequence, 382 resulted in significant alignments. All 382 *H. contortus* sequences aligned to 4 loci of the *B. taurus* genome.

Table 3.4: Representation of 5,906 *H. contortus* single cluster EST sequences within the genomic sequence databases hc_allreads_112607.fasta and combined_worms_contigs_ 121107.fasta.

Name of database	hc_allreads_112607.fasta	combined_worms_contigs_121107.fasta
Number of sequences in database	1,126,320	107,948
Average length of sequences in database (bp)	668	1,985
Number of EST sequences showing a significant alignment	5,657	5,269
% of EST sequences showing a significant alignment	**95.8**	**89.2**

5,906 single cluster *H. contortus* EST sequences were used to search the genomic sequence databases hc_allreads_112607.fasta and combined_worms_contigs_121107.fasta. Hits were found for 95.8% and 89.2% of the query sequences, respectively.

Table 3.5: Coverage of 19 single cluster *H. contortus* EST sequences in shotgun sequence read database hc_allreads_112607.fasta

19 single cluster *H. contortus* EST sequences were used to BLAST search 1,126,320 shotgun genomic sequence reads from the on-going genome sequencing project. Putative homologous sequences were analysed for overlapping sequences and subsequently used to calculate the percentage coverage and identity for each EST sequence.

Single cluster *H.contortus* EST sequence	Length of *H.contortus* single cluster EST (bp)	Number of *H.contortus* sequence reads in BLASTN searching	Final number of Contig Express sequences aligning to EST	% coverage	% identity
HCC00054	1317	16	4	87.6	98.7
HCC00055	1249	1	1	41.6	95
HCC00057	1037	5	1	99.5	91
HCC00058	801	11	1	87	92.5
HCC00121	708	6	1	63	92
HCC00241	741	5	3	53.6	96.6
HCC00250	758	7	2	92.7	99.8
HCC00262	752	14	4	88.7	99.5
HCC00357	676	9	1	64.9	99.2
HCC04212	544	12	1	95.6	96.4
HCC04224	560	11	3	94.1	94.7
HCC04666	677	6	1	78.6	94.8
HCC04679	671	13	2	97.9	96.3
HCC04734	617	3	2	92.5	96.6
HCC04743	626	5	1	87.7	95.7
HCC04744	563	10	2	87	92.9
HCC04753	596	9	2	93.6	96
HCC04754	640	13	2	80.3	99.4
HCC06548	741	5	2	56.3	95.7
TOTAL 19	14274	161	36	1542.2	1822.8
AVERAGE	**751.2**	**8.5**	**1.9**	**81.2**	**95.9**

Table 3.5 Coverage of 19 *H. contortus* single cluster EST sequences within the genomic sequence database hc_allreads_112607.fasta

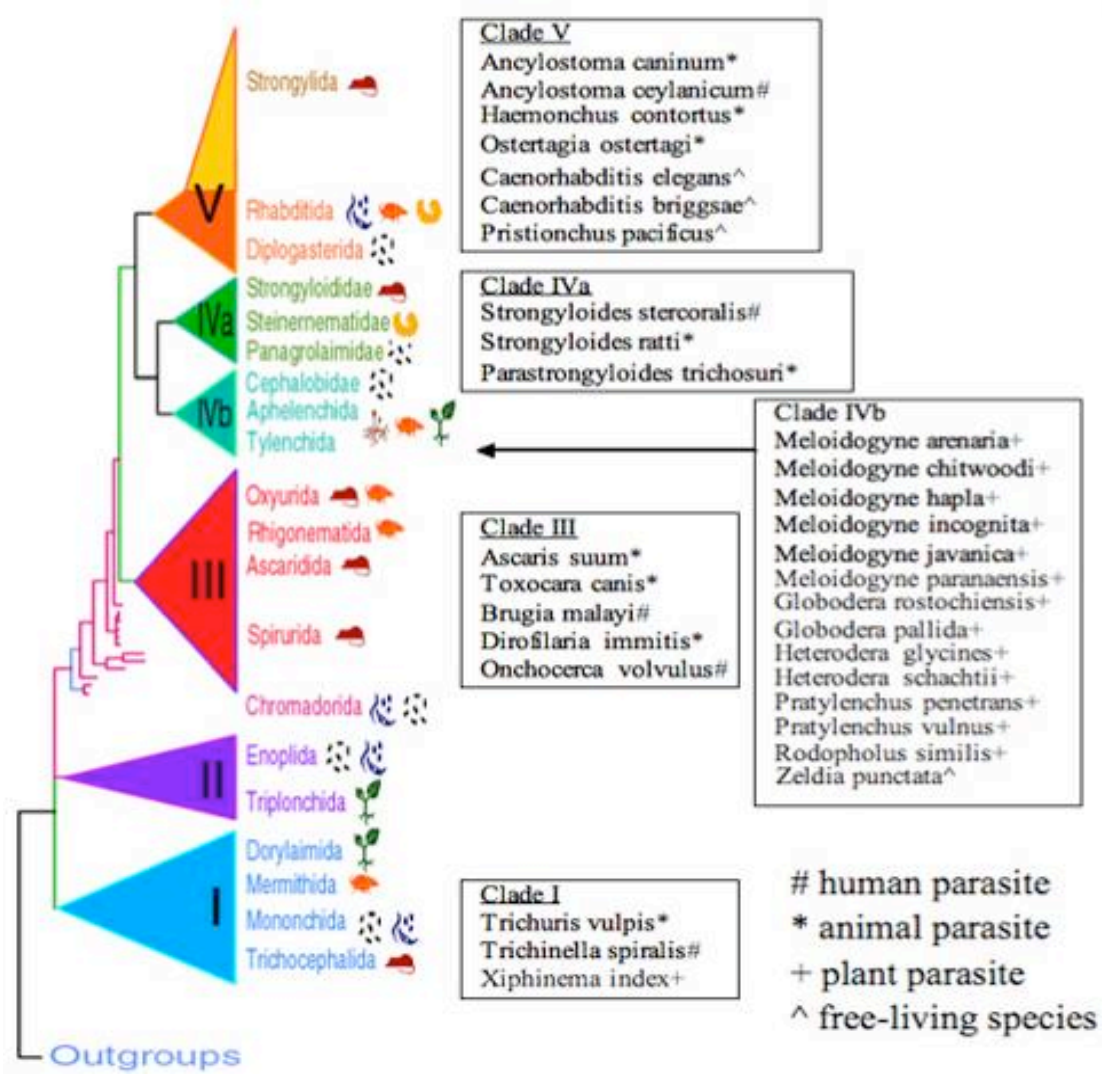


Figure 3.1: The phylum Nematoda is split in to five distinct Clades (Blaxter *et al*., 1998).

Chapter 4: Identification of putative homologues of *Caenorhabditis elegans* classical RNA_i pathway associated proteins

4.1 Introduction

RNA interference (RNAi) is a reverse genetic technique commonly employed to investigate gene function in a number of organisms including protozoan (Ullu *et al*., 2004), amphibian (Oelgeschlager *et al*., 2000), insect (Kennerdell and Carthrew, 1998) and mammalian (Hannon, 2002) species. This technique was first characterised in *C. elegans* (Fire *et al*., 1998). Since being discovered, the molecular mechanism underpinning RNAi in *C. elegans* has been extensively investigated. Through detailed investigations of RNAi deficient strains of *C. elegans*, the majority of the essential proteins involved in the mechanism are now known (Grishok, 2005). Briefly, the classical RNAi pathway is induced by endogenous dsRNA. This dsRNA is then processed into siRNAs by the Dicer complex, which comprises the *C. elegans* DCR-1, RDE-4 and RDE-1 proteins. It is the siRNAs that are incorporated into the RNA induced silencing complex (RISC), by the RDE-1 and DRH-1 proteins. RISC molecules are those which bind to the target mRNA, which is processed for mRNA degradation. A simplified model of the classical *C. elegans* RNAi mechanism is shown in Figure 4.1.

Although RNAi is a well defined and optimised reverse-genetic method of gene function for *C. elegans*, the success of this technique to investigate most other species of helminth is variable. This is most notable by the fact that no other species of the *Caenorhabditis* genus seems to be as amenable to this technique as *C. elegans* (Winston *et al*., 2007). The technique is also variable across the majority of the parasitic helminth species for which the method has been attempted (Table 4.1). The variability of this technique is speculated to be due to the loss of, or differences in, one or more members of the classical *C. elegans* RNAi pathway (Geldhof *et al*., 2007; Figure 4.1). The protein products of genes known to be involved in the *C. elegans* classical RNAi pathway can be split into two main areas; those involved in the cell autonomous, and those involved in the cell non-autonomous pathways. These groups are namely; DCR-1, DRH-1, RDE-1,

RDE-3 and RDE-4 for cell-autonomous and RDE-2, SID-1 and SID-2 for non-autonomous RNAi. The results described in this chapter shall be discussed in reference to both routes of the classical RNAi pathway.

4.1.1 Aims of this chapter

There are three main aims for this chapter:

1. Investigate the available genomic sequence databases for the nematode species *C. briggsae*, *P. pacificus*, *B. malayi* and *D. melanogaster* for the presence of homologues to the 10 core genes involved in the classical *C. elegans* RNAi pathway

2. Carry out a detailed investigation of the available *H. contortus* genomic sequence databases for the presence of homologues to 17 genes involved in the classical *C. elegans* RNAI pathway

3. Conduct a comparative analysis of results obtained

4.2 Results

4.2.1 Searching the Caenorhabditis briggsae, Drosophila melanogaster, Pristionchus pacificus and Brugia malayi sequence datasets for putative homologues to Caenorhabditis elegans classical RNA$_i$ pathway associated proteins

4.2.1.1 Caenorhabditis briggsae

C. briggsae is a free living nematode also belonging to the *Caenorhabditis* genus (Blaxter, 1998). The haploid genome sequence of *C. briggsae* is complete and freely available (Stein *et al.*, 2003). The annotation of this genome is at an advanced stage (http://www.wormbase.org/db/seq/gbrowse/c_briggsae/). Based on neutral mutation rate analysis it has been predicted that *C. elegans* and *C. briggsae* shared a common ancestor <30 Mya (Cutter, 2008). If putative homologous proteins for the *C. elegans* classical RNAi pathway could not be identified in another nematode from the *Caenorhabditis* genus by BLAST homology searching (Altschul *et al*, 1990) it would be unlikely that identification in other species would be possible using sequence homology alone. BLASTp searching of the *C. briggsae* protein database release WS195 (http://www.wormbase.org/db/searches/blast_blat) was able to identify putative homologues to all *C. elegans* proteins that were used as query sequences (Table 4.2). In all cases the putative homologous *C. briggsae* protein was used as a probe in BLASTp searching of the *C. elegans* protein database release WS150 (http://130.209.234.35/blast/blast.html) to ensure that the putative homologous protein was the top hit in the reciprocal search.

RNAi in *C. elegans* works by soaking the worms in media that contains dsRNA, feeding the worms on bacteria that express dsRNA or by micro-injection of the dsRNA into the syncytical arm of the gonad of individual *C. elegans* worms (Strange, 2006). However, RNAi works only by the micro-injection method when using *C. briggsae* (Winston *et al.*, 2007). This is indicative that *C. briggsae* has the basic machinery necessary for RNAi but lacks the components necessary for

dsRNA transport/uptake from the gut/environment. It has been demonstrated that transgenic *C. briggsae* which express the *C. elegans* trans-membrane SID-2 protein are capable of uptake of dsRNA from the environment (Winston *et al.*, 2007). In my analysis, and that of Winston and colleagues, a putative *C. briggsae* SID-2 protein was identified (Table 4.2). The amino acid sequences of two proteins show 41 % identity and 59 % similarity to one another. This is the lowest identity and similarity values shown between two putative RNAi mechanism involved proteins from these species. Comparative genomic investigations work on the premise that when homologous sequences are identified in species that are closely related, function as well as sequence will have been conserved. However this example of SID-2 serves as a reminder that this is not always the case. To summarise, all components of the RNAi pathway, with the exception of SID-2, are well conserved between *C. elegans* and *C. briggsae*.

4.2.1.2 Drosophila melanogaster

The fruit fly *D. melanogaster* is a genetic model organism with a completed and fully annotated haploid genome sequence available (Adams *et al.*, 2000; http://flybase.org). *D. melanogaster* belongs to the Diptera Clade of eukaryotic phylogeny and therefore represents a classical out-group in my analysis. Using BLASTp analysis, putative homologous proteins were identified for seven of the ten *C. elegans* classical RNAi associated proteins used as query sequences to search the protein dataset of *D. melanogaster*. Putative homologous sequence for only the RRF-3, SID-1 and SID-2 proteins could not be identified (http://flybase.org/blast/; Table 4.2). Each identified *D. melanogaster* protein was used to search the *C. elegans* Wormpep protein database release WS150 (http://130.209.234.35/blast/blast.html) to ensure that the top hit of each reciprocal BLAST hit was the original *C. elegans* protein searched for. The inability to putatively identify *D. melanogaster* proteins homologous to the *C. elegans* proteins SID-1 and SID-2 is unsurprising as systemic RNAi does not function efficiently in *D. melanogaster* (Misquitta and Paterson, 1998; Kennerdell and Carthew, 1998). In addition it has been reported that *D. melanogaster* S2 cells which express the *C. elegans* SID-1 protein are able to exhibit comparable RNAi phenotypes as control, untransfected cell lines, using up to 10,000-fold lower concentrations of dsRNA (Feinberg and Hunter, 2003). The fact that sufficient sequence homology was found for seven of the ten *C.*

elegans proteins used to search the *D. melanogaster* protein dataset suggests that, if present, sufficient homology should exist between *C. elegans* proteins and putative homologues within the datasets of all other nematode species investigated for this analysis, due to their closer phylogenetic relationships.

4.2.1.3 Pristionchus pacificus

P. pacificus resides within Clade V of nematode phylogeny and is predicted to have last shared a common ancestor with *C. elegans* 100-200 million years ago (Blaxter, 1998; Pires da Silva and Sommer, 2004). There is only one brief recording of successful RNAi for this organism. This report used the micro-injection method of dsRNA delivery into the organism which was targeted against the sex determining gene *tra-1* (Pires da Silva, 2006). The draft sequence of the 169Mb haploid genome for the nematode *P. pacificus* was searched for putative homologous sequences to ten *C. elegans* proteins known to function within the classical RNAi pathway (Dieterich *et al.*, 2008; Table 4.2). tBLASTn searching of genome assembly version 'P.pac California assembly 2' was performed using the online BLAST web page available on the *P. pacificus* genome website (http://www.pristionchus.org/cgi-bin/blast_iframe.pl). Putative homologous sequences were identified for six of the ten *C. elegans* proteins used as query sequence (Table 4.2). Reciprocal BLAST analysis was carried out to confirm that the original identifying *C. elegans* protein was the most similar protein from this species to the identified *P. pacificus* sequence(s) (release WS150 (http://130.209.234.35/blast/blast.html).

No *P. pacificus* sequence putatively homologous to the *C. elegans* RDE-4 protein could be identified. This suggests that the classical RNAi pathway machinery is either non-existent or is different to that of the *C. elegans* pathway, as this protein is essential for RNAi in *C. elegans* (Tabara, 1999). As RNAi is not a reliable silencing technique in *P. pacificus*, a different reverse genetic technique of antisense morpholino oligonucleotide injection is the technique of choice for investigating gene function for this nematode species (Pires da Silva and Sommer, 2004). Morpholinos exert a reverse genetic effect by blocking translation of their target RNAs and as a result rely on a different set of proteins to those that function within the classical RNAi pathway (Heasman, 2002). However, as successful RNAi for this organism has been reported, the absence

from the *P. pacificus* genome of an identifiable putative homologue to the *C. elegans* protein RDE-4 is surprising (Pires da Silva, 2006).

In addition, sequence putatively homologous to the *C. elegans* RDE-2 or SID-2 proteins could not be identified in this analysis. Similar to RDE-4, the *C. elegans* RDE-2 protein is necessary for functional RNAi investigation (Tabara *et al.*, 1999). The absence of a recognisable SID-2 homologue suggests that systemic RNAi is non-functional in this organism (Winston *et al.*, 2007). With these four proteins suspected to be missing from the genome of *P. pacificus*, it is puzzling that functional RNAi has been reported for this organism (Pires da Silva, 2006). However, what is unknown from the literature is the number of unsuccessful attempts that have been made to use RNAi as a technique to investigate *P. pacificus* gene function.

4.2.1.4 Brugia malayi

Based on sequence analysis of small subunit RNA genes, *B. malayi* is placed within Clade III of nematode phylogeny (Blaxter *et al.*, 1998). RNAi was shown to silence a number of genes in adult stage worms, including the house keeping genes *Bm-tub-1*, *Bm-ama-1* and the microfilarial sheath protein encoding *Bm-shp-1* gene (Aboobaker and Blaxter, 2003; Table 4.1). However, since this initial success there have been no subsequent reports of RNAi using this species (E. Devaney, University of Glasgow, personal communication). The draft genome sequence of *B. malayi* has been sequenced and is freely available (Ghedin *et al.*, 2007; http://www.tigr.org/tdb/e2k1/bma1/). The annotation of this genome sequence is at an advanced stage and as a result a protein database is available (http://blast.jcvi.org/er-blast/index.cgi?project=bma1). BLASTp searches of this database using ten *C. elegans* proteins known to function within the classical RNAi pathway identified putative homologous proteins to nine of the ten proteins investigated (Table 4.2). No *B. malayi* match was found for the *C. elegans* protein RDE-2 within the *B. malayi* dataset. This protein is responsible for siRNA accumulation in the *C. elegans* classical RNAi pathway (Grishok, 2005). Interestingly the *B. malayi* protein hypothesised to be homologous to the *C. elegans* SID-2 protein is more closely related to the putative *C. briggsae* SID-2 homologue with amino acid identity and similarity values of 14 and 32 % respectively. However, it should be noted that the *B. malayi* sequence identified

only covers 10 % of the *C. elegans* protein. The *C. briggsae* and *C. elegans* SID-2 proteins are known to differ in function (Winston *et al.*, 2007). As both RDE-2 and SID-2 are central to a functional systemic RNAi pathway in *C. elegans* these results are suggestive of a non-functional, or different, systemic RNAi pathway for *B. malayi* (Grishok, 2005). All identified *B. malayi* proteins gave the identifying *C. elegans* protein as the top hit in the reciprocal BLASTp search of *C. elegans* protein dataset (release WS150; http://130.209.234.35/blast/blast.html).

4.2.2 Searching the Haemonchus contortus genome sequence databases for putative homologues to Caenorhabditis elegans classical pathway associated proteins

H. contortus resides within Clade V of nematode phylogeny (Blaxter, 1998). It is therefore well placed for comparative genomic investigations using the wealth of data available for the genetic model organism *C. elegans* (Gilleard, 2004). Experimental analysis of gene function for *H. contortus* is currently limited to the use of heterologous systems due to the difficulty in establishing both forward and reverse genetic manipulating techniques in this parasite (Gilleard, 2004). As discussed in the previous chapter the genome project of *H. contortus* is on-going although assembly of the sequence is proving problematic. However, as detailed in Chapter 3, it has been suggested that sequence information for >80% of the transcribed loci from the haploid genome of this organism resides within the current genomic sequence databases. Therefore, using the information within these shotgun genomic sequence reads it is possible to assemble gene families and pathways.

Successful knock-down of *H. contortus* mRNA transcripts using RNAi has been reported (Table 4.1). Although this is the case, the use of RNAi is nowhere near as effective in this organism as it is in *C. elegans*, and instead appears to be target dependant (Table 4.1; Kotze and Bagnall, 2005; Geldhof *et al.*, 2006; B. Samarasinghe and C. Britton, University of Glasgow, personal communication). One possible reason for this is that the machinery required for a functional classical RNAi pathway may not be conserved within the haploid genome sequence of *H. contortus*. Searching for the presence of proteins associated with

the *C. elegans* classical RNAi pathway within the *H. contortus* genomic sequence database hc_reads110305.fasta successfully identified putative homologous sequence for 12 of the 17 *C. elegans* proteins searched for. These were DCR-1, DRH-1, RDE-1, RDE-3, SID-1, ERI-1, RRF-1, TSN-1, VIG-1, RRF-1, EGO-1 and MUT-7 (Table 4.3; Geldhof *et al.*, 2007). Each of the *H. contortus* sequence reads identified was used as query sequence in BLASTx searching of the *C. elegans* Wormpep protein database release WS150 (http://130.209.234.35/blast/blast.html) to ensure that the original identifying protein was the closest sequence match from this organism.

The shotgun genomic sequence read database hc_reads110305.fasta contains 547,477 entries. This database was the furthest advanced at the time of investigation. Since the completion of this work, tBLASTn searches for each of the five *C. elegans* proteins for which no significant match was found within database hc_reads110305.fasta have been repeated against all 1,056,873 shotgun genomic sequence reads that are now available (http://www.sanger.ac.uk/Projects/H_contortus/; http://130.209.234.35/blast/blast.html). No significant alignments have been found for the *C. elegans* RDE-2, RDE-4, RSD-2, RSD-6 or SID-2 proteins.

There is no apparent homologue to the *C. elegans* protein RDE-4 within any of the current *H. contortus* genomic sequence databases (Table 4.3; data not shown). *C. elegans* which lack a functional RDE-4 are classical RNAi defective (Tabara *et al.*, 1999). This protein is responsible for linking with the Dicer complex to direct RNAi in N2 worms (Tabara *et al.*, 2002) It is possible that the sequence of the *H. contortus* homologous protein has diverged from that of the *C. elegans* so much that it is no longer possible to infer a putative homologous relationship based on sequence homology alone. It can be reasonably hypothesised that protein sequence differentiation of such magnitude could also incur functional differentiation. This may be the defining reason for variability in the application of RNAi to this species. However, it might be speculated that even in the absence of RDE-4, siRNAs may function to result in gene silencing. This has been attempted using *H. contortus*, although no reduction in target transcript level was recorded in this study (Geldhof et al., 2006).

The majority of the *C. elegans* proteins for which no *H. contortus* homologous sequence was putatively identified are involved in cell non-autonomous RNAi (Table 4.3; Geldhof *et al.*, 2007; Grishok, 2005;). This suggests that this pathway is non-functional for *H. contortus*. This would mean that the effects of RNAi would only be seen in tissues that are directly accessible to the dsRNA such as the outer surface, gut and reproductive tract. To date successful RNAi knockdown of mRNA transcript using *H. contortus* has been recorded for a total of six genes (Geldhof *et al.*, 2006; B. Samarasinghe and C. Britton, University of Glasgow, personal communication). Interestingly, analysis of the expression pattern of the putative *C. elegans* homologue to each of these *H. contortus* genes reveals no indication that this is the case, as all are expressed in a variety of tissues, not all of which are readily accessible (Table 4.4; www.wormbase.org).

4.3 Discussion

RNAi is a powerful method in the armoury of gene function investigation for the free living nematode *C. elegans* (Fire *et al.*, 1998). The variability in the success of this reverse genetic technique in a number of nematode species remains intriguing. A potential reason for the variability seen is that the required machinery for the classical RNAi pathway is not conserved amongst species for which RNAi investigation of gene function gives variable results. I have investigated protein and genomic datasets for the nematode species *C. briggsae*, *P. pacificus*, *B. malayi* and *H. contortus* for the presence of putative homologous sequence to *C. elegans* proteins known to function in the classical RNAi pathway (Tables 4.2 and 4.3; Grishok, 2005). The protein dataset of the genetic model organism *D. melanogaster* was searched and used as a classical out-group in my analysis (Table 4.2; http://flybase.org/blast/).

Putative homologous sequences were found for all *C. elegans* proteins used to search the available protein dataset for the closely related nematode *C. briggsae* (Table 4.3; Cutter *et al.*, 2008). This implies that based on sequence homology, the machinery required for classical RNAi is present within the haploid genome sequence of this species and that RNAi should function both cell-autonomously and systemically, as it does in *C. elegans* (Fire *et al.*, 1998). However, experimental investigation shows that this is not the case. Although cell-autonomous RNAi for *C. briggsae* is possible, systemic RNAi is successful only when using strains that express the *C. elegans* protein SID-2 (Winston *et al.*, 2007). This result is important when considering the results of this chapter. The results presented here identify homologous relationships based on sequence homology alone. It is crucial that such relationships are investigated experimentally for evidence of functional conservation before anything other than putative homology is insinuated. An important consideration here is that the reverse is also true. In cases where putative homologous relationships are not able to be found by sequence homology, this does not mean that the protein is missing from the haploid genome of the investigated species. It could be the case that the sequence of the protein has diverged from that of its functional homologue so that it is no longer possible to infer such relationships based on sequence homology. The species investigated in the work presented in this

chapter represent Clades of Eukaryotic phylogeny that span millions of years of evolution. Therefore, such sequence differentiation may not only be postulated, but expected. In addition, no match may be found for *C. elegans* proteins within the dataset of the investigated species if it has not been sequenced yet. This is of upmost importance for the analysis reported here. In the case of *P. pacificus*, *B. malayi* and *H. contortus* I was using sequence datasets for draft or incomplete haploid genome sequencing projects (Dieterich *et al.*, 2008; Ghedin *et al.*, 2007; http://www.sanger.ac.uk/Projects/H_contortus/). Although such caveats must be considered, I hope that by investigating a number of species, the utmost has been done in order to limit the impact of these influences on my analysis. For my conclusions I am looking for trends across a number of species instead of direct one to one comparisons.

Across all investigated species, sequence homology suggests that the machinery required for the classical RNAi pathway is broadly conserved (Tables 4.2 and 4.3). For my analysis I would like to break the *C. elegans* classical RNAi pathway into two groups of proteins based on the function derived by experimental investigations: those involved in cell-autonomous RNAi and those involved in non-autonomous RNAi (Grishok, 2005). These groups are namely; DCR-1, DRH-1, RDE-1, RDE-3 and RDE-4 for cell-autonomous and RDE-2, SID-1 and SID-2 for non-autonomous RNAi (Table 4.2). Both methods of gene function investigation are active when using the nematode *C. elegans* (Fire *et al.*, 1998).

In the case of the proteins involved in cell autonomous RNAi, the only notable exclusion across the investigated species is the apparent lack of homologous sequence to the *C. elegans* RDE-4 protein from the datasets of nematodes *P. pacificus* and *H. contortus*. This is the lowest conserved polypeptide sequence of all the cell-autonomous RNAi associated proteins with identities to *C. elegans* of only 55, 18 and 10% in the putative homologous protein sequences of *C. briggsae*, *B. malayi* and *D. melanogaster*, respectively. We are therefore left with several explanations for the poor conservation or lack of an apparent RDE-4 sequence homologue in the haploid genome sequences of the investigated species, two of which are:

1. The RNAi pathway within theses species does not require RDE-4.

2. The polypeptide sequence of the *C. elegans* RDE-4 protein has sufficiently diverged in these species so that it is no longer recognisable based solely on sequence homology.

RDE-4 is essential for functional RNAi in *C. elegans* (Tabara, 1999). It could be therefore construed that a similar functioning protein would be necessary for the RNAi pathway to be functional at all in any other species. The findings presented in this chapter show that no putative homologue for the *C. elegans* Rde-4 protein could be found in the available datasets of the nematode species *P. pacificus* or *H. contortus*. Successful RNAi has been reported for both of these species and, in the case of *H. contortus*, specific mRNA transcript knockdown was shown (Pires da Silva, 2006; Geldhof *et al.*, 2006). It is unlikely that this knockdown effect was achieved by employing a different set of proteins to those investigated in this analysis. It is more likely that there is a functional RDE-4 homologue within all of these investigated species, and all others, where specific mRNA transcript knockdown has been reported. This speculative conclusion is also reinforced by the low identity and similarity percentages shown by putative homologous sequences from the datasets of *C. briggsae*, *B. malayi* and *D. melanogaster* to the *C. elegans* RDE-4 protein.

Draft genome sequence has recently become available for the plant root knot parasitic nematode *M. hapla* (Opperman *et al.*, 2008). Interestingly, homologous sequence to the *C. elegans* RDE-4 protein could not be found within this haploid genome sequence. Successful RNAi has been reported as a method for both gene function investigation and control of root knot parasitic nematode species (Gleason *et al.*, 2008; Fairbairn *et al.*, 2007). This adds further weight to my speculative conclusion that the RDE-4 protein sequence has sufficiently diverged over evolutionary time frames so that it is no longer recognisable by sequence homology in a number of species. Homologous sequence was found for all other machinery required for cell-autonomous RNAi in the investigated species (and *M. hapla*) (Tables 4.2 and 4.3; Opperman *et al.*, 2008). This is something that was expected prior to investigation as successful RNAi has been reported at least once for all species investigated in this analysis (Winston *et al.*, 2007; Pires da Silva, 2006; Aboobaker and Blaxter, 2003; Geldhof *et al.*, 2006; Fairbairn *et al.*, 2007).

Analysis of the conservation of *C. elegans* proteins involved in cell non-autonomous RNAi suggests that it is in the uptake and/or spreading of the RNAi signal that is the main reason for the variability of RNAi effects across the investigated species. *D. melanogaster* cell lines that express the *C. elegans* SID-1 protein are amenable to gene function investigation via RNAi using concentrations of dsRNA up to 10,000 times less as those that do not (Feinberg and Hunter, 2003). However, RNAi investigation of gene function using intact *D. melanogaster* specimens remains unachievable using any other technique than micro-injection to administer the dsRNA to the organism (Misquitta and Paterson, 1998; Kennerdell and Carthew, 1998). This is the same situation as in *C. briggsae* where it has been shown experimentally that the *C. elegans* and *C. briggsae* SID-2 proteins differ in function so that the dsRNA cannot be taken-up from the environment or spread from the lumen of the gut in *C. briggsae* worms as it is in *C. elegans* (Winston *et al.*, 2007). This work lead the authors to hypothesise a pathway for systemic RNAi where the SID-2 protein was responsible for the uptake of dsRNA from the environment or spreading from the gut and the SID-1 protein functioned in transferring the signal from cell to cell.

The putative lack of a functional homologue of the *C. elegans* SID-2 protein may explain the variability of RNAi in the investigated species. From this analysis, *B. malayi* is the only organism for which a putative homologue to the *C. elegans* SID-2 protein could be recognised (Table 4.2). Interestingly this protein is more similar to the putative *C. briggsae* SID-2 than it is to that of *C. elegans* (Table 4.2). *C. elegans* SId-2 is a transmembrane protein that is expressed in the intestine and body wall muscle nuclei (Winston *et al.*, 2007). A functional SID-2 is required for the uptake of dsRNA from the external environment and/or the spreading of the dsRNA from the gut of the worm to other tissues (Winston *et al.*, 2007). If an mRNA transcript is not expressed in the tissues that are immediately accessible (e.g. outer surface, gut, reproductive tract, etc.) then it could be hypothesised that the transcripts are unreachable in organisms lacking a functional SID-2 protein.

In order to further investigate this hypothesis it is important to look at the genes for which RNAi has been reported as successful and non-successful for the parasitic nematode species *P. pacificus*, *B. malayi* and *H. contortus* (Table 4.4; Pires da Silva, 2006; Aboobaker and Blaxter, 2003; Geldhof *et al.*, 2006; Kotze

and Bagnall, 2006). The genes investigated by RNAi for *C. briggsae* and *D. melanogaster* can be discounted for this analysis. All *C. briggsae* investigations use either transgenic strains or micro-injection techniques, both of which are not possible, or practical, for the parasitic nematodes (Winston *et al.*, 2007). *D. melanogaster* RNAi investigations report cell-autonomous RNAi using specific cell types in culture, or the micro-injection methodology of RNAi, therefore the dsRNA does not need to be transported from the gut, or taken-up from the environment, to see a decrease in mRNA transcript of the target gene (Misquitta, and Paterson, 1998; Kennerdell and Carthew, 1998). From this, there are no obvious conclusions that can be drawn from the comparative analysis of the expression pattern associated with the putative *C. elegans* homologue for each of the parasitic genes investigated by RNAi (Table 4.4). Each is expressed in various tissues, and not all are accessible to the environment.

The techniques used to assess RNAi in parasitic nematodes are also important. Only in the cases of *B. malayi* and *H. contortus* was specific mRNA transcript knockdown reported (Aboobaker and Blaxter, 2003; Geldhof, 2006). In the case *P. pacificus* and other nematode parasites (including *M. hapla*), phenotypic variations are used to assess for successful or non-successful RNAi experiments (Pires da Silva, 2006; Gleason *et al.*, 2008; Fairbairn *et al.*, 2007). Such analysis does not reflect specific target mRNA reduction and can confound a complete understanding of RNAi efficiency across species.

From comparative analysis of the results within this chapter the concluding hypothesis is that the variability of RNAi across the investigated species is due to inability in effective transportation of the dsRNA from the outer environment or gut to the other tissues, and that this difficulty is due to functional differentiation, or absence, of a homologue to the *C. elegans* SID-2 protein. To investigate this hypothesis further, detailed reportings of unsuccessful, rather than only successful RNAi attempts, would be beneficial. There are only a small number of genes from the investigated species where this is the case (Table 4.4; Aboobaker and Blaxter, 2003; Geldhof *et al.*, 2006). Analysis of a far greater number of genes may lead to more solid conclusions and hypotheses. In addition to this, employing the use of transgenic techniques using the nematode *C. elegans* as a heterologous expression system (or expression investigations directly within the parasite) to investigate parasitic nematode gene expression

patterns may allow for a more detailed understanding of RNAi variability. Here I have hypothesised that the expression patterns of the genes targeted by RNAi for *P. pacificus*, *H. contortus* and *B. malayi* are mirrored by those of the putative *C. elegans* homologue in each case (Table 4.4). Detailed investigation of the expression pattern of each parasite gene, using *C. elegans* as a heterologous expression system, may elucidate patterns of expression common to genes that are amenable to functional investigation using RNAi. Finally, evidence of specific knockdown, showing unequivocal reduction in mRNA transcript level following RNAi treatment is also important. RNAi investigations which rely on phenotypic evidence in assessing successful RNAi do not imply specific transcript silencing and may limit our understanding of RNAi variability not only in the species investigated here, but all eukaryotes.

4.4 Tables and Figures

Table 4.1: Summary of RNAi studies in parasitic helminth species showing methods of delivery and genes targeted.

(Printed with permission from Geldhof, P., *et al.*, 2006.)

Species	Life stage	Target	Method	Reference
Schistosoma mansoni	cercaria	cathepsin B	soaking	Skelly, P.J., *et al.* (2003)
	miracidium - sporocyst	facilitated-diffusion glucose transporter	soaking	Boyle, J. P., *et al.* (2003)
		glyceraldehydes-3-phosphate dehydrogenase		
		CD36-like class B scavenger receptor	soaking	Dinguirard, N., *et al.* (2006)
	schistosomulum	cathepsin B	electroporation	Correnti, J.M., *et al.* (2005)
		gynecophoral canal protein	soaking	Cheng, G.F., *et al.* (2005)
Nippostrongylus brasiliensis	adult	acetylcholinesterases	soaking	Hussein, A.S., *et al.* (2002)
Brugia malayi	adult	beta-tubulin, RNA polymerase II large subunit, microfilarial sheath protein	soaking	Aboobaker, A.A., and Blaxter, M.L., (2003)
Onchocerca volvulus	L3	cathepsin L-like protease, Cathepsin Z-like protease	soaking	Lustigman, S., *et al.* (2004)
		serine protease inhibitor	soaking	Ford, L., *et al.* (2005)

Trichostrongylus colubriformis	L1	ubiquitin, tropomyosin	feeding-soaking-electroporation	Issa, Z., *et al.* (2005)
Ascaris suum	L3	inorganic phyrophosphatase	soaking	Islam, M.K., *et al.* (2005)
Haemonchus contortus	L3-L4-adult	beta-tubulin genes	soaking	Kotze, A.C., and Bagnall, N.H. (2006)
	L1-L3	beta-tubulin, COPII component, Ca2+ binding protein, heat shock protein HSP70, vacuolar ATPase, cathepsinL, paramyosin, Cu-Zn superoxide dismutase, intermediate filament, type IV collagen, GATA transcription factor	electroporation - soaking	Geldhof, P., *et al.* (2006)
Ostertagia ostertagi	L3	Tropomyosin, beta-tubulin, ATP-synthetase, Superoxide dismutase, polyprotein allergen, ubiquitin, transthyretin-like protein, 17 kDa ES protein	electroporation - soaking	Visser, A., *et al.* (2006)
Litosomoides sigmodontis	Adult	actin	soaking	Pfarr, K., *et al.* (2006)

Meloidogyne species	J2	cathepsin L	soaking	Shingles. J., *et al*. (2006)
		dual oxidase	soaking	Bakhetia, M., *et al*. (2005)
		splicing factor, integrase	plant delivery	Yadav, B.C., *et al*. (2006)
	No details	Calreticulin, polygalacturonase	soaking	Rosso, M.N., *et al*. (2005)
		Parasitism gene *16D10*	Soaking, plant delivery	Huang, G., *et al*. (2006)
	eggs	Chitin synthase gene	soaking	Fanelli, E., *et al*. (2005)
Heterodera glycines	J2	aminopeptidase	soaking	Lilley, C.J., *et al*. (2005)
		csteine proteinases, C type lectin, major sperm protein	soaking	Urwin, P.E., *et al*. (2002)
Globodera pallida	J2	cysteine proteinases	soaking	Urwin, P.E., *et al*. (2002)
Globodera rostichiensis	No details	Beta-1,4 endoglucanase, *gr-ams-1*	No details	Chen, Q., *et al*. (2005)

Table 4.1: Summary of RNAi studies in helminth species.

Table 4.2: Identification of putative homologues of *C. elegans* proteins known to function in the classical RNAi pathway from *C. briggsae*, *P. pacificus*, *B. malayi* and *D. melanogaster*.

The 10 *C. elegans* protein sequences were taken from Wormbase (www.wormbase.org) and used to search the sequence databases of the nematode species *C. briggsae*, *P. pacificus*, *B. malayi* and the fruit fly genetic model organism *D. melanogaster*. Protein datasets are available for *C. briggsae*, *B. malayi* and *D. melanogaster*. BLASTp searches were carried out for each of the three species using the appropriate online BLAST servers at http://www.wormbase.org/db/searches/blast_blat, http://blast.jcvi.org/er-blast/index.cgi?project=bma1 and http://flybase.org/blast/ respectively. tBLASTn searching of the *P. pacificus* genome assembly version 'P.pac California assembly 2' was performed using the online BLAST web page available on the *P. pacificus* genome website (http://www.pristionchus.org/cgi-bin/blast_iframe.pl). In all cases, the percentage identity/similarity values for each sequence when compared to its putative *C. elegans* homologue were calculated as an average of all high-scoring segment pair (HSP) alignment values given on the BLAST output page following each search.

C. elegans gene name	Size (aa)	Protein function	Reference for putative homologue		% Identity	% Similarity
dicer	1845	dsRNase	*C. briggsae*	CBG22974	75	88
			P. pacificus	supercontig 1:contig 40	58.5	74.7
			B. malayi	14979.m04594	49	68
			D. melanogaster	CG4792	36	52
drh-1	1037	Dicer related helicase	*C. briggsae*	CBG05400	64	79
			P. pacificus	supercontig 1:contig 608	49	67.7
			B. malayi	13637.m00283	32	53
			D. melanogaster	CG7922	11	28
rde-1	1020	Takes siRNAs from initiator complex to rest of pathway	*C. briggsae*	CBG19426	64	78
			P. pacificus	supercontig 1:contig 238	42.5	63.7
			B. malayi	14971.m02816	28	49
			D. melanogaster	CG6671	20	40
rde-2	578	siRNA accumulation	*C. briggsae*	CBG21903	42	62
			P. pacificus	No hit		
			B. malayi	No hit		
			D. melanogaster	CG18362	10	25
rde-3	441	Required for siRNA accumulation	*C. briggsae*	CBG12597	58	74
			P. pacificus	supercontig 1:contig 537	50	62.7
			B. malayi	14990.m07896	18	39
			D. melanogaster	CG5732	19	39

rde-4	385	RNA binding in dicer complex	*C. briggsae*	CBG09811	55	72
			P. pacificus	No hit		
			B. malayi	13278.m00103	18	35
			D. melanogaster	CG1620	10	28
eri-1	448	RNAi antagonist (exonuclease)	*C. briggsae*	CBG13539	72	87
			P. pacificus	supercontig 1:contig 11	35.6	50.3
			B. malayi	14977.m04896	38	57
			D. melanogaster	CG6393	25	46
rrf-3	1780	RNA antagonist	*C. briggsae*	CBG00730	77	87
			P. pacificus	supercontig 1:contig 696	41.3	58.2
			B. malayi	14920.m00401	34.8	53.3
			D. melanogaster	No hit		
sid-1	776	Systemic RNAi	*C. briggsae*	CBG08682	62	78
			P. pacificus	supercontig 1:contig 184	32	49
			B. malayi	14992.m10831	30.5	54
			D. melanogaster	No hit		
sid-2	311	Systemic RNAi	*C. briggsae*	CBG18280	41	59
			P. pacificus	No hit		
			B. malayi	13175.m00133	10	31
			D. melanogaster	No hit		

Table 4.2: Identification of putative homologous proteins from the nematode species *C. briggsae*, *P. pacificus*, *B. malayi* and the fruit fly genetic model organism *D. melanogaster* for *C. elegans* proteins known to function within the classical RNAi pathway.

Table 4.3: Analysis of the *H. contortus* shotgun genomic sequence reads database hc_reads110305.fasta for genes encoding putative homologues of *C. elegans* proteins known to function in the classical RNAi pathway.

The sequences of 17 *C. elegans* proteins known to function in the RNAi pathway taken from Wormbase (www.wormbase.org) and used to carry out tBLASTn searches against the *H. contortus* shotgun genomic sequence reads database hc_reads110305.fasta using the online Sanger Institute BLAST server (www.sanger.ac.uk/cgi-bin/blast/submitblast/h_contortus). The *H. contortus* sequences identified in each BLAST analysis were used in the ContigExpress programme of Vector NTI (Invitrogen) to identify overlapping reads (>98 % identity over >50 bp). The individual and/or consensus sequences were subsequently used to calculate the percentage identity/similarity to its putative *C. elegans* homologue and percentage of *the C. elegans* protein covered.

(Printed with permission from Geldhof *et al.*, 2006.)

C. elegans gene name	Size (aa)	Protein function	Number of *H. contortus* sequence reads in TBLASTN	% Coverage of *C. elegans* protein	% Identity	% Similarity
dicer	1845	dsRNase	40	53	59	72
drh-1	1037	Dicer related helicase	12	47	39	54
rde-1	1020	Takes siRNAs from initiator complex to rest of pathway	13	19	52	72
rde-3	441	Possibly engaged in the target	7	38	62	75
sid-1	776	Systemic RNAi	1	7	48	67
eri-1	448	RNAi antagonist (exonuclease)	9	47	45	59
rrf-3	1780	RNA antagonist	8	27	49	68

tsn-1	914	RNA binding (RISC)	9	48	55	79
vig-1	378	RNA binding (RISC)	2	13	49	51
rrf-1	1601	RNA dependent RNA polymerase (RdRP)	10	63	55	70
ego-1	1632	RNA dependent RNA polymerase (RdRP)	13	47	58	77
mut-7	910	Predicted RNA-binding protein	9	19	47	67
rsd-6	689	Systemic RNAi	3 (highly repetitive sequence)			
sid-2	311	Systemic RNAi	No hit			

rde-2	578	siRNA accumulation	No hit
rde-4	385	RNA binding in dicer complex	No hit
rsd-2	1265	Systemic RNAi	No hit

Table 4.3: Analysis of the *H. contortus* shotgun genomic sequence reads database hc_reads110305.fasta for putative homologous sequences to *C. elegans* proteins known to function within the classical RNAi pathway.

Table 4.4: Genes investigated by RNAi for the parasitic nematode species *P. pacificus*, *B. malayi* and *H. contortus* and details of the expression pattern for the putative *C. elegans* homologue in each case.

List of all genes that have been investigated by RNAi for the parasitic nematode species *P. pacificus*, *B. malayi* and *H. contortus* (Pires da Silva, 2006; Aboobaker and Blaxter, 2003; Geldhof *et al.*, 2006; Kotze and Bagnall, 2006). Expression pattern data for the putative *C. elegans* homologue in each case was taken from Wormbase release WS190 (www.wormbase.org) or data presented in this thesis.

Species	Gene	Successful RNAi?	Putative *C. elegans* homologue	
			Gene name	Expression pattern
P. pacificus	*Ppa-tra-1*	Yes	*tra-1*	Intestine, gonad, nervous system, head neurons
B. malayi	*Bm-tub-1*	Yes	*ben-1*	Gut, head and body neurons
	Bm-ama-1	Yes	*ama-1*	Ubiquitous until 550 cell stage, Gut at bean to pretzel stage
	Bm-shp-1	Yes	n/a	N/A
H. contortus	*Hc-iso-1*	Yes	*ben-1*	Gut, head and body neurons
	Hc-sod-1	Yes	*sod-1*	N/A
	Hc-sec-23	Yes	*sec-23*	Hypodermis
	Hc-cpl-1	Yes	*cpl-1*	Hypodermal regions, posterior bulb of pharynx, vulval regions, gut, intestine
	Hc-M03	No	*calu-1*	Pharynx, intestine, rectal epithelial cells, reproductive system, vulva, hypodermis
	Hc-hsp	No	*hsp-1*	Germline, pharynx, intestine, depressor and body wall muscle, excretory cells
	Hc-vha-10	No	*vha-10*	N/A
	Hc-unc-15	No	*unc-15*	Pharyngeal musculature, vulval and body wall muscle cells, anal depressor and sphincter, intestinal and sex muscles
	Hc-mua	No	*mua-6*	Hypodermis
	Hc-let-2	No	*let-2*	N/A
	Hc-elt-2	No	*elt-2*	Gut, intestine

Table 4.4: Genes investigated by RNAi for the parasitic nematode species *P. pacificus*, *B. malayi* and *H. contortus* and details of the expression pattern of the putative *C. elegans* homologue in each case.

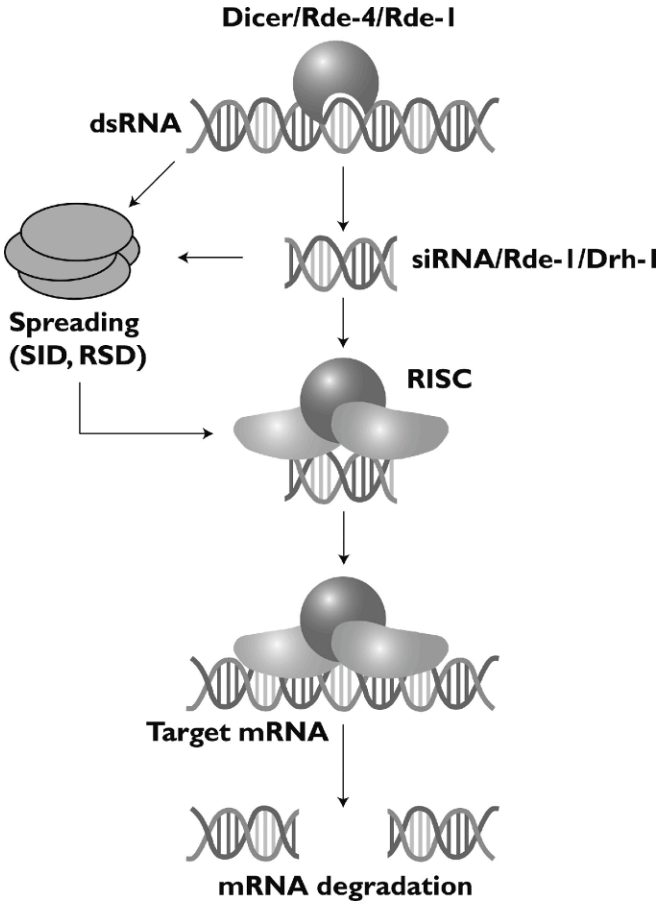


Figure 4.1: Simplified schematic model of the RNAi mechanism in *C. elegans*.

Double stranded RNA (dsRNA) is processed into small inhibitory RNAs (siRNAs) by Dicer. Processed dsRNA is thought to be transported systematically in the worm by the SID and RSD proteins and siRNAs are subsequently bound by the RNA induced silencing complex (RISC). The activated RISC complex will subsequently bind and degrade the homologous mRNA.

(Printed with permission from Geldhof *et al.*, 2007.)

Chapter 5: The *Haemonchus contortus* β-Tubulin gene family

5.1 Introduction

Infection with parasitic nematodes causes serious disease to both human and animals (Bird, 1991; Jackson, 1993). Treatment of these infectious diseases is primarily through the use of chemotherapeutic anthelmintic drugs. Since the 1960s the majority of treatments for infection with intestinal nematodes has been through the use of the benzimidazole (BZ) subclass of anthelmintic drugs (Jackson, 1993). Resistance to BZ drugs is widespread and has been documented for a number of nematode species of veterinary importance and evidence is also beginning to accumulate of the rise of BZ resistance in human nematode infections (Jackson, 1993; Kwa *et al*., 1993; Prichard, 2007).

The major effect of BZ drugs is believed to be in disrupting the polymerisation of tubulin-dimers which comprise microtubules (Lacey, 1988; Prichard, 2001; Robinson *et al*., 2004). Microtubules are responsible for a variety of cellular functions and their disruption has been shown to be the causative effect of many drug classes (Lacey, 1988; Wilson *et al*., 1999). BZ drugs have been shown to bind with high affinity to tubulin *in vitro* however the molecular detail of the tubulin-BZ interaction within nematodes has not been fully resolved, with different potential models being proposed (Prichard, 2002; Robinson *et al*., 2002). Although the particular BZ-β-tubulin binding site is a contentious issue, strong experimental evidence has been generated for a number of nematode species which correlates mutations in β-tubulin genes with BZ resistance (Kwa, *et al*., 1993; Prichard, 2001; Wolstenholme *et al*., 2004).

Direct evidence of the significance of β-tubulin mutations have been demonstrated through the use of the free-living nematode *C. elegans*, where a Phe-to-Tyr substitution at position 200 of the polypeptide encoded by the β-tubulin *ben-1* gene was shown to confer significant resistance to BZ drugs (Driscoll *et al*., 1989). *C. elegans* has also been used as a heterologous system to show that the same change in the protein encoded by the *H. contortus* β-

tubulin gene *isotype-1* can also confer a resistance phenotype. This implies that the same mechanism is employed by both organisms in conferring BZ resistance (Kwa *et al*., 1995).

A major difference in the BZ resistance mechanisms for *C. elegans* and *H. contortus* is that *ben-1* is the only β-tubulin locus from the free-living nematode species that has been implicated in resistance, whereas specific mutations in two β-tubulins from the parasitic nematode have been correlated with this phenotype. Deletions of a second *H. contortus* β-tubulin locus, *isotype-2*, have been shown to confer a heightened level of BZ resistance to the *H. contortus* population (Kwa *et al*., 1993; Beech *et al*., 1994). This implies that although, on the whole, the mechanism of BZ resistance for the parasitic and free-living parasitic nematodes may be very similar, subtle differences exist.

There remains a significant level of ambiguity concerning the relationships that the *Hc-isotype-1* and *Hc-isotype-2* genes share with the *C. elegans* β-tubulin gene family. There are reports that *Hc-isotype-1* is orthologous to the *Ce-ben-1* gene (Geldhof *et al*., 2005). However, such claims are based solely on the ability of both genes to confer significant resistance to BZ drugs and are therefore somewhat unsubstantiated.

5.1.1 Aims of this chapter

1. Investigate the genomic sequence data of the parasitic nematode *H. contortus* for additional members of the β-tubulin gene family, and to relate each sequence to that of the *C. elegans* β-tubulin gene family via detailed sequence analysis.

2. Conduct a comparative genomic investigation on the evolutionary basis behind the β-tubulin gene family across the phylum Nematoda using available nematode β-tubulin polypeptide and genomic sequence databases.

5.2 Results

5.2.1 Determining the size of the Haemonchus contortus β-tubulin gene family

5.2.1.1 Identification and characterisation of β-tubulin sequences within *Haemonchus contortus* EST datasets

The *C. elegans* β-tubulin gene family contains 6 members. The polypeptides encoded by these genes are over 60% identical and 77% similar to one another. Homology based searching of available *H. contortus* EST sequences using each of these six *C. elegans* β-tubulin polypeptides identified a total of 27 sequences as containing β-tubulin sequence (http://130.209.234.35/blast/blast.html). All identified EST sequences contained regions encoding for the signature sequence pattern which distinguishes α, β and γ tubulin family members 'GGGTGSG' at residues 140-146 and/or the characteristic β-tubulin auto-regulation recognition element 'MREI' at codon positions 1-4 of the translated protein, in cases where EST sequence was available that encoded for either of these regions (Hesse *et al* 1987; Cleveland, 1988).

Three full length *H. contortus* cDNA sequences which encode for β-tubulin polypeptides have been previously submitted to Genbank (Geary *et al.*, 1992). Detailed analysis has shown that one of these sequences represents the *isotype-1* and two the *isotype-2*, β-tubulin loci from the *H. contortus* genome (Geary *et al.*, 1992). 20 of the 27 identified β-tubulin encoding *H. contortus* EST sequences show >98% identity to the fully sequenced cDNA sequence encoding the ISO-1 protein. Similarly, 7 of these 27 sequences show >97% identity to cDNA sequences encoding for the *H. contortus* ISO-2 protein. This high level of identity provides evidence that sequence information for only the two previously identified β-tubulin loci from the *H. contortus* genome is present in the 14 Mb of EST sequence data that has been generated for this organism.

5.2.1.2 Sequencing of the *Haemonchus contortus* β-tubulin *isotype-2* genomic locus

The *isotype-2* β-tubulin genomic locus was PCR amplified and sequenced since the genomic sequence has not yet been reported (Sequence detailed in Appendix 3). A unique 5.2 kb product was obtained by PCR amplification from *H. contortus* MHco3 (ISE) L3 genomic DNA template using primers designed to be specific to the β-tubulin *isotype-2* locus based on the available cDNA sequence (Primer sequences in Appendix 2). The gene model of this *H. contortus* β-tubulin was deduced by pair-wise alignment of the 5.2 kb genomic sequence to the available *isotype-2* cDNA sequence (accession number M76491; Geary *et al.*, 1992; AlignX, Invitrogen). The resulting model has 8 introns and 9 exons and is very similar to the available gene model for *H. contortus* β-tubulin *isotype-1*, differing only in intron sizes and the fact that the final exon of the *isotype-2* gene model is split into two compared to *isotype-1* (Figure 5.1).

5.2.1.3 Identification and sequencing of a β-tubulin locus from a *Haemonchus contortus* Bacterial Artificial Chromosome

A β-tubulin locus resides in the centre of a 408,911bp continuous stretch of *H. contortus* genomic sequence. This stretch was fully sequenced and finished at the Pathogen Sequencing Unit of the Sanger Institute and it represents the largest continuous fragment of *H. contortus* genomic sequence currently available. The gene structure for this β-tubulin locus was manually curated and confirmed by full-length cDNA sequencing. This gene model covers 11,596 bp and consists of 13 exons with 12 introns. The resulting structure is very different to that of *H. contortus* β-tubulin *isotype-1* or *2*, confirming it as a third *H. contortus* β-tubulin genomic locus, *H. contortus* β-tubulin *isotype-3* (Figure 5.1).

5.2.1.4 Investigation of similarity between the gene models of *Caenorhabditis elegans* and *Haemonchus contortus* fully sequenced β-tubulin loci

Generally, the gene models of the *H. contortus* β-tubulin family cover a larger genomic region than those of the *C. elegans* family (Figures 5.1 and 5.2). This is largely due to the increase in intronic sequence, in both number and size, within these gene models when compared with those of the free-living species. A pair-wise comparison of the *C. elegans* and *H. contortus* β-tubulin gene models reveals little conservation of exon/intron break-points (Figures 5.1 and 5.2). This may be expected when not only the evolutionary distance between the two species is considered, but also that I have compared gene models from a primarily asexually reproducing organism, in *C. elegans*, with a sexually reproductive species in *H. contortus*. The first exception to this is the similarity in models of the *Ce-ben-1*, *Hc-isotype-1* and *Hc-isotype-2* genes. The exon/intron boundaries of the *Ce-ben-1* gene structure are perfectly conserved in both the parasitic nematode models, although, both the parasite gene models have more introns than that of *Ce-ben-1*. This conservation of gene structure allows the hypothesis that these three genes are evolutionary related to one another. A similar pattern is also seen when the *Ce-mec-7* and *Hc-isotype-3* models are compared.

5.2.1.5 Conservation of macro and micro synteny between the *Haemonchus contortus* β-tubulin *isotype-3* and *Caenorhabditis elegans* β-tubulin *mec-7* loci

Investigation of the inheritance of six micro-satellite genetic markers along the continuous 408,911 bp *H. contortus* genomic sequence fragment has demonstrated that this sequence is part of the X-chromosome (Redman *et al.*, 2008). Although each of the markers was shown to be polymorphic in the population, genotyping each of the markers from 56 male worms showed that all six were monomorphic in every case. BLASTx searching of the *C. elegans* Wormpep dataset (release WS150; http://130.209.234.35/blast/blast.html) using this 408,911 bp sequence identified a total of 17 putative homologous genes (Redman *et al.*, 2008). 14 of the 17 *C. elegans* genes for which putative

homologues were identified, reside on the X-chromosome of this species. This represents over 80% of all putative homologues found and is indicative of conservation of synteny between the *C. elegans* and *H. contortus* X-chromosomes (Figure 5.3).

Experimental investigations detailed elsewhere in this thesis provide evidence that the *C. elegans mec-7* and *H. contortus isotype-3* represent an orthologous gene pair. The *C. elegans* gene *ZK154.1* is located closest to the β-tubulin *mec-7* locus on the X-chromosome of this species. The putative *H. contortus* homologue of the *C. elegans* gene *ZK154.1* was positioned closest of all identified putative homologues to the β-tubulin *isotype-3* locus. In addition, the orientations of these genes are conserved on the sequence of the respective species (Figure 5.3). This is an example of conservation of micro-synteny between these β-tubulin loci from the X-chromosomes of *H. contortus* and *C. elegans*.

5.2.1.6 Identification of β-tubulin sequences within *Haemonchus contortus* genomic shotgun sequence read datasets

tBLASTn searching of all *H. contortus* shotgun genomic sequence reads with each of the six *C. elegans* β-tubulin polypeptides identified 88 reads that contained sequence putatively encoding for β-tubulin polypeptide sequence (Table 5.2, http://130.209.234.35/blast/blast.html). The conserved protein sequence motif which distinguishes α, β and γ tubulin family members and/or the characteristic β-tubulin autoregulation recognition element was present in all reads that contained DNA which putatively encoded for either region (Cleveland, 1988; Hesse *et al.*, 1987).

Each of the 88 *H. contortus* genomic shotgun sequence reads was aligned to the available genomic loci for β-tubulins *isotype-1*, *2* and *3*. Using my parameters, 23 of these reads were designated as representing the *isotype-1*, 26 *isotype-2* and 30 the *isotype-3* loci (Table 5.2). Nine of the 88 *H. contortus* shotgun genomic sequence reads could not be assembled into any of these three β-tubulin loci.

5.2.1.7 Assembling nine uncharacterised *Haemonchus contortus* β-tubulin genomic shotgun sequence reads into one locus

The nine *H. contortus* β-tubulin-containing genomic shotgun sequence reads which could not be designated as representing the fully sequenced loci of *isotype-1*, *2* or *3* were searched for regions of common homology (ContigExpress, Invitrogen). Three contiguous sequences were assembled by this method which encompassed eight of the nine *H. contortus* genomic shotgun sequence reads (Table 5.3). PCR primers were designed to amplify the genomic sequence gaps between each of the three contigs and one outstanding genomic shotgun sequence read (primer sequences Appendix 2). Distinct products were amplified in all cases, linking these nine β-tubulin-containing sequences to one *H. contortus* genomic locus (Figure 5.4). This locus was designated *H. contortus* β-tubulin *isotype-4*.

5.2.1.8 PCR amplification, cloning and sequencing of *Haemonchus contortus* β-tubulin *isotype-3* and *-4* cDNA sequences

PCR primers were designed to amplify the cDNA encoding the complete predicted protein of the *H. contortus* β-tubulin *isotype-3*, and the available sequence of *isotype-4* (primer sequences Appendix 2). Unique products were amplified from *H. contortus* MHco3 (ISE) adult cDNA for both these loci (sequences detailed in Appendix 3). Pair-wise alignment of the cDNA sequences for both of these novel *H. contortus* isotypes to the available cDNA sequences of the two previously characterised β-tubulin loci confirms that all four loci are expressed and that each of the transcribed products are distinctly different from one another (Table 5.4).

The translated protein sequences of *H. contortus* β-tubulin ISO-1, 2, 3 and 4 are highly conserved both in identity and similarity of amino acid residues (Table 5.5). A similar trend occurs both within and between β-tubulin gene families of all species for which β-tubulin sequence is available (Njue and Prichard, 2003). The protein sequences for *H. contortus* β-tubulin ISO-1, 2 and 3 contain the conserved sequence motif which distinguishes α, β and γ tubulin family members at position 140-146 (Hesse *et al.*, 1987; Figure 5.5). In

addition, the β-tubulin characteristic auto-regulatory motif is present as the first four amino acids of these protein sequences (Cleveland *et al.*, 1988; Figure 5.5). No *isotype-4* cDNA sequence is available to deduce the amino acid sequence for this translated protein product at these regions. Throughout the course of this project many attempts were made using cDNA/spliced-leader amplification, 5' RACE and bioinformatic scaffolding to identify this region, but despite best efforts, the 5' region of the cDNA for the *H. contortus* β-tubulin *isotype-4* locus remains unknown (data not shown). I have characterised this sequence as β-tubulin due to reciprocal BLASTp analysis against the Wormpep dataset (release WS150, http://130.209.234.35/blast/blast.html) and the high level of identity it shares with other β-tubulin polypeptide sequences from a wide range of eukaryotic species (data not shown).

The C-terminal region of *H. contortus* β-tubulin ISO-1-4 are hyper-variable (Figure 5.5). This is a common feature of β-tubulin polypeptides, both within and between many other eukaryotic species (Njue and Prichard, 2003). This region of the β-tubulin polypeptide interacts with the microtubule associated proteins (MAPs) and deletion of this region has been shown to interrupt the correct assembly of microtubules (Dominguez *et al.*, 1990). There are only three residues at which the available polypeptide sequences of all four *H. contortus* β-tubulin Isotypes differ, and all are within the hyper-variable region (Figure 5.5).

5.2.2 Using sequence information for an evolutionary evaluation of the β-tubulin gene family across the phylum Nematoda

5.2.2.1 Investigating the evolutionary relationships between β-tubulin loci from the nematode species *Caenorhabditis briggsae*, *Caenorhabditis remanei*, *Caenorhabditis japonica*, *Caenorhabditis brenneri* and *Pristionchus pacificus*

5.2.2.1.1 Identification of β-tubulin loci within the genomes of free-living nematodes from the genus Caenorhabditis

The completed genome sequence for the free-living nematode *C. elegans* was published in 1998 and paved the way for genome-wide evolutionary based investigation (*C. elegans* sequencing consortium, 1998). In 2003, four other free-living nematode species closely related to *C. elegans*, namely *C. briggsae*, C. *remanei*, *C. japonica* and *C. brenneri*, were selected for complete genome sequencing (Sternberg *et al.*, 2003). The genome sequencing projects for these species are now at a well advanced stage, with BLAST searchable databases and manually curated gene models available for all four via the Wormbase web-site (Stein *et al.*, 2003; http://www.wormbase.org/).

Wormbase dataset and BLAST (BLASTp and tBLASTn) based sequence homology searching of the available databases for each of the free-living nematodes *C. briggsae*, C. *remanei*, *C. japonica* and *C. brenneri* identified homologues in each of the four species for most of the six *C. elegans* β-tubulin loci (Table 5.6). No homologous sequence to the *C. elegans* TBB-6 protein was found in any of the databases. The only other exclusions were the lack of putative TBB-2 and TBB-4 homologues encoded within the *C. brenneri* genomic sequence. *C. briggsae* is the only one of these four free-living nematodes with a fully sequenced available genome sequence (Stein *et al.*, 2003). Therefore, it is possible that not all β-tubulin loci from each of the organisms may yet have been sequenced. I am therefore looking for trends in my analysis and not definitive conclusions.

5.2.2.1.2 Identification of β-tubulin encoding loci from the genome of the necromenic nematode Pristionchus pacificus

The 169 Mb draft genome sequence of the necromenic nematode *P. pacificus* contains 23,500 predicted genes (Dieterich *et al.*, 2008). This species is the most closely related to the *Caenorhabditis* genus of all nematode species with an available draft genome sequence. BLAST searchable sequence databases and manually curated gene models are available for this species via the Wormbase server (http://www.wormbase.org/). Searching these datasets for *P. pacificus* β-tubulin loci revealed three manually curated genes (Table 5.6). Based on reciprocal BLAST analysis, these three loci are homologues of the *C. elegans ben-1*, *mec-7* and *tbb-4* genes.

5.2.2.1.3 Phylogenetic analysis of putative β-tubulin protein sequences from five species of the Caenorhabditis genus and Prisitionchus pacificus

A phylogenetic tree was constructed, using the Maximum Parsimony algorithm, to investigate the relationships between hypothesised translated protein sequences from β-tubulin loci identified for the nematode species *C. elegans*, *C. briggsae*, C. *remanei*, *C. japonica*, *C. brenneri* and *P. pacificus* (Figure 5.6). Sequences were named based on the best BLASTp alignment from *C. elegans* Wormpep (release WS150, http://130.209.234.35/blast/blast.html; Table 5.6). Defined clusters, with bootstrap values ranging from 72-98, are evident for the MEC-7, TBB-4, BEN-1 and TBB-1/2 protein sequences. Intra-clustering bootstrap values are very low and there is little maintenance of topology at a species level (Sternberg *et al.*, 2003).

All proteins that showed the *C. elegans* BEN-1 protein as top alignment in reciprocal BLAST searching of *C. elegans* Wormpep (release WS150; http://130.209.234.35/blast/blast.html) segregate into one phylogenetic cluster (Table 5.6 and Figure 5.6). Although reciprocal BLAST, phylogenetic and syntenic analyses provide evidence of an orthologous relationship between each member of this cluster, it is an important consideration that no experimental evidence exists which tests these relationships. This is perhaps of most importance for the BEN-1 cluster where specific mutations in only the *C. elegans* β-tubulin *ben-1* locus have been shown to confer resistance to BZ drugs (Driscoll *et al.*, 1989).

The failure to separate the TBB-1 and 2 protein sets into distinct clusters is intriguing. This is not an artefact of sequence naming as no matter what combination of two β-tubulin proteins from one species were selected to represent TBB-1 and 2, it was impossible to fully resolve this cluster (data not shown). TBB-1 and 2 are the only two of the six β-tubulin proteins from *C. elegans* that have been shown to be functionally redundant (Chenggang *et al.*, 2004). The gene models of the loci encoding these two proteins are very similar and both reside on *C. elegans* Chromosome III. This indicates that *tbb-1* and *tbb-2* represent a paralogous gene pair. *C. brenneri* is more closely related to *C. elegans* than C. *remanei* or *C. briggsae* (Sternberg *et al.*, 2003).

However, a homologous locus to only *tbb-1* and not *tbb-2* was found for *C. brenneri*, whereas homologues to both were found within the available datasets for both C. *remanei* and *C. briggsae*. *P. pacificus* is more distantly related to *C. elegans* than any member of the *Caenorhabditis* genus (Dieterich *et al.*, 2008). *P. pacificus* sequence homologous to either of the *C. elegans* TBB-1 or 2 proteins could not be identified via reciprocal BLAST or phylogenetic analysis (Table 5.6 and Figure 5.6). A possible hypothesis is that the *C. brenneri* paralog of *tbb-1* or *2* has been lost or that the sequence of the protein encoded by this locus has sufficiently diverged so that it is no longer recognisable, using the methods employed for this analysis. This hypothesis could be coupled with the *P. pacificus* functional homologue(s) of TBB-1 or 2 (or both) no longer being recognisable based solely on sequence similarity. A second potential hypothesis for the trend seen is that, as a caveat of working with incomplete or draft genome sequences, these loci have not yet been sequenced. The settling of these hypotheses is beyond the scope of research for this thesis but may be investigated in the future, and may become clear as more data emerges for the genome sequencing projects of the species in question.

The *C. elegans* TBB-6 protein branches alone. This is consistent with my reciprocal BLAST analysis, indicating that none of the genomes of the investigated species contain a β-tubulin gene homologous to that of *C. elegans tbb-6*.

5.2.2.2 Phylogenetic analysis of the β-tubulin gene families of five members from the *Caenorhabditis* genus, *Prisitionchus pacificus* and *Haemonchus contortus*

C. elegans, *C. briggsae*, *C. japonica*, C. *remanei*, *C. brenneri*, *P. pacificus* and *H. contortus* all reside within Clade V of nematode phylogeny (Blaxter, 1998). These are the only species from this Clade of the phylogeny for which sufficient genomic sequence data currently exists in order to investigate gene families with confidence. A Maximum Parsimony phylogenetic tree was constructed to investigate the phylogenetic relationships between protein sequences hypothesised to be transcribed from the β-tubulin gene families

from each of these species (Figure 5.7). Bootstrap values ranging from 48-82 split the tree into four distinct clusters, with *C. elegans* TBB-6 and the root of the tree distinctly branching alone. Within clusters, species topology is poorly maintained and bootstrap values are commonly very low (De Ley and Blaxter, 2004).

Experimental results contained within this thesis Indicate that the *H. contortus isotype-3* and *C. elegans mec-7* are an orthologous gene pair (Chapter 6). As expected, the protein sequences transcribed from these genes branch into the MEC-7 cluster of this phylogenetic tree. The available translated protein sequence of the *H. contortus isotype-4* β-tubulin gene branches with the TBB-4 cluster. Although incomplete, this sequence includes the hyper-variable C-terminal region of the protein. The orthologous relationships inferred by the MEC-7/ISO-3 and TBB-4/ISO-4 clusters are constructed with high confidence with bootstrap values of 82 and 65 respectively.

Both the *H. contortus* β-tubulin *isotype-1* and *-2* loci have been implicated in giving rise to BZ resistance (Kwa *et al.*, 1993; Kwa *et al.*, 1995). In this respect it is logical that both proteins cluster with the only other BZ resistance linked protein sequence included in this phylogenetic tree, *C. elegans* BEN-1 (Figure 5.7). This implies that both the *H. contortus* β-tubulin ISO-1 and -2 are homologous to the free-living and *P. pacificus* nematode BEN-1 polypeptides. It is intriguing that neither of these *H. contortus* protein sequences clusters with the TBB-1/2 cluster of the phylogenetic tree. These polypeptides have essential roles in *C. elegans* embryonic development and both meiotic and mitotic spindle formation (Chenggang *et al.*, 2004). Deletions in the *C. elegans ben-1*, *tbb-1* or *tbb-2* genomic loci individually have no deleterious effect of the viability of *C. elegans* worms (Driscoll *et al.*, 1989; Chenggang *et al.*, 2004). However, an embryonic lethal phenotype is associated when there is deletion of both the *C. elegans tbb-1* and *tbb-2* loci (Chenggang *et al.*, 2004). This is indicative that the *tbb-1* and *-2* genes are functionally redundant and that the *ben-1* gene is non-essential.

Another aspect of this analysis, based on this phylogenetic tree and genetic

distance figures, is that the *H. contortus* ISO-1 and -2 polypeptides are more related to one another than they are to any other used in this analysis (Figure 5.7, data not shown). This indicates that these loci are paralogous.

5.2.2.3 Phylogenetic analysis of the β-tubulin protein sequences from the *Caenorhabditis* genus, *Pristionchus pacificus* and *Haemonchus contortus* and other Trichostrongyloid nematodes

A total of 5 unique cDNA sequences encoding full length β-tubulin polypeptide sequences from species of the Trichostrongyloid superfamily of nematode phylogeny were identified using the Uniprot database (http://www.uniprot.org/). The translated protein sequences from these cDNA sequences along with those of the β-tubulin families of *C. elegans*, *C. briggsae*, *C. japonica*, C. *remanei*, *C. brenneri* and *P. pacificus* were constructed into a phylogenetic tree using the Maximum Parsimony algorithm (Figure 5.8).

The tree is very similar to that of Figure 5.7. The clusters of the TBB-1/2, MEC-7/ISO-3 and TBB-4/ISO-4 proteins all contain the same members. Relationships within clusters are different to those in Figure 5.6 and, again, are associated with poor conservation of topology and low bootstrap values (De Ley and Blaxter, 2004). Unlike the phylogenetic tree of Figure 5.7, the *H. contortus* ISO-1 and -2 proteins cluster independently and separately from that of BEN-1. The bootstrap values that separate this Clade into clusters are relatively low at 23 and 33 (Isotype-2 cluster value of 76 apart), respectively. This is an indication of the high level of similarity shared between these sequences at the amino acid level (Duff and Nickrent, 1998).

The ISO-1 and ISO-2 clusters both include Trichostrongylid β-tubulin protein sequences which have been implicated in giving rise to BZ resistance (Samson-Himmelstjerna *et al.*, 2007). There are 24 known amino acid differences between the *H. contortus* ISO-1 and -2 protein sequences (Njue and Prichard, 2003). All sequences within either the ISO-1 or -2 clusters show a higher identity to the *H. contortus* protein that the sequence clusters with at these residues (Figure 5.9). This is evidence to support the ISO-1 and -2 clusters formed by this phylogenetic tree, and is indicative that they represent

orthologous gene groups.

5.2.2.4 Phylogenetic and sequence analysis of β-tubulin protein sequences from Clade V parasitic nematodes

All twenty full length β-tubulin protein sequences from Clade V parasitic nematodes added to the previous tree branch into two clusters (Figure 5.10). Each cluster was named based on the presence of either *H. contortus* ISO-1 or -2. There is poor conservation of species topology within all clusters of this tree, with ISO-2 being the only exception, where sequences for the Strongyloidae and Trichostrongyloidae Super-families branch independently within this cluster.

There are low supporting bootstrap percentages for the formation of the ISO-1 and -2 clusters (Figure 5.10). 24 specific amino acid residues were selected from each of the sequences which branched into either the ISO-1 or -2 clusters. The 24 residues are the positions at which the *H. contortus* ISO-1 and -2 proteins differ (Njue and Prichard, 2003). Using the sequences of these residues, each cluster member was more related to both the *H. contortus* sequences with which it clustered, and the rest of the sequences from the same cluster, than to any sequence from the opposing cluster (Figure 5.11). This evidence justifies the formation of the ISO-1 and -2 clusters, despite low bootstrap support.

5.2.2.5 Phylogenetic analysis of the β-tubulin protein sequences from Clade I-IV parasitic nematodes

The application of phylogenetics to establish orthologous relationships between β-tubulin sequences from species which branch in different Clades of nematode phylogeny appears to be limited in its use (Figure 5.12; Blaxter, 1998). The only Clade I sequence used in this analysis clusters alone. This sequence is the closest to the root of the tree of all investigated sequences. Both Clade IV sequences cluster with the *C. elegans* TBB-6 protein. This protein is the transcribed product from a *C. elegans* lineage specific duplicated β-tubulin locus (M. Han, Wormbase, personal communication). The Clade III sequences used in constructing this tree are largely more related to

one another than to any other sequence used in this analysis. The only exception to this is the *Onchocerca gibsoni* β-tubulin polypeptide which branches with the MEC-7 cluster, with high bootstrap support.

5.3 Discussion

The polypeptide sequences of β-tubulin proteins are highly similar, both within and between families from a wide range of phylogentically diverse species (Njue and Prichard, 2003). β-tubulin proteins are the known target of many drug classes, and mutations within nematode specific β-tubulin proteins have been shown to correlate with a rise in BZ resistance (Kwa *et al.*, 1993; Prichard, 2001; Wolstenholme *et al.*, 2004).

The *H. contortus* shotgun genomic sequence read databases were tBLASTn searched with each of the six *C. elegans* β-tubulin protein sequences. This revealed a total of 88 reads which contained β-tubulin sequence which were assembled into four distinct loci (Tables 5.2 and 5.3). This expands the gene family members, from the two previously characterised, to four (Geary *et al.*, 1992). I have defined the two additional loci as *H. contortus* β-tubulin *isotypes 3* and *4*. The coverage of three of these four loci within the ~800 Mb of *H. contortus* shotgun genomic sequence data is very similar. The exception to this is the *isotype-4* locus which is represented by only nine genomic shotgun sequence reads. This locus is the only one of the four identified which does not contain sufficient sequence data to cover the complete transcribed region. The fact that the *Hc-isotype-4* locus is so poorly represented in the current *H. contortus* genomic sequence databases is intriguing. The >800 Mb of data available for this genome sequencing project has all been sequenced using the Sanger shotgun sequencing methodology. A well established caveat of this sequencing technique is 'cloning bias', in which certain genomic regions are difficult to clone into plasmid vectors. This results in under representation of these loci in the sequenced data. Whether the *Hc-isotype-4* resides within such a region remains undefined. However, with the next generation sequence technologies currently being applied to vastly increase the wealth of *H. contortus* genomic sequence data, it should be possible to address this irregularity in the near future as no cloning bias exists.

cDNA sequence was amplified, cloned and sequenced for the complete *isotype-3* locus and all but the predicted initial ~600 bp of *isotype-4*. Complete cDNA sequences for the *isotype-1* and -2 loci have previously been obtained

and submitted to Genbank (Geary *et al.*, 1992). The transcribed cDNA sequences from the four loci are 74-79 % identical to one another (Table 5.4). This level of differentiation at the cDNA level confirms that each is expressed from a different locus of the *H. contortus* genome. The β-tubulin polypeptide sequences translated from these cDNA sequences are between 89-94 % identical and 94-97 % similar to each another (Table 5.5). These polypeptide sequences are distinct only at the C-terminal region. This is the only hyper-variable region of all β-tubulin polypeptides which have been sequenced from any organism (Chen *et al.*, 2008).

The β-tubulin gene family of the free-living nematode *C. elegans* has six members (Figure 5.2). Using these six polypeptides in BLAST (tBLASTn/BLASTp) of the available sequence databases for the free-living nematodes *C. briggsae*, *C. japonica*, C. *remanei* and *C. brenneri* identified between 3-5 β-tubulin loci within the genome of each organism. The same approach identified three β-tubulin loci in the draft genome sequence of the necromenic nematode *P. pacificus* (Dieterich *et al.*, 2008). The only completed and finished genome sequence project within this group is that of *C. briggsae* (Stein *et al.*, 2003). Only continuous sequence databases were searched for β-tubulin loci from these species, as piecing together gene families from core shotgun genomic sequence read data is a time consuming and laborious task. Clear orthologous relationships were established between the identified β-tubulin proteins from each of these organisms with those of *C. elegans* using phylogenetic, reciprocal BLAST and syntenic analysis (Figure 5.6). None of the identified polypeptides from these organisms formed an orthologous relationship with that of *C. elegans* TBB-6. This is consistent with my reciprocal BLAST analysis, indicating that none of the genomes of the investigated species contain a β-tubulin gene homologous to that of *C. elegans tbb-6*.

Phylogenetic analysis of the β-tubulin proteins from the free-living nematodes *C. elegans*, *C. briggsae*, *C. japonica*, C. *remanei*, *C. brenneri* plus those from the necromenic species *P. pacificus* form four distinct clusters with high bootstrap support, inferring orthologous relationships between proteins. This tree formed BEN-1, TBB-1/2, TBB-4 and MEC-7 clusters. Addition of the four *H. contortus* proteins to this phylogenetic tree resulted in ISO-3 and -4 grouping

with the MEC-7 and TBB-4 clusters respectively, with high bootstrap support (Figure 5.7). The addition of the *H. contortus* proteins to these clusters is also re-enforced by reciprocal BLASTp analysis and, in the case of ISO-3, syntenic investigation (Figure 5.3).

Both the *H. contortus* ISO-1 and -2 proteins clustered with the BEN-1 proteins from *C. elegans*, *C. briggsae*, *C. japonica*, C. *remanei*, *C. brenneri* and *P. pacificus* (Figure 5.7). Mutations in both these β-tubulin loci from the *H. contortus* genome have been implicated in BZ resistance (Kwa *et al.*, 1993; Kwa *et al.*, 1995). This is in conflict with the free-living nematodes where only one, the *C. elegans ben-1* locus, has been shown to confer resistance to BZ drugs (Driscoll *et al.*, 1989). This suggests that the mechanism by which both species obtain the resistance phenotype is similar but not identical. Reciprocal BLAST and genetic distance scores imply that the *Hc-isotype-1*, *Hc-isotype-2* and *Ce-ben-1* gene are paralogous. As a result, neither of these parasitic genes can be proven to be the direct orthologue of the *Ce-ben-1* locus. This paralogous relationship makes sense when each gene is considered alongside its involvement in the BZ resistance phenotype. Only the *ben-1* locus of *C. elegans*, but both *isotypes-1* and *-2* of *H. contortus,* have been implicated in this resistance mechanism (Driscoll *et al.*, 1989; Kwa *et al.*, 1993). The *H. contortus* ISO-1 and -2 proteins are more related to each other than any other protein included in this analysis. In addition, *H. contortus* worms with a deleted *isotype-2* locus are perfectly viable, with only the phenotype being heightened resistance to BZ drugs when present in conjunction with known BZ resistance conferring mutations in the *isotype-1* locus (Kwa *et al.*, 1995; Beech *et al.*, 1994). Specific point-mutations, and not deletions, within the *isotype-1* gene have been shown to confer resistance to BZ drugs (Samson-Himmelstjerna *et al.*, 2007). However, specific and repeatable knock-down of the *isotype-1* transcript by RNAi has been successful; with no associated phenotype at least in L3 stage larvae (B. Samarasinghe, University of Glasgow, personal communication). This indirect experimental evidence implies that the *H. contortus* ISO-1 and -2 proteins are functionally redundant. However specific experimental investigation is required to test this hypothesis directly. This paralogous relationship is further supported by similarity in the gene model structures of the three genes (Figures 5.1 and 5.2). Such conservation

of exon/intron boundary break-points is shown only in this example and that of *Hc-isotype-3*/*Ce-mec-7*.

Addition of Trichostrongyloidae Super-family and Clade V β-tubulin polypeptides to this phylogenetic tree provides another dimension to this hypothesis. Each of the sequences from parasitic nematode species branch into one of two distinct clusters (Figure 5.8). Although bootstrap support is poor, sequence analysis justifies the separation of these clusters. As none of these parasite sequences branch with any sequence from the free-living or *P. pacificus* sequences, it suggests that the duplication of the *isotype-1* locus to form that of *isotype-2* was after the divergence of the parasitic nematode lineage from that of the free-living nematode/*P. pacificus*. However, a crucial bias to this analysis must be considered here. The β-tubulin sequences added to this phylogenetic tree from all Clade V parasitic species, with the exception of *H. contortus*, were identified as BZ resistance linked loci, or through the use of degenerate primers designed to amplify *isotype-1* or *-2* like sequences. It is therefore possible that β-tubulin loci more similar to *C. elegans ben-1* than those presented in this data are present within the genome sequences of the investigated species. There is a current focus on vastly increasing the available genomic resources for parasitic nematode species. As such genomic sequences become available, this hypothesis can be investigated more directly.

From the results in this chapter, the use of sequence analysis to generate hypothesises on the orthologous relationships between β-tubulin polypeptide sequences from nematodes in different Clades is limited in its application. The only exception to this is an *O. gibsoni* β-tubulin sequence which clusters with the MEC-7 polypeptides of five free living species, as well as that of *P. pacificus*, with high bootstrap support (Figure 5.12). This is not a surprising result given the complexity of hypothesis for the evolution of the Clade V β-tubulin gene family.

In summary the *H. contortus* genome contains four distinct β-tubulin loci. Analysis of these loci and the proteins expressed provide evidence that the *isotype-3* and *-4* are direct orthologues of the *C. elegans mec-7* and *tbb-4* loci.

The ISO-1 and -2 protein sequences are both more similar to *C. elegans* BEN-1 than any other *C. elegans* β-tubulin. Phylogenetic analysis of the available β-tubulin polypeptide sequences from Clade V species suggests that these three loci are paralogous and that the *Hc-isotype-1* and *-2* genes have arisen from duplication of an ancestral *H. contortus* β-tubulin gene (Figure 5.10). Similarly, as no parasite sequence branches into the TBB-1/-2 cluster and as the *ben-1*, *tbb-1* and *tbb-2* genes are located on the same chromosome in both *C. elegans* and *C. briggsae*, it is hypothesised that these three loci are all duplications of one another. Based on the sequences from the species investigated in this chapter the concluding hypothesis is that all these duplication events happened after the divergence of the parasitic from the free-living/*P. pacificus* nematode lineages.

5.4 Tables and Figures

Table 5.1: Details of all unique nematode full-length β-tubulin protein sequences identified in the Uniprot database

Accession Number	Species	Super-family	Clade	Length of protein sequence (AA)
O44388	*Trichuris trichiuria*	Trichuridae	I	444
A8NQI2	*Brugia malayi*	Filarioidea	III	452
A8PHP7	*Brugia malayi*	Filarioidea	III	448
P18241	*Brugia pahangi*	Filarioidea	III	448
O61357	*Dirofilaria immitis*	Filarioidea	III	448
O61358	*Onchocerca volvulus*	Filarioidea	III	448
P41387	*Onchocerca gibsoni*	Filarioidea	III	444
Q4KS26	*Strongyloides ratti*	Panagrolaimoidea	IV	447
Q4KS25	*Strongyloides stercoralis*	Panagrolaimoidea	IV	447
A3FKG1	*Ancylostoma duodenale*	Ancylostomatoidea	V	448
B5AX19	*Ancylostoma caninum*	Ancylostomatoidea	V	448
Q2QJK8	*Ancylostoma caninum*	Ancylostomatoidea	V	448
A3FKG2	*Necator americanus*	Ancylostomatoidea	V	448
Q5ZPR9	*Cylicostephanus goldi*	Strongyloidae	V	448
Q7KBQ5	*Cyathostomum pateratum*	Strongyloidae	V	448
Q7KBP8	*Cylicocyclus radiatus*	Strongyloidae	V	448
Q7KBP9	*Cylicocyclus insigne*	Strongyloidae	V	448
Q7KBQ4	*Cyathostomum coronatum*	Strongyloidae	V	448
Q8MV61	*Cylicocyclus elongatus*	Strongyloidae	V	448
Q6BCK8	*Cylicocyclus nassatus*	Strongyloidae	V	450
A9LNP9	*Cylicocyclus nassatus*	Strongyloidae	V	448
Q9GT34	*Cyathostomum catinatum*	Strongyloidae	V	448
A8WEX2	*Cyathostomum catinatum*	Strongyloidae	V	450
Q5ZPS0	*Cylicostephanus longibursatus*	Strongyloidae	V	448
B0LJ60	*Dictyocaulus viviparus*	Trichostrongyloidae	V	449
Q6X2D7	*Cooperia oncophora*	Trichostrongyloidae	V	448
Q6X2D8	*Cooperia oncophora*	Trichostrongyloidae	V	448
Q26900	*Trichostrongylus colubriformis*	Trichostrongyloidae	V	448
Q26901	*Teladorsagia circumcincta*	Trichostrongyloidae	V	448

Table 5.2 79 of the 88 *H. contortus* β-tubulin-containing shotgun sequence reads significantly align to the genomic locus of either *isotype-1*, *2* or *3*

H. contortus genomic shotgun sequence read	% Identity in Pair-wise aligning to the genomic locus of:			Amino Acids regions contained on read
	isotype-1	*isotype-2*	*isotype-3*	
haem-550k18.p1k	96	-	-	1-19
haem-1852d17.q1k	93	-	-	1-19
haem-190g22.p1ka	97	-	-	1-19
haem-888h23.q1k	98	-	-	1-19
haem-945d19.p1k	94	-	-	19-54
haem-2246e20.q1k	93	-	-	19-56
haem-53i19.q1k	98	-	-	19-56
haem-811a05.q1k	92	-	-	19-56
haem-550k18.q1k	92	-	-	19-56
haem-606f07.p1k	93	-	-	36-92
haem-748e02.p1k	92	-	-	26-87
haem-1754g06.q1k	95	-	-	47-189
haem-1530n24.p1k	95	-	-	52-222
haem-2422h18.p1k	95	-	-	140-356
haem-971a20.q1k	95	-	-	148-324
haem-284j15.p1k	92	-	-	182-246
haem-606f07.q1k	95	-	-	244-346
haem-609c22.p1k	93	-	-	293-324
haem-172f18.p1k	94	-	-	293-377
haem-1109k18.p1k	94	-	-	295-389
haem-1967a22.q1k	94	-	-	318-389
haem-2248p04.p1k	97	-	-	404-437
haem-2220f03.p1k	98	-	-	425-448
haem-1477g11.p1k	-	98	-	1-19
haem-1305l22.q1k	-	98	-	1-19
haem-1296b22.q1k	-	96	-	1-19
haem-1049c22.q1k	-	100	-	1-19
haem-206f16.q1ka	-	92	-	1-19
haem-1429c07.p1k	-	95	-	19-131
haem-1770g16.q1k	-	96	-	19-131
haem-2598c20.p1k	-	97	-	19-75
haem-2334l06.p1k	-	91	-	19-170
haem-2155f20.q1k	-	99	-	37-131
haem-2528o08.q1k	-	92	-	103-164
haem-1217k06.p1k	-	97	-	174-223
haem-2155f20.p1k	-	95	-	174-228
haem-2639h23.q1k	-	96	-	174-228
haem-1966c18.p1k	-	98	-	174-228
haem-306d04.q1k	-	94	-	229-285
haem-948g21.p1k	-	92	-	229-324
haem-1477g11.q1k	-	90	-	229-324
haem-1595i07.q1k	-	90	-	229-324
haem-2463e18.p1k	-	95	-	267-390
haem-2639h23.p1k	-	96	-	295-324

haem-230j12.p1k	-	**97**	-	325-411
haem-1036m08.p1k	-	**98**	-	325-438
haem-2510h23.p1k	-	**99**	-	325-390
haem-1890a22.p1k	-	**98**	-	425-448
haem-2482m04.p1k	-	-	**95**	1-24
haem-2575h22.p1k	-	-	**97**	1-24
haem-557i09.q1k	-	-	**98**	1-24
haem-1395o16.q1k	-	-	**98**	1-24
haem-1543e12.q1k	-	-	**99**	19-59
haem-2590b24.q1k	-	-	**93**	19-62
haem-2543p05.p1k	-	-	**91**	19-62
haem-2424g23.q1k	-	-	**90**	19-62
haem-1198i05.q1k	-	-	**100**	49-62
haem-1218j17.p1k	-	-	**94**	62-86
haem-1395o16.p1k	-	-	**98**	62-94
haem-1060e05.q1k	-	-	**95**	62-94
haem-1543e12.p1k	-	-	**98**	62-131
haem-1754f16.q1k	-	-	**98**	90-180
haem-1938c12.q1k	-	-	**98**	90-221
haem-1942c03.q1k	-	-	**92**	93-234
haem-1796o02.q1k	-	-	**91**	93-131
haem-1796p11.q1k	-	-	**90**	93-131
haem-2575h22.q1k	-	-	**99**	95-205
haem-536h02.p1k	-	-	**98**	131-246
haem-2424g23.p1k	-	-	**91**	245-324
haem-1218j17.q1k	-	-	**98**	281-324
haem-2526g19.q1k	-	-	**98**	281-324
haem-1198i05.p1k	-	-	**99**	321-338
haem-2543p05.q1k	-	-	**90**	318-350
haem-519c21.p1k	-	-	**100**	350-389
haem-2256o17.q1k	-	-	**98**	350-383
haem-1754f16.p1k	-	-	**99**	350-389
haem-2493c12.p1k	-	-	**90**	382-426
haem-1942c03.p1k	-	-	**92**	382-441

Table 5.3 Nine *H. contortus* β-tubulin-containing shotgun genomic sequence reads that do not significantly align to the fully sequenced genomic loci of *isotype-1*, *2* or *3* assemble into three contiguous sequences and one unassembled read

H. contortus genomic shotgun sequence read	β-tubulin Amino Acid regions contained on sequence
haem-111c22.q1k	70-89
Contiguous sequence 1	
haem-2316m19.q1k	90-131
haem-576l15.p1k	90-204
haem-1208p22.p1k	179-246
Contiguous sequence 2	
haem-478h19.p1k	245-282
haem-1225b24.q1k	245-282
Contiguous sequence 3	
haem-2316m19.p1k	355-386
haem-57615.q1k	362-444
haem-2287g18.p1k	408-426

Table 5.4 Percentage pair-wise identity of the transcribed cDNA sequences for four identified *H. contortus* β-tubulin loci

cDNA sequence	cDNA pair-wise comparison identity (%)			
	***isotype-1* (M76493)**	***isotype-2* (M76491)**	***isotype-3***	***isotype-4 ** **
***isotype-1* (M76493)**	-	77	74	79
***isotype-2* (M76491)**	77	-	75	77
isotype-3	74	75	-	75
***isotype-4 ** **	79	77	75	-

* The cDNA sequence of *H. contortus* β-tubulin *isotype-4* is incomplete. Based on alignments with β-tubulin polypeptides from a number of species, sequence encoding the predicted initial 183 amino acid residues is missing. Despite best efforts using a variety of molecular methods, this region has been unable to be amplified during the course of this project (data not shown).

Table 5.5 Pair-wise alignment percentage values for the translated protein sequences for four *H. contortus* β-tubulin loci

A. Identical residues

Protein sequence	Protein sequence pair-wise comparison identity (%)			
	Isotype-1 (M76493)	Isotype-2 (M76491)	Isotype-3	Isotype-4 *
ISO-1 (M76493)	-	94	89	89
ISO-2 (M76491)	94	-	89	90
ISO-3	89	89	-	90
ISO-4 *	89	90	90	-

B. Residues with similar properties

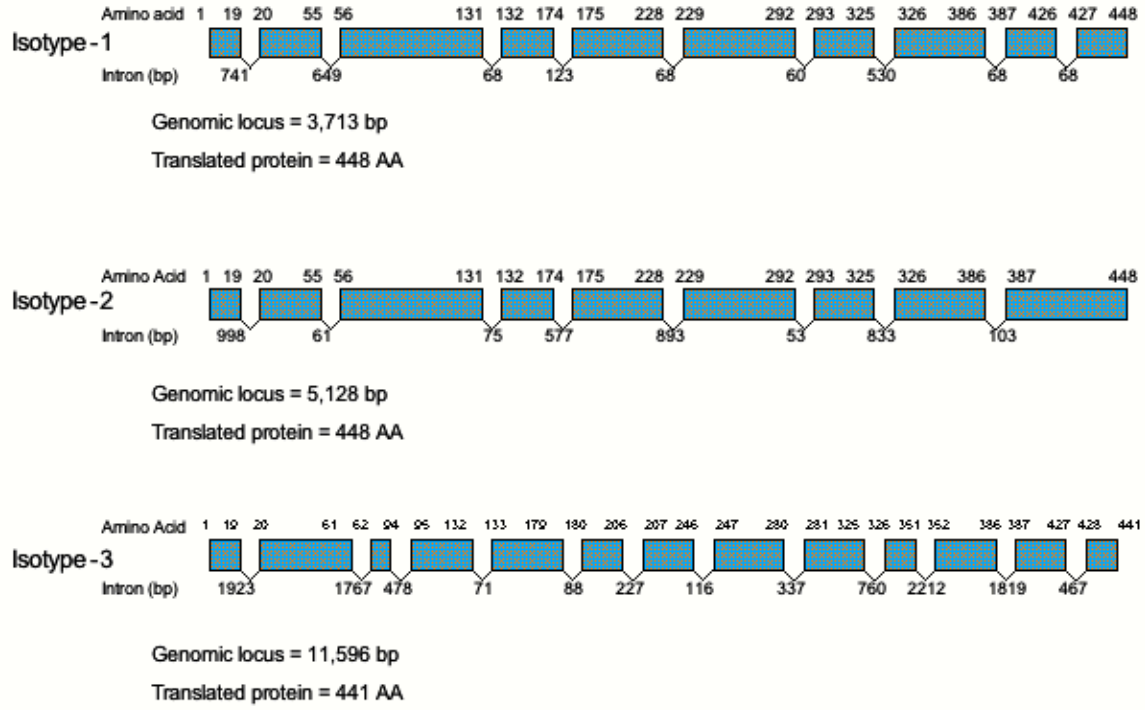
Protein sequence	Protein sequence pair-wise comparison similarity (%)			
	Isotype-1 (M76493)	Isotype-2 (M76491)	Isotype-3	Isotype-4 *
ISO-1 (M76493)	-	97	94	95
ISO-2 (M76491)	97	-	95	96
ISO-3	94	95	-	96
ISO-4 *	95	96	96	-

* The cDNA sequence used to predict the protein sequence of *H. contortus* β-tubulin locus *isotype-4* is incomplete. Based on alignments with β-tubulin polypeptides from a number of species, sequence encoding the predicted initial 183 residues is missing.

Table 5.6: Characterisation of putative translated β-tubulin protein sequences for the nematode species *C. briggsae*, C. *remanei*, *C. japonica*, *C. brenneri* and *P. pacificus* based on top BLASTp alignments from *C. elegans* Wormpep

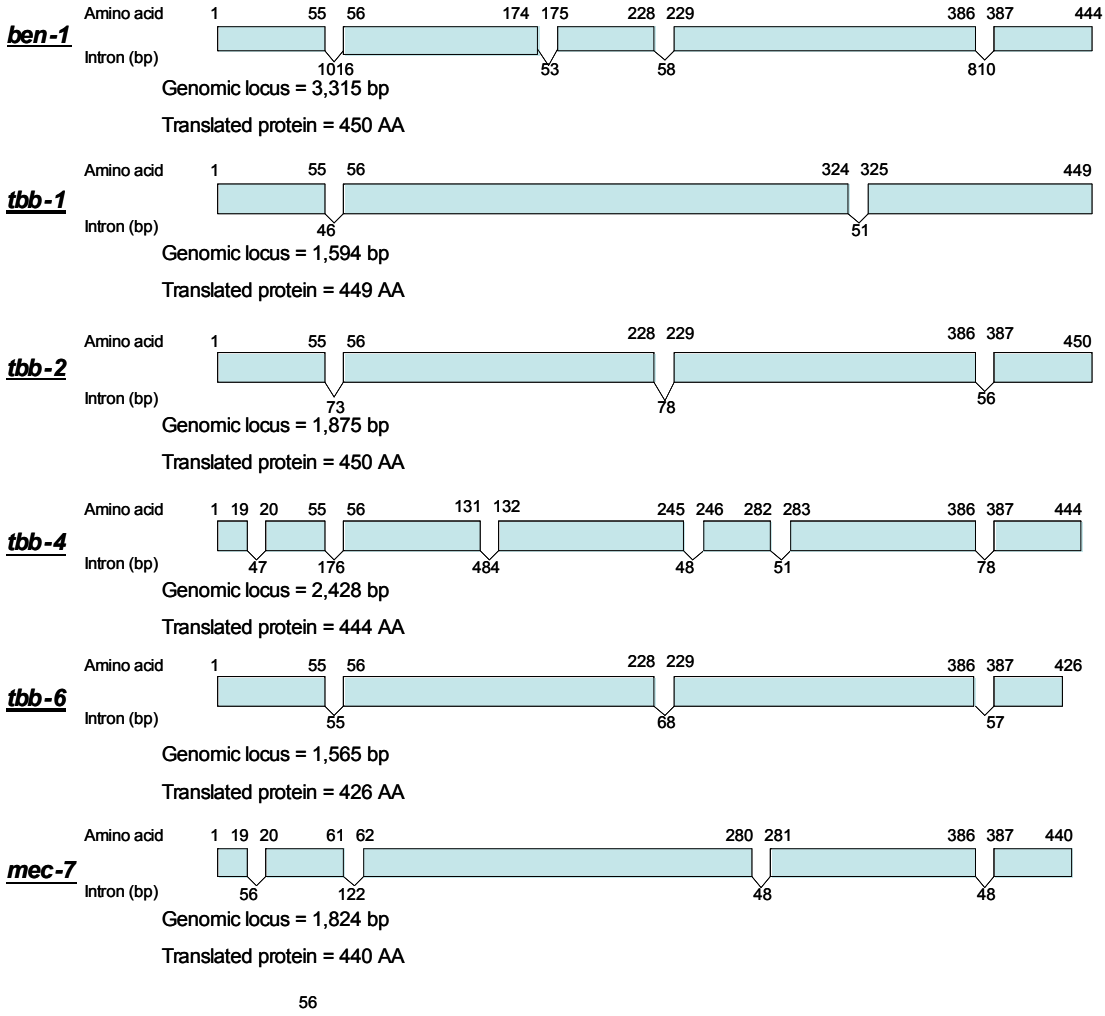
Species		Length of putative translated β-tubulin protein	Top *C. elegans* BLASTp alignment
C. briggsae	1	271	BEN-1
	2	441	MEC-7
	3	446	TBB-1
	4	449	TBB-2
	5	330	TBB-4
C. remanei	1	699	BEN-1
	2	441	MEC-7
	3	459	TBB-1
	4	449	TBB-2
	5	456	TBB-4
C. japonica	1	447	BEN-1
	2	441	MEC-7
	3	448	TBB-1
	4	450	TBB-2
	5	445	TBB-4
C. brenneri	1	212	BEN-1
	2	441	MEC-7
	3	449	TBB-1
P. pacificus	1	448	BEN-1
	2	439	MEC-7
	3	448	TBB-4

Figure 5.1 Deduced gene models for *H. contortus* β-tubulin *isotype-1*, *2* & *3*



The annotated genomic sequence for *H. contortus* β-tubulin *isotype-1* is available through Genbank (accession number: X67489). The *isotype-2* and *3* models were deduced by manual curation by aligning the full-length cDNA to the genomic sequences. The *isotype-1* and *-2* models are very similar, differing only in intron lengths and the fact that that the final exon of *isotype-2* is split into two in that of *isotype-1*. The gene model for *isotype-3* is vastly different. It contains far more introns than either *isotype-1* or *-2* and encodes for a smaller polypeptide. The *isotype-3* gene model is more than double the size of the two previously characterised *H. contortus* β-tubulin loci. Full DNA and translated sequences are shown in Appendix 3.

Figure 5.2 Gene models of the *C. elegans* β-tubulin gene family



The annotated gene models for all six of the *C. elegans* β-tubulin loci were obtained from the *C. elegans* Wormbase server (http://www.wormbase.org/). Exon/intron boundaries are conserved across the *Ce-ben-1/Hc-isotype-1/Hc-isotype-2* and *Ce-mec-7/Hc-isotype-3* gene models.

Figure 5.3 Conservation of macro and micro-synteny between the *H. contortus isotype-3* and *C. elegans mec-7* loci

17 loci homologous to *C. elegans* proteins were identified on a contiguous 409 Kb stretch of *H. contortus* X-chromosomal genomic sequence. 14 of the *C. elegans* proteins are encoded by genes on the X-chromosome, showing conservation of synteny between these genomic sequences. Micro-synteny is also evident. The putative homologous pair of ZK154.1 genes are both positioned closest to the β-tubulin *isotype-3* locus on the X-chromosome of the respective species.

Positions of the *C. elegans* genes along the 17 Mb X-chromosome are given in purple. Positions of *Hc-isotype-3* and *Ce-mec-7* are shown in blue.

(Adapted from Redman *et al.*, 2008)

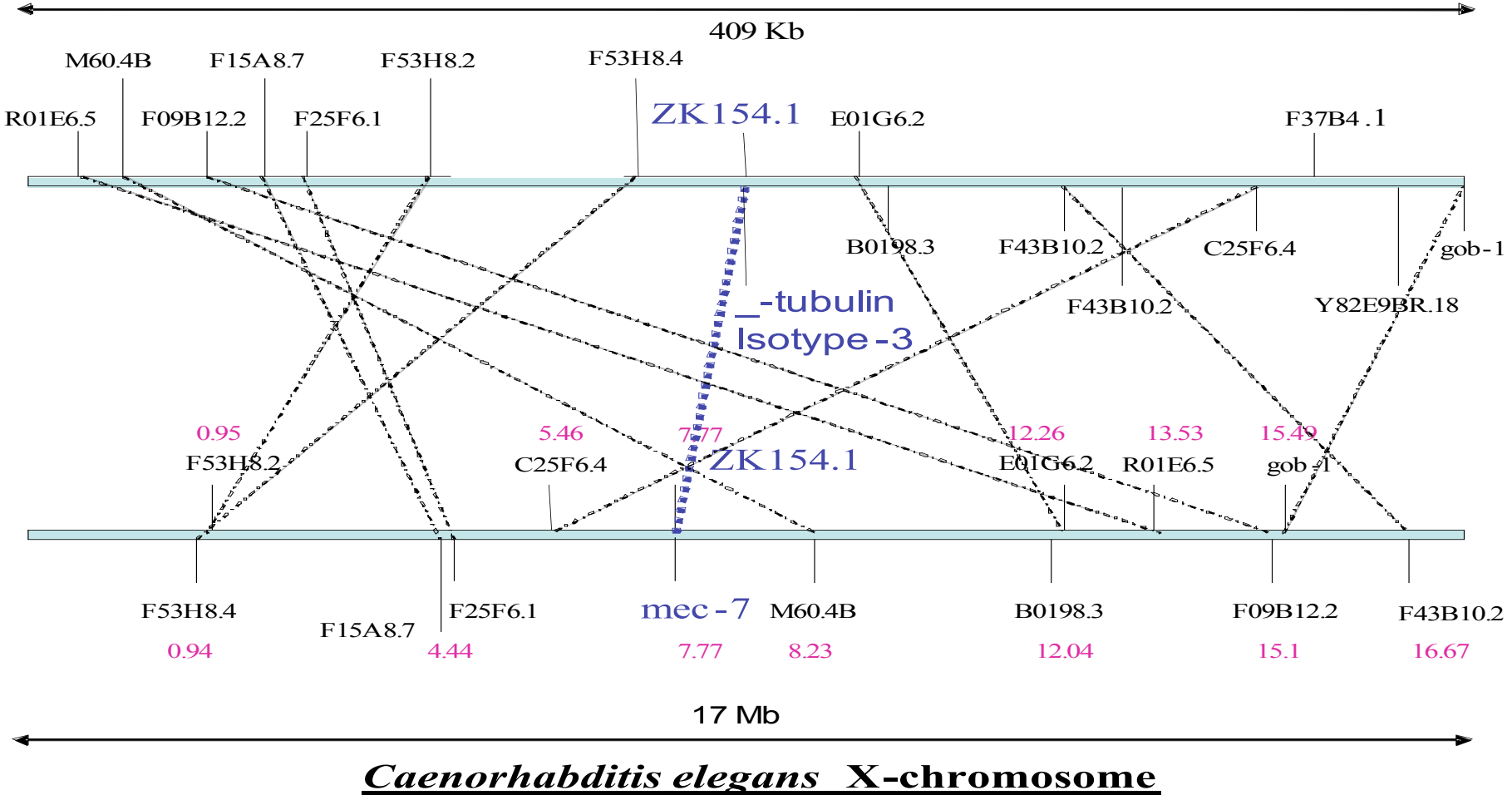

Haemonchus contortus contig
409 Kb
M60.4B
F15A8.7
F53H8.2
F53H8.4
R01E6.5
F09B12.2
F25F6.1
ZK154.1
E01G6.2
F37B4.1
B0198.3
F43B10.2
C25F6.4
gob-1
F43B10.2
Y82E9BR.18
_-tubulin
Isotype-3
0.95
5.46
7.77
12.26
13.53
15.49
F53H8.2
C25F6.4
ZK154.1
E01G6.2
R01E6.5
gob-1
F53H8.4
F15A8.7
F25F6.1
mec-7
M60.4B
B0198.3
F09B12.2
F43B10.2
0.94
4.44
7.77
8.23
12.04
15.1
16.67
17 Mb
Caenorhabditis elegans X-chromosome

Figure 5.4 Four *H. contortus* consensus sequences consisting of nine uncharacterised β-tubulin containing shotgun genomic sequence reads can be linked together via PCR amplification of gap regions

A. Schematic figure detailing the PCR amplification reactions attempted to link all 4 uncharacterised β-tubulin containing *H. contortus* genomic sequences to one locus. The dashed lines represent PCR products which are shown in figures B and C.

B. PCR amplification products showing linkage of haem-111c22.q1k to Consensus sequence 1. Numbers underneath the gel for each lane correlate to the link on schematic Panel A which was amplified. Positive control (PCR+) was carried out with primers Iso-3-2Kb F + R (Primer sequences detailed in Appendix 2). Negative control (PCR-) was the positive control mix minus template.

C. PCR products linking Consensus sequences 1, 2 and 3. Four possible primer sets were tested to try and link each of the consensus sequences. There are blank lanes on this gel as not all combinations amplified a product. Numbers underneath each lane of the gel correlate to the links in schematic Panel A. Positive control primers were Iso-1-2Kb F + R. Negative control PCR was the positive control mix minus template.

The primer sequences that were used to amplify these regions can be found in Appendix 2.

This figure demonstrates that all nine of the novel *H. contortus* β-tubulin containing shotgun genomic sequence reads can be assembled into one locus. I have named this locus *H. contortus* β-tubulin *isotype-4*.

Figure 5.4 Four *H. contortus* consensus sequences consisting of nine uncharacterised β-tubulin containing shotgun genomic sequence reads can be linked together via PCR amplification of gap regions

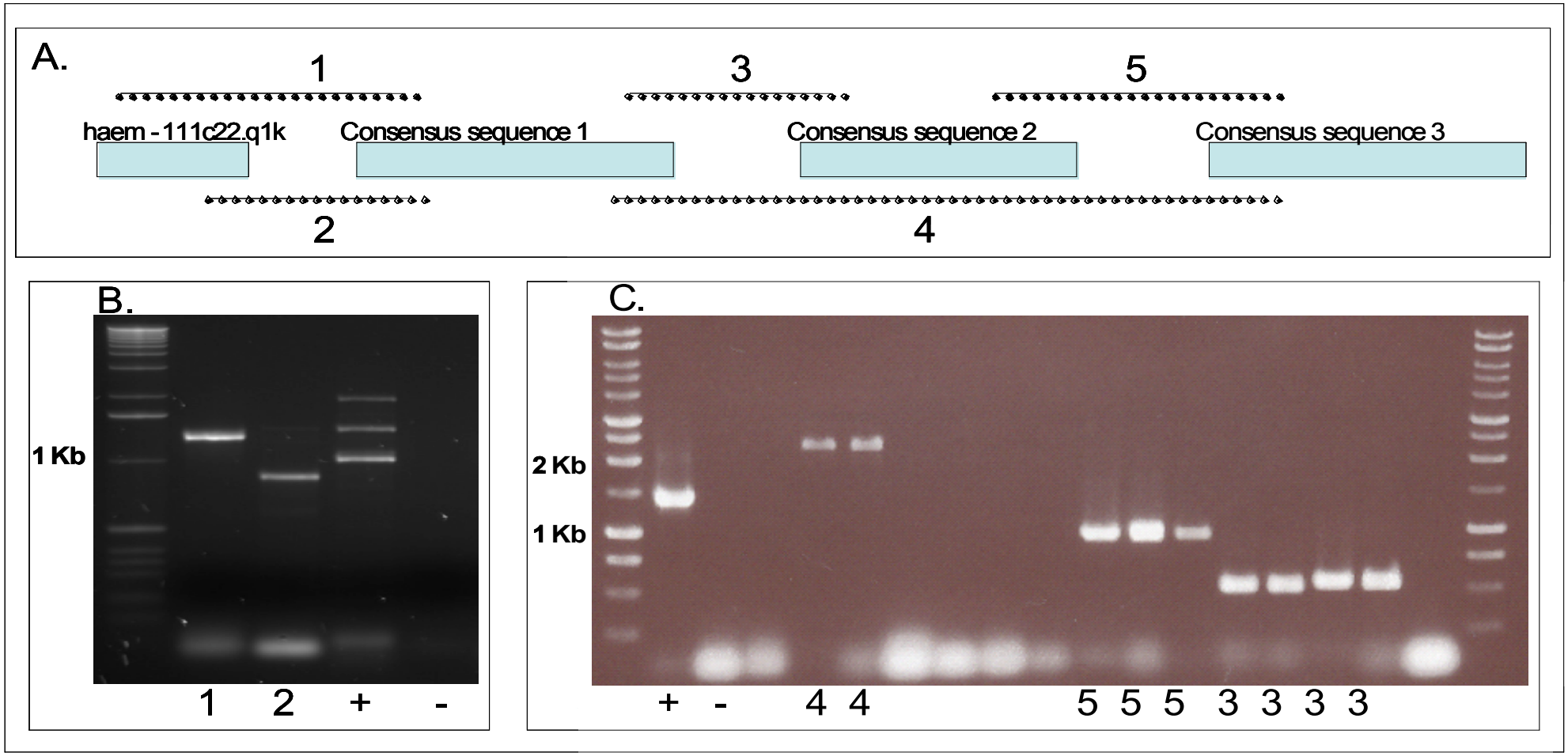
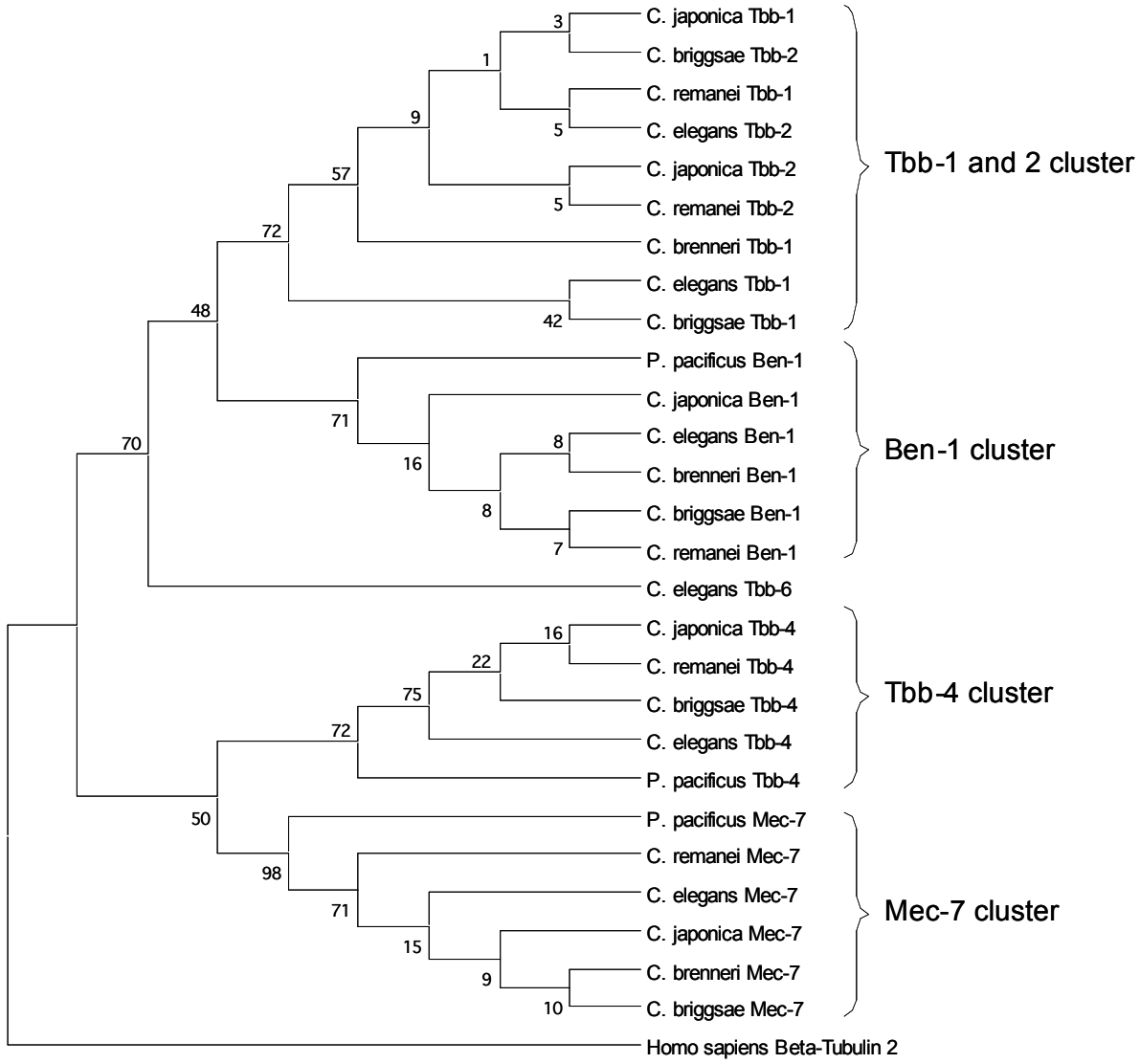


Figure 5.5 Alignment of the available translated protein sequences for the *H. contortus* β-tubulin Isotypes 1, 2, 3 and 4

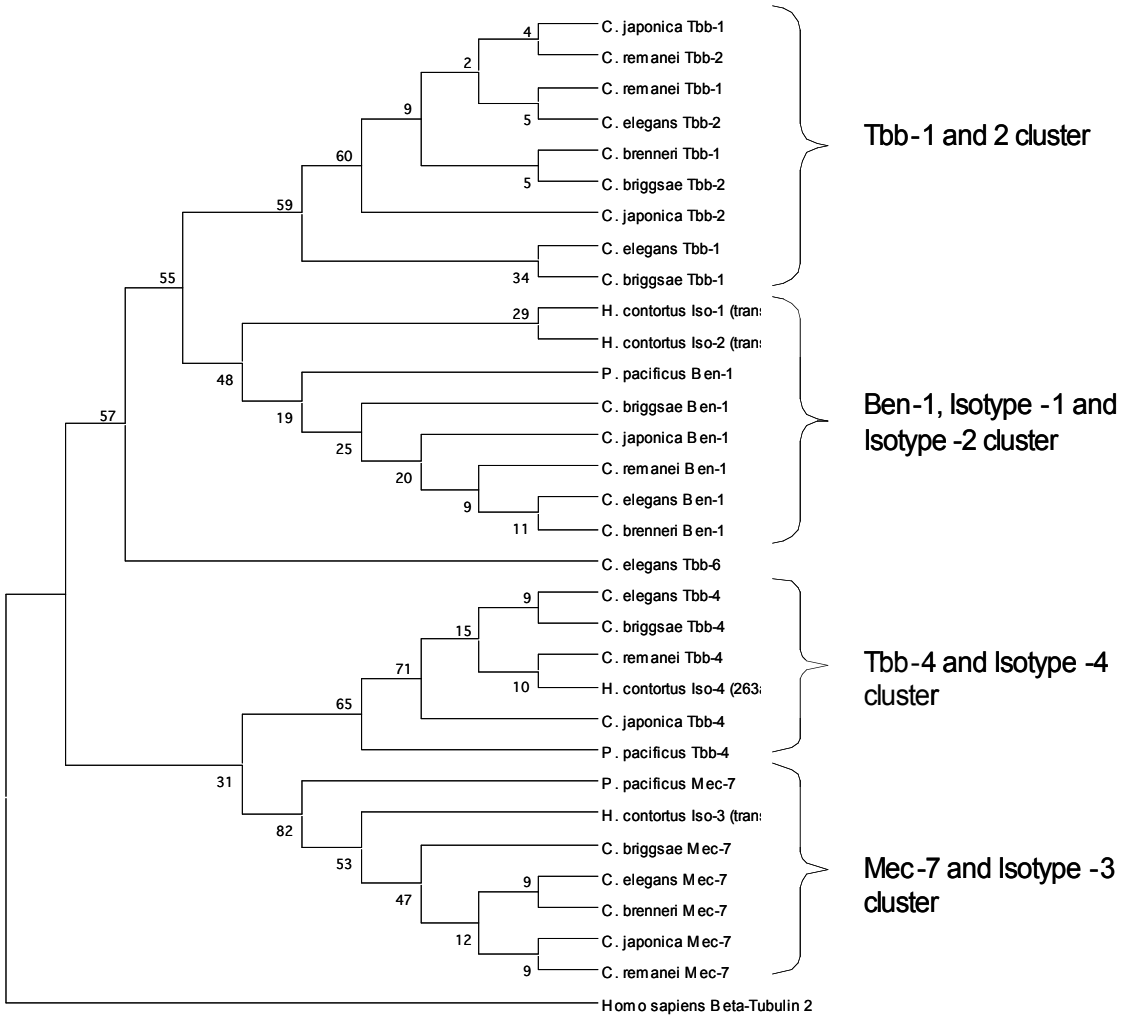
This alignment shows that the translated protein sequences for the *H. contortus* β-tubulin Isotypes 1, 2, 3 and 4 are highly conserved both in identity and similarity of amino acid residues. The conserved sequence motif which distinguishes α, β and γ tubulin family members at position 140-146 and the β-tubulin characteristic auto-regulatory motif at position 1-4 of the protein sequences for ISO-1-3 are highlighted by blue lines. There are three instances where the residues of all four *H. contortus* β-tubulin aligned sequences differ. All are contained within the hyper-variable C-terminal region of the polypeptides and are indicated by filled circles.

Figure 5.6: Phylogenetic tree of putative translated protein products for all annotated β-tubulin loci from the nematode species *C. elegans*, *C. briggsae*, C. *remanei*, *C. japonica*, *C. brenneri* and *P. pacificus*



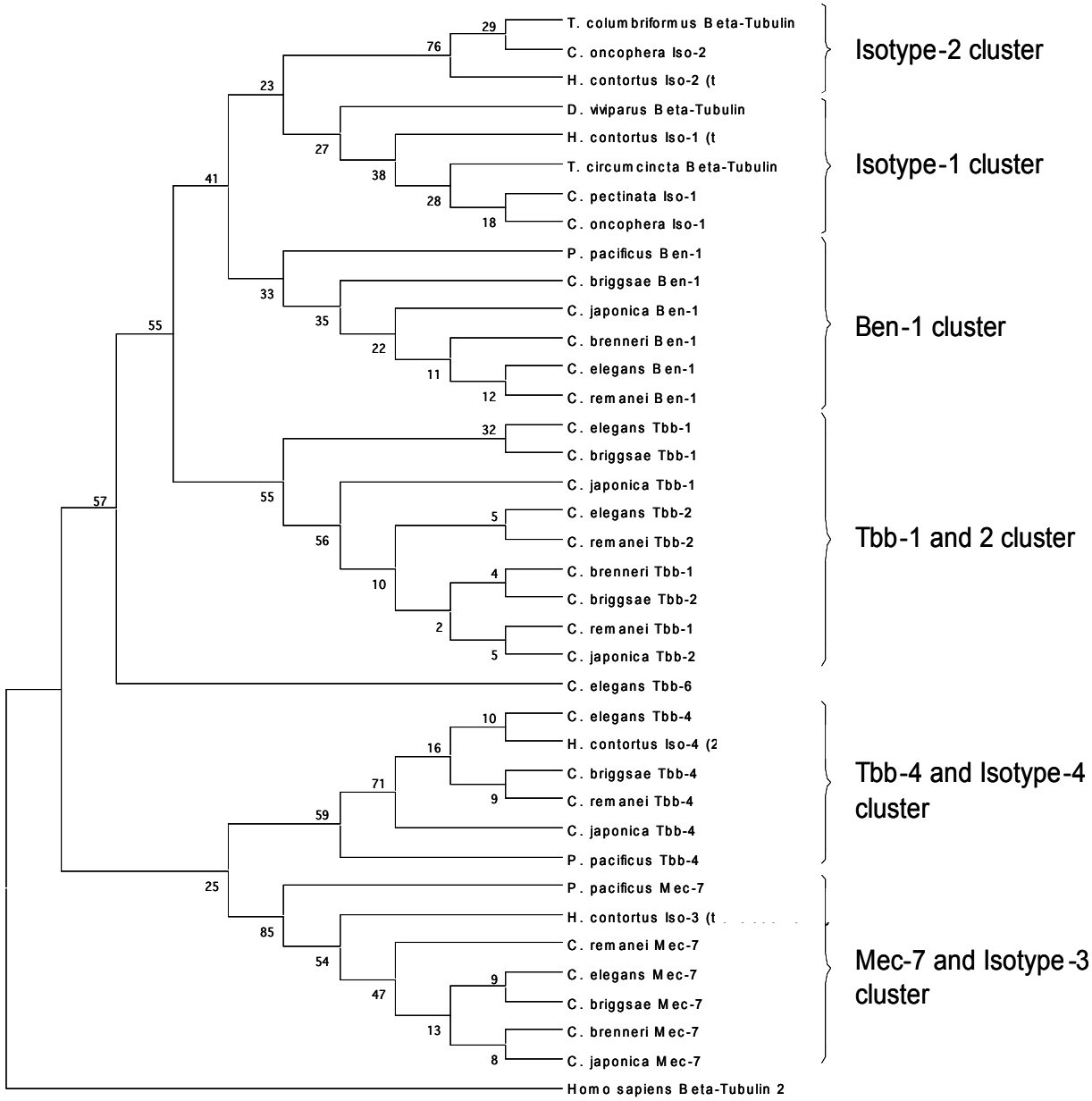
This tree was constructed using the Maximum Parsimony algorithm and bootstrapped for significance using 500 replicates. Sequences are named based on the best BLASTp alignment from *C. elegans* Wormpep (Table 5.6). Defined clusters, with bootstrap values ranging from 71-98, are evident for the MEC-7, TBB-4, BEN-1 and TBB-1/2 protein sequences. Intra-clustering bootstrap values are very low and there is little maintenance of topology at a species level. The *C. elegans* TBB-6 protein branches alone. This is consistent with my reciprocal BLAST analysis, indicating that none of the investigated genomes contain a β-tubulin gene homologous to this *C. elegans* locus.

Figure 5.7: Phylogenetic tree of putative translated protein products for all annotated β-tubulin loci from the nematode species *C. elegans*, *C. briggsae*, C. *remanei*, *C. japonica*, *C. brenneri*, *P. pacificus* and *H. contortus*



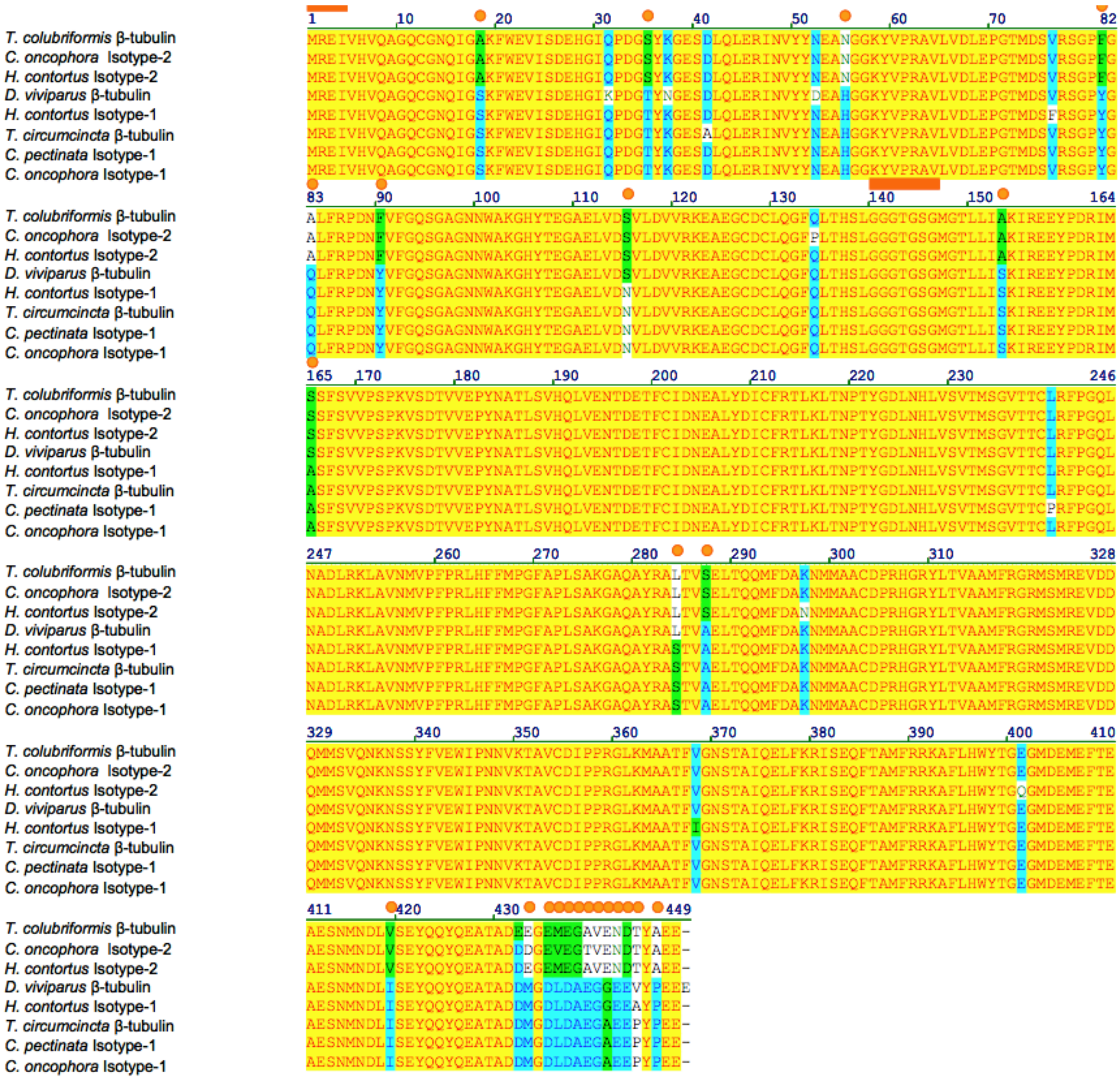
This tree was constructed using the Maximum Parsimony algorithm and bootstrapped for significance using 500 replicates. The tree forms four clusters, all of which are built with high confidence bootstrap values raging from 48-82. Within clusters species topology is poorly maintained and bootstrap values are commonly low. This tree hypothesises that both *H. contortus* ISO-1 and -2 proteins are homologous to *C. elegans* BEN-1. Similarly, orthologous relationships are hypothesised between *H. contortus* Isotype-3/-4 and the *C. elegans* MEC-7/TBB-4 proteins, respectively.

Figure 5.8: Phylogenetic analysis of the β-tubulin protein sequences of five members from the *Caenorhabditis* genus, *P. pacificus* and *H. contortus* plus all full-length sequences from other nematode species from the Trichostrongyloidae super-family



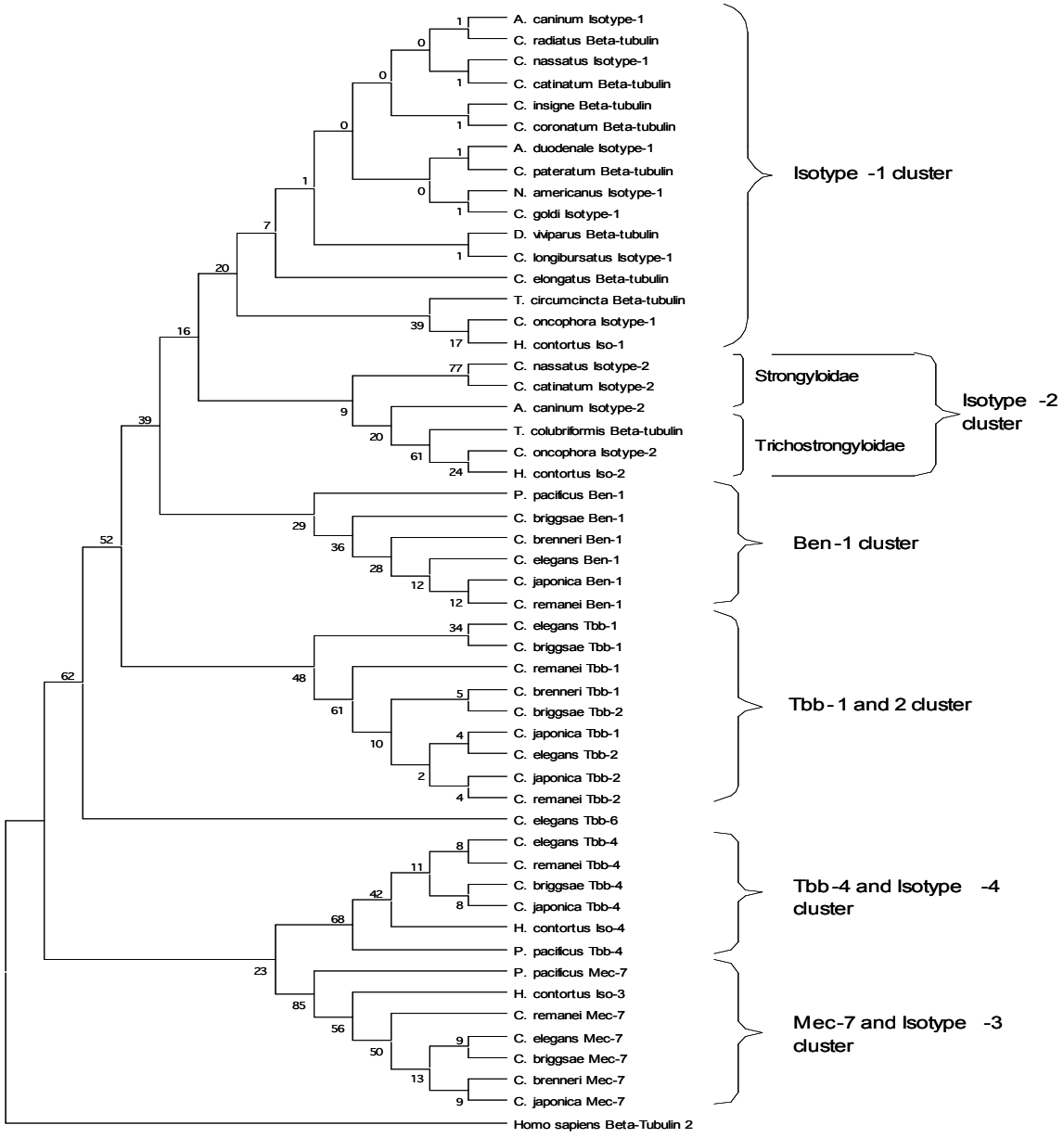
This tree was constructed using the Maximum Parsimony algorithm and bootstrapped for significance using 500 replicates. All additional sequences for β-tubulin proteins from the super-family Trichostrongyloidae branch into either the Isotype-1 or -2 clusters. The Isotype-2 cluster is constructed with high confidence bootstrap of 76. The separation of the Isotype-1 cluster from that of Ben-1 is supported with relatively low bootstrap analysis. Supporting evidence for the formation of this cluster is presented in Figure 5.9.

Figure 5.9: Alignment of all polypeptide sequences that branch within both the Isotype-1 and -2 clusters of the phylogenetic tree in Figure 5.8



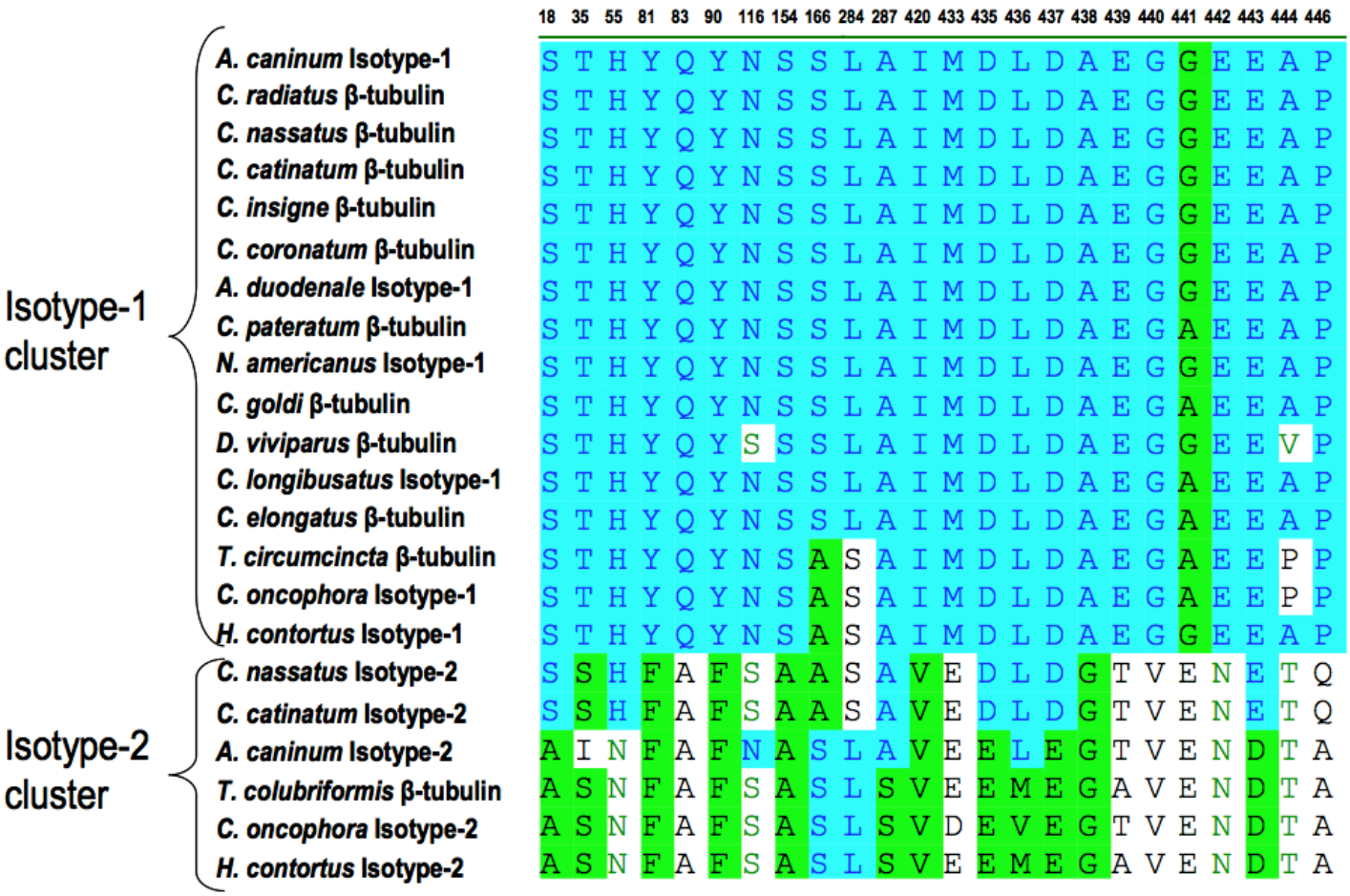
indicated by a filled orange circle. The auto-regulation signal at position 1-4 and tubulin signature sequence at position 140-146 are indicated by an orange line. This alignment shows that each of the sequences within the Isotype-1 or -2 clusters of the tree constructed in Figure 5.7 show a higher similarity to every other member of the same cluster than they do to any member of the other cluster.

Figure 5.10: Phylogenetic analysis of β-tubulin protein sequences of five members from the *Caenorhabditis* genus, *P. pacificus* and *H. contortus* plus all unique full-length sequences from other Clade V nematode species



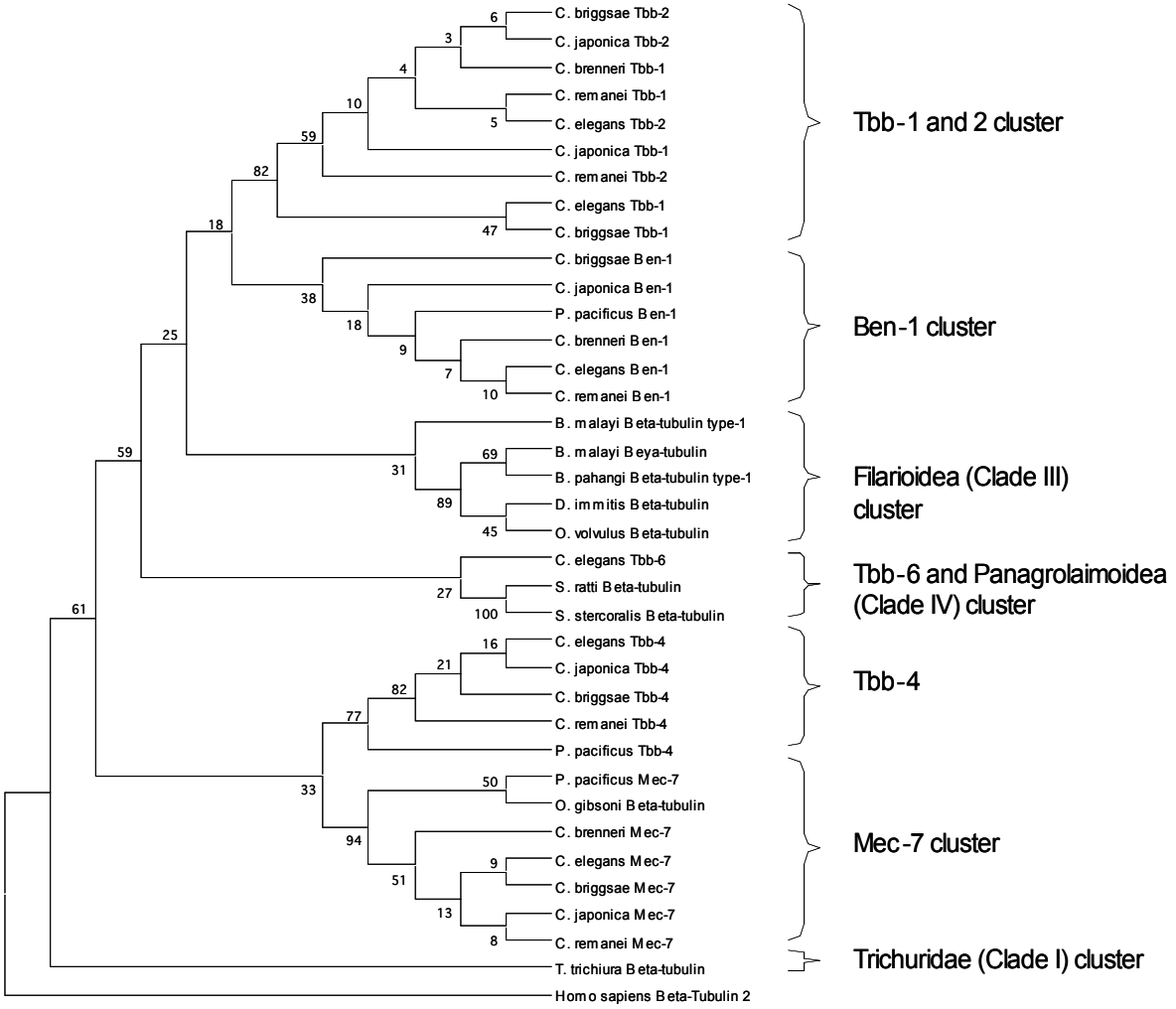
This tree was constructed using the Maximum Parsimony algorithm and bootstrapped for significance using 500 replicates. All β-tubulin sequences from parasite species from Clade V of nematode phylogeny branch within either the Isotype-1 or -2 clusters. Although the formation of these clusters is supported with low bootstrapping values, sequence analysis confirming their justification is presented in Figure 5.10. There is little conservation of species topology within this tree. The exception to this is the Isotype-2 cluster, where sequences from the Strongyloidae and Trichostrongyloidae Super-families cluster independently.

Figure 5.11: Alignment of 24 residues from the polypeptide sequences that branch within the Isotype-1 and -2 clusters of the phylogenetic tree in Figure 5.9



This is an alignment of the amino acid residues at the 24 positions unique to the *H. contortus* Isotype-1 and -2 proteins from all protein sequences that branch into either the Isotype-1 or -2 cluster of the phylogenetic tree in Figure 5.9. Members of each cluster are more identical to other members from the same cluster than they are to any sequence from the other cluster. This provides justification for the Isotype-1 and -2 cluster formation in Figure 5.9.

Figure 5.12: Phylogenetic analysis of β-tubulin protein sequences of five members from the *Caenorhabditis* genus, *P. pacificus* and unique full-length sequences from other nematode species from Clades I-IV of nematode phylogeny



This Maximum Parsimony tree was bootstrapped for significance using 500 replicates. On the whole, this tree is not very useful for establishing orthologous relationships. The only sequence from the Trichuridae Super-family branches independently from the rest of the investigated sequences, and closest to the root. Both sequences from the Panagrolaimoidea Super-family branch with Tbb-6, a *C. elegans* specific β-tubulin duplication. Each sequence from the Filarioidea Super-family is more related to one another than any other sequence used to construct this tree. The only exception to this, an example of where an orthologous relationship can be established, is with the *O. gibsoni* β-tubulin. This protein sequence branches with the Mec-7 cluster, with high bootstrap support.

Chapter 6: Comparative expression pattern analysis of *Haemonchus contortus* and *Caenorhabditis elegans* β-tubulin genes and proteins

6.1 Introduction

H. contortus is an economically important pathogen of small ruminants. Control of this parasitic nematode by the classical method of anthelmintic dosing is becoming increasingly redundant due to soaring levels of resistance to all three classes of anthelmintic drug (Jackson, 1993). Specific mutations in two *H. contortus* β-tubulin genes have been correlated with a rise in resistance to the Benzimidazole (BZ) subclass of anthelmintic (Kwa *et al*., 1993). Investigations detailed elsewhere in this thesis show that the *H. contortus* genome contains four distinct β-tubulin genes (Chapter 5). This compares with six genes within the genome sequence of the closely related free-living genetic model organism *C. elegans* (Blaxter, 1998). Relationships between the *Hc-isotype-1*/*Hc-isotype*-2/*Ce-ben-1*, *Hc-isotype-3*/*Ce-mec-7* and *Hc-isotype-4*/*Ce-tbb-4* genes have been established based on phylogenetic, syntentic and genetic distance scores (Chapter 5). Only the *isotype-1* and *-2* β-tubulin genes of *H. contortus* and the *ben-1* gene of *C. elegans* have been implicated in giving rise to BZ resistance (Driscoll *et al*., 1989; Kwa *et al*., 1993; Kwa *et al*., 1995; Beech *et al*., 1994). It is the current hypothesis that neither the *isotype-3* nor *-4* genes of *H. contortus*, or any of the other *C. elegans* β-tubulin genes, have a role in the BZ resistance phenotype (Chapter 5).

The relative ease with which *C. elegans* can be transformed with heterologous DNA has highlighted the relevance of this organism as a research tool to study the function, expression and regulation of parasitic nematode genes (Gilleard *et al*., 2005). There have been more reports of *H. contortus* gene sequences being investigated in this manner than there have been for any other parasitic nematode species (Gilleard, 2004). However, although there have been many successful examples, there remains much debate over how representative a system *C. elegans*

is to study gene sequences from an organism with which it last shared a common ancestor an estimated 450 Mya.

To date, the expression patterns for a total of 12 parasitic nematode genes have been published using *C. elegans* as a heterologous expression vector (Chapter 1; Table 1.2). However, although an expression pattern has been described for these investigations, how this pattern relates to the true gene expression in the native host is an important consideration, and a question that should be investigated as thoroughly as possible. For some genes, confirmation of site of expression in the parasite may be obtained from tissue-specific transcriptome data. For example, for *H. contortus* and *A. suum* analysis and comparison of intestine-enriched genes has recently been carried out (Yin *et al.*, 2008). However, as interest lies not only in where the gene is expressed but also in where the encoded protein functions, it is also informative to examine the localisation of the gene product through immunolocalisation studies.

6.2 Aims of this chapter

The aim of the work outlined in this chapter was to determine the expression patterns of the different *H. contortus* B-tubulin genes and compare them with those of their closest *C. elegans* homologues based on the sequence comparisons presented in Chapter 5. The purpose of this was to investigate the extent of conservation of the gene families between the two species and to provide information on the likely functional relationship of family members. I have taken a dual approach to investigate these expression patterns:

1. I have assayed the tissue specificity of the gene promoters by making reporter gene fusions and examining expression patterns in transgenic *C. elegans*

2. I have generated antibodies specific for the *C. elegans* MEC-7 and BEN-1 along with the *H. contortus* ISO-1, -2 and -3 polypeptides and used them to examine their endogenous expression pattern

This dual approach not only allowed me to undertake a comparative analysis of the expression patterns in the two species but also allows us to test the fidelity of the reporter gene system. I have undertaken these studies on the *C. elegans* genes *ben-*

1 and *mec-7* and their respective closest homologues in *H. contortus*; *isotype-1*, *-2* and *-3*. I was unsuccessful in identifying 5' regulatory or specific cDNA sequence for the *H. contortus* β-tubulin *isotype-4* gene, and so could not include this gene in my analysis. Finally, I have carried out initial analysis of the relative expression levels of the β-tubulin family members in *H. contortus* and *C. elegans*.

6.3 Results

6.3.1 Reporter gene studies

6.3.1.1 *Caenorhabditis elegans mec-7* reporter gene expression pattern

Hamelin *et al*. demonstrated that a translational *Ce-mec-7-lacZ* fusion gene injected into *C. elegans* N2 wild-type animals resulted in *lacZ* expression primarily in the touch neurons (Hamelin *et al*., 1992). This fusion construct consisted of 850 bp 5' *mec-7* flanking region, 792 bp encoding the amino-terminal 206 aa of the Ce-MEC-7 protein and the lacZ coding region and 1.3 kb of the *unc-54* gene 3' UTR. In order to determine if the 5' regulatory sequences of this gene alone could faithfully reproduce the expression pattern, I constructed a transcriptional promoter-reporter PCR product consisting of 850 bp *mec-7* 5' flanking region fused to the GFP coding region of Fire vector pPD95.67. This construct was injected into the gonad of live *C. elegans* N2 worms. Three transgenic lines were obtained (Table 6.1). All lines were examined, and although the expression pattern was variable between worms of the same line, the same expression pattern was seen across all three transgenic lines. This variability in expression pattern of worms within transgenic lines is a known feature of expression from extra-chromosomal arrays in *C. elegans*, presumably due at least in part to mitotic segregation in somatic lineages (Mello *et al*., 1992).

The expression of the GFP reporter gene was seen primarily in the six touch receptor cells and associated neurons (Figure 6.1). Faint staining was also associated with what is likely to correspond to the FLPL/R and ALNL/R neurons (data not shown). The precision of identifying these neurons was aided by the Wormatlas web page (http://www.wormatlas.org/cellid/cellID.htm) and comparisons with the pattern published by Hamelin *et al*. (Hamelin *et al*., 1992). The only difference between my obtained pattern and that of Hamelin *et al*. was that the PVD neuron, reported as staining in <1 % of the worms with the previously reported *mec-7-lacZ* fusion gene, was not seen in my promoter-reporter fusion transgenic lines. It could be that the amino 206 aa of the *mec-7* coding region and 792 bp of downstream region contains the regulatory elements required to direct this aspect of the expression pattern. Another possible explanation is the lower sensitivity of GFP as a reporter gene, compared to Lac –Z, fails to detect a lower level of expression in the PVD neuron.

Overall, the expression pattern that I obtained was highly similar to that of Hamelin *et al*. (Hamelin *et al*., 1992). This result demonstrates that the use of a simple transcriptional fusion faithfully reproduces the majority of the expression pattern of the *Ce-mec-7* gene. Hence, I decided to take the same approach to investigate the *H. contortus isotype-3* gene, the putative orthologue of *Ce-mec-7*.

6.3.1.2 *Caenorhabditis elegans ben-1* reporter gene expression pattern

Specific mutations in the *C. elegans ben-1* gene have been shown to confer significant resistance to BZ drugs (Driscoll *et al*., 1989). This is the only locus from this organism that has been implicated in giving rise to this resistance phenotype. Mutations in β-tubulin genes from many other nematode species have also been correlated with BZ resistance (Kwa *et al*., 1993; Prichard, 2001; Wolstenholme *et al*., 2004). The direct mechanism of how mutations in β-tubulin genes are able to confer this phenotype remains unknown. However, BZ drugs are known to lethally disrupt the intestine of the nematode worm *H. contortus* (Jasmer *et al*., 1999). I wished to determine the expression pattern of *ben-1* to provide further information as to the mode of action of BZ drugs and to be able to undertake comparative studies with the *H. contortus isotype-1* and *isotype-2* β-tubulin genes. The sequence analysis in Chapter 5 has suggested that these two genes have a paralogous relationship with *ben-1* and so expression pattern data may help determine if one or both of these genes are the functional homologue of this *C. elegans* gene.

A promoter-reporter transcriptional fusion construct consisting of 1.971 kb 5' flanking region of the *C. elegans ben-1* gene fused with the GFP coding region of Fire vector pPD95.67 was injected into the syncytical arm of live *C. elegans* wild-type N2 worms. Three transgenic lines were obtained (Table 6.1). Similar to my *mec-7* fusion construct, variation was seen between worms within the same transgenic line, but the same pattern of expression was seen across all three lines. The majority of expression was seen in the posterior gut of the worms (Figure 6.2, Panel A). GFP was also localised in head neurons of >80 % of worms, which I have putatively identified as the BDU neurons based on comparisons with Wormatlas (http://www.wormatlas.org/cellid/cellID.htm) and various publications of GFP expression that localised to this tissue (http://gfpworm.org/index; Figure 6.2, Panel B). GFP expression was first seen in the gut and neurons of 3-fold embryos (Figure 6.2, Panel C). However, due to the compressed nature of the worms, the

identification of the neurons fluorescing at this stage was markedly more difficult that during larval development or adult stages. The same expression pattern is seen throughout the development of the worm therefore it is my hypothesis that it is the BDU neurons that fluoresce during the embryogenesis stages of *C. elegans* development.

6.3.1.3 Identification, cloning and sequencing of three allele lengths of 5' regulatory region of the *Haemonchus contortus isotype-3* locus

The combination of primers designed to amplify 2.056 kb 5' of the *H. contortus isotype-3* locus repeatedly amplified three distinct products of 2.056, 1.488 and 1.071 kb when genomic DNA extracted from a starting population of >10,000 Mhco3 (ISE) L3 worms were used as template (Figure 6.3). As this locus resides on the *H. contortus* X-chromosome, male worms only carry a single copy (Redman *et al.*, 2008). Consequently, a number of PCR amplifications were performed on single male worm lysates. Only a single allele size was amplified from each single male worm confirming the allelic nature of the different sized PCR products (Figure 6.3). Single worms with products of each allele length were selected for amplification. The products of each length were cloned into vector pCR-4 (Invitrogen) and sequenced in triplicate (MWG) for three individual plasmid clones and a consensus sequence was deduced. Each of the three consensus sequences were aligned against the *H. contortus* 408, 911 bp stretch of X-chromosomal sequence and BLAST searched against one another (Figure 6.4). This analysis confirmed that the 1.082, 1.488 and 2.056 kb sequences are allelic variants of one another, which can be explained by insertion, deletion and re-arrangement events.

6.3.1.4 Three distinct 5' regulatory sequence alleles flanking the *Haemonchus contortus* β-tubulin *isotype-3* gene all direct GFP expression primarily to the touch neurons of *Caenorhabditis elegans*

The polypeptide encoded by the *H. contortus* β-tubulin *isotype-3* gene clusters tightly with the MEC-7 polypeptide of five species from the *Caenorhabditis* genus and that of the necromenic nematode species *P. pacificus* (Figure 5.6). In addition, I have been able to show a conservation of synteny between the *Hc-isotype-3* and *Ce-mec-7* loci (Figure 5.2; Redman *et al.*, 2008). This detailed analysis has allowed us to conclude that these genes are orthologous. Three promoter alleles, of 1.082, 1.488

and 2.056 kb in length, of the 5' flanking region of the *Hc-isotype-3* locus were each amplified, cloned and fused to the GFP coding region of Fire vector pPD95.67 (Figures 6.3 and 6.4). Each construct was injected into the gonad of *C. elegans* in order to determine the expression of the reporter gene under the control of each of these regulatory regions. As I was investigating regulatory sequence from *H. contortus*, in this case *C. elegans* was being utilised as a heterologous expression system. At least three transgenic lines were obtained for each of the 1.082, 1.488 and 2.056 kb promoter sequence lengths. Again the expression pattern varied between worms of the same transgenic line but was identical across all transgenic lines of the same promoter allele (Table 6.1). In addition to this, identical expression of the GFP reporter gene was seen whether under the control of the 1.082, 1.488 or 2.056 kb promoter sequence of the *Hc-isotype-3* locus. The expression pattern for the 1.082 kb promoter reporter construct is described in detail below.

GFP expression was seen in all six touch neurons at a high frequency (Figure 6.5). Tail PLML/R neurons were seen at the highest frequency (Panel C) followed by the ALML/R, PVM and AVM neurons (Panels B and C). Each tissue was seen fluorescing in at least 10 % of worms of each transgenic line for each promoter allele length. The frequency of tissues seen fluorescing also reflects the strength of the GFP signal. Expression was also seen in neurons putatively described as the FLPL/R and ALNL/R at a much lower frequency and intensity. This is the same GFP expression pattern as observed with 850 bp of the *C. elegans mec-7* promoter sequence. Identical to the *Ce-mec-7*::GFP construct, no GFP expression was seen in the PVD neuron under the control of any of the three promoter sequences of the *Hc-isotype-3* locus. The expression of the GFP reporter gene under the control of the *H. contortus isotype-3* regulatory sequences was much fainter when compared to the *Ce-mec-7* promoter sequence. This could reflect true expression signals of the respective genes, or more likely, is a caveat of using a heterologous host to investigate *H. contortus isotype-3* regulatory sequences. The highly similar nature of the expression patterns directed by their respective promoters provides further support for the orthologous relationship of the *Hc-isotype-3* and *Ce-mec-7* genes established in the previous chapter (Chapter 5).

6.3.1.5 *Haemonchus contortus isotype-1* reporter gene expression pattern

The *H. contortus* β-tubulin *isotype-1* gene shares, along with *isotype-2*, a paralogous relationship with *C. elegans ben-1*. Through the use of *C. elegans* as a heterologous system, specific alleles of the *H. contortus* β-tubulin *isotype-1* gene have been shown rescue the BZ resistant phenotype of *ben-1* mutants (Driscoll *et al.*, 1989). Hence, I wanted to determine if the *Hc-isotype-1* shared the same reporter gene expression pattern with that of *Ce-ben-1*. A transcriptional promoter-reporter construct consisting of 1.971 kb *Hc-isotype-1* 5' flanking sequence fused to the GFP coding region of Fire vector pPD95.67 was injected into the gonad of *C. elegans*. Four transgenic lines were generated which contained this promoter-reporter expression construct (Table 6.1). As before, variation in expression pattern was seen between worms of the same transgenic line, but the pattern was the same across all four lines. The 1.971 kb of 5' flanking sequence directed GFP expression primarily to the posterior gut of the *C. elegans* worms (Figure 6.6). In addition, staining was seen in what are putatively described as the BDU neurons in the head of the worm. Expression was first detected at the three-fold stage of embryogenesis and remained constant throughout the development of the worm. This expression pattern is the same as that associated with the *C. elegans ben-1* promoter-reporter fusion construct, shown in Figure 6.2.

The *Hc-isotype-1* promoter construct also directed expression of GFP in head neuron structures additional to those fluorescing under the control of the *Ce-ben-1* gene promoter (Figure 6.6, Panel C). From comparisons with Wormatlas and published reports of expression in the same tissues I postulate that this expression corresponds to the amphid neurons (http://www.wormatlas.org/cellid/cellID.htm; http://gfpworm.org/index). Expression in these structures was seen in ~10 % of worms, which was far lower than the frequency of the posterior gut or BDU neuron fluorescence.

6.3.1.6 *Haemonchus contortus isotype-2* reporter gene expression pattern

In conjunction with specific polymorphisms in the *Hc-isotype-1* gene, deletions of the *H. contortus* β-tubulin *isotype-2* locus have been correlated with a heightened resistance to BZ drugs, although no functional studies with the *isotype-2* gene have yet been performed (Kwa *et al.*, 1993; Beech *et al.*, 1994). The *Hc-isotype-2* locus

shares, along with the *Hc-isotype-1* gene, a paralogous relationship with the *Ce-ben-1* locus (Chapter 5). A transcriptional promoter-reporter gene fusion was constructed for this gene by fusing 802 bp 5' of the *Hc-isotype-2* initiating codon with the GFP coding region of Fire vector pPD95.67, and the construct was injected into the gonad of *C. elegans*. Two transgenic lines were obtained (Table 6.1). Very faint GFP expression in both lines was localised to the posterior gut of ~30 % of the worms from each line (Figure 6.7). Fluorescence was first detected at the L1 stage of *C. elegans* development, and maintained throughout the rest of the lifecycle of the worm. No expression in any other cell type was observed.

It is uncertain if this fluorescence represents true GFP expression. The signal is detected at the early larval stages of *C. elegans* development and is therefore unlikely to be a result of gut granule auto-fluorescence. However, most regulatory sequences reside within the first 1-2 kb of 5' flanking region for the vast majority of nematode genes (Gilleard, 2004). The promoter length used in this example is only 802 bp and therefore much of the regulatory sequence may be missing. Furthermore, faint posterior gut expression is a noted artefact associated with Fire expression vectors. This is thought to be as a result of the inclusion of the *unc-54* 3'-UTR in most Fire kit vectors, including pPD95.67 (http://www.wormbook.org/chapters/www_reportergenefusions/reportergenefusions.html).

6.4 Indirect Fluorescence Antibody (IFA) protein localisation patterns for members of the *Caenorhabditis elegans* and *Haemonchus contortus* β-tubulin families

6.4.1 Caenorhabditis elegans MEC-7 protein localisation

β-tubulin polypeptide sequences for all species, for which β-tubulin sequence data exists, show a very high level of identity and similarity. The main differences between the different family members are seen at the last few amino acids in the carboxy terminus (Njue and Prichard, 2003). Antiserum generated against the last ten carboxy-terminal amino acids of the Ce-MEC-7 polypeptide has previously been

reported to localise to the six touch cell receptors and associated axons of *C. elegans* N2 worms (Savage *et al.*, 1994). In addition, lower levels of expression were seen in the FLP, PVD and BDU neurons. An aliquot of the same antiserum was a kind gift from the Chalfie laboratory (Columbia University). In using this antiserum at the same concentrations as in the published reportings I obtained exactly the same antigen localisation pattern against both freeze-cracked and Ruvkun fixed N2 specimens, (Table 6.2; Figure 6.8). Identification of the fluorescing tissues was aided by direct comparisons with the published findings and the *C. elegans* Wormatlas website (http://www.wormatlas.org/cellid/cellID.htm). Detection of the signal was first possible at the three-fold stage of embryogenesis and remained constant throughout the lifecycle of the worm (Figure 6.8). A high level of background fluorescence was seen in all IFAT analyses when using *C. elegans* specimens prepared by the Ruvkun fixation protocol (Table 6.2). As a result, slides prepared via the freeze-crack protocol were preferentially used to establish the MEC-7 protein localisation pattern. Due to the consistent pattern between my localisation pattern and those previously published for the MEC-7 protein, my technique and protocols used to establish this pattern were verified.

6.4.2 Caenorhabditis elegans BEN-1 protein localisation

IgG antibody was immunoaffinity purified from antiserum directed against a thirteen-amino acid region from the carboxy-terminus of the *C. elegans* BEN-1 protein sequence (Chapter 2). This is the only hyper-variable region of the *C. elegans* β-tubulin protein family. As with the MEC-7 antiserum, all IFAT analysis using *C. elegans* Ruvkun fixed specimens resulted in a very high level of background fluorescence, therefore freeze-cracked samples were used to determine the localisation pattern of this antibody. The IgG antibody strongly stained the nucleus of the vast majority, possibly all, cells at all stages of *C. elegans* N2 freeze-cracked specimens (data not shown). A *ben-1*$^{-}$ deletion strain of *C. elegans*, tm234, was obtained from the Tokyo Women's Medical University as a gift of Dr Shohei Mitani. Pre-absorbing the purified BEN-1 IgG antibody against acetone powdered samples of this *C. elegans* strain prior to staining N2 worms markedly reduced this nuclear staining, and instead resulted in predominant antibody localisation to the neuronal circuitry of the worm (Figure 6.9). The dorsal/ventral cord was seen most prominently in these worms, as this cord consists of many axonal processes. At

higher magnification many more axonal structures could be seen fluorescing (Figure 6.9, Panels B and C). From the complexity of the neuronal staining in all structures of the worms, and in comparisons with the *C. elegans* Wormatlas (http://www.wormatlas.org/cellid/cellID.htm), it is postulated that this antibody binds to a large majority of, if not the entire, *C. elegans* N2 neuronal circuitry.

In using the same pre-absorbed purified antibody to stain freeze-cracked *C. elegans* tm234 specimens as negative controls, only very faint general nuclear staining was noted, with none of the neuronal staining described above seen (Figure 6.10). From these results I conclude that this antibody specifically binds to the vast majority of the neuronal circuitry of *C. elegans* N2 freeze-cracked worm preparations.

6.4.3 Haemonchus contortus Isotype-3 protein localisation

IgG antibody immunoaffinity purified from antiserum directed against the carboxy-terminal 11 amino acids of the *H. contortus* β-tubulin Isotype-3 protein sequence primarily localised to the six touch cell receptors (ALML/R, PLML/R, AVM and PVM) of exsheathed L3 MHco3 (ISE) worms (Figure 6.11). Faint expression was also seen in tissues postulated to correspond to the FLP, PVD and BDU neurons. The same localisation pattern was seen in worms prepared by either the freeze-crack or Ruvkun fixation method (Table 6.3). Fluorescing tissues were identified based on *C. elegans* morphology as the complete neuronal circuitry of *H. contortus* has not been described (http://www.wormatlas.org/cellid/cellID.htm). The positions of all six touch receptors in the *H. contortus* worm were very similar to those of *C. elegans*. In addition, the direction and length of associated axonal processes was also very similar.

This antibody localisation pattern directly mirrors the expression pattern associated with the *isotype-3*::GFP promoter-reporter fusion construct injected into *C. elegans* using this heterologous system to determine the putative expression pattern of this gene (Figure 6.5). This similarity demonstrates that the promoter regions used in the transgenic *C. elegans* worms faithfully reproduces the true endogenous expression pattern of the *Hc-isotype-3* gene. This validates the use of *C. elegans* as heterologous expression system to investigate the *H. contortus isotype-3* gene.

This antibody localisation pattern also directly reflects that associated with the *C. elegans* MEC-7 antibody (Figure 6.8). An orthologous relationship between the *H. contortus* β-tubulin *isotype-3* and *C. elegans mec-7* genes has been proposed by phylogenetic, syntetic, expression pattern and now antibody localisation pattern investigation.

6.4.4 Haemonchus contortus Isotype-1 protein localisation

H. contortus exsheathed L3 MHco3 (ISE) worms stained with purified IgG antibody specific for the Isotype-1 antigen putatively stained the majority of the neuronal circuitry of the worms (Chapter 2; Table 6.3; Figure 6.12). Complex staining was seen in the head and tails of the worms. Fluorescing structures were seen in all tissues where neuronal axons would be, using the *C. elegans* neuronal circuitry as template. I putatively postulate this result as only worms prepared by the freeze-crack method showed internal staining. The fragmented nature of worms prepared by freeze-cracking made specific neuron identification a difficult task. In addition, the complete neuronal circuitry of *H. contortus* is yet to be fully detailed, further hampering neuron identification. However, analysis of the complexity of the staining and detailed comparisons with the neuronal circuitry of *C. elegans* (http://www.wormatlas.org/cellid/cellID.htm), has led to this conclusion.

Specific staining of the majority of the neuronal circuitry of the worm is the same as the localisation pattern associated with the *C. elegans* BEN-1 antibody (Figure 6.9). Sequence analysis detailed in the previous chapter and similarities in reporter-promoter expression patterns have allowed the conclusion that these genes share a paralogous relationship. This result adds another dimension to this analysis suggesting that the protein expression patterns are also similar for Ce-BEN-1 and Hc-ISO-1.

6.4.5 Haemonchus contortus Isotype-2 protein localisation

When *H. contortus* MHco3 (ISE) worms were incubated with immunoaffinity purified IgG antibody directed against the 14 carboxy-terminal amino acids of the Isotype-2 protein, the localisation pattern was highly similar to that associated with the Isotype-1 antibody (Table 6.3; Figure 6.12; Figure 6.13). Complex neuronal axonal staining was seen in the head and tail structures of the worm. The dorsal/ventral

nerve cords were seen most clearly, as these cords have many axonal processes associated. Again, using the *C. elegans* neuronal circuitry as template, axonal processes were seen in all expected tissues (http://www.wormatlas.org/cellid/cellID.htm). In addition, specific staining was again seen only with *H. contortus* exsheathed L3 specimens prepared by the freeze-crack method, not by the Ruvkun method (Table 6.3; Figure 6.13, Panel C). My conclusion is that this antibody is also specific to the majority of the neuronal circuitry of these worms. The *H. contortus isotype-2* gene shares a paralogous relationship with those of *Hc-isotype-1* and Ce-*ben-1* (Chapter 5). This result suggests that the localisation pattern of the protein product of each of the three genes is also very similar. No obvious differences between the three staining patterns were seen. Instead, many instances of fluorescing tissues common to all three antibody localisation patterns were noted. These include the dorsal/ventral nerve cords, the touch neuronal circuitry, and the phasmid neurons.

6.5 Analysis of gene expression levels for the *Caenorhabditis elegans* and *Haemonchus contortus* β-tubulin gene families

6.5.1 Utilisation of EST and SAGE data to evaluate expression levels of the six Caenorhabditis elegans β-tubulin genes

The expression levels of many genes have been established by the number of gene specific tags present within the *C. elegans* SAGE datasets (Patino *et al., 2002*; http://elegans.bcgsc.ca/home/sage.html). In searching all available datasets for tags specific to transcripts from the six *C. elegans* β-tubulin genes, a total of 451 SAGE tags were identified across all *C. elegans* life-cycle stages (Table 6.4). *tbb-2* was the highest represented, with almost half the entire β-tubulin tags accounting for this gene. The complete order of genes starting with the most represented was *tbb-2* (205 tags), *tbb-1* (117 tags), *mec-7* (103 tags), *tbb-4* (14 tags), *ben-1* (9 tags) and finally *tbb-6* (3 tags).

A total of 367, 044 *C. elegans* EST sequences are available for homology based BLAST searching (http://www.wormbase.org/db/searches/blast_blat). In searching this

sequence database for *C. elegans* β-tubulin specific sequences, a total of 481 ESTs were identified (Table 6.5). The *tbb-2* transcript accounts for almost 70% of these EST sequences. When the six *C. elegans* β-tubulin genes are arranged in order of EST data, starting with the largest dataset, the same list is found as for the SAGE analysis. Both these methods of gene expression level investigation suggest that the order of expression, starting with the highest, is *tbb-2*, *tbb-1*, *mec-7*, *tbb-4*, *ben-1* and finally *tbb-6*.

6.5.2 Utilisation of EST data and RT-PCR analysis to evaluate expression levels of the four Haemonchus contortus β-tubulin genes

There are two searchable *H. contortus* EST sequence datasets publicly available which between them contain in excess of 30,000 sequences, consisting of more than 14 Mb of sequence data (http://xyala.cap.ed.ac.uk/services/blastserver/; http://130.209.234.35/blast/blast.html). Investigation detailed in the previous chapter of this thesis shows that there are 27 β-tubulin sequences within these datasets (Chapter 5). Aligning these sequences to the available cDNA sequences transcribed from the four *H. contortus* β-tubulin genes shows that 20 of these sequences represent the *isotype-1* and 7 the *isotype-2* locus from the *H. contortus* genome. Work detailed in this thesis reports that I have been able to amplify, clone and sequence transcripts from both the *isotype-3* and *-4* β-tubulin loci of the *H. contortus* genome (Chapter 5). It therefore seems that the absence of these transcripts from current *H. contortus* EST databases is due to the size of these current datasets, which are not a true indication of gene expression.

To comparatively evaluate the expression levels of the four *H. contortus* β-tubulin genes, primers were designed for RT-PCR analysis (Primer sequences in Appendix 2). An initial screen of cDNA reverse transcribed from male and female specific total RNA suggests that the *isotype-2* locus is the most highly expressed, followed by *isotype-1*, both of which were expressed at a much greater level than the *isotype-4* gene. My primer set used to investigate *isotype-3* expression failed on both the male and female samples. Amplifications were controlled by the inclusion of the constitutively expressed *H. contortus* gene *ama-1*. In addition, linear regression

analysis shows that all amplifications were in the exponential stages and that the primer set efficiencies were very similar.

From this combined EST and RT-PCR data, my early hypothesis is that the *Hc-isotype-1* and *-2* genes are the most highly expressed of all four *H. contortus* β-tubulin genes. I would further postulate that the expression level of these loci is far greater than that of the *H. contortus* β-tubulin *isotype-3* and *-4* genes. However, the early nature of this analysis must be stressed. Bio-replicates are required to fully conclude the expression levels of the four genes.

6.6 Discussion

β-tubulin genes are one of the few validated nematode drug targets. Specific mutations in the *isotype-1* and *-2* genes of *H. contortus* have been correlated with resistance to the BZ class of anthelmintic drug (Kwa *et al.*, 1993; Kwa *et al.*, 1995; Beech *et al.*, 1994). In addition, the *H. contortus isotype-1* locus has also been implicated as involved in the macro-cyclic lactone resistance mechanism (Eng *et al.*, 2006). Although the involvement of these genes has been associated with such phenotypes for in excess of ten years (in the case of BZ), little is known of the direct mechanism by which these mutations are able to confer such resistances. Work presented in this chapter investigated the expression and localisation patterns for the β-tubulin gene families and proteins of *C. elegans* and *H. contortus*. These results have allowed us to test the relationships established by sequence analysis presented in the previous chapter, and help illuminate the possible functions of the different polypeptides.

Sequence analysis in the previous chapter led to the conclusion that the *H. contortus* β-tubulin *isotype-3* and *C. elegans mec-7* genes share an orthologous relationship. Previously published work has shown that the *C. elegans mec-7* gene is primarily expressed in the six touch cells and associated neurons (Savage *et al.*, 1994). I was able to replicate this pattern of GFP expression, using a *mec-7*::GFP transcriptional promoter-reporter fusion construct (Figure 6.1). Furthermore, this GFP expression pattern was also replicated when driven from three distinct alleles of the 5' flanking region of the *Hc-isotype-3* gene (Figure 6.5).

In addition to the GFP expression pattern, a Ce-MEC-7 antibody has previously been reported as localising primarily to the touch sensory network of *C. elegans* worms (Savage *et al.*, 1994). Again, I was able to fully replicate this localisation pattern using an aliquot of the same antibody and my specimen preparation methods (Table 6.1, Figure 6.8). Using an antibody specifically generated against the C-terminus of the Hc-ISO-3 polypeptide, resulted in the localisation to the touch sensory neuronal network of *H. contortus* MHco3 (ISE) worms (Table 6.1, Figure 6.1).

Therefore, the antibodies targeted to the *C. elegans* MEC-7 and *H. contortus* Isotype-3 proteins generated staining patterns that correlated precisely with those

of the promoter-reporter gene fusions (Figures 6.9 and 6.12). One of the main aims of this chapter was to investigate the use of *C. elegans* as a tool to examine the *H. contortus* β-tubulin family. These results fully verify the use of the free-living nematode as a transgenic tool to investigate the *Hc-isotype-3* gene.

In addition, the results presented in this chapter show that neither the *Ce-mec-7* of *Hc-isotype-3* genes are expressed at high levels in the native organism. In the case of *mec-7*, EST and SAGE tag analysis shows that this gene is expressed at much lower levels than that of the *tbb-2* and *-1* genes (Tables 6.4 and 6.5). There are no EST sequences that are specific to the *Hc-isotype-3* gene in any of the publicly available datasets, whereas sequence specific to both the *isotype-1* and *-2* loci can be found. The *Ce-mec-7* and *Hc-isotype-3* genes therefore represent an example of exquisite conservation between the *C. elegans* and *H. contortus* genomes. Both are expressed in a highly specified set of neurons, and appear to be expressed at relatively similar levels. These results therefore imply that the function of these proteins in the native system are likely to be very similar.

This experimental analysis overlaps with my sequence investigations in the previous chapter. From the work presented in this thesis, characterisation of the *H. contortus isotype-3* gene provides a perfect example of exploiting the wealth of data known for the genetic model organism *C. elegans* in investigating genes from the *H. contortus* genome. Finally, there has been no evidence reported linking mutations within either the *Ce-mec-7* or the *Hc-isotype-3* genes to anthelmintic resistance phenotypes. This analysis strongly suggests that neither of these loci have any involvement in such phenotypes and, in addition, further strengthens their orthologous relationship.

Despite *ben-1* being the only locus from the *C. elegans* genome within which mutations are able to confer significant resistance to BZ drugs, there are no published investigations of the expression pattern of this gene. A transcriptional promoter-reporter gene fusion injected into *C. elegans* worms resulted in GFP expression primarily in the posterior gut of the worm, with BDU neurons in the head seen fluorescing at a lesser intensity and at a lower frequency (Figure 6.2). This is very similar to the pattern of GFP expression seen under the control of the 5' regulatory regions of the *H. contortus isotype-1* and *-2* β-tubulin loci (Figures 6.6 and 6.7). Each of the three sequences directed GFP expression to the posterior gut

of the *C. elegans* worms. Interestingly, this may give an indication of how mutations within these genes are able to confer significant resistance to BZ drugs. BZ anthelmintics are known to lethally disrupt the intestine of *H. contortus* worms (Jasmer *et al.*, 1999). In addition, BZ drugs are known to bind with high-affinity to nematode microtubules (Oxberry *et al.*, 2000). With the expression of the reporter gene in the posterior of the gut, it could be hypothesised that specific mutations within these genes have a direct effect in preventing the drug from binding to its target. However, it must be stressed that this hypothesis is purely speculative. Although this work may give a possible indication of the way mutations in the β-tubulin genes can give rise to resistance, the data here does not allow this conclusion to be drawn.

One of the main differences in these GFP patterns was that only *Hc-isotype-1* directed reporter gene expression to the amphid neurons of the *C. elegans* worms. Interestingly, of all *C. elegans* and *H. contortus* β-tubulin gene family members, only *Hc-isotype-1* has been implicated in giving rise to macro-cyclic lactone resistance (Eng *et al.*, 2006). My transcriptional promoter-reporter gene expression pattern generated for this gene gives a possible hypothesis of why this may be the case. Physical distortions of amphid neurons, particularly shortening of the axonal processes, in *H. contortus* worms resistant to this class of anthelmintic drug has been reported (Guerrero and Freeman, 2004). It could be that mutations within the *Hc-isotype-1* gene are responsible for this axonal shortening. This result potentially links mutations in this gene with amphid neuron defects and how this may give rise to macro-cyclic lactone resistance. Again, however, the work presented here is not sufficient to conclude this, but may provide the focus of future investigations.

The *Hc-isotype-2* transcriptional GFP reporter gene expression pattern was limited only to the posterior gut of the *C. elegans* worms. Furthermore, the expression levels of GFP seen in this tissue were far less intense than associated with the patterns driven by either the *Ce-ben-1* or *Hc-isotype-1* regulatory sequences. It is an important consideration that only 802 bp of *Hc-isotype-2* 5' sequence was used to generate this expression pattern. This compares with ~2 kb used to generate both the *Ce-ben-1* and *Hc-isotype-1* patterns. It is therefore a reasonable hypothesis that the regulatory region that directs BDU neuronal staining is contained within the *Hc-isotype-2* 5' flanking region, but was not included in my reporter construct. A second

reasonable hypothesis for this result is that the GFP signal was not strong enough to detect from BDU neurons in strain GIS18. It is a known caveat that GFP is not the most sensitive of reporter genes available. As a result of these caveats, further investigation into the *Hc-isotype-2* expression pattern in transgenic *C. elegans* worms could be suggested, using up to 2 kb 5' flanking sequence to encompass more of the regulatory region, and using a more sensitive reporter gene (e.g. l*ac-Z*) as a method to detect low levels of expression.

Unlike the cases for Ce-MEC-7 and Hc-ISO-3, the antibody localisation patterns generated for the Ce-BEN-1, Hc-ISO-1 and Hc-ISO-2 proteins do not replicate those generated by the reporter-promoter gene fusions (Figures 6.9, 6.12 and 6.13). Each of the antibodies is putatively described as binding to the vast majority of, if not the entire, neuronal circuitry of the respective worm, and no gut staining was seen. In the case of Ce-BEN-1, axonal processes were seen in all tissues that axons are known to exist (http://www.wormatlas.org/cellid/cellID.htm). However, the complexity of the staining made the identification of individual neurons extremely difficult. A similar scenario was found with the staining patterns of both the Hc-ISO-1 and Hc-ISO-2 antibodies against MHco3 (ISE) worms. However, the identification of specific neurons was further hampered in these cases as the complete neuronal circuitry of *H. contortus* remains unknown. In addition, as the antibodies could only be used on freeze-cracked worms, it was difficult to follow neuronal axons due to the fragmented nature of the worms. As a result, subtle differences may exist between the localisation patterns that were generated by these three antibodies that were undetectable via the methods used in this analysis.

The failure of any of these antibodies to bind to microtubules outwith the neuronal processes is somewhat surprising. Microtubules are ubiquitous and essential for the viability of all eukaryotic cells. A potential hypothesis to explain this result is that the techniques used to fix the nematode specimens for IFAT analysis were not harsh enough to maintain the microtubule structures. Microtubule concentrations are known to be very high within the axonal processes of neurons; therefore, it may not be too surprising that these are the only microtubule structures that I was able to maintain using my fixing techniques (Siddiqui *et al.*, 1989).

A major source of concern in analysing the results within this chapter is that specificity of the antibodies used in this analysis remains untested. Western blot

analysis, ensuring that each antibody recognises only the one targeted antigen should be performed in evaluating these IFAT staining results. Homology based BLAST database searching was employed for each of the peptides that were selected to generate my antibodies. However, given that I was working with an incomplete genomic sequence database, in the case of *H. contortus*, it could be the case that genes generating proteins with similar peptide sequences may have been missed. Such anomalies would be resolvable via crude extract Western blot analysis. An added complication is that the β-tubulin protein family is very similar both within and between species. As the proteins targeted with antibodies in this work are all at least 90% identical at the amino acid level, it would be very difficult to separate them on a Western blot. Instead, testing the specificity of the antibodies used to ensure that each recognises only one β-tubulin protein would require the generation of recombinant proteins for each, purification of these proteins and specific antibody recognition tag Western blot analysis. Although complex and intensive, such investigation should be considered when discussing the IFAT staining patterns observed in this chapter.

Sequence analysis in the previous chapter dictates that the *Ce-ben-1*, *Hc-isotype-1* and *Hc-isotype-2* are paralogous in that they are derived from a common ancestor. However, this does not imply that the proteins encoded by these genes are functional homologues. Furthermore, investigation into the expression levels of these three genes suggests that they are not. EST representation and SAGE tag analysis depicts that the *C. elegans ben-1* gene is expressed at a far lower level than the *tbb-1* and *-2* genes (Tables 6.4 and 6.5). Extensive investigation into the *Ce-tbb-1* and *-2* genes has shown that they are functionally redundant and that, importantly, an embryonic lethal phenotype is observed when both genes are knocked-out (Lu *et al.*, 2004). This compares with only the phenotype of BZ resistance associated with *ben-1*$^{-}$ *C. elegans* worms (Driscoll *et al.*, 1989). In comparison, the expression levels of the four *H. contortus* β-tubulin genes investigated both by EST representation and RT-PCR suggests that both the *isotype-1* and *-2* are the highest expressed of this gene family. As microtubules are ubiquitous and essential for the viability of all eukaryotic cells, it is probably the case that one, if not both, the *H. contortus isotype-1* and/or *-2* genes are functional homologues of the *Ce-tbb-1* and/or *-2* genes. Indirect evidence further supports this hypothesis. It has been shown possible to reduce the expression of the *Hc-isotype-1* gene by RNAi

with no obvious detrimental phenotype (Gelhof *et al*., 2006). Furthermore, the deletion of the *Hc-isotype-2* locus has been reported as giving only the phenotype of heightened resistance to BZ drugs, when in combination with known *Hc-isotype-1* mutations which also give BZ resistance (Kwa *et al*., 1993; Beech *et al*., 1994). No examples have been reported where both genes have been able to knocked-down or deleted in *H. contortus* worms. Although this indirect evidence does not allow the conclusion that either the *Hc-isotype-1* or *-2* genes are essential for the viability of *H. contortus* worms, it does allow this hypothesis to be strongly suggested.

Overall, the results presented in this chapter allow us to hypothesis the functions of the *H. contortus* β-tubulin *isotype-1 -2* and *-3* genes in relation to the *C. elegans* gene family. In addition I have been able to provide results adding to knowledge of the mechanisms underlying the involvement of specific mutations in β-tubulin genes in anthelmintic drug resistance phenotypes. However, as discussed, the results presented in this chapter may require further investigation to allow full conclusions to be drawn.

6.7 Tables and Figures

Table 6.1 Summary of *C. elegans* transgenic lines described

Strain name	Description
GIS11	N2 injected with *Ce-ben-1*::GFP and rol-6
GIS20	N2 injected with *Ce-mec-7*::GFP and rol-6
GIS14	N2 injected with *Hc-isotype-1*::GFP and rol-6
GIS18	N2 injected with *Hc-isotype-2*::GFP and rol-6
GIS8	N2 injected with 1.082 kb *Hc-isotype-3*::GFP and rol-6
GIS6	N2 injected with 1.488 kb *Hc-isotype*-3::GFP and rol-6
GIS1	N2 injected with 2.056 kb *Hc-isotype*-3::GFP and rol-6

Table 6.2 Summary of *C. elegans* β-tubulin IFAT investigations

Strain	Antibody	Concentration	Method	Result
N2	MEC-7	1:1000	Freeze-crack	6 touch neurons and axons (excluding nucleus) Additional faint staining in the FLP, PVD and BDU neurons
N2	Pre-absorbed BEN-1	1:100	Freeze-crack	Faint nuclear staining; neuronal staining (putatively all neurons)
tm234	MEC-7	1:1000	Freeze-crack	6 touch neurons and axons (excluding nucleus) Additional faint staining in the FLP, PVD and BDU neurons
tm234	Pre-absorbed BEN-1	1:100	Freeze-crack	Faint nuclear staining
N2	MEC-7	1:1000	Ruvkun Fixed	6 touch neurons and axons (excluding nucleus) Additional faint staining in the FLP, PVD and BDU neurons + High level of background
N2	Pre-absorbed BEN-1	1:100	Ruvkun Fixed	Neuronal staining (putatively all neurons) and faint nuclear staining + High level of background
tm234	MEC-7	1:1000	Ruvkun Fixed	6 touch neurons and axons (excluding nucleus) Additional faint staining in the FLP, PVD and BDU neurons + High level of background
tm234	Pre-absorbed BEN-1	1:100	Ruvkun Fixed	Faint nuclear staining + High level of background

Table 6.3 Summary of *H. contortus* β-tubulin IFAT investigation

Strain	Antibody	Concentration	Method	Result
MHco3 (ISE)	Isotype-3	1:100	Freeze-crack	6 touch neurons and axons (excluding nucleus) Additional faint staining in the FLP, PVD and BDU neurons
MHco3 (ISE)	Isotype-1	1:20	Freeze-crack	Neuronal staining (putatively all neurons)
MHco3 (ISE)	Isotype-2	1:20	Freeze-crack	Neuronal staining (putatively all neurons)
MHco3 (ISE)	Isotype-3	1:20	Ruvkun Fixed	6 touch neurons and axons (excluding nucleus) Additional faint staining in the FLP, PVD and BDU neurons
MHco3 (ISE)	Isotype-1	1:1	Ruvkun Fixed	Non-specific cuticle staining
MHco3 (ISE)	Isotype-2	1:1	Ruvkun Fixed	Non-specific cuticle staining

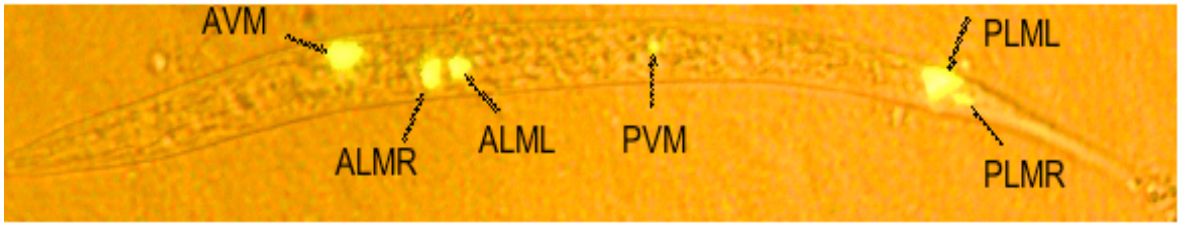
Table 6.4: β-tubulin tag data from *C. elegans* SAGE datasets

Gene name	Number of tags in each *C. elegans* life-cycle stage SAGE database						
	Embryo	L1	L2	L3	L4	Young adult	**Total (%)**
tbb-2	30	48	24	18	30	55	**205 (45)**
tbb-1	16	23	14	13	22	29	**117 (25)**
mec-7	10	47	27	9	6	4	**103 (23)**
tbb-4	1	0	1	11	1	0	**14 (3)**
ben-1	2	3	1	0	2	1	**9 (2)**
tbb-6	2	0	1	0	0	0	**3 (0.7)**

Table 6.5: Representation of each of the six *C. elegans* β-tubulin genes in all EST datasets

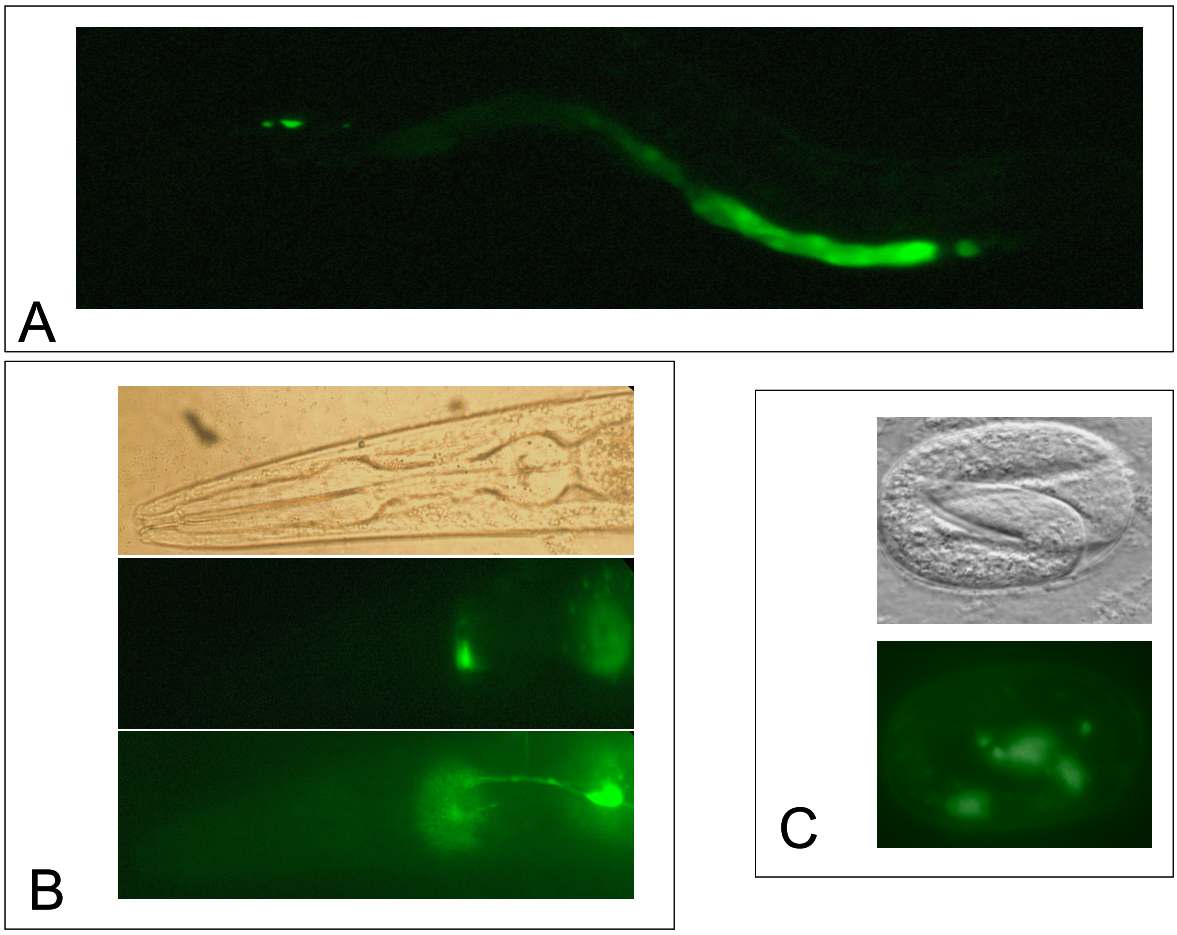
Gene name	Number of ESTs	% of total ESTs
tbb-2	**336**	**70**
tbb-1	**103**	**21**
mec-7	**29**	**6**
tbb-4	**9**	**2**
ben-1	**3**	**0.6**
tbb-6	**1**	**0.2**

Figure 6.1 Expression of *Ce-mec-7*::GFP promoter-reporter gene fusion in transgenic *C. elegans*



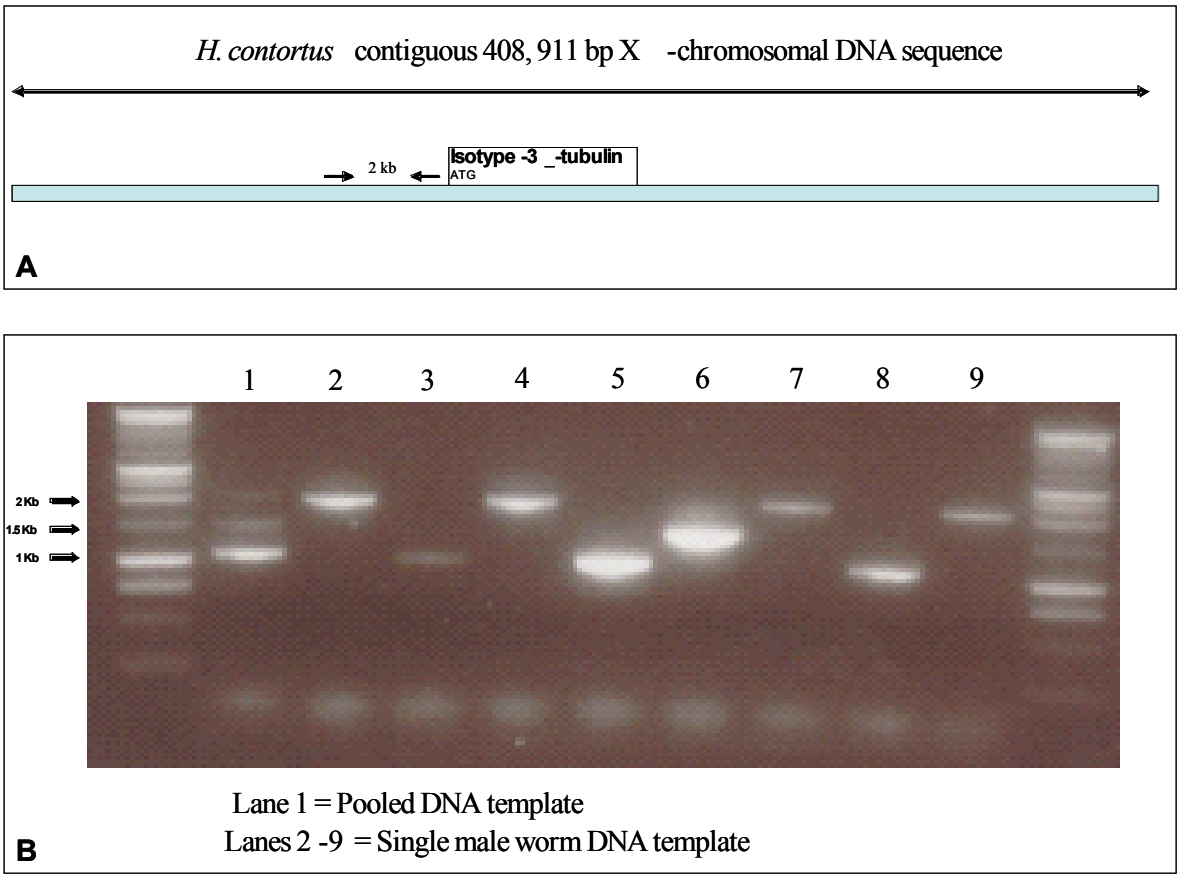
Expression of GFP is detected in all six touch neurons and associated axon structures of strain GIS20 (Table 6.1). Each cell is indicated by an arrow. These cells were identified based on their positions and the direction in which the axonal structures extended based on comparisons with Wormatlas (http://www.wormatlas.org/cellid/cellID.htm) images and the *mec-7-lacZ* fusion gene expression pattern published by Hamelin *et al.* (Hamelin *et al.*, 1992).

Figure 6.2 Expression of *Ce-ben-1*::GFP promoter-reporter gene fusion in transgenic *C. elegans*



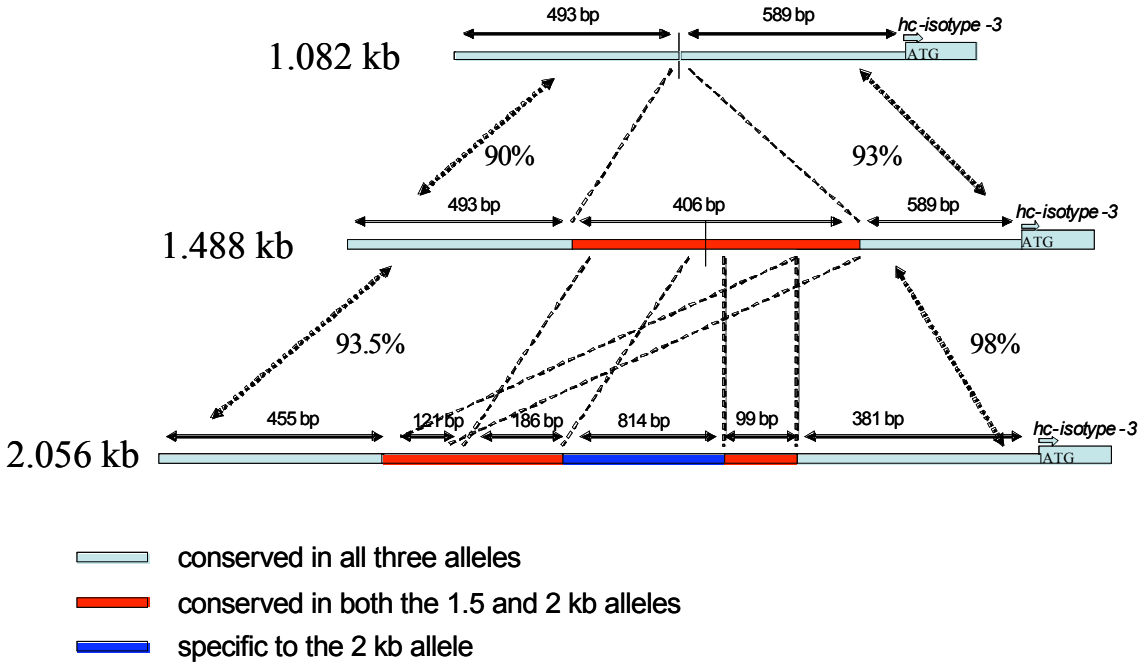
Expression is mainly localised to the posterior gut and head neurons of transgenic *C. elegans* strain GIS11 (Table 6.1; Panel A). Head neurons are putatively described as the BDU neurons based on comparisons with Wormatlas (http://www.wormatlas.org/cellid/cellID.htm) and other published GFP patterns (http://gfpworm.org/index) that localise to these neurons (B). Expression is first seen at the 3-fold embryo stage and the pattern is postulated to localise to the same tissues in all *C. elegans* life stages from this point (C).

Figure 6.3 Amplification of three 5' flanking region alleles of the *H. contortus* β-tubulin *isotype-3* locus



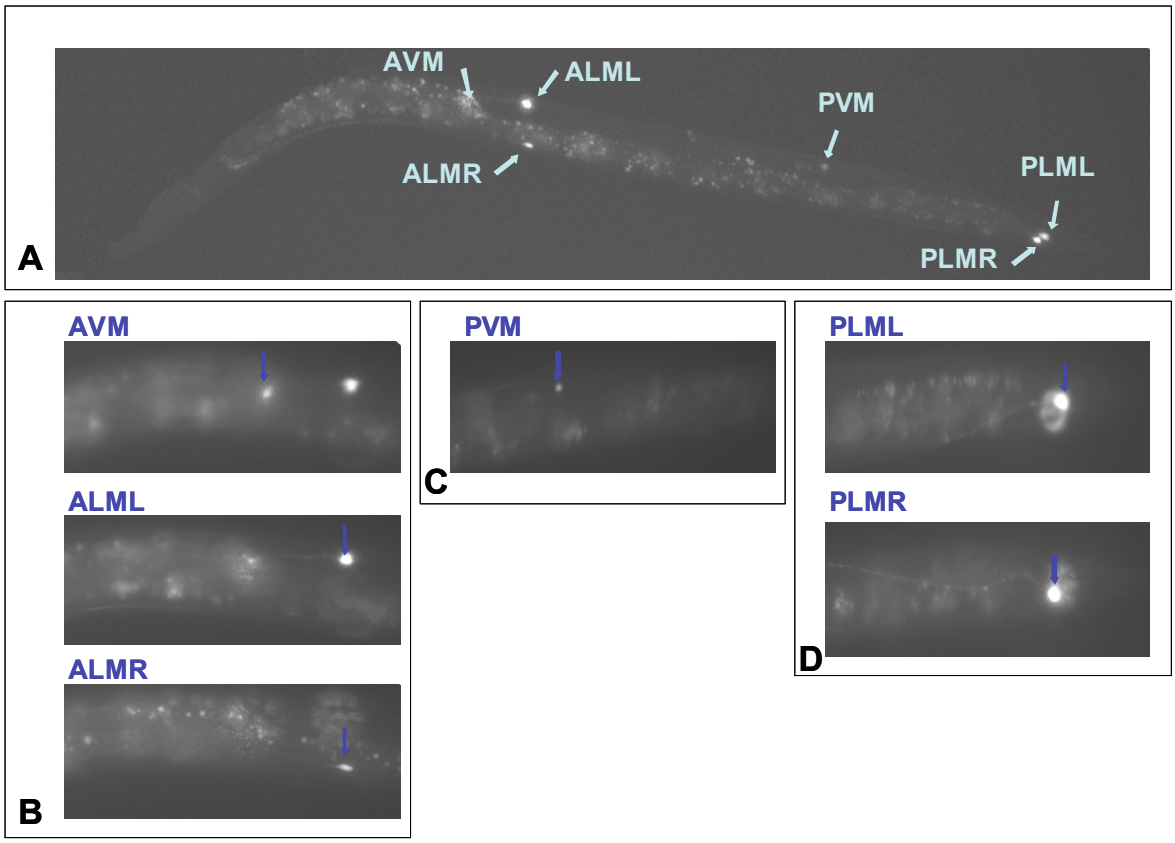
Primers were designed to amplify 2.056 kb 5' flanking region of the *H. contortus* β-tubulin *isotype-3* locus, which is fully contained on a 408,911 bp contiguous stretch of X-chromosomal DNA (A; primer sequences in Appendix 2). This primer set repeatedly amplified three products of 1.082, 1.488 and 2.056 kb when pooled *H. contortus* MHco3 (ISE) genomic DNA was used as template. Amplification from single *H. contortus* MHco3 (ISE) male lysates produced each allele individually (B).

Figure 6.4 Comparative analysis of consensus sequences for three allele lengths of 5' flanking region of the *H. contortus* β-tubulin *isotype-3* locus



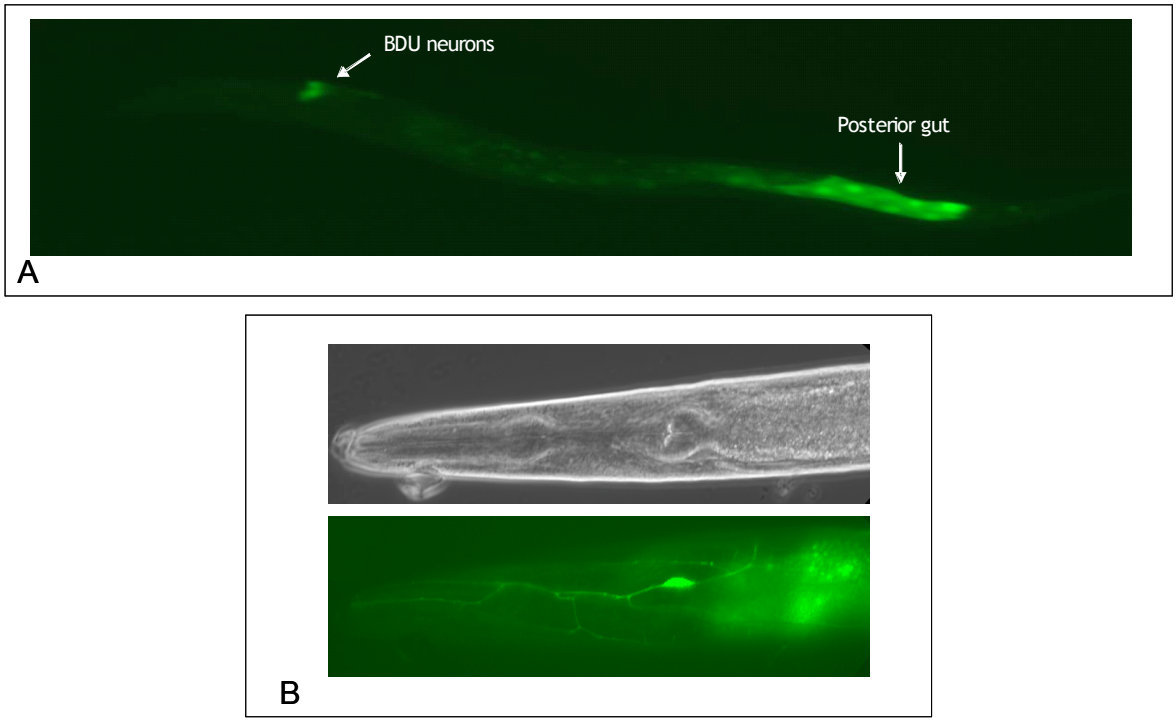
This schematic figure compares three alleles of 5' flanking region for the *H. contortus* β-tubulin *isotype-3* locus, showing that all are related to one another. There is an insertion of a 406 bp fragment from the 1.082 kb to the 1.488 kb alleles (red block). This insertion is maintained in the 2.056 kb allele, where the insert region closest to the translational start site has been translocated (indicated by dotted lines). The 2.056 kb allele has an insertion of an 841 bp fragment that is present in the 1.082 or 1.488 kb allele sequences (royal blue block). The percentage identities of the sequences common to all three alleles are shown.

Figure 6.5 Expression of *Hc-iso*-3::GFP promoter-reporter gene fusion in transgenic *C. elegans*



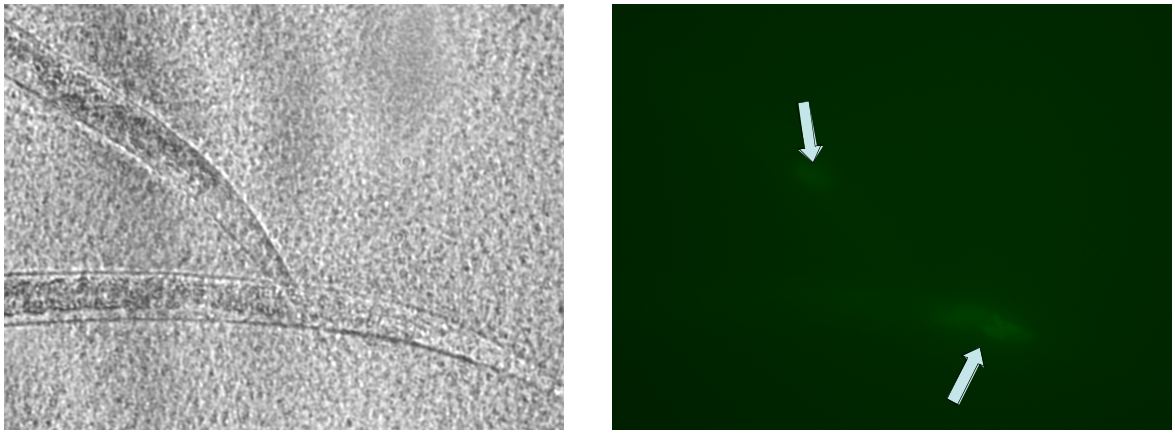
All images of GFP expression in transgenic *C. elegans* (strain GIS8, Table 6.1) for this figure were driven by the 1.082 kb allele promoter of the *Hc-isotype-3* gene. Images were taken by a non-colour camera and not pseudo-coloured so that axons could be seen. A full *C. elegans* worm with the six touch cells fluorescing is shown in panel A. Panels B, C and D are images of these touch neurons at a higher magnification (x 40), allowing the axons to be seen for most. The blue arrow in each panel indicates the neuronal cell fluorescing. Cells were identified based on their positions and directions of associated axons in conjunction with comparison of the pattern associated with the *mec-7*::GFP promoter-reporter fusion construct. Generally the level of GFP fluorescence seen from the touch receptors in this figure is far lower than that observed with the *Ce-mec-7* promoter sequence. Auto-fluorescence from the *C. elegans* gut granules can be seen in the majority of images within this figure. This auto-fluorescence is distinguishable down the microscope from GFP expression since it is a yellow colour as opposed to green.

Figure 6.6 Expression of *Hc-isotype*-1::GFP promoter-reporter gene fusion in transgenic *C. elegans*



GFP expression from transgenic *C. elegans* strain GIS14 is controlled by 1.971 kb of *Hc-isotype-1* 5' flanking sequence (Table 6.1). GFP fluorescence is localised to the posterior gut and head neurons of *C. elegans*, annotated with arrows (Panel A). Head neurons are putatively described as the BDU neurons based on comparisons with Wormatlas (http://www.wormatlas.org/cellid/cellID.htm); other published GFP patterns that localise to this tissue (http://gfpworm.org/index) and the GFP expression pattern associated with the *Ce-ben-1* promoter-reporter fusion construct (Figure 6.2). Images were taken by a non-colour camera and pseudo-coloured. Panel B shows head neurons seen fluorescing at higher magnification (x 40). These neurons are putatively described as the amphid neurons based on comparisons with Wormatlas (http://www.wormatlas.org/cellid/cellID.htm) and other published GFP patterns (http://gfpworm.org/index) that localise to this tissue.

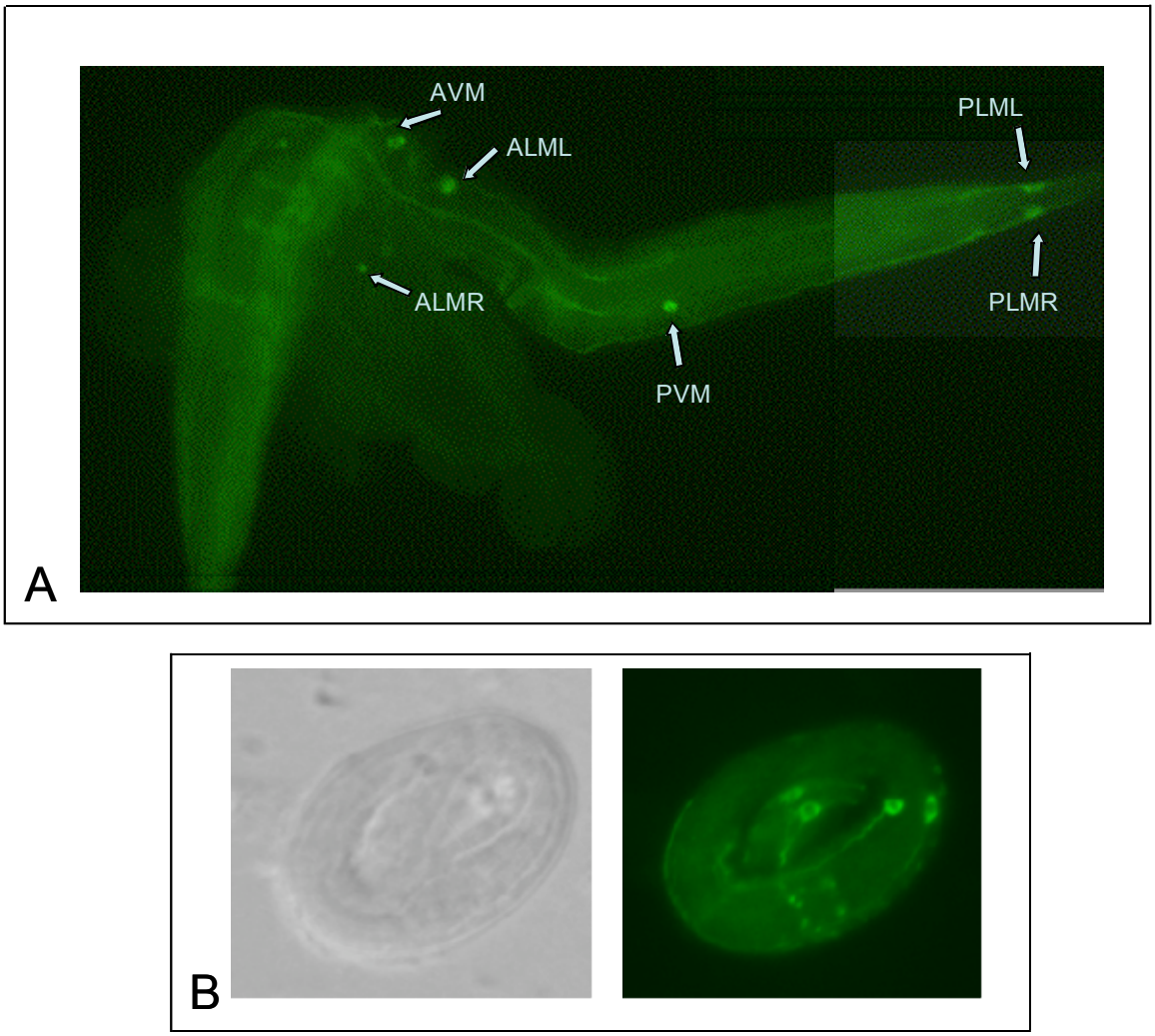
Figure 6.7 Expression of *Hc-isotype*-2::GFP promoter-reporter gene fusion in transgenic *C. elegans*



Images are of *C. elegans* transgenic strain GIS18 (Table 6.1). Very weak GFP fluorescence is detected in the posterior gut of two *C. elegans* worms during the early larval stages of development. Regions of fluorescence are indicated by blue arrows. Faint posterior gut expression is a known caveat of the Fire expression vectors, thought to be as an artefact of the *unc-54* 3'-UTR which is included in most vectors (http://www.wormbook.org/chapters/www_reportergenefusions/reportergenefusions.html). It remains uncertain if this fluorescence represents true GFP expression.

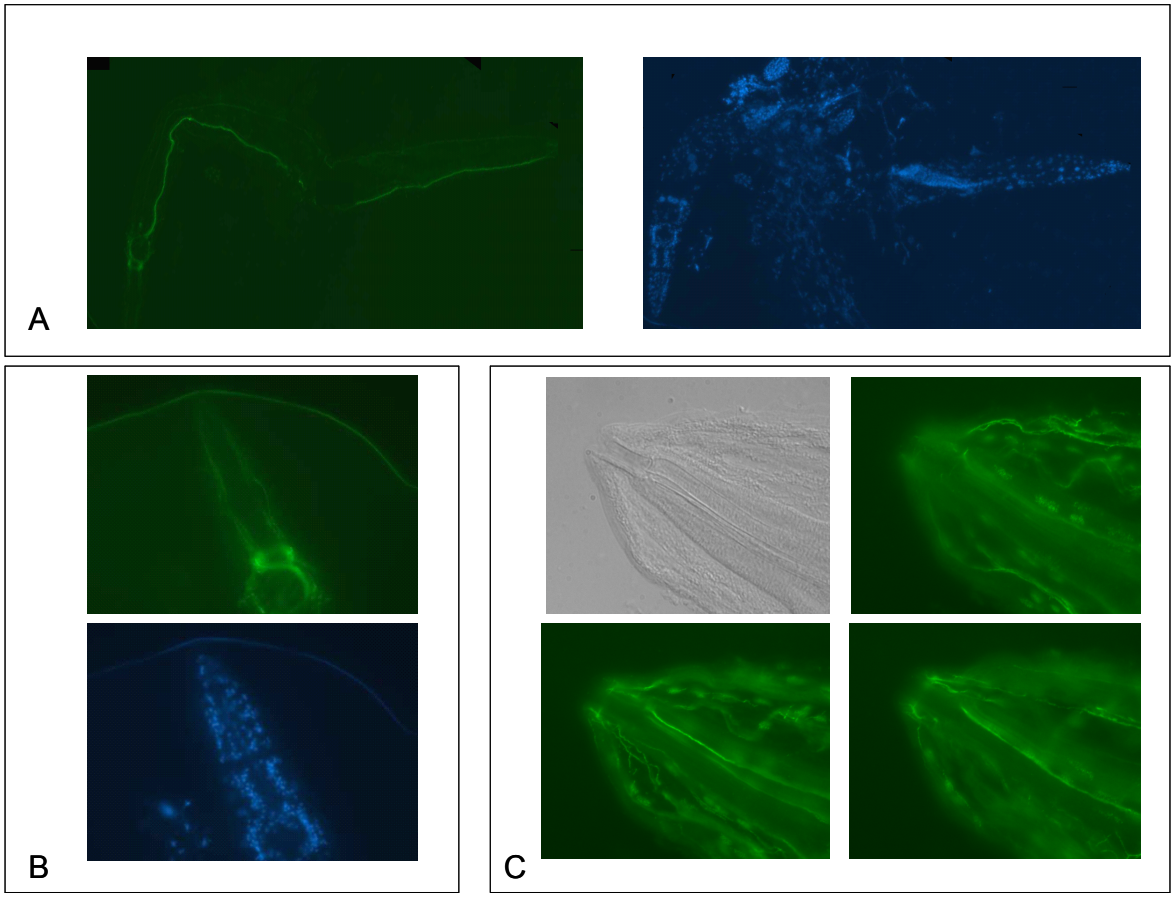
`

Figure 6.8 A MEC-7 specific antibody localises to the touch cell receptors of *C. elegans* N2 worms



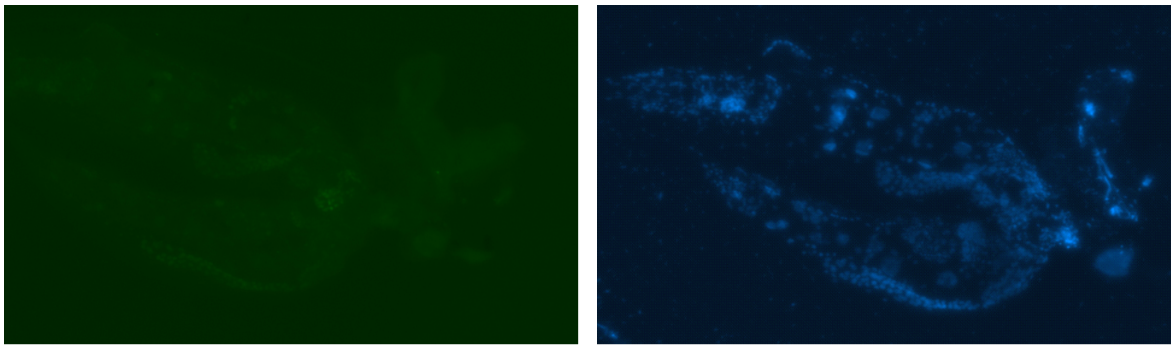
This Ce-MEC-7 specific antibody was a gift of the Chalfie laboratory. Detection of the MEC-7 protein is shown in all six touch receptors and associated axonal processes of *C. elegans* N2 worms (Panel A). Detection was first noted at the three-fold stage of *C. elegans* development (Panel B). This is the same localisation pattern as previously reported when using the same antibody at this concentration against *C. elegans* N2 worms (Savage *et al.*, 1994). This allowed verification of my IFAT preparation and staining protocols.

Figure 6.9 A BEN-1 antibody stains the neuronal circuitry and nucleus of all cells of *C. elegans* N2 worms



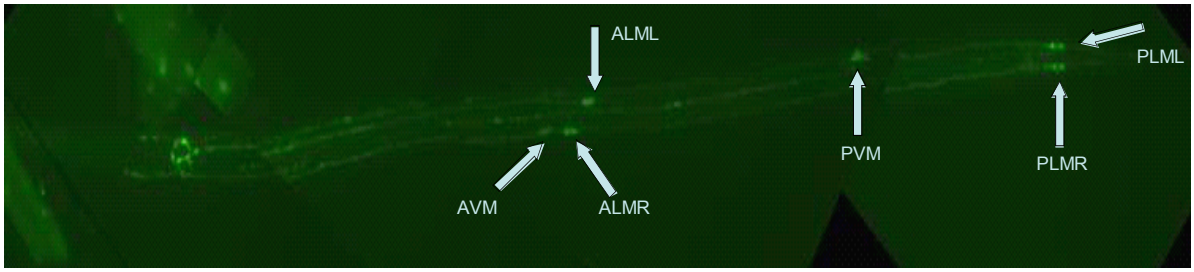
A BEN-1 specific antibody (Table 2.1) localises to the majority of the neuronal circuitry of N2 worms when first pre-absorbed against *ben-1*$^{-}$ strain tm234. The dorsal/ventral cord of *C. elegans* contains many axonal processes and is seen most prominently (Panel A). At higher magnification the complexity of the neuronal staining can be seen (Panels B and C). Faint background nuclear staining is still associated with this localisation pattern, and can be seen at x100 magnification (Panel C). Panels A and B show DAPI comparison, whereas panel C shows a Bright Field image comparison.

Figure 6.10 Background staining of *C. elegans* strain tm234 (*ben-1* mutant) seen with pre-absorbed Ce-BEN-1 antibody



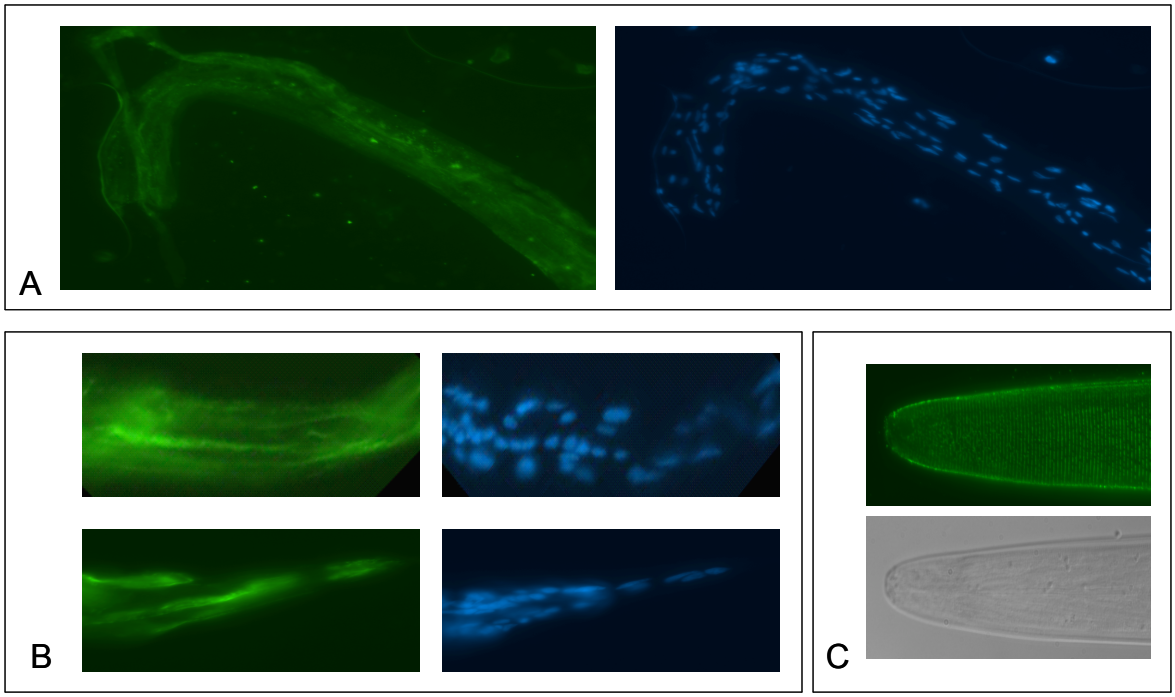
The BEN-1 antibody (Table 2.1) faintly stains the nucleus of many, possibly all, cells of tm234 worms when pre-absorbed against tm234 acetone powders (Panel A). However no neuronal staining is seen in tm234 worms with the pre-absorbed antibody indicating that the neuronal staining seen in wild type worms with the Ce-BEN-1 purified IgG antibody is specific (Figure 6.9). A DAPI stained comparison is shown to the right of the FITC image.

Figure 6.11 An ISO-3 specific antibody primarily stains the touch cell circuitry of *H. contortus* MHco3 (ISE) specimens



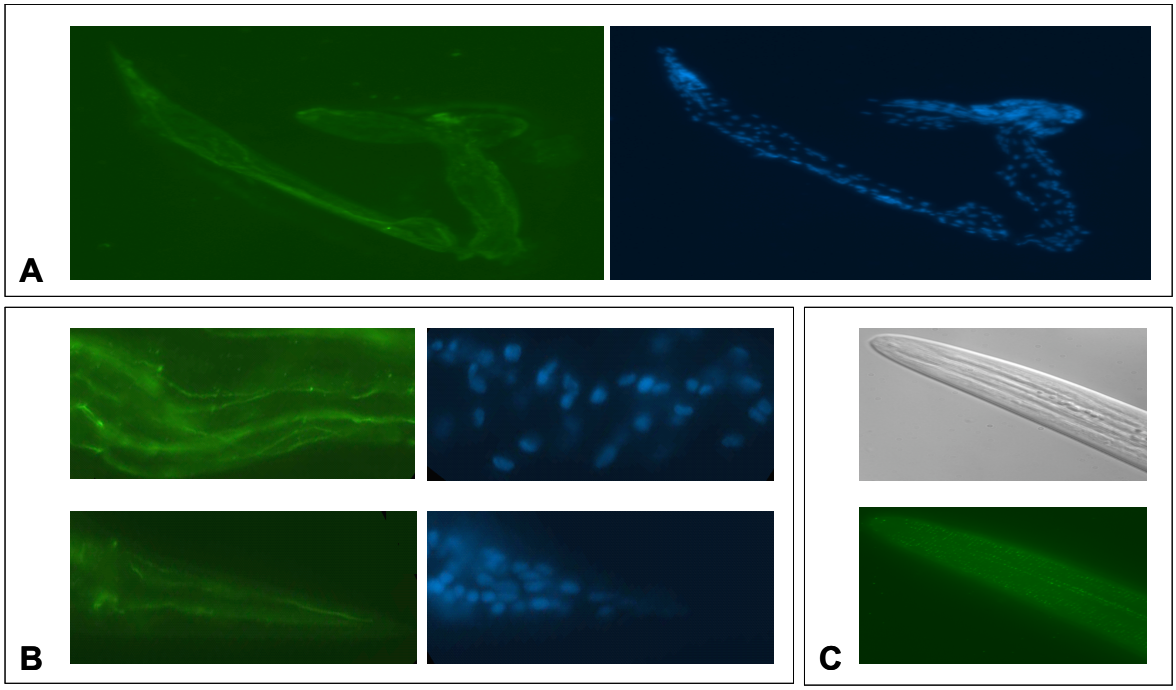
Exsheathed L3 *H. contortus* MHco3 (ISE) worms fixed by the Ruvkun method and stained with purified IgG directed against the *H. contortus* Isotype-3 polypeptide (Table 2.1). The touch cell circuitry primarily fluoresces. Positions of each of the six touch cells are indicated by the blue arrows. These tissues were identified based on similarities with the *C. elegans* neuronal circuitry (http://www.wormatlas.org/cellid/cellID.htm).

Figure 6.12 Hc-ISO-1 antibody staining of *H. contortus* MHco3 (ISE) worms



Purified IgG antibody from antiserum directed against the final thirteen amino acids of the *H. contortus* Isotype-1 protein localises to the majority of, possible the entire, neuronal circuitry of *H. contortus* exsheathed L3 MHco3 (ISE) worms prepared by the freeze-crack method (Table 2.1, Panels A and B). The complexity of axonal processes stained with the antibody can be seen at high magnification. The dorsal/ventral cord is brightest in both figures of panel B, as this neuron cord has many axonal processes associated. However, surrounding this cord a multitude of other axonal processes show antibody binding. Worms prepared by the Ruvkun fixation method show only non-specific binding of antibody to the cuticle of the worm due to the very high concentrations used, but no internal fluorescence, when stained with the same antibody (Panel C). Panels A and B show a DAPI stained image comparison, whereas a Bright Field comparison image is shown in Panel C.

Figure 6.13 Hc-ISO-2 antibody staining of *H. contortus* MHco3 (ISE) worms



Purified IgG antibody from antiserum directed against the final fourteen amino acids of the *H. contortus* Isotype-2 protein localises to the majority of, possible the entire, neuronal circuitry of *H. contortus* exsheathed L3 MHco3 (ISE) worms prepared by the freeze-crack method (Table 2.1, Panels A and B). A number of axonal processes can be seen by IFA at high magnification (Panel B). Identification of specific neurons was aided by comparisons with the fully characterised neuronal circuitry of the nematode genetic model organism *C. elegans* (http://www.wormatlas.org/cellid/cellID.htm). Worms prepared by the Ruvkun fixation method show only non-specific binding of the antibody to the cuticle of the worm due to the very high concentrations used, but no internal fluorescence, when stained with the same antibody (Panel C). Panels A and B show a DAPI stained image comparison, whereas a Bright Field comparison image is shown in Panel C.

Chapter 7: General Discussion

Significant effort is being focussed on improving the impoverished state of genomic resources for parasitic nematode species (http://www.sanger.ac.uk/Projects/Helminths/; http://genomeold.wustl.edu/genome_group.cgi?GROUP=6). The advent of next generation sequencing techniques is beginning to produce vast amounts of genomic sequence at very affordable prices. However, the generation of sequence data is just the first step in the process of having assembled and annotated genome data for a species. The *H. contortus* genome sequencing project first started in 2004 and currently more than 800 Mb of genomic sequence is available (http://www.sanger.ac.uk/Projects/H_contortus/). Although this data can be readily searched for genes of interest, identification and comparison of full gene sequence and gene organisation is currently difficult. This is due mainly to the difficult challenge of assembling the sequence data into large contiguous fragments. Based on the results presented in this thesis, and those recently published (Chapter 3; Redman *et al.*, 2008), the most likely explanation for the problems of assembly is the high level of natural genetic variation between *H. contortus* haplotypes. Although progress on the *H. contortus* genome sequencing project has, therefore, not been as straightforward as may have been anticipated, the difficulties encountered and approaches taken to aid assembly will be of significant importance for genome sequencing projects undertaken for similar species. This will have most relevance to other Trichostrongylid parasitic nematode species, as the evidence suggests that it is within populations of this class of nematode that the highest levels of natural genetic variation exist (Blouin, 1998). Therefore the lessons learned from the *H. contortus* genome sequencing project are not only beneficial to the *H. contortus* research community, but also to researchers working on similar species and planning sequencing projects for their organism of interest.

Another factor which seems to be contributing to the problem of assembling the *H. contortus* sequence data is genome size. The *H. contortus* haploid genome has previously been estimated at ~53 Mb, based on flow cytometry (Leroy *et al.*, 2003). Data presented in this thesis on percentage coverage of the genome, as well as analysis of gene size and organisation suggests that this is likely to be a gross underestimate (Chapter 3; R.N. Beech, McGill University, R. Laing, University of

Glasgow, and J.S. Gilleard, University of Calgary, personal communication). Ongoing efforts to generate more sequence data should allow more accurate determination of the *H. contortus* genome size as well as being important to further aid the assembly of the data.

One of the major reasons that *H. contortus* was selected as one of the first parasitic nematode genome sequencing projects to be initiated was the size and strength of the research community. This species is the organism of choice for a number of research groups investigating a diverse range of topics including drug resistance, vaccine production and evolutionary biology to name but three (Gilleard, 2006). With the genome project now at a far advanced stage, we are beginning to see the impact that the completed genome sequence will have on this research community. For example, the available *H. contortus* genomic sequence databases have been used to piece together the DEG-3 subfamily of nicotinic acetylcholine receptors from the genome of this species. This work has allowed specific mutations in these genes which result in reduced sensitivity to the amino-acetonitrile derivatives class of novel anthelmintic drug be characterised, and ultimately a detailed understanding of the mechanism by which such compounds exert their anthelmintic effects be understood (Rufener, 2009).

Very recently, the first genome sequence data was published for the human platyhelminth parasites *Schistosoma mansoni* and *Schistosoma japonicum* (Berriman *et al.*, 2009; *Schistosoma japonicum* Genome Sequencing and Functional Analysis Consortium *et al.*, 2009). These published genome sequences (363 Mb for *S. mansoni* and 397 Mb for *S. japonicum*) now pave the way for genome wide investigations for this class of species. Employing such resources allows poignant, niche questions to be investigated, and data to be readily and rapidly generated. Schistosome parasites cause significant morbidity in tropical regions, and the available data will have major importance to the field of drug discovery. At the present time there is only a single drug, praziquantel, available to treat schistosomiasis. However, this genome sequence data will make rational, target-based approaches possible and a number of candidate targets for new therapeutic drugs have already been identified through bioinformatic approaches. These include a family of nuclear receptors and a thirodexin glutathione reductase redox enzyme, as well as novel targets identified through analysis of metabolic chokepoints and similarity to known druggable proteins (Berriman *et al.*, 2009). These high throughput target-based drug discovery

approaches are enabled by a complete genome sequence, and it is very likely that such resources will have similar application to the field of parasitic nematode research, particularly with resistance to available drugs increasing.

The availability of draft genome sequence of two closely related platyhelminth species gives this research community the additional power of genome wide comparative investigation. The power of such analysis is already having an impact on the future directions of platyhelminth research. For instance, a new class of micro-exon gene has been identified within the genomes of both species, lending themselves not only to be novel drug targets, but also amenable to sophisticated analysis of the evolution of the *Schistosoma* spp. Therefore, the important impact that genomic resources can have on a research community, and how this can change the direction of future research is already becoming evident in this case.

Although the advantages of having a complete, assembled genome sequence for organisms of interest is clear, the results presented in this thesis also highlight the impact that available genomic sequence fragments are able to provide. Parasitic nematode species are difficult experimental subjects as, for the vast majority, there are no *in vitro* life cycle propagation methods or any transgenic or reverse genetic investigative tools. One approach that would be extremely advantageous for functional genomic studies of parasitic nematode genes would be RNAi. This technique allows researchers to investigate the function of their gene of interest by reducing specific gene expression and looking for visible phenotypic differences in these worms compared to non-RNAi treated worms. While RNAi is widely used in the free-living nematode species *C. elegans* (Kamath and Ahringer, 2003; Wang and Barr, 2005; Kim *et al.*, 2005), the findings from RNAi studies in other nematode species are extremely variable and very inconsistent (Winston *et al.*, 2007). The variable effects of RNAi on *H. contortus* gene expression have been highlighted (Geldhof *et al.*, 2007). A more recent study supports the inefficiency of RNAi in parasitic nematodes (Lendner *et al.*, 2008). I have searched the available *H. contortus* genomic DNA sequence datasets for the presence of genes required for viable RNAi in *C. elegans* (Grishok, 2005) to hypothesise why this may be the case. Although bioinformatic investigations do not provide conclusive evidence for the observed inefficiency of parasite RNAi, they do provide a platform for further investigation.

Using sequence homology based investigative techniques, I found that genes encoding two proteins necessary for functional RNAi in *C. elegans*, RDE-2 and RDE-4, appear to be missing from the *H. contortus* genomic sequence databases. In addition, I failed to identify three genes, *rsd-2*, *rsd-6* and *sid-2*, encoding proteins involved in systemic RNAi in the free living nematode in the current *H. contortus* sequence databases. However, there are many caveats associated with this analysis. As discussed earlier, the *H. contortus* genome sequencing project is currently incomplete, therefore, loci homologous to these *C. elegans* proteins may reside within the *H. contortus* genome but not yet have been sequenced. However, my analysis suggests that this is unlikely to be the case for the majority of *H. contortus* loci, as we know that 96% of sequenced *H. contortus* single cluster EST sequences are represented in current databases and the current estimate of the coverage of each is >80% (Chapter 3). A second reason that loci homologous to these *C. elegans* proteins may not have been identified by sequence analysis could be that *H. contortus* loci are functional homologues to these *C. elegans* proteins, but not necessarily identifiable based on sequence homology. This is of critical relevance when it is considered that current estimates are that the *H. contortus* and *C. elegans* lineages diverged hundreds of millions of years ago.

Despite these caveats, comparative genomic investigations can be useful starting points. As stated previously, parasitic nematode genomic resources are in an impoverished state. There are completed genome sequences for a number of nematodes; although very few are for parasitic species (Thomas, 2008). Interestingly, the application of RNAi to investigate gene function is variable across the phylum Nematoda (Winston *et al.*, 2007; Geldhof *et al.*, 2007). Therefore, it is reasonable to hypothesis that the RNAi machinery may be well-conserved between species for which RNAi functions robustly, but less conserved in species where RNAi is less effective. From the results presented in this thesis, most interesting in this regard is the SID-2 protein. Sequence homologous to this protein could not be identified within the available sequence datasets for *P. pacificus* or *H. contortus*. Furthermore, the identified putative *B. malayi* SID-2 protein is more similar to *C. briggsae* SID-2 than it is to the *C. elegans* protein. Functional and sequence differences have been found between the *C. elegans* and *C. briggsae* SID-2 proteins (Winston *et al.*, 2007), and this has been demonstrated, by transgenic studies, to explain the differences in susceptibility to systemic RNAi between the two species.

Within the *C. elegans* pathway, the SID-2 protein is hypothesised to facilitate the uptake of dsRNA from the environment. As systemic RNAi is not routinely functional in *C. briggsae*, *P. pacificus*, *H. contortus* or *B. malayi*, my results suggest that the this could be due to lack of a homologous *sid-2* locus or differences in function when compared with *C. elegans* SID-2 protein. Although the bioinformatic analysis does not confirm the suggested hypothesis, it provides support for this and may aid future studies aimed at improving dsRNA uptake and accessibility for the phylum Nematoda.

The results presented in this thesis also use the available *H. contortus* sequence databases to investigate the increasingly significant problem of anthelmintic drug resistance (Bartley *et al.*, 2003; Sargison *et al.*, 2001; SAC report, 2000; Jackson *et al.*, 1992). Resistance to all three classes of anthelmintic used to control *H. contortus* infection is widespread and is of major global concern (Gilleard, 2006). Of the three drug types, resistance to the BZ class is most common and widespread (Table 1.1). This was the first class of anthelmintic drug developed and was released during the 1960s, with resistance first reported only a few years later. The mechanism which underpins BZ resistance has been the most intensively studied and is the best understood (Von Samson-Himmelstjerna, G., 2007). Mutations within β-tubulin genes are known to confer resistance to BZ drugs in a wide range of species, including nematodes (Von Samson-Himmelstjerna, G., 2006), although other genes may also be involved in resistance (Blackhall *et al.*, 2008). In the case of *C. elegans*, specific mutations at only one locus, that of the β-tubulin *ben-1* gene, have been shown to confer the BZ resistance phenotype (Kwa, M.S., *et al.*, 1995). However, mutations in two *H. contortus* β-tubulin loci, *isoype-1* and *isotype-2,* have been implicated in the BZ resistance mechanism for this species (Kwa, M.S., *et al.*, 1993; Beech, R.N., *et al.*, 1994). This implies that although these resistance mechanisms are similar, they are not identical. The results presented in this thesis (Chapter 5) have used comparative genomic techniques to investigate the evolution of the β-tubulin gene families of *C. elegans* and *H. contortus* since the organisms diverged hundreds of millions of years ago. I have been able to show that the three β-tubulin loci from the two organisms involved in the respective BZ resistance mechanisms (*C. elegans ben-1* and *H. contortus isotype-1 and isotype-2*) share a paralogous relationship. This partially explains the involvement of only one *C. elegans* but two *H. contortus* β-tubulin loci in these BZ resistance mechanisms (Chapter 6). My

results suggest that neither the β-tubulin *isotype-3* nor *isotype-4* loci from the *H. contortus* genome are involved in this mechanism. However, it is important to consider that this hypothesis is based on sequence analysis only. Experimental validation of this hypothesis should be undertaken in order to prove whether this is the case. This is perhaps of most significance for the *Hc-isotype-4* locus, where the complete sequence of the polypeptide encoded by the gene remains unknown.

This example of a comparative genomic approach to understand the mechanisms underpinning BZ resistance paves the way for future investigations into the mechanisms underpinning both the LEV and IVM resistance phenotypes. The loci involved in the LEV and IVM resistance mechanisms remain largely unknown and those identified thus far have been based on knowledge of similar mechanisms in other species (Gilleard, 2006). However, it is an important consideration that the BZ resistance mechanism is very likely to be the simplest of all three anthelmintic classes. This is estimated from the time that resistance started emerging in the field once the drug was used to treat infections, and the ease with which laboratory organisms are able to be mutated in order to obtain this phenotype. As I have shown, subtle differences do exist between the BZ resistance mechanisms of even the very closely related species *C. elegans* and *H. contortus*. It can therefore be feasibly predicted that greater differences may exist between the mechanisms by which these species develop resistance to the LEV and/or IVM class of anthelmintic drug. Such analysis will be aided significantly by availability of genome, as well as transcriptome data, as discussed below.

The advent of next generation sequencing technologies has not only seen an increase in the number of genome sequencing projects that have been initiated for parasitic nematodes, but also in the number of transcriptome sequencing projects. For example, recently, comparative transcriptome analysis of three stages of the *O. ostartagi* life cycle and investigation of the *B. malayi* L3 life-cycle stages were analysed and published (Abubucker *et al.*, 2009; Li *et al.*, 2009). These datasets allow for a rapid improvement in our understanding of the biology of these complex organisms. In the case of *B. malayi*, for example, identifying genes that are differentially expressed in the L3 lifecycle stage allows insights into the infectivity, immunogenicity and survival when the filarial larvae are transmitted by arthropod vectors to initiate infection in mammals. This study identified 771 loci that were

differentially expressed between the parasites in the mosquito and mammalian hosts, greatly extending current knowledge of the genes involved in these processes. In conjunction with genome sequencing, transcriptome sequencing projects have recently been initiated for *H. contortus* (R. Laing, University of Glasgow, M. Berriman, Wellcome Trust Sanger Institute, and J.S. Gilleard, University of Calgary, personal communication). A comparative transcriptomic approach is planned using wild-type worms or strains resistant to one, two or three classes of anthelmintic drug to help further understand mechanisms of drug resistance in *H. contortus*. A major advantage of such investigations is that, as they are genome wide, novel loci can be implicated in resistance mechanisms with no prior knowledge, or hypothesis, needed. Examples of such loci that have been identified thus far are those of the P-glycoprotein and Cytochrome-P450 families from the *H. contortus* genome (R.N. Beech, McGill University, R. Laing, University of Glasgow, and J.S. Gilleard, University of Calgary, personal communication).

Although the availability of genomic resources is likely to have a major impact on parasitic nematode research, the experimental validation of results obtained is difficult in the absence of reliable transgenic or reverse genetic tools. However, *C. elegans* can be exploited as a suitable tool applicable to parasitic nematode research (Gilleard, 2004). Genetic modification of this free-living nematode species is now fairly routine and there are many examples where *C. elegans* has been employed as a transgenic system to investigate parasitic nematode gene expression and function (Table 1.2). Data presented in this thesis (Chapter 6) demonstrate the suitability of *C. elegans* to study the relationships and expression of the *H. contortus* β-tubulin gene family. Importantly, the expression pattern of *H. contortus* β-tubulin genes in transgenic *C. elegans* was complemented by immunolocalisation studies directly in the parasite, validating this approach.

The extent to which *C. elegans* can be exploited as a transgenic tool for parasitic nematode research is an important question. The results presented here provide another example of the successful use of this system to investigate genes from a Trichostrongylid nematode species. Data is starting to accumulate, however, both in publication and presentation, that the use of *C. elegans* to investigate genes from more divergent nematode species may be difficult (Gillan *et al.*, 2009; E. Devaney and A.D Winter, University of Glasgow, personal communication) and may depend on

the gene under investigation. It is therefore likely that the use of *C. elegans* as a tool to investigate parasitic nematodes will be most relevant to species most closely related to *C. elegans*, such as the Trichostrongylid nematodes. When it is considered that there are examples of species within this Clade that diverged from a common ancestor hundreds of millions of years ago, this may not be surprising.

Overall, this thesis presents the findings of the first detailed analysis of the data generated from the *H. contortus* genome sequencing project. The approach has exploited *C. elegans* sequence data and experimental techniques to study the *H. contortus* genome, in particular helping to identify and characterise *H. contortus* RNAi pathway genes and the β-tubulin gene family. The bioinformatic and experimental investigations support the use of *C. elegans* in parasitic nematode research, particularly for Clade V nematodes, and the findings provide new comparative genomic driven hypotheses on the variability of RNAi across the phylum Nematoda, and the subtle differences in BZ resistance mechanisms across Trichostrongylid nematodes.

Appendices

8.1 Standard reagents

AbA Buffer	1X PBS, 1 % BSA (fraction V), 0.5 % Triton-X-100, 0.05 % Na Azide, 1mM EDTA
AbB Buffer	1X PBS, 0.1 % BSA (fraction V), 0.5 % Triton-X-100, 0.05 % Na Azide, 1mM EDTA
Ampicillin	100 mg/ml ampicillin (Sigma) in sterile distilled H_2O. Filter sterilised and stored at -20^oC.
BO_3 Buffer (20X)	1M H_3BO_3, 0.5M NaOH
DEPC H_2O	0.1% (v/v) diethylpyrocarbonate (Sigma) in sterile distilled H_2O mixed overnight and autoclaved. Stored at room temperature.
EDTA	Ethylenediaminetetra-acetic acid in sterile distilled H_20. Stock solution of 0.5 M, pH 8.0. Autoclaved and stored at room temperature.
Ethidium bromide	8 mg/ml in sterile distilled H_2O. Stored at room temperature.
Freezing solution	5.85g NaCl, 6.8g KH_2PO_4, 300g Glycerol, 5.6 ml 1 M NaOH in 1 L dH_2O. Autoclaved and 3ml of 0.1 M $MgSO_4$ added.
L-broth	1% bacto tryptone (Difco), 0.5% yeast extract (Difco), 0.5% NaCl in sterile distilled H_2O, pH 7.0 NaOH. Autoclaved and stored at room temperature.
LB-agar	L-broth + 15 g/L bacto agar (Difco). Autoclaved and stored at room temperature.
Loading buffer (X5)	2 ml Ficoll-EDTA (100mM EDTA) pH 7.5 of 0.5 M stock, 22g 22% Ficoll, 5g 0.005% Bromophenol Blue.

M9 buffer	3% KH_2PO_4, 6% Na_2HPO_4, 5% NaCl, 10 mM $MgSO_4$. 10 X stock autoclaved and stored at room temperature.
Mounting media	50% glycerol, 0.5 X PBS, 2.5% DABCO. 100 ng/ml of 4', 6-diamidino-2-phenylindole (DAPI) was added for nuclei staining if, and when, required.
NGM-agar	0.3% NaCl, 1.7% agar (Difco), 0.25% peptone (Difco), 0.0003% cholesterol (1 ml/L of 5 mg/ml stock in ethanol), in sterile distilled H_2O. Autoclaved and 1 ml/L 1M $CaCl_2$, 1 ml/L 1M $MgSO_4$ and 25 ml/L KPO_4 (pH 6.0) added.
PBS	Phosphate Buffered Saline: 7.31 g NaCl, 2.36 g Na_2HPO_4, 1.31 g Na_2HPO_4 2H2O in 1 L, pH 7.2. Autoclaved and stored at room temperature.
PBST	1 x PBS + 0.2% Tween-20 (Sigma).
Proteinase K	20 mg/ml proteinase K (Roche) in sterile distilled H_2O. Stored at -20^{o}C.
Recovery Buffer	2% glucose (Sigma) in 1 X M9.
Ruvkun Buffer	0.16M KCl, 0.04M NaCl, 0.02M Na_2EGTA, 0.01M spermidine-3 HCL, 0.03M NaPipes in 50 % methanol.
Single worm lysis buffer	50 mM KCl, 10 mM Tris (pH 8.3), 2.5 mM $MgCl_2$, 0.45% Nonidet P-40, 0.45% Tween 20, 0.01% (w/v) gelatine and proteinase K at 200 µg/ml (Williams, 1995).
SDS	Sodium dodecyl sulphate (Sigma) in H_2O. Stock solution 10% stored at room temperature.
TAE	50 X stock, 2 M Tris Base, 5.7% Acetic acid, 50 mM EDTA. pH 8.0, in sterile distilled H_2O. Stored at room temperature.
Tris-Triton Buffer	0.1M Tris-HCl (pH 7.4), 1 % Triton-X-100, 0.001M EDTA

Tween-20	Polyoxyethylenesorbitan monolaurate (Sigma).
Worm lysis buffer	50 mM Tris-HCl (pH 8.0), 100 mM NaCl, 50 mM EDTA, 1% SDS, 30 mM 2-mercaptoethanol, 100 µg/ml Proteinase K (Roche).

8.2 Primer sequences

Sequencing primers for vector backbone:

M13uni-21	5’ TGT AAA ACG ACG GCC AGT 3’
M13rev-29	5’ CAG GAA ACA GCT ATG ACC 3’

Hc-isotype-1 5’

Hc-iso-1F2	5’ TAA TCA ATT CTG AAA GTT AGC TG 3’
Hc-iso-1R2	5’ TCA TCG TAC CCT ATT TAT AAT TC 3’

Hc-isotype-2 5’

haem-1009d15.qF1	5’ AGC TCT CGT GTT GGT AGT ACA TC 3’
Hc-Iso2ATGR1	5’ TCT GCT GAT GGA CTG GTG TGA TT 3’

Hc-isotype-3 5’

Hc-Iso3HindIIIF2	5’ TAC ACT TCC ATG AGA TTA CCA CG 3’
Hc-Iso3BamHIR1	5’ GGT GTC ACA CAC ACA CAA CTG GGA CAG GCC GTC T 3’

Transcriptional promoter-reporter fusion constructs:

Ce-Ben-1A	5’ TCT CTG ATT CAA ACA TCG TAG TG 3’
Ce-Ben-1A*	5’ AAC ATG CGA TAG TGC TAT TTC TC 3’
Ce-Ben-1B	5’ CCT TTG GGT CCT TTG GCC AAT CCC TTT TTC CAA GTT CTT TTT GAA 3’
Ce-Mec-7A	5’ AGT AAT CTA GAA ATG TAA ACC TG 3’
Ce-Mec-7A*	5’ ATT TCT GTG TAT TTC AGC ACC GC 3’
Ce-Mec-7B	5’ CCT TTG GGT CCT TTG GCC AAT CCC GTT GCT TGA AAT TTG GAC CCG AC 3’
Hc-Iso-1A	5’ TCA GAC AAA ATC GCG AAG TAA AG 3’
Hc-Iso-1A*	5’ GTA AAA AGG TAG GAA GGA GTA TG 3’
Hc-Iso-1B	5’ CCT TTG GGT CCT TTG GCC AAT CCC GAT TGT AGT CGA GGA GAG 3’
Hc-Iso-2A	5’ AGC TCT CGT GTT GGT AGT ACA TC 3’

Hc-Iso-2A*	5’ AGA ACT TTC CTC TGA CAA TCA AG 3’
Hc-Iso-2B	5’ CCT TTG GGT CCT TTG GCC AAT CCC TCT GCT GAT GGA CTG GTG TGA TT 3’
Hc-Iso-3A	5’ TAC ACT TCC ATG AGA TTA CCA CG 3’
Hc-Iso-3A*	5’ AGG GCA GAG ATT AGG TAA GCT CC 3’
Hc-Iso-3B	5’ CCT TTG GGT CCT TTG GCC AAT CCC GGT GTC ACA CAC ACA CAA CTG GGA CAG GCC GTC T 3’
Ppd95.67C	5’ GGG ATT GGC CAA AGG ACC CAA AGG 3’
Ppd95.67D	5’ AAG GGC CCG TAC GGC CGA CTA GTA GG 3’
Ppd95.67D*	5’ GGA AAC AGT TAT GTT TGG TAT ATT GGG 3’

Genomic loci sequencing:

haem-206f16.qkaF2	5’ TTC TCT TTC TTG ATC ACC CAA GG 3’
Hc-Iso2locusseqR4	5’ TTA CTC CTC TGC ATA CGT GTC G 3’

cDNA sequencing:

Hc-Iso-322bp5’ATG	5’ CCA GTT GTG TGT GTG TGA CAC C 3’
LM-Hc-Iso3cDNAR1	5’ TCA CTC AGC TTC GAA GG 3’

Hc-Iso4cDNAF2	5’ CAA TGC AAC CTT ATC TGT TCA TC 3’
Hc-Iso4cDNA3'R1	5’ TTA CTC CTC GAC GTC TTG GTC G 3’

PCR+ primers

Hc-iso-1-2kbF	5’ ACT GGT CGA CTC AGA CAA AAT CGC GAA GTA AAG 3’
Hc-iso-1-2kbR	5’ ACT GGG ATC CGA TTG TAG TCG AGG AGA GTT CTA C 3’

8.3 DNA sequences

Hc-isotype-1 gDNA (X67489)

atttctcatccaaacttccgcgagagcccatttcattctaatggagaacattccggtatttctaatgaagttggagggatt
tatcgaggctcgttagagagaaatgccctttgcaccagaatactttatcgggaacctgacattgcgaaatcctaagttc
tactttaaatccagttcccttatgcagcacgatgagttttggtagttttgatgtaaatatgaatacatggtggggtcactc
actactagaatgactctattgttacctaatattggtagttttagcacaattctcagctcagacgaaagctttctattgcca
agcctccgtgacccagttgaaatccgcgccccacggccaacttttagatttgcaagcacgccgtctactgtattttctgg
ttcgatgtcccaaatttcgcggcgtacgtattcactctttgctcgccgtgattgcgtgtggttggttgtgagccatcgtcgt
tatttctagctacagattgtaaattgtagaactctcctcgactacaatcatgcgtgaaatcgttcatgtgcaagccggtc
aatgcggcaaccagatcggatcaaaggtaagacttttttgagtgtagattctttaggcaaatattgggctgctcagtat
ctaactggctccaatcaataggtttcgacttccgattcctttcaagcgagagtactggattgggctaatagtgggcagcg
atgagtcatcgtagtattcgttgcagcgctgcactggcgccacatcaatgaacgccgactatataaatttcgtagacga
aaaactattacggagcgatggaatccaaaaagtaatatgggagcgatttcgaaatgcgatatgccgtaatcgttctccc
agttgaccgtaatgtttacgcgtagaagtatttaggtcatcatattggtactggaatcaatactttacggtggaaacaac
tgctatttgtgaagtggatttcgaaagataatggcaaaagaaccctgctctgagtctgagtgactcatgggaataagcct
tgttaattgcgggtgtcactagtggagaatctgtgaaaggctgaggggggggggggtaacaagtgctgcattgtcttacc
agaaggctgctacagtcgttagcagagaattccattggtttctctgcctcttgagtatcaggagtgtggtcaacaccgag
gggcactttgagatcacggtgccatcactttactactccctacagaaaatagctctgcatcaagtaactgaaggaactg
tcgaagaccagaactatttggtcagccaacaacattggcatatttgcagttctgggaagtgatctctgatgagcacggta
tccagcccgatggaacatacaaaggagaatcagatctgcaattagaaaggatcaatgtgtactacaatgaagcacatg
gtaagcttgattcgcattccttgtttatgtttttaatgtttaagtgcagtagtatcttgctttttgattagcactgaactac
ctgcacacgttacgtggctatacagttatatcgttatatgaaagcatggtcattccgaaaaatctacgaaaacggcaag
attctatagttcgccgcgttccatcacgttgtctgcagtcaaggagcacgccgttactactattgccgaaacctgtgctg
gaaatcgaaaatatgaatttcgtctagacgtaggcgtgcatgtctggtttttcgattggtgagcaagcaaatatgagcc
ataaaccgtgctgaggtgaagcgacgctcgttgcctatatggcaataagtagtgctgtgtattatttgtttcaaggcaat
gcgttccacattcaattttatgcggcctcggataggtagagggacattgtgtgtcagattaatgtgcaaaggacagatcg
ccgtatatatacacgtggtaacttgaatgggtagtgacggctgtactaaaaactttgtggtaattgaacttaaactttcg
agagtggcgtggtaagtggtggtttctggcattcgtttgatcacccatgtttctttattggaaatgtatgtgcggtcgttct
tcaggaggcaagtatgttccacgtgctgttcttgttgatctcgagcctggaacgatggactccgttcgttctggaccgtat
ggacagcttttccgtccagataattacgtgtttggccagtcaggagcgggtaacaattgggcgaagggccactatactg
agggagccgagctagttgataacgtattagacgttgtccgcaaagaagctgaaggttgtgattgccttcaggtactgact
tcatcaacaatttacagcttcaactttgatgtgtgaatacatttcaattcgtgctcagggcttccaattgacgcattcact
tggaggaggcactggatctggaatgggcactttgtaaatttcaaaaattcgtgaagagtaccctgatagaattatggctt

cgttctccgttgttccatcacccaaggtgagatcgtgttaatctttgctttttttcctaaattgtgtatttgaattacttatcc
tcatgaagatccaagttgaaataagtctcaccacctgtaaacatgtgaaaggaagatgttttaaggtatccgacactgt
cgtagaaccctacaatgctacccctttccgtccatcaactggtagagaacaccgatgaaacatactgtattgacaacgaa
gctctgtatgatatctgcttccgcactttgaaactcacaaatccaacctatggagatctcaaccaccttggtaattgttat
tacacttactaaagtatacttagacctttttttcatgctgaaaatgtgcaattgaagtgtctgtcacaatgtctggtgtcac
gacctgccttcgattccctggacagctgaatgctgatcttcgcaagttagccgtgaacatggttccattccctcgtcttca
cttcttcatgcccggttttgctccactgtctgcaaagggtgctcaagcatatcgcgcttcgacagttgctgagcttacaca
gcaagtacgcccattactctttatagcatagtgattcaaaattacagccacccatattcttagatgttcgatgcaaagaa
catgatggctgcctgtgatcctcgccatggacgttatcttacggtcgctgctatgttccgtggtcgtatgagcatgcgagt
gagtatttttaaacgtgtttttttcatcaagtaaatgatgaatactggtggagctggaaatgagtcagaagtagtgttctat
gtgagctattcacaaaaacttgtatctcactgtgggccaagacttctacatggcaagccagatgttttataaatacgcgc
tatgagtagcaaccatatatagtttctataagatcggctctaagattattgaaatgttaccggtataactagttactggtt
acttatgcccatgaaggagtagtaattgatgcagggatgtcatctgctataattctaatcgcaataatattgtagctttgg
agagacctgcagcattcctcgctaccgtagcccgcatagatttttttaagcgaaacgggtaggctattgtggtttttgggg
gatttttctgaaatgctgcagtcaattgtaaaaccgcgatgtgattgtgtgatcggttcattgcagtatttttgatcgtaga
gataagaaaatatcatgtggaaggtcaactgcttcatttcaggaagtagatgatcagatgatgtccgtgcagaacaaga
actcatcatatttcgtggaatggattccaaacaacgttaagactgccgtttgtgacattcctcctcgtggactgaaaatg
gcggctaccttcgtgggtaactcgactgctatccaggagctgttcaagcgtatttcggagcaattcactggtttgttttga
tttatgagctgtagtacatgagtgcgtcttgattccaacactcttcaccgcttcagccatgttccgacgcaaagctttcct
tcattggtacactggtgagggtatggacgaaatggagttcacagaagctgagtcgaacatgaatgaccttatctccgaa
taccagcagtaccaggtccgtattatcttattccagctatgtcttagcgtgtcgtaagacgtttagtctttaatatttccag
gaagctaccgctgacgatatgggcgatctcgatgcagaaggtggagaagaggcatatcccgaggagtaatgatccaca
aagttgtgctctttttcctgtgtcaatgcgaaatacacattggttgcgttgtgtatggtgttattaaagcttgctaattgat
aaagttaatgaagtttacaagatattcgataacgaggtgatcaatgggttcgaaattaatcgattgatggactcttaaga
aggggaagttcgtggaaaatgacagggtttaggtccttgtttcggtacaagaccatcctcatcattagtctccatgtcat
gcgctgtgtgtagaggttccgggaacgaggtct

Hc-isotype-1 cDNA (M76493)

gaattcccgagaacacaatcatgcgtgaaatcgttcatgtgcaagccggtcaatgcggcaaccagatcggatcaaagt
tctgggaagtgatctctgatgagcacggtatccagcccgatggaacatacaaaggagaatcagatctgcaattagaaa
ggatcaatgtgtactacaatgaagcacatggaggcaagtatgttccacgtgctgttcttgttgatctcgagcctggaacg
atggactcctttcgttctggaccgtatggacagcttttccgtccagataattacgtgtttggccagtcaggagcgggtaac
aattgggcgaagggccactatactgagggagccgagctagttgataacgtattagacgttgtccgcaaagaagctgaa
ggttgtgattgccttcagggcttccaattgacgcattcacttggaggaggcactggatctggaatgggcactttgttaatt
tcaaaaattcgtgaagagtaccctgatagaattatggcttcgttctccgttgttccatcacccaaggtatccgacactgt

cgtagaaccctacaatgctacccttccgtccatcaactggtagagaacaccgatgaaacattctgtattgacaacgaa
gctctgtatgatatctgcttccgcactttgaaactcacaaatccaacctatggagatctcaaccaccttgtgtctgtcac
aatgtctggtgtcacgacctgccttcgattccctggacagctgaatgctgatcttcgcaagttagccgtgaacatggttc
cattccctcgtcttcacttcttcatgcccggttttgctccactgtctgcaaagggtgctcaagcatatcgcgcttcgacag
ttgctgagcttacacagcaaatgttcgatgcaaagaacatgatggctgcctgtgatcctcgccatggacgttatcttacg
gtcgctgctatgttccgtggtcgtatgagcatgcgagaagtagatgatcagatgatgtccgtgcagaacaagaactcat
catatttcgtggaatggattccaaacaacgttaagactgccgtttgtgacattcctcctcgtggactgaaaatggcggct
accttcattggtaactcgactgctatccaggagctgttcaagcgtatttcggagcaattcactgccatgttccgacgcaa
agctttccttcattggtacactggtgagggtatggacgaaatggagttcacagaagctgagtcgaacatgaatgacctt
atctccgaataccagcagtaccaggaagctaccgctgacgatatgggcgatctcgatgcagaaggtggagaagaggca
tatcccgaggagtaatgatccacaaagttgtgctcttttcctgtgtcaatgcgaaatacacattggttgcgttgtgtatg
gtgttattaaagcttgctaattgataaagttaatgaagtttaaaaaaaaaaaaaaaagaa

Translation of cDNA (M76493)

MREIVHVQAGQCGNQIGSKFWEVISDEHGIQPDGTYKGESDLQLERINVYYNEAHGGKYVPRAVLV
DLEPGTMDSFRSGPYGQLFRPDNYVFGQSGAGNNWAKGHYTEGAELVDNVLDVVRKEAEGCDCL
QGFQLTHSLGGGTGSGMGTLLISKIREEYPDRIMASFSVVPSPKVSDTVVEPYNATLSVHQLVENTD
ETFCIDNEALYDICFRTLKLTNPTYGDLNHLVSVTMSGVTTCLRFPGQLNADLRKLAVNMVPFPRLH
FFMPGFAPLSAKGAQAYRASTVAELTQQMFDAKNMMAACDPRHGRYLTVAAMFRGRMSMREVDD
QMMSVQNKNSSYFVEWIPNNVKTAVCDIPPRGLKMAATFIGNSTAIQELFKRISEQFTAMFRRKAFL
HWYTGEGMDEMEFTEAESNMNDLISEYQQYQEATADDMGDLDAEGGEEAYPEE

***Hc-isotype-2* gDNA**

atgcgtgagatcgtccacgttcaggccggacagtgcggaaaccagatcggagccaaggtgagtggagatggggacctt
tgtgtagcatgtttgactagcgataatttaatttatgtcatttcgttccactacatgaatgatcctgtcctatatgaggaac
cgctctgttgaagaatttaccctggcacattcaaaaaaacacccatgggacagcatcaatcgtctcgattctctttcttg
atcacccaaggattttacgagtccattcaaggtgtgccctaaaattgcatgacctacggggcttcagatgggtcttgtcg
tcatgtcttttgaagctcgaaagcaccctcaggagctcttcaggaaccttgcatttttgcagtctgcccacggtaaccac
taattgatctacctgggaattttcgttgggggtcccggaatcgtaagagaacttactcaaccgactgctaggaaactttc
aatgtagagtaaaagtacatcccgtccattgatgagtggggtgtggaacggtagagaaatcgtagaactgctgtaattc

actttttatatgcctgaaacggttgtggaaagtagcctagtcactaggcactagttatgttgattacgatctgttcgggaaa
aaacctttggcagattgattttgagggtgttagatttgttatggatttctgggaatttttcgaatttcatcttctgaagaga
ttttcgcgaatttctgggtttatgtgtcagacgttcgggtaaattctgagctccaattcatggccacttcctgattgtgctc
gatttgccagcgctccccgtcataccttggtgcccgttcaatggttaccctgtttttatcgtctattcacaagtggtgattt
gaaggttaagtgtaattacggtggaacgaggtgttttagatgaaaattggaagcaccagaaacgattcccaaaattcct
ggtctttctttttctgagctagaaggacgtctgggcaacatgagttcaatcccaagcctttcattagcctccgttctcgga
tttcggtcaaaaagtgccctcgaggtcaccgcggtagacccaggtcgtcgtgtgcgcggtgactgaccccgtaagcaaa
aattgtcataataaaggcgacgtccgatggctcctaacccctacccactttactcgtaatgatgagggtggaaaactca
tgtgagtgcttctcctaaaccatttaacttttttaactcgaaaaaagtggaacaattattacccctttattgcccatttttcag
ttctgggaggttatctctgacgaacatggcatccaacccgatggatcatacaaaggagagagcgatctacaactagaa
cgaatcaatgtctactacaatgaggctaatggtaacatttagaagaagttcgttgtacccaaaatcgttcttgaaacaa
ccattttaggaggaaagtatgttccacgagctgtacttgttgatctggaacccggcacgatggactctgttcgctccgga
ccttttggtgctctatttcgcccggacaacttcgtctttgggcaatccggagctagtaataactgggcaaaaggtcacta
cacagaaggagctgaacttgttgatagcgtccttgatgttgttcgcaaagaggcagaagggtgcgattgtcttcaggtca
gtgtatttgaacagtattgacaattttagggagtgaaatgaattttgttattctatggttataatttttagggtttccaactc
acgcactcgcttggaggaggcacaggatcaggaatgggaactctgttgattgctaagatccgtgaagaatatcctgata
gaataatgtcatccttctcagtggtaccttcaccaaaggtcagtgttcgacagttccatgtcgttcgtactggtttagcttt
gaagcagatgaatccctcacttttacatcccccagcagtcttctggcgctaggggcacgaacccagagagggagaggt
agccttcgtgataacgctccctaagcaggaatagtaagatagatagcgccctactgcctacccttttaagaacagatga
ttcagcggtcagtttagtgctcgattccctcatcccacctttttttcctcgggatgatcactcatttggtgctcctccaaatg
agtgatcatcctaagccttagtattcgcttgatcaaaggtttaactgtattcccctagctaaagccgtttttttgctgttccc
aagcccaatcacaagtctgcagagctggctacttctccctctctgcaccttgacgtctggtgtcagagggtatgcttacc
caaactacatttttacctgaatttgcttagtattttcttctatgcaattgcttatccacaacggatttttagacaatttacc
tgtcatattttatggcaacggttccagcattctcaaagcatatgttcaggtttccgatactgtcgtggaaccttacaatgc
caccctatctgtccatcaactggtggaaaatactgacgaaacgttctgtattgacaacgaagcactctacgacatttgct
tccgtacactgaaacttactaacccaacttatggagatctcaaccatctgggtacgaaaattcgtttgtattgtaatattt
taccctgcgagtttgggtaattgaaaaatgagtggcgtttttgttggcgtgcatattcgatacgcgagtgctggaagtgag
tcatggctggtaggatgggtattgcgtaatgtccgttgcaagcaatgcaacgaataggctgagggtgaagctgtcgtga
cttgcctttgggcagcaccattttctgcactcgtgcaagggaatttgcgtgggtgatgcatgcgtagagaacacgttcgg
tcaatattcgactctttggcaatgatggacccaattagttgggttgccagatcattcaacatctaggcgagatatgtggg
ttccctgttcccactgtttctagcgcataaagtccgtagcgattatacggggagggacttttgaatttagtccgaattatc
cgatcaatgtgtcgcgtgagcctgactggccttctcaatttaatgggacaatcgagggaataggttgccgaagatagat
gcgtccagtgttagcgaatcgattcgtgtccgccttttatgttggttggcattatcctcaggaagagacagccattgcgt
ctgaccagatgaaattgtcattgcccttggattgacgttatactaagccaatatagtggagacgaaacattaaggaata
ctgcaaggttgcaaagttaatgtgcgaagtgaacgtacgcttgaactgccagatcttttcggcttatccccgagtagtat

ccactgtttgggaaatgcgtgtgggctcgtaagcagccaattgttgtaataaagtcgttgaacaagtggatcaatgaaa
gtaaatttctagtggagtggagctttttcgatagagaatgtgccatatgtctttgtatttaagtatccgtcactatgtctgg
agtaacaacatgcttgcgattccctggacaactaaatgctgatctacgtaaactggctgtcaacatggtgcccttccctc
ggctacatttcttcatgcctggattcgcaccactatcagcaaagggagctcaagcctaccgagcactaactgtttctgag
ctcacgcagcaggtgagatagttcacaaaatgtcctgattcgtttataaaacacttttttgccttagatgttcgacgcaa
agaacatgatggccgcttgtgatccacgacacgggcggtaccttactgtcgcggcaatgttccgtggtcgtatgagcat
gcgtgtgagttatatccctttcccttcctcaaaaccaatgaggtgattagagcttcttttgagcgctccatgtgtcttca
tagatctaatactagagttgtgccccgttttaaagggagttcggtttgatgggtcagattgcaagtgctaatgctcgaagt
taggttggcatgttaggatccttcgtaatcaatctgttaaggcacagttcatgataatttggtaatccagattatgatgaa
atggaagacaagatcgaaaatgccgtaaggggctggacgagaatgagaccagaaaaatgaagaaaaagtgggcgtt
cagacgagctatacaatgcatctcattgcgcgcgtgtgggtcctggtgcgaataacagataacggtaagcgaaaattca
ttcacttataagcataagtaaacaaaaaaccctgataacagcaaaagtagctcacctttcagcctctaggagtagga
atttaaataagggattgcggcagcaaaaagaaaaacagaggaaaaaagaatgggagaaataataatcatcgcctcat
cactcggagacagctgattgtgtgtgggtcctaaaacgaatagttgcaatctgttgtctcgacactcgtctgacgtgatg
acgtgacatcgatcgagaggcgggaaagccattgacgtccccgcccattgaattatcgtgaactgtgcctttaaagcga
agtatcccgacgtggttaacggatatttaatagttcttgtatgtcaataatgccgccgtttgcgacccctataatatttct
ctggtacttttgggtttctcaataaaagatctataatttaggaggttgatgaccagatgatgtctgtgcaaaataagaatt
cgtcctactttgtcgaatggattcccaacaatgtaaaaacagcagtatgcgatattccaccacgcggtttgaaaatggct
gccacctttgttggaaactcgacagcaatccaagaactgttcaagcgtatatcagagcagttcactggtatgtttcttttt
ccttcagcagtttgtggtgagccttcagctatcgtagcacttttttcctttagctatgttccgtcgtaaagccttcttacatt
ggtacactggcgaaggtatggacgaaatggaattcactgaagccgaatcgaatatgaatgatctggtatccgagtacca
acaatatcaagaggcaactgctgatgatgagggtgaaatggaaggtgccgttgaaaacgacacgtatgcagaggagta
a

Hc-isotype-2 cDNA (M76491)

gaattccgcagaatgcgtgagatcgtccacgttcaggccggacagtgcggaaaccagatcggagccaagttctgggag
gttatctctgacgagcatggcatccaacccgatggatcatacaaaggagagagcgatctacaactagaacgaatcaat
gtttactacaatgaggctaatggaggaaagtatgttccacgggcggtacttgttgatctggaacccggcacgatggact
cagttcgctccggacctttggtgctctattccgcccggacaactttgtcttcgggcaatctggagccggtaataactggg
caaaaggtcactacacagaaggagctgaacttgttgatagtgtccttgatgttgttcgcaaagaggcagaagggtgcga
ttgtcttcagggtttccaactcacgcactcgcttggaggaggcacaggatcaggaatgggaactctgttgattgctaaga
tccgtgaagaatatcctgatagaattatgtcatccttctcagtggtaccttcaccaaaggtttccgatactgtcgtggaa
ccttacaatgccaccctatctgtccatcaactggtggaaaatactgacgaaacgttctgtattgataacgaagcactct
acgacatttgcttccgtacactgaaacttactaacccaacttatggagatctcaaccatctggtatccgtgactatgtct
ggagtaacaacatgcttgcgattccctggacaactaaatgctgatctacgtaaactggctgtcaacatggtgcccttcc

ctcggctacatttcttcatgcctggattcgcaccactatcagcaaagggagctcaagcctaccgagcactaactgtttct
gagctcacgcagcagatgttcgacgcaaataacatgatggccgcttgtgatccacgacacgggcgctatcttactgttg
cggcaatgttccgtggtcgtatgagcatgcgtgaggttgatgaccagatgatgtctgtgcaaaataagaattcgtcctac
tttgtcgaatggattcccaacaatgtaaaaacagcagtatgcgatattccaccacgcggtttgaaaatggctgccacctt
tgttggaaactcgacagcaatccaagaactgttcaagcgtatatcagagcagtttactgctatgttccgtcgtaaagcct
tcttacattggtacactggccaaggtatggacgaaatggaattcactgaagccgaatcgaatatgaatgatctggtatct
gagtaccaacaatatcaagaggcaactgctgatgatgagggtgaaatggaaggtgccgttgaaaacgacacgtatgca
gaggagtaaaggaagcatctgcccaactcatttatgcatctcattcttgcggctattgtattgcgcgtactg

Translation of cDNA (M76491)

MREIVHVQAGQCGNQIGAKFWEVISDEHGIQPDGSYKGESDLQLERINVYYNEANGGKYVPRAVLV
DLEPGTMDSVRSGPFGALFRPDNFVFGQSGAGNNWAKGHYTEGAELVDSVLDVVRKEAEGCDCL
QGFQLTHSLGGGTGSGMGTLLIAKIREEYPDRIMSSFSVVPSPKVSDTVVEPYNATLSVHQLVENTD
ETFCIDNEALYDICFRTLKLTNPTYGDLNHLVSVTMSGVTTCLRFPGQLNADLRKLAVNMVPFPRLH
FFMPGFAPLSAKGAQAYRALTVSELTQQMFDANNMMAACDPRHGRYLTVAAMFRGRMSMREVDD
QMMSVQNKNSSYFVEWIPNNVKTAVCDIPPRGLKMAATFVGNSTAIQELFKRISEQFTAMFRRKAFL
HWYTGQGMDEMEFTEAESNMNDLVSEYQQYQEATADDEGEMEGAVENDTYAEE

***Hc-isotype-3* gDNA**

attaccgtactcccgccaatcgataggacgcggaatttttccttgactggcttgctttttcattggattttatcgttttttct
tctgtttctttttgagttttcatgtttattagacttgattaattcaaatgactcctcctatataaacgacgatcagaatcgca
ttggaactgaagagagcggccggctaattcaaatgtttaaaggactccccaccacacaaattgcgcgctccaatcttcc
atcgcaatcctgcgataggcggcgggtgcggcgtgtgcgacgccgtgttaggggagaagttcccggaccccagtaatc
agacagaggatatcttctcaactactgtaccactgaatacggaacttggacattcagtgaattcctcccttcttccagaa
aattctttaatcatacctcctttgcaatggttcgaagttgtttttttatttttttatttttttgtatttgtgcactttactatgaatg
aggctattgcacattcacagaatctatgaggggaaagggcgacattgtcgttaggcctgtccaggccatatgtccatca
ggctcctccctcaggaaagactgaacgtcgaccagagtcagatggccatagcatgtatagtggtgagtgagcagagtg
attcagactcccatggtgggactcatcttcaccatcgtcgaagggagtgagctcatagcagacacgcagacatacccc
ccattcgggactcaaaaattcatggacatggcccagtgacgacaagagacggcctgtcccagttgtgtgtgtgtgacac
catgcgcgagatcgttcacatccaagccggtcagtgtggtaaccagatcggatccaaggtaagtgaagtcgtattgtcg

caattcacttgttattctctttctgatggtgaaaaagggtgtaggaaacaaaattaccaccttatttgagaggtcacgaa
ttttctgacccaccagttatggggatcgcgaagttatcagcctgcagatggccgtttccaagctacggaattcctgatct
ggcctgagtttgcaggcttcttattcaacgcatcacctgaatcccacagggtcagaatctgtaaatcctggtgaatagaa
cagtcgggtccactaggtggggtaccctcagcaaaaacaaaagaggcccttctgaaactgccgcatgggtctgcacat
cttggattctgggagcgtttcggctctcaacaaagaagtagccaaccactcgccttcaacctttcggatacccgatcct
ccactgtatgacccaagtaagaccccttaaaaactctcaaacccctgagttgcttttcgacacccttactttcgatcatt
ggaaaagcttcaggggtccgtggggcttaagactcattttagctcgcccattgattgacgccattaacttgagtgataac
tctcattgagacttttctcgagggaattcattcaattaatgtggttgcaccactctttcgaagattggtaacagcaaagaa
aaccatgacattgacgtcgtttcctggacgctctcaggaaatacgactacctcgattcaactggtaatctcagggcttaa
actctggcttacccgaatgcgtaatccagaccacagggcatcggcgtctcggtgtttttggtgaagccagtaattagct
atatgaatggcatttcgtacatctgagcccattttcttaagaaacatggataccatgggaaatccttcatacattaccta
ttcatgcaacaacttacgtatttggtaagttgccaagaagccttggggtgcttaagttgttaaatatgttattacttatcat
ttacatctccagcaatctggtaaaagtagggctacttccacagtcgtggtccgagaatagaacctcgcctagagcagct
gagcctcgtctcgtccttgttgatgccggacatttttagcaaatatagaatcatgggccctgtcatgggcatctaccacc
caccgcatggtcctgaaatggaatcccacataaatcaactgaggcttcgcccttgttcgtcatgaggatctcccaggcc
tctttggcagatgttgtctttgtcataggcatagccatcatagcatgaactatgtacatatgtaggtctgcaggcgttcaa
atgcgactaccattctgaacaggaatcggatgcacggcttactctcaggagaggccctacatagtaaagttctctatct
actatcttactactccttcttctgagccaatttttgggattttgaagggtttttcaaattttgagctcccatgacctgttctt
ttggtgtaggggtggcaacgatagccataccaagtagtctgacttgtaggttttcttaattcatcatatcctgtcagggct
gataaaggaggtagcttgacccagacaaaaaggtggttttaactggtaaaccaggttctagtacagattccaatgttttc
cacgttttcaaagagaggtcggcggtgtagaaccgagtgataattagcttcacaggctttcatttttttttgtccgagttga
tcttccggccaattcacctttcgaggggctcgactaattcgataatgaaaaagtaatgggcccagtcaaagggacttcc
tttttaccatgttgagctgctctccaagtcatgccatgtcgtctagaagggatattactatgacattactactactccaac
ttaaactgctgttgccaaagtaaagttgaaaatccttccgtaattctgaaatttcagttctgggaagtaatatcggatga
gcatggcattgatcctacaggccagtacgtcggtgattcagatcttcagcttgaacgtatctccgtctactacaacgaag
ccggacacaacaaatatgttcccaggttagaatttacttttactcgaggaatcgaaacacgaaagtcgaggctatgaag
tagaagaatccaaaaataagaatgatgaacctaatgacctagtcgatcctaggtcaagggtgaaaaccgtgcgagttt
cgttttctctttaatcgacttaaatgagcttctgaaattcagagattttgtagtaaactttgtatagagataggaatgcgtc
caggggacctgtgaaattcacatgggattgcccagacccggactttgggtttcactggatttttcaaaccttcgttatact
agccaccacccaggctagaccacgtcagctcaagccttctttgttgaactatctattgaaaaatcagtctatctaggtga
aaaatctgtaaaagactaaaactttgttattggtccgttttttaaaaacatttttaattactggtgtttgaggtgtctgtcg
tgatggcaatctacttatttttcacggatattttaaattcaaatccccttctaaataccgtactcacggcattttcacattt
tttctcatgaaccaatcaaaattttcaaaaataacttttctgaagagtaagtatactttgtatatgcgggaatgtgtgaat
acgatcggactcatttgtagggccagtggtccaaccatctgtttttactcctagactcgtactcaatcccattctcaaatt
tctccaaaattttgatcaaccccattctagaacttctctgaattttccaagaactcaacaaaaaaaagcaaaaatatag

tccgaagttacttgattctagtgtactttctatcagttctccgcgttttactgtatggatttcacacgctgacatttttacaa
acaccatcaaaactcgttttctgagtcatagcaatctattatagttcttcaaacctctaaaatattactattacatatata
ttatatgtacatcttcaaagctcattacataaaacagttccggttacatctggccctgaattcgacaacacgcagctgtc
cctacccagtagccccttgtcggagtagagctggatcccaaacgtatctgacccgaggtgggtgcttaaagctgctgct
gggtaaagtgcgcttttagaaatgaacacgttgttattgattctagagacgaaaataaatggtccttagtgcattggtgt
cgattggttttcaatgaaaaatcacccctcacatcattgacagacagcttgcgttcgtcttctgcacaacccacaacca
ataaataaacaaataaacccaaaaaaactctatctgagcactgacatggtgatcaccttcgttcggaaaataaccaaa
cgttttaaggaattgttggttttctgtctgagccatggcgatccacttatttgagagtttgttaaaagcactatgaagaga
aaggtctccgaatttctggtggtggcattcgttcagaacctgctaccttgcaatttttttgatcatactcaggaatgcatag
aatggaagctcatgaggatcaggcacaagccatgaatgcttaataaagcccgctgcctgtccacggcctagccacccg
gccagtatctcattcaattgcatgtcaaacatgtatggagtaacagctgctcaacaggcctaaatggcaaaccgaaaa
actgtttgctacgagcgcaactcaattagttctcagcacaactagtaggataaagttgagaatgcactagggaacaaga
acgaagtctgaagactgtcctgtttcagggctgtactagtggatctcgaacctggtaccatggactcggtacgcgcagg
accatttggacagttgtttcgtccagataattatgtttttggtaagttgtacaaatgcgaaatgagttgattcttaagttaa
agacacagttaacgataaccctgagatatttagttcgagtttaaataatggaggttagcaaagaacgcaaaaatgggg
aaggaagtacaaaaaaatagtttgaaaagtgtgctatacgagattatgtgtctgtaattgtctgggtgcctcccatataa
gcgtactattcctcggggccttaatgctttactgtgaatggtgcaggctagcaaggccgtcataacacgacctttttttgg
gtgagttcatcggcgaagccggtgttgaagcccttatacgacagtgctatccttgggagccacaggatgtcttgaagga
ggaatttcagaagtcacttttgtcgtgaactgtgccttcaaacggtttactcggggctgcagacagctgctcctcactag
aaaaaagctctactgtaacagtaagcactgtttattctaggccaatcaggcgcgggaaacaactgggcaaagggtcac
tacacagaaggtgcagaacttgtcgacagtgttctcgatgtggtcaggaaagaagctgagggttgtgattgccttcaggt
tggtacttgtccaaaaacgatcaaaatttgtgcgactatgctttaggagttattcgaaaagtacttttcagggctttcaaa
tgacgcactctcttggtggtggcaccggttcgggtatgggtactttgttgatctcaaaaattcgtgaagaatatcctgatc
gaattatgaacacgttttccgtggttcctagcccgaaagtctctgatacggtaattaatcatctacggataaacctctga
gatactttgttctgatgcattcgagctttcgagttgtctacgtgctttttatttttccaggtagtcgaaccctacaacgccac
cctctcagttcaccagcttgtcgagaacaccgacgagaccttctgcatcgacaacgaggtattcacttcctgaagccga
aaaatatagaagcgcggtgaccctggctctattctcattggctgtttaatgagctacgctaggagtccaccaataagga
gggcgctaccagctccacccatttgcagagcgcccagccattatgcacaacttttttttgtgaatttctctgcagggagag
ttaaattatgttgtttataaactatatagcaaagaaccatctttccaggctttgtatgatatttgcttccgcacgctgaag
cttactaccccaacctatggtgatctgaaccatttggtatcagccaccatgagtggtgtcacgacatgtcttcgcttccct
ggacaggtaatacggaacaaacttactgattcccacgaaaactttttgattctcatgattggataactctcaaggttact
ggcttgattgttcaagtttccagagttcacagtaattttacagcttaatgccgacctgcgtaaactagccgtcaatatggt
gccgttcccacgtttgcatttcttcatgcccggtttcgcaccacttacgtcaagaagcaatcagcaggtacgttgcgtca
acataggctgaaagggtgcactgaaagaatgtaggcggagtcaatgcggccatcaaccaatgagacgagctcgagca
cccagtcctatttgtgcgtggtccgtcagccccaccctccagattgcagcgagttctgctcatttttttttaattactggat

ataatcacgttaagtgaagcacgatcgcaggcaatttgggaagatgggaagttccccttagtagcttgcagtagttaac
tcctactttgagttgaaagttttttcatccgatacctgtttcttcattaatgattaacaaaattgaatggtattccgagtttt
ctagtatcgcgccgtatcggtatccgaattgacacaacagtgtttcgatgccaagaatatgatggcagcatgcgatcca
cgtcatggcagatatttgacagcagctgcaattttccgtggccgaatgagtatgaaggtctgttactacaatttaattact
gaatgttactggttaaagaagaactcgttattaatagctaaactggttctctatgagctcatgattgctccagaagtatc
atgtttaacagaatccaaaaatcaaaccaaaaatttggtgttgctcattttgttgacagaccttaacaacaccttcagtg
ttcgaataagccttgattgcgaaaacgccctttgattattagtttttgaggtcctctagcataacactcccttacagttgg
ctttggttccctattagcaactctacacattcgaatcaagccgtatgtagtaagcacacattttcgagtatccagatgac
gtcattctcaggaaccatttaatcaataactaagtcattgtgtatatatatatgaaaaatttgtttcgttatgaatttttttt
tgccggtttttgccggagcgggatcgaagcgtcacgcgaactgcagtagacccgacgaattgctaagctgtcactggaa
ctgaactgtagtagacccgacaaacttctacaactgaagaagtcctcggcttggaaactaccctatttcgagcgtttctt
tgggagaaccctcggaaagcaatctcaaggaattgttgattgtttggctgtgacccatttatttgggagtggacgaatca
agagagattatcgaattgtgctagagtgaaattcattctatccagaattatgaataattgttgaggtcgtttacattgttc
gatgcttgttcaggaagtggacgaacagatgctgaacattcaaaacaaaaattcatcgtatttcgtcgattggatcccg
aacaacgtaaaggtaaagctttgagcgctattacattacttttaccgtctgaacaattcatggttcctatgaggttcaa
gttatgagctggaatgagtagaatctatgtattggtattggtaaacaattttgaaagctccaaactattagggtgtccca
aaagttttgagaaaattttggtcatgtatttcgcaaccatgctgcgcaagaaagcgccgaggcagagccactgagaaa
aggtcagtaagaggtcacgcttagcagatcatcataacaagagaacagtattgcactttatctttggaagtgagttatca
acatttgaattattttcgactgacacgactacaggcgccttctgcttcattactaccgatctggtcactccgccactcaa
gagcggaatgaattgcagtcgtcgatgggcagtactgcaccatccctcgccacatgctataggtggtacggccgcttct
ctagagagagtttcacttcgacgaagtgccacggagcggacgtccacgttccaccaagaccgaaattgtcttagcctcc
gtccaaagcaatccttctcaaagcttcagaggaatggagaggacgacttctgctccccgttccaccacccacgacatcc
tacgcagacgtcgctttcgcgcagcaattcttgcaatcatcccccacaccttgactgagggcgagcgtcaggtgcgggt
cgatgtgtgtcggaagctactcgaccgcaagcgaacggtggattggacgagcttcatcatcgctcaggatgagaagtgg
gttccttatgataatccatacaggaagttgcagtggctaccaatcgtcataaggccagaagccgttgccaagcgtgggg
aacacgtcgaaaagaatacgatctcgttcttcttacgttccgccggctttttctactatgagatcctacctgagggtactg
ctgtgactgtaagggtattctgcgagcagctgcgggagatggtgtctggagccccggtttcacaccgaatgcagggaaa
gtttctgattttgatgagtcacgctcgaccacatcatgctagggcgacggctgaggagctggaaaggcttgggattacct
ggctgccccacccccatattcacctgatcttagtccttgtgactaccatgctttcagaagcctggaggcctacacgaaa
ggaaaaaaattcaacaatcgtcaggaggtcaaacatgaagtggatgaatggatcgcttcacgcccttcttctttctgga
tgaatggcatcagtgctctgctcagtcgatggagggacgtcgtggctaaagagggccactatctcgattactgatctcac
ctttttgttgtagtatttttgtgttgaacaaaccaaataaattttctcaaaacttttgggacaccctaatagttaaggctgg
atagtgttctatagtttggaagtgaaaacggatcaaaaagatggccgccgtggaacgactcgaaattttgtgcttcaaat
agtttagaaccactgaatttttgctcttcgtatttggcaggtgacgatgtagtcacagtactcacgactattaacaaactg
tattaacattccattaatggctacgtttattaaaatcttttcgaagaactagtactgatactgaagaggtaccgatttggt

gctcttagttgagtttcagagcctagggatcatgattgctctaataactccgcccactgcaacgactccgccctagaagt
agatggggatgcttgcaattaatctcaataaccgtgggtacaacagaaactctgacattgaaatggtctttcacagaga
aaggaagttgtgcagaatataaaatgccaataacgagcaaggactcatatacatagagtgtcacgtcttccgatagaa
tctcgtcctaagtgggcggtgaaaatgggagctggcctactaacgatttcctcagatttttaggaaagtttttggaagcc
attaagtggtacttccacacttcatcttgatcgcaacttcgtacctctaggccattgaaaagtgcttcccgctgtttcgaa
gattccgattataaaaaacttcaataccctactgaaccgtaaaagggtaaagtcgtttgaccctgatcaatcccatttga
ggaatgcgccccagcgttcaccttcgaattaactaagctatatttatgaaaaaggagtgaattcactttcagactgccgt
ttgcgatatacctccacgtggcttgaaaatgtcagcgacgttcatcggtaactcgaccgccattcaagaactcttcaag
aggatctcagaacaatttactggttagttcctgtaaaatcaacaacatcgttacaatactcaagtaatagatgagcaat
agacagcgtgaattgcatgtctgtcgtagcctgtgtaagcaggtggtccataatgattcaaattatttttaatttgtctac
ttctgtattttgtaaggggattgcaacccatatattcagttatgagattgagctaaaatttcaaaacgagtgggaccgaa
cccaaacgggtgctttttaagggttgtacttatccattgcacgaacattcaaagacatggcccatttgcaacagtaatgt
ttcatgcatgcgataatcccctcctcgtcgcctccactactcatctattccacgcaaccgtccgaagatcagcacaagt
ccggggacatctagccaaaagacttttcgttcacttcacagcgttacccgttgactacgtcaacccttgggataaaatc
attcttcacggatctaaagcggcacgaatggtccttgtaggtgggaggctgtgtgctccagtcctgctcatatgaccaga
atctacccgcatacccgtggtcctactgtccactggaacgctttggcctatgcgttccgccgttaccgcaatagcgccac
cattatgcccttcggttgtagatcttgttgccttcccagctgcttttcctagctttgcgtccccgtgtgagcgtttcctatag
gttaataccttcgtaactaaggtaggtcagttaactctcaaaatttctctttcacgtttctggctggaaagtgtgttcaaa
tttccctaaccatgcttgccccagataggatttgtcaaggctcccgctgggttgtatctcgagtaaccaacctcccagga
acccccggggtcaagcacccatcgaaacagtggaatccgaaggagacgttgataaagtcgctcagaactgtaacgata
aaatgaatagctcttatattggaatttcaaagtagtttttgggctgagagagcagtaacaggaaagtttccaatgttctg
aatcagatggtttaagtagttccaaactaaatagcttcacatcttaaaactcataagtcgttggagtagtgtggggactc
gaagagattcttggctaaatctcgctgcaccaatatgacggagacgaagccggtcccgattgggttcaaaggggcaac
ctgggcctcagtagttatagaacactctccaaggtggcgattgaggctgtgtactattcatcgagtgtggtacgttagga
acgaaatggtcgcgtcgcggtgagaggggcgtcgcgcgcgaagggaagtggcaggatgcagcagtaggtaagaaata
caagaacaaaagtactaatagtagaattatgcataaaataatgatcgcatatcgattgagctgtatgaagaacgctag
agctgtaataacggcgaatgagtggtccgcgacagccgcgtgcggcgaatcgatcacccaaaatcgttcctaacgccc
gtctcccctattcatcactcttgccctccccaataagttcggtgcaaatgaaccaattttaggggcatcaattactcggtt
ggcctgggtggatcttgggtcagattttgtaaccataccttacgttatccttctttcagtaaacgatccgaaattatgaa
gtattatcgatgaagtcattcaaaataccaaaattgcccatttcggtaattcgatttccggtttctcaattctcaagcgca
taaccatctctctgttgtgattatcattcgattgttattggttttccgacgatatcgagtaaaggctaaatttttgcattttt
tcagcgatgttccgtcgaaaagccttccttcactggtacactggtgaaggtatggacgagatggagtttactgaagctga
atcaaatatgaatgacctggtttctgaatatcaacaatatcaggtattgatttctccatgaaactcagacacatccgatc
gtaatgatagagttatcgctccatctgttttgaatagctttgtgatgagtaagaaacaaaagaaatttctttgaggtgacg
ctatagtaatttcgctcaaactattagaccaggcagatgtgggctgaagatcctcttctgggcaaggtctccccttgaag

ccggtcctgaaacgggtttcctgtcttggacttttgaaaaatatcgcccgagggcaggagtacggtagtaaaccgaatc
ttagactatgaaaatatgcattacttttcatggcccacagtaagcgtgacttcaaattcagagattccaaaatcgctaat
aattatttgcgtaaaaattaccctcagatcaaactattacctcgtactactcgttaatgtgatgatggtctgagttatata
agtttaacttcaatcaattcagtaattacaggaggcggctgccgacgaggaaccagccgaaaccttcgaagctgagtg
aaatcaatggagtttatcgatcgaaatgtcatctatttgcttctttcgttgcctagctttctgcttcattgccatgaattat
agggggaataaacgctttgtttaatctattttattacatttcatccaagacacacccgcaaccagtaggcgggatcgcct
gggctattgcgccgctgcgattggtgtatccgttccattaagttatgatgggtaagttcatcgatcaatttcttataaggt
atctccgcccatacaggaattctaacgttcagtaggagtcgactgtacgttcactgactttgtatcccagacaatcccag
gatcgctctttagtatgcgtataagtaacctagagctttagtg

***Hc-isotype-3* cDNA**

ccagttgtgtgtgtgacaccatgcgcgagatcgttcacatccaagccggtcagtgtggtaaccagatcggatccaag
ttctgggaagtaatatcggatgagcatggcattgatcctacaggccagtacgtcggtgattcagatcttcagcttgaacg
tatctccgtctactacaacgaagccggacacaacaaatatgttcccagggctgtactagtggatctcgaacctggtacc
atggactcggtacgcgcaggaccatttggacagttgtttcgtccagataattatgtttttggtcaatcaggcgcgggaaa
caactgggcaaagggtcactacacagaaggtgcagaacttgtcgacagtgttctcgatgtggtcaggaaagaagctga
gggttgtgattgccttcagggctttcaaatgacgcactctcttggtggtggcaccggttcgggtatgggtactttgttgatc
tcaaaaattcgtgaagaatatcctgatcgaattatgaacacgttttccgtggttcctagcccgaaagtctctgatacggt
agtcgaaccctacaacgccaccctctcagttcaccagcttgtcgagaacaccgacgagaccttctgcatcgacaacga
ggctttgtatgatatttgcttccgcacgctgaagcttactaccccaacctatggtgatctgaaccatttggtatcagccac
catgagtggtgtcacgacatgtcttcgcttccctggacagcttaatgccgacctgcgtaaactagccgtcaatatggtgc
cgttcccacgtttgcatttcttcatgcccggtttcgcaccacttacgtcaagaagcaatcagcagtatcgcgccgtatcg
gtatccgaattgacacaacagtgtttcgatgccaagaatatgatggcagcatgcgatccacgtcatggcagatatttga
cagcagctgcaattttccgtggccgaatgagtatgaaggaagtggacgaacagatgctgaacattcaaaacaaaaatt
catcgtatttcgtcgattggatcccgaacaacgtaaagactgccgtttgcgatatacctccacgtggcttgaaaatgtca
gcgacgttcatcggtaactcgaccgccattcaagaactcttcaagaggatctcagaacaatttacagcgatgttccgtc
gaaaagccttccttcactggtacactggtgaaggtatggacgagatggagtttactgaagctgaatcaaatatgaatga
cctggtttctgaatatcaacaatatcaggaggcggctgccgacgaggaaccagccgaaaccttcgaagctgagtga

Translation of cDNA

MREIVHIQAGQCGNQIGSKFWEVISDEHGIDPTGQYVGDSDLQLERISVYYNEAGHNKYVPRAVLVD
LEPGTMDSVRAGPFGQLFRPDNYVFGQSGAGNNWAKGHYTEGAELVDSVLDVVRKEAEGCDCLQ
GFQMTHSLGGGTGSGMGTLLISKIREEYPDRIMNTFSVVPSPKVSDTVVEPYNATLSVHQLVENTDE
TFCIDNEALYDICFRTLKLTTPTYGDLNHLVSATMSGVTTCLRFPGQLNADLRKLAVNMVPFPRLHF

FMPGFAPLTSRSNQQYRAVSVSELTQQCFDAKNMMAACDPRHGRYLTAAAIFRGRMSMKEVDEQM
LNIQNKNSSYFVDWIPNNVKTAVCDIPPRGLKMSATFIGNSTAIQELFKRISEQFTAMFRRKAFLHWY
TGEGMDEMEFTEAESNMNDLVSEYQQYQEAAADEEPAETFEAE

***Hc-isotype-4* partial cDNA**

ttcaatgcaaccttatctgttcatcaactggttgaaaataccgacgagactttctgtatcgacaatgaggctctctatgat
atctgcttccggacactgaaactaacgacgccaacttatggagatctcaatcatttagtgtctatgactatgagcggagt
gacgacgtgcttacgattccctggacagttaaatgctgatttacgtaaattggcagtaaatatggttccatttcctcgtct
acacttttcatgcctggctttgctcctctcacttcacgaggttcccagcagtatcggtcccttaccgtacctgagctgac
ccaacagatgtttgacgcaaaaaatatgatggccgcttgtgatccacggcatggacggtaccttacggtggcggcgatg
ttcagagggcggatgagcatgaaggaagttgacgagcagatgctgaacgtccagaataagaattcgtcctattttgttg
aatggataccgaataacgtcaaaacagccgtatgcgatattccaccacgaggagtgaagatggccgctacattcgtcg
gcaactcgactgccatccaagagctgttcaagagaatatccgaacaattcacggcgatgttccgtcgtaaggcgttcct
tcactggtataccggtgagggtatggatgaaatggaattcactgaagcggaatcgaacatgaacgatctggtgtccgag
tatcaacaatatcaggaggctacagcagacgatgagggagagtttgatgatcgcgaccaagacgtcgaggagtaa

Translation of partial cDNA

NATLSVHQLVENTDETFCIDNEALYDICFRTLKLTTPTYGDLNHLVSMTMSGVTTCLRFPGQLNADL
RKLAVNMVPFPRLHFFMPGFAPLTSRGSQQYRSLTVPELTQQMFDAKNMMAACDPRHGRYLTVAA
MFRGRMSMKEVDEQMLNVQNKNSSYFVEWIPNNVKTAVCDIPPRGVKMAATFVGNSTAIQELFKRI
SEQFTAMFRRKAFLHWYTGEGMDEMEFTEAESNMNDLVSEYQQYQEATADDEGEFDDRDQDVEE

References

Abad, P., Gouzy, J., Aury, J. M., Castagnone-Sereno, P., Danchin, E. G., Deleury, E., Perfus-Barbeoch, L., Anthouard, V., Artiguenave, F., Blok, V. C., Caillaud, M. C., Coutinho, P. M., Dasilva, C., De Luca, F., Deau, F., Esquibet, M., Flutre, T., Goldstone, J. V., Hamamouch, N., Hewezi, T., Jaillon, O., Jubin, C., Leonetti, P., Magliano, M., Maier, T. R., Markov, G. V., McVeigh, P., Pesole, G., Poulain, J., Robinson-Rechavi, M., Sallet, E., Segurens, B., Steinbach, D., Tytgat, T., Ugarte, E., van Ghelder, C., Veronico, P., Baum, T. J., Blaxter, M., Bleve-Zacheo, T., Davis, E. L., Ewbank, J. J., Favery, B., Grenier, E., Henrissat, B., Jones, J. T., Laudet, V., Maule, A. G., Quesneville, H., Rosso, M. N., Schiex, T., Smant, G., Weissenbach, J., & Wincker, P. 2008, "Genome sequence of the metazoan plant-parasitic nematode Meloidogyne incognita", *Nat.Biotechnol.*, vol. 26, no. 8, pp. 909-915.

Aboobaker, A. A. & Blaxter, M. L. 2003, "Use of RNA interference to investigate gene function in the human filarial nematode parasite Brugia malayi", *Mol.Biochem.Parasitol.*, vol. 129, no. 1, pp. 41-51.

Abubucker, S., Zarlenga, D. S., Martin, J., Yin, Y., Wang, Z., McCarter, J. P., Gasbarree, L., Wilson, R. K., & Mitreva, M. 2009, "The transcriptomes of the cattle parasitic nematode Ostertagia ostartagi", *Vet.Parasitol.*, vol. 162, no. 1-2, pp. 89-99.

Adams, M. D., Celniker, S. E., Holt, R. A., Evans, C. A., Gocayne, J. D., Amanatides, P. G., Scherer, S. E., Li, P. W., Hoskins, R. A., Galle, R. F., George, R. A., Lewis, S. E., Richards, S., Ashburner, M., Henderson, S. N., Sutton, G. G., Wortman, J. R., Yandell, M. D., Zhang, Q., Chen, L. X., Brandon, R. C., Rogers, Y. H., Blazej, R. G., Champe, M., Pfeiffer, B. D., Wan, K. H., Doyle, C., Baxter, E. G., Helt, G., Nelson, C. R., Gabor, G. L., Abril, J. F., Agbayani, A., An, H. J., Andrews-Pfannkoch, C., Baldwin, D., Ballew, R. M., Basu, A., Baxendale, J., Bayraktaroglu, L., Beasley, E. M., Beeson, K. Y., Benos, P. V., Berman, B. P., Bhandari, D., Bolshakov, S., Borkova, D., Botchan, M. R., Bouck, J., Brokstein, P., Brottier, P., Burtis, K. C., Busam, D. A., Butler, H., Cadieu, E., Center, A., Chandra, I., Cherry, J. M., Cawley, S., Dahlke, C., Davenport, L. B., Davies, P., de Pablos, B., Delcher, A., Deng, Z., Mays, A. D., Dew, I., Dietz, S. M., Dodson, K., Doup, L. E., Downes, M., Dugan-Rocha, S., Dunkov, B. C., Dunn, P., Durbin, K. J., Evangelista, C. C., Ferraz, C., Ferriera, S., Fleischmann, W., Fosler, C., Gabrielian, A. E., Garg, N. S., Gelbart, W. M., Glasser, K., Glodek, A., Gong, F., Gorrell, J. H., Gu, Z., Guan, P., Harris, M., Harris, N. L., Harvey, D., Heiman, T. J., Hernandez, J. R., Houck, J., Hostin, D., Houston, K. A., Howland, T. J., Wei, M. H., Ibegwam, C., Jalali, M., Kalush, F., Karpen, G. H., Ke, Z., Kennison, J. A., Ketchum, K. A., Kimmel, B. E., Kodira, C. D., Kraft, C., Kravitz, S., Kulp, D., Lai, Z., Lasko, P., Lei, Y., Levitsky, A. A., Li, J., Li, Z., Liang, Y., Lin, X., Liu, X., Mattei, B., McIntosh, T. C., McLeod, M. P., McPherson, D., Merkulov, G., Milshina, N. V., Mobarry, C., Morris, J., Moshrefi, A., Mount, S. M., Moy, M., Murphy, B., Murphy, L., Muzny, D. M., Nelson, D. L., Nelson, D. R., Nelson, K. A., Nixon, K., Nusskern, D. R., Pacleb, J. M., Palazzolo, M., Pittman, G. S., Pan, S., Pollard, J., Puri, V., Reese, M. G., Reinert, K., Remington, K., Saunders, R. D., Scheeler, F., Shen, H., Shue, B. C., Siden-Kiamos, I., Simpson, M., Skupski, M. P., Smith, T., Spier, E., Spradling, A. C., Stapleton, M., Strong, R., Sun, E., Svirskas, R., Tector, C., Turner, R., Venter, E., Wang, A. H., Wang, X., Wang, Z. Y., Wassarman, D. A., Weinstock, G. M., Weissenbach, J., Williams, S. M., WoodageT, Worley, K. C., Wu, D., Yang, S., Yao,

Q. A., Ye, J., Yeh, R. F., Zaveri, J. S., Zhan, M., Zhang, G., Zhao, Q., Zheng, L., Zheng, X. H., Zhong, F. N., Zhong, W., Zhou, X., Zhu, S., Zhu, X., Smith, H. O., Gibbs, R. A., Myers, E. W., Rubin, G. M., & Venter, J. C. 2000, "The genome sequence of Drosophila melanogaster", *Science*, vol. 287, no. 5461, pp. 2185-2195.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. 1990, "Basic local alignment search tool", *J.Mol.Biol.*, vol. 215, no. 3, pp. 403-410.

Anderson, T. J., Blouin, M. S., & Beech, R. N. 1998, "Population biology of parasitic nematodes: applications of genetic markers", *Adv.Parasitol.*, vol. 41, pp. 219-283.

Bakhetia, M., Charlton, W., Atkinson, H. J., & McPherson, M. J. 2005, "RNA interference of dual oxidase in the plant nematode Meloidogyne incognita", *Mol.Plant Microbe Interact.*, vol. 18, no. 10, pp. 1099-1106.

Bartley, D. J., Jackson, E., Johnston, K., Coop, R. L., Mitchell, G. B., Sales, J., & Jackson, F. 2003, "A survey of anthelmintic resistant nematode parasites in Scottish sheep flocks", *Vet.Parasitol.*, vol. 117, no. 1-2, pp. 61-71.

Beech, R. N., Prichard, R. K., & Scott, M. E. 1994, "Genetic variability of the beta-tubulin genes in benzimidazole-susceptible and -resistant strains of Haemonchus contortus", *Genetics*, vol. 138, no. 1, pp. 103-110.

Bernat, J. A., Crawford, G. E., Ogurtsov, A. Y., Collins, F. S., Ginsburg, D., & Kondrashov, A. S. 2006, "Distant conserved sequences flanking endothelial-specific promoters contain tissue-specific DNase-hypersensitive sites and over-represented motifs", *Hum.Mol.Genet.*, vol. 15, no. 13, pp. 2098-2105.

Berriman, M., Haas, B. J., LoVerde, P. T., Wilson, R. A., Dillon, G. P., Cerqueira, G. C., Mashiyama, S. T., Al Lazikani, B., Andrade, L. F., Ashton, P. D., Aslett, M. A., Bartholomeu, D. C., Blandin, G., Caffrey, C. R., Coghlan, A., Coulson, R., Day, T. A., Delcher, A., DeMarco, R., Djikeng, A., Eyre, T., Gamble, J. A., Ghedin, E., Gu, Y., Hertz-Fowler, C., Hirai, H., Hirai, Y., Houston, R., Ivens, A., Johnston, D. A., Lacerda, D., Macedo, C. D., McVeigh, P., Ning, Z., Oliveira, G., Overington, J. P., Parkhill, J., Pertea, M., Pierce, R. J., Protasio, A. V., Quail, M. A., Rajandream, M. A., Rogers, J., Sajid, M., Salzberg, S. L., Stanke, M., Tivey, A. R., White, O., Williams, D. L., Wortman, J., Wu, W., Zamanian, M., Zerlotini, A., Fraser-Liggett, C. M., Barrell, B. G., & El Sayed, N. M. 2009, "The genome of the blood fluke Schistosoma mansoni", *Nature*, vol. 460, no. 7253, pp. 352-358.

Bird, J. The antiparasitics market. supplement . 1991.
Ref Type: Generic

Blackhall, W. J., Pouliot, J. F., Prichard, R. K., & Beech, R. N. 1998, "Haemonchus contortus: selection at a glutamate-gated chloride channel gene in ivermectin- and moxidectin-selected strains", *Exp.Parasitol.*, vol. 90, no. 1, pp. 42-48.

Blackhall, W. J., Prichard, R. K., & Beech, R. N. 2003, "Selection at a gamma-aminobutyric acid receptor gene in Haemonchus contortus resistant to avermectins/milbemycins", *Mol.Biochem.Parasitol.*, vol. 131, no. 2, pp. 137-145.

Blackhall, W. J., Prichard, R. K., & Beech, R. N. 2008, "P-glycoprotein selection in strains of Haemonchus contortus resistant to benzimidazoles", *Vet.Parasitol.*, vol. 152, no. 1-2, pp. 101-107.

Blaxter, M. L., De Ley, P., Garey, J. R., Liu, L. X., Scheldeman, P., Vierstraete, A., Vanfleteren, J. R., Mackey, L. Y., Dorris, M., Frisse, L. M., Vida, J. T., & Thomas, W. K. 1998, "A molecular evolutionary framework for the phylum Nematoda", *Nature*, vol. 392, no. 6671, pp. 71-75.

Blouin, M. S. 1998, "Mitochondrial DNA diversity in nematodes", *J.Helminthol.*, vol. 72, no. 4, pp. 285-289.

Blouin, M. S. 2000, "Neutrality tests on mtDNA: unusual results from nematodes", *J.Hered.*, vol. 91, no. 2, pp. 156-158.

Bordenstein, S. R., Uy, J. J., & Werren, J. H. 2003, "Host genotype determines cytoplasmic incompatibility type in the haplodiploid genus Nasonia", *Genetics*, vol. 164, no. 1, pp. 223-233.

Borneman, A. R., Zhang, Z. D., Rozowsky, J., Seringhaus, M. R., Gerstein, M., & Snyder, M. 2007, "Transcription factor binding site identification in yeast: a comparison of high-density oligonucleotide and PCR-based microarray platforms", *Funct.Integr.Genomics*, vol. 7, no. 4, pp. 335-345.

Boyle, J. P. & Yoshino, T. P. 2003, "Gene manipulation in parasitic helminths", *Int.J.Parasitol.*, vol. 33, no. 11, pp. 1259-1268.

Brenner, S. 1974, "The genetics of Caenorhabditis elegans", *Genetics*, vol. 77, no. 1, pp. 71-94.

Britton, C., Redmond, D. L., Knox, D. P., McKerrow, J. H., & Barry, J. D. 1999, "Identification of promoter elements of parasite nematode genes in transgenic Caenorhabditis elegans", *Mol.Biochem.Parasitol.*, vol. 103, no. 2, pp. 171-181.

Britton, C. & Murray, L. 2002, "A cathepsin L protease essential for Caenorhabditis elegans embryogenesis is functionally conserved in parasitic nematodes", *Mol.Biochem.Parasitol.*, vol. 122, no. 1, pp. 21-33.

Britton, C. & Murray, L. 2006, "Using Caenorhabditis elegans for functional analysis of genes of parasitic nematodes", *Int.J.Parasitol.*, vol. 36, no. 6, pp. 651-659.

Budine, R. D. & Stern, M. J. Easy RNA isolation from *C.elegans*: a TRIZOL based method. Worm Breeders Gazette p.14. 1996.
Ref Type: Generic

C.elegans genome sequencing consortium 1998, "Genome sequence of the nematode C. elegans: a platform for investigating biology", *Science*, vol. 282, no. 5396, pp. 2012-2018.

Chelala, C., Hahn, S. A., Whiteman, H. J., Barry, S., Hariharan, D., Radon, T. P., Lemoine, N. R., & Crnogorac-Jurcevic, T. 2007, "Pancreatic Expression database: a generic model for the organization, integration and mining of complex cancer datasets", *BMC.Genomics*, vol. 8, p. 439.

Chen, K., Huzil, J. T., Freedman, H., Ramachandran, P., Antoniou, A., Tuszynski, J. A., & Kurgan, L. 2008, "Identification of tubulin drug binding sites and prediction of relative differences in binding affinities to tubulin isotypes using digital signal processing", *J.Mol.Graph.Model.*, vol. 27, no. 4, pp. 497-505.

Chen, Q., Rehman, S., Smant, G., & Jones, J. T. 2005, "Functional analysis of pathogenicity proteins of the potato cyst nematode Globodera rostochiensis using RNAi", *Mol.Plant Microbe Interact.*, vol. 18, no. 7, pp. 621-625.

Cheng, G. F., Lin, J. J., Shi, Y., Jin, Y. X., Fu, Z. Q., Jin, Y. M., Zhou, Y. C., & Cai, Y. M. 2005, "Dose-dependent inhibition of gynecophoral canal protein gene expression in vitro in the schistosome (Schistosoma japonicum) by RNA interference", *Acta Biochim.Biophys.Sin.(Shanghai)* , vol. 37, no. 6, pp. 386-390.

Chenggang, L. Mutations of a Redundant a-Tubulin Gene Affect Caenorhabditis elegans Early Embryonic Cleavage via MEI-1/Katanin-Dependent and -Independent Pathways. Mains, P. E. 2004.
Ref Type: Generic

Clark, A. G., Eisen, M. B., Smith, D. R., Bergman, C. M., Oliver, B., Markow, T. A., Kaufman, T. C., Kellis, M., Gelbart, W., Iyer, V. N., Pollard, D. A., Sackton, T. B., Larracuente, A. M., Singh, N. D., Abad, J. P., Abt, D. N., Adryan, B., Aguade, M., Akashi, H., Anderson, W. W., Aquadro, C. F., Ardell, D. H., Arguello, R., Artieri, C. G., Barbash, D. A., Barker, D., Barsanti, P., Batterham, P., Batzoglou, S., Begun, D., Bhutkar, A., Blanco, E., Bosak, S. A., Bradley, R. K., Brand, A. D., Brent, M. R., Brooks, A. N., Brown, R. H., Butlin, R. K., Caggese, C., Calvi, B. R., Bernardo, d. C., Caspi, A., Castrezana, S., Celniker, S. E., Chang, J. L., Chapple, C., Chatterji, S., Chinwalla, A., Civetta, A., Clifton, S. W., Comeron, J. M., Costello, J. C., Coyne, J. A., Daub, J., David, R. G., Delcher, A. L., Delehaunty, K., Do, C. B., Ebling, H., Edwards, K., Eickbush, T., Evans, J. D., Filipski, A., Findeiss, S., Freyhult, E., Fulton, L., Fulton, R., Garcia, A. C., Gardiner, A., Garfield, D. A., Garvin, B. E., Gibson, G., Gilbert, D., Gnerre, S., Godfrey, J., Good, R., Gotea, V., Gravely, B., Greenberg, A. J., Griffiths-Jones, S., Gross, S., Guigo, R., Gustafson, E. A., Haerty, W., Hahn, M. W., Halligan, D. L., Halpern, A. L., Halter, G. M., Han, M. V., Heger, A., Hillier, L., Hinrichs, A. S., Holmes, I., Hoskins, R. A., Hubisz, M. J., Hultmark, D., Huntley, M. A., Jaffe, D. B., Jagadeeshan, S., Jeck, W. R., Johnson, J., Jones, C. D., Jordan, W. C., Karpen, G. H., Kataoka, E., Keightley, P. D., Kheradpour, P., Kirkness, E. F., Koerich, L. B., Kristiansen, K., Kudrna, D., Kulathinal, R. J., Kumar, S., Kwok, R., Lander, E., Langley, C. H., Lapoint, R., Lazzaro, B. P., Lee, S. J., Levesque, L., Li, R., Lin, C. F., Lin, M. F., Lindblad-Toh, K., Llopart, A., Long, M., Low, L., Lozovsky, E., Lu, J., Luo, M., Machado, C. A., Makalowski, W., Marzo, M., Matsuda, M., Matzkin, L., McAllister, B., McBride, C. S., McKernan, B., McKernan, K., Mendez-Lago, M., Minx, P., Mollenhauer, M. U., Montooth, K., Mount, S. M., Mu, X., Myers, E., Negre, B., Newfeld, S., Nielsen, R., Noor, M. A., O'Grady, P., Pachter, L., Papaceit, M., Parisi, M. J., Parisi, M., Parts, L., Pedersen, J. S., Pesole, G., Phillippy, A. M., Ponting, C. P., Pop, M., Porcelli, D., Powell, J. R., Prohaska, S., Pruitt, K., Puig, M., Quesneville, H., Ram, K. R., Rand, D., Rasmussen, M. D., Reed, L. K., Reenan, R., Reily, A., Remington, K. A., Rieger, T. T., Ritchie, M. G., Robin, C., Rogers, Y. H., Rohde, C., Rozas, J., Rubenfield, M. J., Ruiz, A., Russo, S., Salzberg, S. L., Sanchez-Gracia, A., Saranga, D. J., Sato, H., Schaeffer, S. W., Schatz, M. C., Schlenke, T., Schwartz, R., Segarra, C., Singh, R. S., Sirot, L., Sirota, M., Sisneros, N. B., Smith, C. D., Smith, T. F., Spieth, J., Stage, D. E., Stark, A., Stephan, W., Strausberg, R. L., Strempel, S., Sturgill, D., Sutton, G., Sutton, G. G., Tao, W., Teichmann, S., Tobari, Y. N., Tomimura, Y., Tsolas, J. M., Valente, V. L., Venter, E., Venter, J. C., Vicario, S., Vieira, F. G., Vilella, A. J., Villasante, A., Walenz, B., Wang, J., Wasserman, M., Watts, T., Wilson, D., Wilson, R. K., Wing, R. A., Wolfner, M. F., Wong, A., Wong, G. K., Wu, C. I., Wu, G., Yamamoto, D., Yang, H. P., Yang, S. P., Yorke, J. A., Yoshida, K., Zdobnov, E., Zhang, P., Zhang, Y.,

Zimin, A. V., Baldwin, J., Abdouelleil, A., Abdulkadir, J., Abebe, A., Abera, B., Abreu, J., Acer, S. C., Aftuck, L., Alexander, A., An, P., Anderson, E., Anderson, S., Arachi, H., & Azer, M. 2007, "Evolution of genes and genomes on the Drosophila phylogeny", *Nature*, vol. 450, no. 7167, pp. 203-218.

Cleveland, D. W. 1988, "Autoregulated instability of tubulin mRNAs: a novel eukaryotic regulatory mechanism", *Trends Biochem.Sci.*, vol. 13, no. 9, pp. 339-343.

Correnti, J. M., Brindley, P. J., & Pearce, E. J. 2005, "Long-term suppression of cathepsin B levels by RNA interference retards schistosome growth", *Mol.Biochem.Parasitol.*, vol. 143, no. 2, pp. 209-215.

Costa, J. C., Lilley, C. J., Atkinson, H. J., & Urwin, P. E. 2009, "Functional characterisation of a cyst nematode acetylcholinesterase gene using Caenorhabditis elegans as a heterologous system", *Int.J.Parasitol.*

Couthier, A. Functional analyisis of a parasitic nematode GATA transcription factor using *C. elegans* as a heterologous expression system. 2004.
Ref Type: Generic

Coyne, M. J. & Smith, G. 1992, "The mortality and fecundity of Haemonchus contortus in parasite-naive and parasite-exposed sheep following single experimental infections", *Int.J.Parasitol.*, vol. 22, no. 3, pp. 315-325.

Crawford, G. E., Davis, S., Scacheri, P. C., Renaud, G., Halawi, M. J., Erdos, M. R., Green, R., Meltzer, P. S., Wolfsberg, T. G., & Collins, F. S. 2006, "DNase-chip: a high-resolution method to identify DNase I hypersensitive sites using tiled microarrays", *Nat.Methods*, vol. 3, no. 7, pp. 503-509.

Cutter, A. D. 2008, "Divergence times in Caenorhabditis and Drosophila inferred from direct estimates of the neutral mutation rate", *Mol.Biol.Evol.*, vol. 25, no. 4, pp. 778-786.

Davis, R. E., Parra, A., LoVerde, P. T., Ribeiro, E., Glorioso, G., & Hodgson, S. 1999, "Transient expression of DNA and RNA in parasitic helminths by using particle bombardment", *Proc.Natl.Acad.Sci.U.S.A* , vol. 96, no. 15, pp. 8687-8692.

De Ley P. A new system for Nematoda: combining morphological characters with molecular trees, and translating clades into ranks and taxa. Blaxter M. 2004.
Ref Type: Generic

Dieterich, C., Clifton, S. W., Schuster, L. N., Chinwalla, A., Delehaunty, K., Dinkelacker, I., Fulton, L., Fulton, R., Godfrey, J., Minx, P., Mitreva, M., Roeseler, W., Tian, H., Witte, H., Yang, S. P., Wilson, R. K., & Sommer, R. J. 2008, "The Pristionchus pacificus genome provides a unique perspective on nematode lifestyle and parasitism", *Nat.Genet.*, vol. 40, no. 10, pp. 1193-1198.

Dinguirard, N. & Yoshino, T. P. 2006, "Potential role of a CD36-like class B scavenger receptor in the binding of modified low-density lipoprotein (acLDL) to the tegumental surface of Schistosoma mansoni sporocysts", *Mol.Biochem.Parasitol.*, vol. 146, no. 2, pp. 219-230.

Dlugosz, E. & Wisniewski, M. 2006, "[Molecular diagnostic of parasites using rRNA gene sequence]", *Wiad.Parazytol.*, vol. 52, no. 4, pp. 263-269.

Dominguez, J. E., Padilla, R., Avila, J., & Carrascosa, J. L. 1990, "Removal of the carboxy terminus of beta-tubulin subunit produces lateral annealing of microtubules with different orientations", *Int.J.Biochem.*, vol. 22, no. 12, pp. 1419-1425.

Driscoll, M., Dean, E., Reilly, E., Bergholz, E., & Chalfie, M. 1989, "Genetic and molecular analysis of a Caenorhabditis elegans beta-tubulin that conveys benzimidazole sensitivity", *J.Cell Biol.*, vol. 109, no. 6 Pt 1, pp. 2993-3003.

Duff, R. J. & Nickrent, D. L. 1997, "Characterization of mitochondrial small-subunit ribosomal RNAs from holoparasitic plants", *J.Mol.Evol.*, vol. 45, no. 6, pp. 631-639.

Eng, J. K., Blackhall, W. J., Osei-Atweneboana, M. Y., Bourguinat, C., Galazzo, D., Beech, R. N., Unnasch, T. R., Awadzi, K., Lubega, G. W., & Prichard, R. K. 2006, "Ivermectin selection on beta-tubulin: evidence in Onchocerca volvulus and Haemonchus contortus", *Mol.Biochem.Parasitol.*, vol. 150, no. 2, pp. 229-235.

Fairbairn, D. J., Cavallaro, A. S., Bernard, M., Mahalinga-Iyer, J., Graham, M. W., & Botella, J. R. 2007, "Host-delivered RNAi: an effective strategy to silence genes in plant parasitic nematodes", *Planta*, vol. 226, no. 6, pp. 1525-1533.

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

Feinberg, E. H. & Hunter, C. P. 2003, "Transport of dsRNA into cells by the transmembrane protein SID-1", *Science*, vol. 301, no. 5639, pp. 1545-1547.

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*, vol. 391, no. 6669, pp. 806-811.

Ford, L., Guiliano, D. B., Oksov, Y., Debnath, A. K., Liu, J., Williams, S. A., Blaxter, M. L., & Lustigman, S. 2005, "Characterization of a novel filarial serine protease inhibitor, Ov-SPI-1, from Onchocerca volvulus, with potential multifunctional roles during development of the parasite", *J.Biol.Chem.*, vol. 280, no. 49, pp. 40845-40856.

Freeman, A. S., Nghiem, C., Li, J., Ashton, F. T., Guerrero, J., Shoop, W. L., & Schad, G. A. 2003, "Amphidial structure of ivermectin-resistant and susceptible laboratory and field strains of Haemonchus contortus", *Vet.Parasitol.*, vol. 110, no. 3-4, pp. 217-226.

Fujimura, M., Oeda, K., Inoue, H., & Kato, T. 1992, "A single amino-acid substitution in the beta-tubulin gene of Neurospora confers both carbendazim resistance and diethofencarb sensitivity", *Curr.Genet.* , vol. 21, no. 4-5, pp. 399-404.

Fukushige, T., Hawkins, M. G., & McGhee, J. D. 1998, "The GATA-factor elt-2 is essential for formation of the Caenorhabditis elegans intestine", *Dev.Biol.*, vol. 198, no. 2, pp. 286-302.

Geary, T. G., Nulf, S. C., Favreau, M. A., Tang, L., Prichard, R. K., Hatzenbuhler, N. T., Shea, M. H., Alexander, S. J., & Klein, R. D. 1992, "Three beta-tubulin cDNAs from the parasitic nematode Haemonchus contortus", *Mol.Biochem.Parasitol.*, vol. 50, no. 2, pp. 295-306.

Geldhof, P., Murray, L., Couthier, A., Gilleard, J. S., McLauchlan, G., Knox, D. P., & Britton, C. 2006, "Testing the efficacy of RNA interference in Haemonchus contortus", *Int.J.Parasitol.*, vol. 36, no. 7, pp. 801-810.

Geldhof, P., Visser, A., Clark, D., Saunders, G., Britton, C., Gilleard, J., Berriman, M., & Knox, D. 2007, "RNA interference in parasitic helminths: current situation, potential pitfalls and future prospects", *Parasitology*, vol. 134, no. Pt 5, pp. 609-619.

Ghedin, E., Wang, S., Spiro, D., Caler, E., Zhao, Q., Crabtree, J., Allen, J. E., Delcher, A. L., Guiliano, D. B., Miranda-Saavedra, D., Angiuoli, S. V., Creasy, T., Amedeo, P., Haas, B., El Sayed, N. M., Wortman, J. R., Feldblyum, T., Tallon, L., Schatz, M., Shumway, M., Koo, H., Salzberg, S. L., Schobel, S., Pertea, M., Pop, M., White, O., Barton, G. J., Carlow, C. K., Crawford, M. J., Daub, J., Dimmic, M. W., Estes, C. F., Foster, J. M., Ganatra, M., Gregory, W. F., Johnson, N. M., Jin, J., Komuniecki, R., Korf, I., Kumar, S., Laney, S., Li, B. W., Li, W., Lindblom, T. H., Lustigman, S., Ma, D., Maina, C. V., Martin, D. M., McCarter, J. P., McReynolds, L., Mitreva, M., Nutman, T. B., Parkinson, J., Peregrin-Alvarez, J. M., Poole, C., Ren, Q., Saunders, L., Sluder, A. E., Smith, K., Stanke, M., Unnasch, T. R., Ware, J., Wei, A. D., Weil, G., Williams, D. J., Zhang, Y., Williams, S. A., Fraser-Liggett, C., Slatko, B., Blaxter, M. L., & Scott, A. L. 2007, "Draft genome of the filarial nematode parasite Brugia malayi", *Science*, vol. 317, no. 5845, pp. 1756-1760.

Gibbs, E. & Oger, J. 2007, "The IgG subclass-specificities of anti-IFNbeta antibodies change with time and differ between the IFNbeta products in relapsing remitting multiple sclerosis patients", *J.Neuroimmunol.*, vol. 190, no. 1-2, pp. 146-150.

Gillan, V., Maitland, K., McCormack, G., Him, N. A., & Devaney, E. 2009, "Functional genomics of hsp-90 in parasitic and free-living nematodes", *Int.J.Parasitol.*, vol. 39, no. 10, pp. 1071-1081.

Gilleard, J. S. 2004, "The use of Caenorhabditis elegans in parasitic nematode research", *Parasitology*, vol. 128 Suppl 1, p. S49-S70.

Gilleard, J. S., Woods, D. J., & Dow, J. A. 2005, "Model-organism genomics in veterinary parasite drug-discovery", *Trends Parasitol.*, vol. 21, no. 7, pp. 302-305.

Gilleard, J. S. & Beech, R. N. 2007, "Population genetics of anthelmintic resistance in parasitic nematodes", *Parasitology*, vol. 134, no. Pt 8, pp. 1133-1147.

Gleason, C. A., Liu, Q. L., & Williamson, V. M. 2008, "Silencing a candidate nematode effector gene corresponding to the tomato resistance gene Mi-1 leads to acquisition of virulence", *Mol.Plant Microbe Interact.*, vol. 21, no. 5, pp. 576-585.

Gomez-Escobar, N., Gregory, W. F., Britton, C., Murray, L., Corton, C., Hall, N., Daub, J., Blaxter, M. L., & Maizels, R. M. 2002, "Abundant larval transcript-1 and -2 genes from Brugia malayi: diversity of genomic environments but conservation of 5' promoter sequences functional in Caenorhabditis elegans", *Mol.Biochem.Parasitol.*, vol. 125, no. 1-2, pp. 59-71.

Grandien, K. & Sommer, R. J. 2001, "Functional comparison of the nematode Hox gene lin-39 in C. elegans and P. pacificus reveals evolutionary conservation of protein function despite divergence of primary sequences", *Genes Dev.*, vol. 15, no. 16, pp. 2161-2172.

Grant, W. N. & Mascord, L. J. 1996, "Beta-tubulin gene polymorphism and benzimidazole resistance in trichostrongylus colubriformis", *Int.J.Parasitol.*, vol. 26, no. 1, pp. 71-77.

Grant, W. N., Skinner, S. J., Newton-Howes, J., Grant, K., Shuttleworth, G., Heath, D. D., & Shoemaker, C. B. 2006, "Heritable transgenesis of Parastrongyloides trichosuri: a nematode parasite of mammals", *Int.J.Parasitol.*, vol. 36, no. 4, pp. 475-483.

Grishok, A. 2005, "RNAi mechanisms in Caenorhabditis elegans", *FEBS Lett.*, vol. 579, no. 26, pp. 5932-5939.

Guerrero, J. & Freeman, A. S. 2004, "Amphids: the neuronal ultrastructure of macrocyclic-lactone-resistant Haemonchus contortus", *Parassitologia*, vol. 46, no. 1-2, pp. 237-240.

Hamelin, M., Scott, I. M., Way, J. C., & Culotti, J. G. 1992, "The mec-7 beta-tubulin gene of Caenorhabditis elegans is expressed primarily in the touch receptor neurons", *EMBO J.*, vol. 11, no. 8, pp. 2885-2893.

Hannon, G. J. 2002, "RNA interference", *Nature*, vol. 418, no. 6894, pp. 244-251.

Heasman, J. 2002, "Morpholino oligos: making sense of antisense?", *Dev.Biol.*, vol. 243, no. 2, pp. 209-214.

Hesse, J., Thierauf, M., & Ponstingl, H. 1987, "Tubulin sequence region beta 155-174 is involved in binding exchangeable guanosine triphosphate", *J.Biol.Chem.*, vol. 262, no. 32, pp. 15472-15475.

Higazi, T. B., Merriweather, A., Shu, L., Davis, R., & Unnasch, T. R. 2002, "Brugia malayi: transient transfection by microinjection and particle bombardment", *Exp.Parasitol.*, vol. 100, no. 2, pp. 95-102.

Hillier, L. W., Miller, R. D., Baird, S. E., Chinwalla, A., Fulton, L. A., Koboldt, D. C., & Waterston, R. H. 2007, "Comparison of C. elegans and C. briggsae genome sequences reveals extensive conservation of chromosome organization and synteny", *PLoS.Biol.*, vol. 5, no. 7, p. e167.

Hobert, O. 2002, "PCR fusion-based approach to create reporter gene constructs for expression analysis in transgenic C. elegans", *Biotechniques*, vol. 32, no. 4, pp. 728-730.

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", *Proc.Natl.Acad.Sci.U.S.A*, vol. 103, no. 39, pp. 14302-14306.

Hussein, A. S., Kichenin, K., & Selkirk, M. E. 2002, "Suppression of secreted acetylcholinesterase expression in Nippostrongylus brasiliensis by RNA interference", *Mol.Biochem.Parasitol.*, vol. 122, no. 1, pp. 91-94.

Islam, M. K., Miyoshi, T., Yamada, M., & Tsuji, N. 2005, "Pyrophosphatase of the roundworm Ascaris suum plays an essential role in the worm's molting and development", *Infect.Immun.*, vol. 73, no. 4, pp. 1995-2004.

Issa, Z., Grant, W. N., Stasiuk, S., & Shoemaker, C. B. 2005, "Development of methods for RNA interference in the sheep gastrointestinal parasite, Trichostrongylus colubriformis", *Int.J.Parasitol.*, vol. 35, no. 9, pp. 935-940.

Jackson, F. 1993, "Anthelmintic resistance--the state of play", *Br.Vet.J.*, vol. 149, no. 2, pp. 123-138.

Jackstadt, P., Wilm, T. P., Zahner, H., & Hobom, G. 1999, "Transformation of nematodes via ballistic DNA transfer", *Mol.Biochem.Parasitol.*, vol. 103, no. 2, pp. 261-266.

Jasmer, D. P., Yao, C., Rehman, A., & Johnson, S. 2000, "Multiple lethal effects induced by a benzimidazole anthelmintic in the anterior intestine of the nematode Haemonchus contortus", *Mol.Biochem.Parasitol.*, vol. 105, no. 1, pp. 81-90.

Jung, M. K., Wilder, I. B., & Oakley, B. R. 1992, "Amino acid alterations in the benA (beta-tubulin) gene of Aspergillus nidulans that confer benomyl resistance", *Cell Motil.Cytoskeleton*, vol. 22, no. 3, pp. 170-174.

Kamath, R. S. & Ahringer, J. 2003, "Genome-wide RNAi screening in Caenorhabditis elegans", *Methods*, vol. 30, no. 4, pp. 313-321.

Kennerdell, J. R. & Carthew, R. W. 1998, "Use of dsRNA-mediated genetic interference to demonstrate that frizzled and frizzled 2 act in the wingless pathway", *Cell*, vol. 95, no. 7, pp. 1017-1026.

Kevin Strange 2006, *C. Elegans: Methods and Applications*.

Kim, J. K., Gabel, H. W., Kamath, R. S., Tewari, M., Pasquinelli, A., Rual, J. F., Kennedy, S., Dybbs, M., Bertin, N., Kaplan, J. M., Vidal, M., & Ruvkun, G. 2005, "Functional genomic analysis of RNA interference in C. elegans", *Science*, vol. 308, no. 5725, pp. 1164-1167.

Kohler, P. 2001, "The biochemical basis of anthelmintic action and resistance", *Int.J.Parasitol.*, vol. 31, no. 4, pp. 336-345.

Kotze, A. C. & Bagnall, N. H. 2006, "RNA interference in Haemonchus contortus: suppression of beta-tubulin gene expression in L3, L4 and adult worms in vitro", *Mol.Biochem.Parasitol.*, vol. 145, no. 1, pp. 101-110.

Krause, S., Sommer, A., Fischer, P., Brophy, P. M., Walter, R. D., & Liebau, E. 2001, "Gene structure of the extracellular glutathione S-transferase from Onchocerca volvulus and its overexpression and promoter analysis in transgenic Caenorhabditis elegans", *Mol.Biochem.Parasitol.*, vol. 117, no. 2, pp. 145-154.

Kumar, S., Tamura, K., & Nei, M. 2004, "MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment", *Brief.Bioinform.*, vol. 5, no. 2, pp. 150-163.

Kwa, M. S., Veenstra, J. G., & Roos, M. H. 1993, "Molecular characterisation of beta-tubulin genes present in benzimidazole-resistant populations of Haemonchus contortus", *Mol.Biochem.Parasitol.*, vol. 60, no. 1, pp. 133-143.

Kwa, M. S., Kooyman, F. N., Boersema, J. H., & Roos, M. H. 1993, "Effect of selection for benzimidazole resistance in Haemonchus contortus on beta-tubulin

isotype 1 and isotype 2 genes", *Biochem.Biophys.Res.Commun.*, vol. 191, no. 2, pp. 413-419.

Kwa, M. S., Veenstra, J. G., & Roos, M. H. 1994, "Benzimidazole resistance in Haemonchus contortus is correlated with a conserved mutation at amino acid 200 in beta-tubulin isotype 1", *Mol.Biochem.Parasitol.*, vol. 63, no. 2, pp. 299-303.

Kwa, M. S., Veenstra, J. G., Van Dijk, M., & Roos, M. H. 1995, "Beta-tubulin genes from the parasitic nematode Haemonchus contortus modulate drug resistance in Caenorhabditis elegans", *J.Mol.Biol.*, vol. 246, no. 4, pp. 500-510.

Lacey, E. 1988, "The role of the cytoskeletal protein, tubulin, in the mode of action and mechanism of drug resistance to benzimidazoles", *Int.J.Parasitol.*, vol. 18, no. 7, pp. 885-936.

Lendner, M., Doligalska, M., Lucius, R., & Hartmann, S. 2008, "Attempts to establish RNA interference in the parasitic nematode Heligmosomoides polygyrus", *Mol.Biochem.Parasitol.*, vol. 161, no. 1, pp. 21-31.

Leroy, S., Duperray, C., & Morand, S. 2003, "Flow cytometry for parasite nematode genome size measurement", *Mol.Biochem.Parasitol.*, vol. 128, no. 1, pp. 91-93.

Li, B. W., Rush, A. C., Mitreva, M., Yin, Y., Spiro, D., Ghedin, E., & Weil, G. J. 2009, "Transcriptomes and pathways associated with infectivity, survival and immunogenicity in Brugia malayi L3", *BMC.Genomics*, vol. 10, p. 267.

Li, X., Massey, H. C., Jr., Nolan, T. J., Schad, G. A., Kraus, K., Sundaram, M., & Lok, J. B. 2006, "Successful transgenesis of the parasitic nematode Strongyloides stercoralis requires endogenous non-coding control elements", *Int.J.Parasitol.*, vol. 36, no. 6, pp. 671-679.

Lilley, C. J., Goodchild, S. A., Atkinson, H. J., & Urwin, P. E. 2005, "Cloning and characterisation of a Heterodera glycines aminopeptidase cDNA", *Int.J.Parasitol.*, vol. 35, no. 14, pp. 1577-1585.

Lindqvist, A., Ljungstrom, B. L., Nilsson, O., & Waller, P. J. 2001, "The dynamics, prevalence and impact of nematode infections in organically raised sheep in Sweden", *Acta Vet.Scand.*, vol. 42, no. 3, pp. 377-389.

Liu, F., Zhou, Y., Wang, Z. Q., Lu, G., Zheng, H., Brindley, P. J., McManus, D. P., Blair, D., Zhang, Q. H., Zhong, Y., Wang, S., Han, Z. G., & Chen, Z. 2009, "The Schistosoma japonicum genome reveals features of host-parasite interplay", *Nature*, vol. 460, no. 7253, pp. 345-351.

Liu, J., Dent, J. A., Beech, R. N., & Prichard, R. K. 2004, "Genomic organization of an avermectin receptor subunit from Haemonchus contortus and expression of its putative promoter region in Caenorhabditis elegans", *Mol.Biochem.Parasitol.*, vol. 134, no. 2, pp. 267-274.

Lok, J. B. & Massey, H. C., Jr. 2002, "Transgene expression in Strongyloides stercoralis following gonadal microinjection of DNA constructs", *Mol.Biochem.Parasitol.*, vol. 119, no. 2, pp. 279-284.

Lu, C., Srayko, M., & Mains, P. E. 2004, "The Caenorhabditis elegans microtubule-severing complex MEI-1/MEI-2 katanin interacts differently with two superficially redundant beta-tubulin isotypes", *Mol.Biol.Cell*, vol. 15, no. 1, pp. 142-150.

Lustigman, S., Zhang, J., Liu, J., Oksov, Y., & Hashmi, S. 2004, "RNA interference targeting cathepsin L and Z-like cysteine proteases of Onchocerca volvulus confirmed their essential function during L3 molting", *Mol.Biochem.Parasitol.*, vol. 138, no. 2, pp. 165-170.

MAFF. Manual of Veterinary Parasitology Laboratory Techniques HMSO, London. 1986.
Ref Type: Generic

McLaren, D. J., Worms, M. J., Laurence, B. R., & Simpson, M. G. 1975, "Micro-organisms in filarial larvae (Nematoda)", *Trans.R.Soc.Trop.Med.Hyg.*, vol. 69, no. 5-6, pp. 509-514.

Mello, C. & Fire, A. 1995, "DNA transformation", *Methods Cell Biol.*, vol. 48, pp. 451-482.

Mello, C. C., Kramer, J. M., Stinchcomb, D., & Ambros, V. 1991, "Efficient gene transfer in C.elegans: extrachromosomal maintenance and integration of transforming sequences", *EMBO J.*, vol. 10, no. 12, pp. 3959-3970.

Miller, D. M. & Shakes, D. C. 1995, "Immunofluorescence microscopy", *Methods Cell Biol.*, vol. 48, pp. 365-394.

Misquitta, L. & Paterson, B. M. 1999, "Targeted disruption of gene function in Drosophila by RNA interference (RNA-i): a role for nautilus in embryonic somatic muscle formation", *Proc.Natl.Acad.Sci.U.S.A*, vol. 96, no. 4, pp. 1451-1456.

Mitreva, M. & Jasmer, D. P. 2008, "Advances in the sequencing of the genome of the adenophorean nematode Trichinella spiralis", *Parasitology*, vol. 135, no. 8, pp. 869-880.

Murray, L., Geldhof, P., Clark, D., Knox, D. P., & Britton, C. 2007, "Expression and purification of an active cysteine protease of Haemonchus contortus using Caenorhabditis elegans", *Int.J.Parasitol.*, vol. 37, no. 10, pp. 1117-1125.

Neveu, C., Charvet, C., Fauvin, A., Cortet, J., Castagnone-Sereno, P., & Cabaret, J. 2007, "Identification of levamisole resistance markers in the parasitic nematode Haemonchus contortus using a cDNA-AFLP approach", *Parasitology*, vol. 134, no. Pt 8, pp. 1105-1110.

Newton-Howes, J., Heath, D. D., Shoemaker, C. B., & Grant, W. N. 2006, "Characterisation and expression of an Hsp70 gene from Parastrongyloides trichosuri", *Int.J.Parasitol.*, vol. 36, no. 4, pp. 467-474.

Njue, A. I. & Prichard, R. K. 2003, "Cloning two full-length beta-tubulin isotype cDNAs from Cooperia oncophora, and screening for benzimidazole resistance-associated mutations in two isolates", *Parasitology*, vol. 127, no. Pt 6, pp. 579-588.

Oelgeschlager, M., Larrain, J., Geissert, D., & De Robertis, E. M. 2000, "The evolutionarily conserved BMP-binding protein Twisted gastrulation promotes BMP signalling", *Nature*, vol. 405, no. 6788, pp. 757-763.

Opperman, C. H., Bird, D. M., Williamson, V. M., Rokhsar, D. S., Burke, M., Cohn, J., Cromer, J., Diener, S., Gajan, J., Graham, S., Houfek, T. D., Liu, Q., Mitros, T., Schaff, J., Schaffer, R., Scholl, E., Sosinski, B. R., Thomas, V. P., & Windham, E. 2008, "Sequence and genetic map of Meloidogyne hapla: A compact nematode genome for plant parasitism", *Proc.Natl.Acad.Sci.U.S.A*, vol. 105, no. 39, pp. 14802-14807.

Patino, W. D., Mian, O. Y., & Hwang, P. M. 2002, "Serial analysis of gene expression: technical considerations and applications to cardiovascular biology", *Circ.Res.*, vol. 91, no. 7, pp. 565-569.

Pfarr, K., Heider, U., & Hoerauf, A. 2006, "RNAi mediated silencing of actin expression in adult Litomosoides sigmodontis is specific, persistent and results in a phenotype", *Int.J.Parasitol.*, vol. 36, no. 6, pp. 661-669.

Pires-daSilva, A. & Sommer, R. J. 2004, "Conservation of the global sex determination gene tra-1 in distantly related nematodes", *Genes Dev.*, vol. 18, no. 10, pp. 1198-1208.

Pires, d. S. 2005, "Pristionchus pacificus genetic protocols", *Wormbook* pp. 1-8.

Pop, M., Kosack, D. S., & Salzberg, S. L. 2004, "Hierarchical scaffolding with Bambus", *Genome Res.*, vol. 14, no. 1, pp. 149-159.

Prichard, R. 2001, "Genetic variability following selection of Haemonchus contortus with anthelmintics", *Trends Parasitol.*, vol. 17, no. 9, pp. 445-453.

Prichard, R. K. 2007, "Markers for benzimidazole resistance in human parasitic nematodes?", *Parasitology*, vol. 134, no. Pt 8, pp. 1087-1092.

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", *Mol.Biochem.Parasitol.*, vol. 96, no. 1-2, pp. 59-67.

Redman, E., Grillo, V., Saunders, G., Packard, E., Jackson, F., Berriman, M., & Gilleard, J. S. 2008, "Genetics of mating and sex determination in the parasitic nematode Haemonchus contortus", *Genetics*, vol. 180, no. 4, pp. 1877-1887.

Redmond, D. L., Clucas, C., Johnstone, I. L., & Knox, D. P. 2001, "Expression of Haemonchus contortus pepsinogen in Caenorhabditis elegans", *Mol.Biochem.Parasitol.*, vol. 112, no. 1, pp. 125-131.

Roberts, B., Clucas, C., & Johnstone, I. L. 2003, "Loss of SEC-23 in Caenorhabditis elegans causes defects in oogenesis, morphogenesis, and extracellular matrix secretion", *Mol.Biol.Cell*, vol. 14, no. 11, pp. 4414-4426.

Robinson, M. W., McFerran, N., Trudgett, A., Hoey, L., & Fairweather, I. 2004, "A possible model of benzimidazole binding to beta-tubulin disclosed by invoking an inter-domain movement", *J.Mol.Graph.Model.*, vol. 23, no. 3, pp. 275-284.

Roos, M. H., Boersema, J. H., Borgsteede, F. H., Cornelissen, J., Taylor, M., & Ruitenberg, E. J. 1990, "Molecular analysis of selection for benzimidazole resistance in the sheep parasite Haemonchus contortus", *Mol.Biochem.Parasitol.*, vol. 43, no. 1, pp. 77-88.

Roos, M. H., Otsen, M., Hoekstra, R., Veenstra, J. G., & Lenstra, J. A. 2004, "Genetic analysis of inbreeding of two strains of the parasitic nematode Haemonchus contortus", *Int.J.Parasitol.*, vol. 34, no. 1, pp. 109-115.

Rosso, M. N., Dubrana, M. P., Cimbolini, N., Jaubert, S., & Abad, P. 2005, "Application of RNA interference to root-knot nematode genes encoding esophageal gland proteins", *Mol.Plant Microbe Interact.*, vol. 18, no. 7, pp. 615-620.

Rosso, M. N., Jones, J. T., & Abad, P. 2009, "RNAi and Functional Genomics in Plant Parasitic Nematodes", *Annu.Rev.Phytopathol.*, vol. 47, pp. 207-232.

Sabo, P. J., Kuehn, M. S., Thurman, R., Johnson, B. E., Johnson, E. M., Cao, H., Yu, M., Rosenzweig, E., Goldy, J., Haydock, A., Weaver, M., Shafer, A., Lee, K., Neri, F., Humbert, R., Singer, M. A., Richmond, T. A., Dorschner, M. O., McArthur, M., Hawrylycz, M., Green, R. D., Navas, P. A., Noble, W. S., & Stamatoyannopoulos, J. A. 2006, "Genome-scale mapping of DNase I sensitivity in vivo using tiling DNA microarrays", *Nat.Methods*, vol. 3, no. 7, pp. 511-518.

Sabo, P. J., Kuehn, M. S., Thurman, R., Johnson, B. E., Johnson, E. M., Cao, H., Yu, M., Rosenzweig, E., Goldy, J., Haydock, A., Weaver, M., Shafer, A., Lee, K., Neri, F., Humbert, R., Singer, M. A., Richmond, T. A., Dorschner, M. O., McArthur, M., Hawrylycz, M., Green, R. D., Navas, P. A., Noble, W. S., & Stamatoyannopoulos, J. A. 2006, "Genome-scale mapping of DNase I sensitivity in vivo using tiling DNA microarrays", *Nat.Methods*, vol. 3, no. 7, pp. 511-518.

Samson-Himmelstjerna, G. 2006, "Molecular diagnosis of anthelmintic resistance", *Vet.Parasitol.*, vol. 136, no. 2, pp. 99-107.

Samson-Himmelstjerna, G. V., Blackhall W.J., McCarthy J.S., Skuce P.J. 2007, "Single nucleotide polymorphism (SNP) markers for benzimidazole resistance in veterinary nematodes", *Parasitology*;134(Pt 8):1077-86.

Sangster, N. C., Bannan, S. C., Weiss, A. S., Nulf, S. C., Klein, R. D., & Geary, T. G. 1999, "Haemonchus contortus: sequence heterogeneity of internucleotide binding domains from P-glycoproteins", *Exp.Parasitol.*, vol. 91, no. 3, pp. 250-257.

Sargison, N., Scott, P., & Jackson, F. 2001, "Multiple anthelmintic resistance in sheep", *Vet.Rec.*, vol. 149, no. 25, pp. 778-779.

Savage, C., Xue, Y., Mitani, S., Hall, D., Zakhary, R., & Chalfie, M. 1994, "Mutations in the Caenorhabditis elegans beta-tubulin gene mec-7: effects on microtubule assembly and stability and on tubulin autoregulation", *J.Cell Sci.*, vol. 107 (Pt 8), pp. 2165-2175.

Scott, A. L. & Ghedin, E. 2009, "The genome of Brugia malayi - all worms are not created equal", *Parasitol.Int.*, vol. 58, no. 1, pp. 6-11.

Seaton, D. S., Jackson, F., Smith, W. D., & Angus, K. W. 1989, "Development of immunity to incoming radiolabelled larvae in lambs continuously infected with Ostertagia circumcincta", *Res.Vet.Sci.*, vol. 46, no. 2, pp. 241-246.

Shi, B., Guo, X., Wu, T., Sheng, S., Wang, J., Skogerbo, G., Zhu, X., & Chen, R. 2009, "Genome-scale identification of Caenorhabditis elegans regulatory elements by tiling-array mapping of DNase I hypersensitive sites", *BMC.Genomics*, vol. 10, p. 92.

Shi, L., Ko, S., Kim, S., Echchgadda, I., Oh, T. S., Song, C. S., & Chatterjee, B. 2008, "Loss of androgen receptor in aging and oxidative stress through Myb protooncoprotein-regulated reciprocal chromatin dynamics of p53 and poly(ADP-ribose) polymerase PARP-1", *J.Biol.Chem.*, vol. 283, no. 52, pp. 36474-36485.

Shingles, J., Lilley, C. J., Atkinson, H. J., & Urwin, P. E. 2007, "Meloidogyne incognita: molecular and biochemical characterisation of a cathepsin L cysteine proteinase and the effect on parasitism following RNAi", *Exp.Parasitol.*, vol. 115, no. 2, pp. 114-120.

Siddiqui, S. S., Aamodt, E., Rastinejad, F., & Culotti, J. 1989, "Anti-tubulin monoclonal antibodies that bind to specific neurons in Caenorhabditis elegans", *J.Neurosci.*, vol. 9, no. 8, pp. 2963-2972.

Simcock, D. C., Joblin, K. N., Scott, I., Burgess, D. M., Rogers, C. W., Pomroy, W. E., & Simpson, H. V. 1999, "Hypergastrinaemia, abomasal bacterial population densities and pH in sheep infected with Ostertagia circumcincta", *Int.J.Parasitol.*, vol. 29, no. 7, pp. 1053-1063.

Simcock, D. C., Lawton, D. E., Scott, I., & Simpson, H. V. 2006, "Abomasal bacteria produce an inhibitor of gastrin secretion in vitro", *Res.Vet.Sci.*, vol. 81, no. 1, pp. 152-157.

Skelly, P. J., Da'dara, A., & Harn, D. A. 2003, "Suppression of cathepsin B expression in Schistosoma mansoni by RNA interference", *Int.J.Parasitol.*, vol. 33, no. 4, pp. 363-369.

Stein, L. D., Bao, Z., Blasiar, D., Blumenthal, T., Brent, M. R., Chen, N., Chinwalla, A., Clarke, L., Clee, C., Coghlan, A., Coulson, A., D'Eustachio, P., Fitch, D. H., Fulton, L. A., Fulton, R. E., Griffiths-Jones, S., Harris, T. W., Hillier, L. W., Kamath, R., Kuwabara, P. E., Mardis, E. R., Marra, M. A., Miner, T. L., Minx, P., Mullikin, J. C., Plumb, R. W., Rogers, J., Schein, J. E., Sohrmann, M., Spieth, J., Stajich, J. E., Wei, C., Willey, D., Wilson, R. K., Durbin, R., & Waterston, R. H. 2003, "The genome sequence of Caenorhabditis briggsae: a platform for comparative genomics", *PLoS.Biol.* , vol. 1, no. 2, p. E45.

Sternberg, P. W. Genome Sequence of Additional Caenorhabditis species: Enhancing the Utility of C.elegans as a Model Organism. 2003.
Ref Type: Generic

Tabara, H., Sarkissian, M., Kelly, W. G., Fleenor, J., Grishok, A., Timmons, L., Fire, A., & Mello, C. C. 1999, "The rde-1 gene, RNA interference, and transposon silencing in C. elegans", *Cell*, vol. 99, no. 2, pp. 123-132.

Tabara, H., Yigit, E., Siomi, H., & Mello, C. C. 2002, "The dsRNA binding protein RDE-4 interacts with RDE-1, DCR-1, and a DExH-box helicase to direct RNAi in C. elegans", *Cell*, vol. 109, no. 7, pp. 861-871.

Taylor, M. A., Coop, R. L., & Wall, R. L. Veterinary parasitology. 2007. Ref Type: Generic

Taylor, M. J. & Hoerauf, A. 1999, "Wolbachia bacteria of filarial nematodes", *Parasitology Today*, vol. 15, no. 11, pp. 437-442.

Thomas, J. H. 2008, "Genome evolution in Caenorhabditis", *Brief.Funct.Genomic.Proteomic.*, vol. 7, no. 3, pp. 211-216.

Ullu, E., Tschudi, C., & Chakraborty, T. 2004, "RNA interference in protozoan parasites", *Cell Microbiol.*, vol. 6, no. 6, pp. 509-519.

Urquhart, G. M., Jarrett, W. F., Jennings, F. W., McIntyre, W. I., & Mulligan, W. 1966, "Immunity to Haemonchus contortus infection: relationship between age and successful vaccination with irradiated larvae", *Am.J.Vet.Res.*, vol. 27, no. 121, pp. 1645-1648.

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.*, vol. 15, no. 8, pp. 747-752.

Visser, A., Geldhof, P., de, M., V, Knox, D. P., Vercruysse, J., & Claerebout, E. 2006, "Efficacy and specificity of RNA interference in larval life-stages of Ostertagia ostertagi", *Parasitology*, vol. 133, no. Pt 6, pp. 777-783.

Wang, J. & Barr, M. M. 2005, "RNA interference in Caenorhabditis elegans", *Methods Enzymol.*, vol. 392, pp. 36-55.

Waterston, R. H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J. F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., An, P., Antonarakis, S. E., Attwood, J., Baertsch, R., Bailey, J., Barlow, K., Beck, S., Berry, E., Birren, B., Bloom, T., Bork, P., Botcherby, M., Bray, N., Brent, M. R., Brown, D. G., Brown, S. D., Bult, C., Burton, J., Butler, J., Campbell, R. D., Carninci, P., Cawley, S., Chiaromonte, F., Chinwalla, A. T., Church, D. M., Clamp, M., Clee, C., Collins, F. S., Cook, L. L., Copley, R. R., Coulson, A., Couronne, O., Cuff, J., Curwen, V., Cutts, T., Daly, M., David, R., Davies, J., Delehaunty, K. D., Deri, J., Dermitzakis, E. T., Dewey, C., Dickens, N. J., Diekhans, M., Dodge, S., Dubchak, I., Dunn, D. M., Eddy, S. R., Elnitski, L., Emes, R. D., Eswara, P., Eyras, E., Felsenfeld, A., Fewell, G. A., Flicek, P., Foley, K., Frankel, W. N., Fulton, L. A., Fulton, R. S., Furey, T. S., Gage, D., Gibbs, R. A., Glusman, G., Gnerre, S., Goldman, N., Goodstadt, L., Grafham, D., Graves, T. A., Green, E. D., Gregory, S., Guigo, R., Guyer, M., Hardison, R. C., Haussler, D., Hayashizaki, Y., Hillier, L. W., Hinrichs, A., Hlavina, W., Holzer, T., Hsu, F., Hua, A., Hubbard, T., Hunt, A., Jackson, I., Jaffe, D. B., Johnson, L. S., Jones, M., Jones, T. A., Joy, A., Kamal, M., Karlsson, E. K., Karolchik, D., Kasprzyk, A., Kawai, J., Keibler, E., Kells, C., Kent, W. J., Kirby, A., Kolbe, D. L., Korf, I., Kucherlapati, R. S., Kulbokas, E. J., Kulp, D., Landers, T., Leger, J. P., Leonard, S., Letunic, I., Levine, R., Li, J., Li, M., Lloyd, C., Lucas, S., Ma, B., Maglott, D. R., Mardis, E. R., Matthews, L., Mauceli, E., Mayer, J. H., McCarthy, M., McCombie, W. R., McLaren, S., McLay, K., McPherson, J. D., Meldrim, J., Meredith, B., Mesirov, J. P., Miller, W., Miner, T. L., Mongin, E., Montgomery, K. T., Morgan, M., Mott, R.,

Mullikin, J. C., Muzny, D. M., Nash, W. E., Nelson, J. O., Nhan, M. N., Nicol, R., Ning, Z., Nusbaum, C., O'Connor, M. J., Okazaki, Y., Oliver, K., Overton-Larty, E., Pachter, L., Parra, G., Pepin, K. H., Peterson, J., Pevzner, P., Plumb, R., Pohl, C. S., Poliakov, A., Ponce, T. C., Ponting, C. P., Potter, S., Quail, M., Reymond, A., Roe, B. A., Roskin, K. M., Rubin, E. M., Rust, A. G., Santos, R., Sapojnikov, V., Schultz, B., Schultz, J., Schwartz, M. S., Schwartz, S., Scott, C., Seaman, S., Searle, S., Sharpe, T., Sheridan, A., Shownkeen, R., Sims, S., Singer, J. B., Slater, G., Smit, A., Smith, D. R., Spencer, B., Stabenau, A., Stange-Thomann, N., Sugnet, C., Suyama, M., Tesler, G., Thompson, J., Torrents, D., Trevaskis, E., Tromp, J., Ucla, C., Ureta-Vidal, A., Vinson, J. P., Von Niederhausern, A. C., Wade, C. M., Wall, M., Weber, R. J., Weiss, R. B., Wendl, M. C., West, A. P., Wetterstrand, K., Wheeler, R., Whelan, S., Wierzbowski, J., Willey, D., Williams, S., Wilson, R. K., Winter, E., Worley, K. C., Wyman, D., Yang, S., Yang, S. P., Zdobnov, E. M., Zody, M. C., & Lander, E. S. 2002, "Initial sequencing and comparative analysis of the mouse genome", *Nature*, vol. 420, no. 6915, pp. 520-562.

Williams, B. D. 1995, "Genetic mapping with polymorphic sequence-tagged sites", *Methods Cell Biol.*, vol. 48, pp. 81-96.

Wilson, L., Panda, D., & Jordan, M. A. 1999, "Modulation of microtubule dynamics by drugs: a paradigm for the actions of cellular regulators", *Cell Struct.Funct.*, vol. 24, no. 5, pp. 329-335.

Winston, W. M., Sutherlin, M., Wright, A. J., Feinberg, E. H., & Hunter, C. P. 2007, "Caenorhabditis elegans SID-2 is required for environmental RNA interference", *Proc.Natl.Acad.Sci.U.S.A*, vol. 104, no. 25, pp. 10565-10570.

Winter, A. D., Myllyharju, J., & Page, A. P. 2003, "A hypodermally expressed prolyl 4-hydroxylase from the filarial nematode Brugia malayi is soluble and active in the absence of protein disulfide isomerase", *J.Biol.Chem.*, vol. 278, no. 4, pp. 2554-2562.

Wolstenholme, A. J., Fairweather, I., Prichard, R., Samson-Himmelstjerna, G., & Sangster, N. C. 2004, "Drug resistance in veterinary helminths", *Trends Parasitol.*, vol. 20, no. 10, pp. 469-476.

Yadav, B. C., Veluthambi, K., & Subramaniam, K. 2006, "Host-generated double stranded RNA induces RNAi in plant-parasitic nematodes and protects the host from infection", *Mol.Biochem.Parasitol.*, vol. 148, no. 2, pp. 219-222.

Yin, Y., Martin, J., Abubucker, S., Scott, A. L., McCarter, J. P., Wilson, R. K., Jasmer, D. P., & Mitreva, M. 2008, "Intestinal Transcriptomes of Nematodes: Comparison of the Parasites Ascaris suum and Haemonchus contortus with the Free-living Caenorhabditis elegans", *PLoS.Negl.Trop.Dis.*, vol. 2, no. 8, p. e269.

Zarlenga, D. S., Rosenthal, B., Hoberg, E., & Mitreva, M. 2008, "Integrating genomics and phylogenetics in understanding the history of Trichinella species", *Vet.Parasitol.*