The cytochrome P450 family in the parasitic nematode *Haemonchus contortus*

Roz Laing BSc (Hons) BVMS

Institute of Infection and Immunity
Faculty of Veterinary Medicine
University of Glasgow

Submitted in fulfilment of the requirements for the degree of Doctor of Philosophy at the University of Glasgow

September 2010

Abstract

Haemonchus contortus, a parasitic nematode of sheep, is unsurpassed in its ability to develop resistance to the anthelmintic drugs used as the mainstay of its control. A reduction in drug efficacy leads to prophylactic and therapeutic failure, resulting in loss of productivity and poor animal welfare. This situation has reached crisis point in the sheep industry, with farms forced to close their sheep enterprises due to an inability to control resistant nematodes.

The mechanisms of anthelmintic resistance are poorly understood for many commonly used drugs. Altered or increased drug metabolism is a possible mechanism, yet has received little attention despite the clear role of xenobiotic metabolism in pesticide resistance in insects. The cytochrome P450s (CYPs) are a large family of drug-metabolising enzymes present in all species. Their expression is induced on exposure to their substrate and over-expression of a single CYP has been shown to confer multi-drug resistance in insects.

The *H. contortus* genome is currently being sequenced and assembled at the Wellcome Trust Sanger Institute, Cambridge. Despite the lack of a completed genome, the public provision of read, contig and supercontig databases has facilitated the identification of 73 partial gene sequences representing a large family of *H. contortus* CYPs. Their constitutive expression is highest in larval stages although adult expression was also detected. The majority of CYPs are most highly expressed in the worm intestine, which is thought to be the main organ of detoxification in nematodes and is consistent with a role in xenobiotic metabolism. A small number of CYPs were more highly expressed in anthelmintic resistant isolates than in an anthelmintic-susceptible isolate and may represent candidate genes for further research. The identification of putative *H. contortus* orthologues of the *Caenorhabditis elegans* nuclear hormone receptors controlling CYP transcription and the cytochrome P450 reductase gene catalysing electron transfer to CYPs suggests that regulatory and functional pathways may be conserved between the species.

Transcriptome analysis using next generation sequencing was undertaken to guide a pilot annotation of 590 Kb genomic sequence. A high degree of

conservation was observed between the conceptual translations of *H. contortus* and *C. elegans* genes, although at a genomic level, *H. contortus* consistently had a larger number and size of introns, which may reflect a larger genome than previously predicted. Gene order was not conserved, although regions of microsynteny were present and a bias for intra-chromosomal rearrangements resulted in putative orthologues frequently residing on the corresponding chromosome in both species. Partial conservation of a number of *C. elegans* operons in *H. contortus* was identified. These findings have important implications for the *H. contortus* genome project and the transcriptome databases provide a valuable resource for future global comparisons of gene expression.

Contents

List of Tables

List of Figures

Acknowledgement

Firstly, I would like to thank my supervisor, John Gilleard, for his support, guidance, encouragement and optimism. I would also like to thank my assessor Brian Shiels and to acknowledge Quality Meat Scotland (QMS), English Beef and Sheep Sectors (EBLEX), Hybu Cig Cymru (HCC) and Pfizer for funding my studentship.

I would like to thank my colleagues and friends at the Moredun Research Institute, Edinburgh, particularly Alison Donnan, Dave Bartley and Frank and Liz Jackson; at The Wellcome Trust Sanger Institute, Cambridge, particularly Anna Protasio, Martin Hunt, Mike Quail and Matt Berriman; and Robin Beech at McGill University, Montreal. All of these people have helped me enormously by providing material and/or expertise and all of whom, without fail, have been supportive, enthusiastic and generous with their time. I am grateful to Professor Doug Jasmer and members of his lab at Washington State University for their provision of *H. contortus* intestinal cDNA.

I would like to thank the staff and students of the Institute of Infection and Immunity at the University of Glasgow, particularly Eileen Devaney and Andy Tait for guidance and advice, Libby Redman and Fiona Whitelaw for microsatellite analysis, William Weir for technical support, Gary Saunders for wisdom on all matters, and my great friends Steven, Vicky and Joana. I would also like to thank the Gilleard family for their hospitality during my stay in Calgary and Val Wood for her hospitality during my stays in Cambridge.

Finally, I must also thank my family, JC and Jack, who will be as happy as I am to see the end of this thesis.

Declaration

The work presented in this thesis was performed entirely by the author except where indicated. This thesis contains unique work and will not be submitted for any other degree, diploma or qualification at any other university.

Roz Laing, September 2010.

Definitions/ Abbreviations

AAD	Amino-acetonitrile derivative
ABZ	Albendazole
ABZ-SO	Albendazole sulphoxide
ABZ-SOO	Albendazole sulphone
AChR	Acetylcholine receptor
ACT	Artemis comparison tool
BAC	Bacterial artificial chromosome
BAM	Binary alignment/map
BLAST	Basic local alignment search tool
BWA	Burrows-Wheeler aligner
BZ	Benzimidazole
CAR	Constitutive androstane receptor
CYP	Cytochrome P450
EHT	Egg hatch test
EST	Expressed sequence tag
FECRT	Faecal egg count reduction test
FLU	Flubendazole
FMO	Flavin-containing mono-oxygenases

GABA	γ-aminobutyric acid
GFP	Green fluorescent protein
GluCl	Glutamate-gated chloride channel
GST	Glutathione S transferase
HP-LC-MS	High performance liquid chromatography and mass spectrometry
IVM	Ivermectin
LDT	Larval development test
LEV	Levamisole
MAQ	Mapping and assembly with qualities
ML	Macrocyclic lactone
MXD	Moxidectin
NAT	N-acetyltransferase
NGS	Next generation sequencing
NR	Nuclear hormone receptor
PB	Phenobarbital
PGP	P-glycoprotein
PXR	Pregnane X receptor
RT-QPCR	Real-time quantitative polymerase chain reaction
SSAHA	Sequence search and alignment by hashing algorithm

SAM	Sequence alignment/map
SL	Spliced-leader
SULT	Sulphotransferase
TBZ	Triclabendazole
Tm	Melting temperature
UGT	Uridine dinuceotide phosphate glucuronosyl transferase
WTSI	Wellcome Trust Sanger Institute

1 General Introduction

Parasitic disease affects humans and livestock worldwide. More than a third of the world's population are infected with parasites and one person in four harbours a hookworm infection (www.who.int/wormcontrol/en/). An estimated 300 million people are severely ill with parasitic disease, but as the infections are often insidious in nature, the impact of chronic ill health is thought to be even greater.

Parasitism in livestock further impacts on human health by resulting in significant production losses for stock keepers throughout the world relying on livestock as a source of income and nutrition (Krecek and Waller 2006). Many parasitic diseases are also zoonoses (www.who.int/zoonoses/en/).

The negative effect of parasitism on animal welfare is also significant. Chronic infections with gastrointestinal (GI) nematodes are common in sheep and cattle worldwide, causing weight loss, diarrhoea, inappetance, depression, anaemia and hypoproteinaemia, with acute infections resulting in death (Holmes 1985).

1.1 Drug resistance in parasites

The control of parasitic disease in livestock generally relies on strategic dosing with anthelmintics. However, resistance to the three most widely used classes of anthelmintic drugs: the benzimidazoles, the imidazothiazoles and the macrocyclic lactones, is now widespread and resistance to the two newer classes, the amino-acetonitrile derivatives (AADs) and paraherquamide derivatives, is expected to follow (Kaminsky 2003, Kaplan 2004, Kaminsky *et al* 2008, Zinser *et al* 2002).

Anthelmintic resistance is currently a major constraint to the sheep industry worldwide (Jackson and Coop 2000, Waller 1997). Although resistance in horses and cattle has not yet reached the level seen in small ruminants, it is an emerging problem (Kaplan 2004). There are only sporadic reports of anthelmintic resistant nematodes in humans, but nonetheless, it is thought to be an important public health concern due to the potential rapidity of its spread (as observed in

nematodes of sheep) and the lack of diagnostic tools and safe effective treatments should it arise (Geerts and Gryseels 2000).

Anthelmintic resistance was first reported in the 1950s, in the sheep nematode *Haemonchus contortus* to the drug phenothiazine (Drudge *et al* 1957). The first benzimidazole (BZ), thiabendazole, was released in 1961 and by 1964 resistance to this drug was reported, again in *H. contortus* (Conway 1964, Drudge *et al* 1964). With the introduction of each new class of anthelmintic, this worrying pattern has continued, with resistance developing within 10 years of each drug being licensed for use (reviewed in Kaplan 2004). Table 1.1 charts the development of resistance to each class of anthelmintics in nematodes of sheep.

Drug class	Drug	Year of approval	First published report of resistance
Benzimidazoles	Thiabendazole	1961	1964
Imidazothiazoles	Levamisole	1970	1979
Macrocyclic lactones	Ivermectin	1981	1988
	Moxidectin	1991	1995

Table 1.1: Anthelmintic resistance in sheep. Adapted from Kaplan (2004)

Resistance to one drug in a class usually confers side-resistance to other members of the same group (Sangster 1999). Sheep nematodes resistant to all three classes have been now reported in the UK (Bartley *et al* 2004, Sargison *et al* 2001, Sargison *et al* 2007, Sargison *et al* 2010). The inability to control such 'triple-resistant' nematodes threatens sheep production globally and triple-resistant *Teladorsagia circumcincta* have already resulted in the closure of a sheep enterprise in South East Scotland (Sargison *et al* 2005).

To confound matters further, once gained, there seems to be no reversion from resistance. In a controlled trial using a PCR genotyping approach, Leignel *et al* (2010) monitored the proportion of BZ-resistant *T. circumcincta* over two years and found this did not alter in fields grazed by lambs treated with levamisole (a non-BZ anthelmintic = counter-selection) or in fields grazed by lambs given no anthelmintic treatment. Further, in fields grazed by lambs treated with BZ alone or with alternate treatments of BZ and levamisole, the proportion of BZ resistant nematodes increased proportionally to the number of BZ treatments, regardless of the regime. These findings suggest neither counter selection nor alternate treatment offer solutions to reverse or halt the evolution of resistance.

1.1.1 Measuring resistance

Anthelmintic resistance can be defined as 'a heritable reduction in the sensitivity of a parasite population to the action of a drug' (Geerts and Gryseels 2000). However, the methods available to accurately detect this reduction in sensitivity are limited to certain parasite species and drug classes (Coles *et al* 2006).

The most widely used method to detect and monitor anthelmintic resistance is the faecal egg count reduction test (FECRT), which compares the number of nematode eggs in the faeces before and after anthelmintic treatment. This test can be used for all classes of anthelmintic drugs but varies in its suitability for different parasite species because the correlation between egg count and parasite burden varies e.g. higher correlation for *H. contortus* but lower for *Trichostrongylus colubriformis, T. circumcincta* and *Nematodirus spp.* (reviewed in Taylor *et al* 2002). But perhaps the main limitation of the FECRT is its lack of sensitivity: studies by Martin *et al* (1989) showed it is only reliable if more than 25% of the worm population is resistant.

Other available assays include the egg hatch test (EHT) for the detection of benzimidazole resistance and the larval development test (LDT) for the detection of benzimidazole and levamisole resistance (Coles *et al* 2006). Similarly to the FECRT, the EHT lacks reliability if less than 25% of the population are benzimidazole resistant (Martin *et al* 1989) and neither the EHT nor LDT are suitable for the detection of macrocyclic lactone resistance (Coles *et al* 2006). More recently, a standardised protocol for a larval migration inhibition test (LMIT) to detect ivermectin resistance has been developed (Demeler *et al* 2010). The test is capable of discriminating between IVM-susceptible and IVM-resistant isolates of *Ostertagia ostertagi, Cooperia oncophora* and *H. contortus* in a reproducible manner, although its suitability for the detection of resistance in mixed populations remains to be evaluated.

Molecular tests are the most sensitive means of detecting anthelmintic resistance, but are currently only available for benzimidazole resistance, as the mechanisms of imidazothiazole and macrocyclic lactone resistance are yet to be fully elucidated (see Section 1.2). Benzimidazole resistance is most commonly

conferred by a phenylalanine to tyrosine mutation at amino acid 200 in β-tubulin isotype-I, which can be detected in resistant but not susceptible populations of *H. contortus* and *T. colubriformis* with allele specific PCR (Kwa *et al* 1994). The frequency of this single nucleotide polymorphism (SNP) in pooled samples of *H. contortus* L3 larvae can also be measured with real-time PCR and pyrosequencing and used to assess the benzimidazole resistance status of different *H. contortus* populations (von Samson-Himmelstjerna *et al* 2009).

Such sensitive molecular tests are desperately needed for the detection of resistance for all classes of anthelmintic, in order to effectively manage and control parasitic disease at an individual farm level and to preserve the efficacy of the anthelmintics globally (Coles *et al* 2006, Gilleard 2006, Gilleard and Beech 2007, Kaplan 2004, Prichard 1990, von Samson-Himmelstjerna and Blackhall 2005, Sangster *et al* 2002, Sargison *et al* 2010). This requires a better understanding of the mode of action of each anthelmintic and knowledge of the underlying mutations that confer resistance.

1.2 Anthelmintics

1.2.1 Benzimidazoles

The benzimidazoles (BZs) are the oldest of the three anthelmintic classes and their widespread use has resulted in the development of BZ-resistance in sheep and horses throughout the world (Kaplan 2004). BZs act by binding to nematode tubulin and disrupting microtubule formation (Prichard 1990). Knowledge of the genetic basis of resistance is most advanced for this group: single nucleotide polymorphisms (SNPs) encoding an amino acid change from phenylalanine to tyrosine at codons 167 and 200 of β-tubulin isotype-I appear to play a major role in the mechanism of BZ resistance (Kwa *et al* 1994). The significance of each SNP may vary between species (von Samson-Himmelstjerna *et al* 2003) and similar mutations in β-tubulin isotype-II may also confer BZ resistance (Prichard, 2001) (as can deletion of β-tubulin isotype-II (Kwa *et al* 1993)), but the TCC to TAC mutation at codon 200 appears to be a reliable marker for BZ-resistance in *H. contortus* and *T. colubriformis* (Kwa *et al* 1994, von Samson-Himmelstjerna *et al* 2009).

1.2.2 Imidazothiazoles

Imidazothiazoles such as levamisole (LEV) were introduced to the market in the 1970s. Their major mode of action is as nicotinic acetylcholine receptor (AChR) agonists, which result in opening of cationic (Na^+ and K^+) channels and depolarisation of nematode muscle cell membranes (Prichard 2001).

Resistance to LEV is thought to be due to changes in the AChR (Sangster *et al* 1998). In *C. elegans*, AChRs are formed by the association of five essential subunits encoded by the genes *unc-38, unc-63, lev-8, unc-29* and *lev-1* (Boulin *et al* 2008). In *H. contortus, T. circumcincta* and *T. colubriformis*, orthologues or paralogues of all genes other than *lev-8* have been identified and a comparison of LEV-susceptible and LEV-resistant nematodes identified the expression of an abbreviated *unc-63* transcript associated with the resistant phenotype in all three species (Neveu *et al* 2010).

1.2.3 Macrocyclic lactones

The macrocyclic lactone (ML) class consists of two drug groups sharing a similar structure and action: the avermectins, such as ivermectin (IVM), and the milbemycins, such as moxidectin (MXD). Drugs from both groups are widely used in veterinary medicine and IVM is the core drug for the treatment and control of onchocerciasis (river blindness) in humans (Taylor and Greene 1989).

MLs are thought to target ligand-gated chloride channels, acting as γ-aminobutyric acid (GABA) receptor antagonists (Holden-Dye and Walker 1990) and glutamate-gated chloride (GluCl) channel agonists (Cully *et al* 1994, Dent *et al* 1997). The drug concentration required for a GABA-antagonistic effect is considerably higher than for a GluCl-agonist effect so the latter may be the clinically relevant mode of action (Yates *et al* 2003).

The effect of IVM-exposure on the parasite is paralysis of the pharynx, body wall and uterus, with the pharyngeal muscles affected at a lower concentration than the somatic muscles (Geary *et al* 1993), which may relate to differing sensitivities and distributions of ligand-gated chloride channels (Prichard 2001). In *C. elegans*, the sensitivity of the pharynx to IVM is thought to be dependant on

the presence of a GluCl α-subunit, *avr-15* (Dent *et al* 1997). High-level IVM resistance is observed in *avr-15* mutants with concurrent mutations in two additional genes encoding GluCl α-subunits, *avr-14* and *glc-1*, but little or no resistance is conferred if mutations are present in only two of the three genes (Dent *et al* 2000).

In *H. contortus*, resistance to the MLs has been associated with selection at both GluCl channel and GABA receptor loci (Blackhall *et al* 1998b,Blackhall *et al* 2003) as well as at the locus of a P-glycoprotein (PGP), a multi-drug efflux pump (Blackhall *et al* 1998a). Increased expression of a PGP has also been detected in IVM-selected strains of the parasite (Xu *et al* 1998). None of these mutations have, however, been found to be widespread in field populations of IVM-resistant *H. contortus* (Gilleard and Beech 2007), so the mechanism of resistance disrupting anthelmintic therapy in the field remains unclear.

1.2.4 New anthelmintics

Subsequent to the emergence of resistance to the three main classes of anthelmintic, two new anthelmintics with differing modes of action have been licensed and marketed for use in small ruminants. The first, monepantel, is an amino-acetonitrile derivative (AAD) and targets the nematode-specific *deg-3* acetylcholine receptor (AChR) family (Kaminsky *et al* 2008, Rufener *et al* 2009). Monepantel is currently effective against nematodes resistant to all other classes of anthelmintic in the field, but *in vitro* experiments have shown *H. contortus* exposed to sub-lethal concentrations of the drug can develop resistance through loss-of-function *Hco-mptl-1* gene mutations (Rufener *et al* 2009).

The second anthelmintic, derquantel, is a paraherquamide derivative and targets nicotinic AChRs. Unlike the imidazothiazoles, which are nicotinic AChR agonists, derquantel has an antagonistic action and preferentially targets the B-subtype of the receptor rather than the L-subtype target of LEV (Qian *et al* 2006, Zinser *et al* 2002). Derquantel has been licensed as a combination product with the ML abamectin, and this formulation is currently effective against nematodes resistant to other classes of anthelmintic (Little *et al* 2010).

1.3 Alternative methods of parasite control

All anti-parasitic drugs will be expected, with time, to succumb to anthelmintic resistance, so it is vitally important that the efficacy of currently available drugs is preserved by developing sustainable programmes for their use (Jackson and Coop 2000, von Samson-Himmelstjerna and Blackhall 2005). Management practices such as reducing the frequency of treatment, avoiding under-dosing and rotating the class of anthelmintic used may slow the development of resistance (Jackson and Coop 2000) but novel methods of parasite control, which do not rely on chemotherapeutics, could offer a valuable alternative or adjunct to anthelmintic treatment (Sayers and Sweeney 2005). These include the breeding of resilient or nematode-tolerant sheep, biological control with nematophagous microfungi (Larsen 2006), and nematode vaccines (Knox *et al* 2003). However, none of these methods are suitable for the treatment of severe life-threatening parasite infection and for the present and near future, the mainstay of treatment is likely to depend on the use of chemotherapeutics (Kaplan 2004).

1.4 The parasitic nematode *H. contortus*

Haemonchus contortus is the most important nematode of small ruminants worldwide. It is a blood-feeding gastro-intestinal nematode causing severe disease and loss of production in sheep, goats and cattle (Urquhart *et al* 1996). As a clade V Strongyloid (see Figure 1.1), it shares a close phylogenetic relationship with a number of the most economically significant parasites of grazing ruminants, with human hookworm species, and with the model nematode *Caenorhabditis elegans* (Blaxter *et al* 1998, Gilleard 2004).

Infection with *H. contortus* is most common in the tropics and subtropics and other regions dominated by summer rainfall, but is becoming increasingly prevalent in Europe (O'Connor *et al* 2006). Larvae can undergo arrested development within the host at the L4 stage to survive cold winters, resulting in sporadic outbreaks of haemonchosis in the UK, France, Denmark, Sweden and the Netherlands, but the impact of milder wetter winters and ineffective worming practices in the face of increasing resistance are thought to have

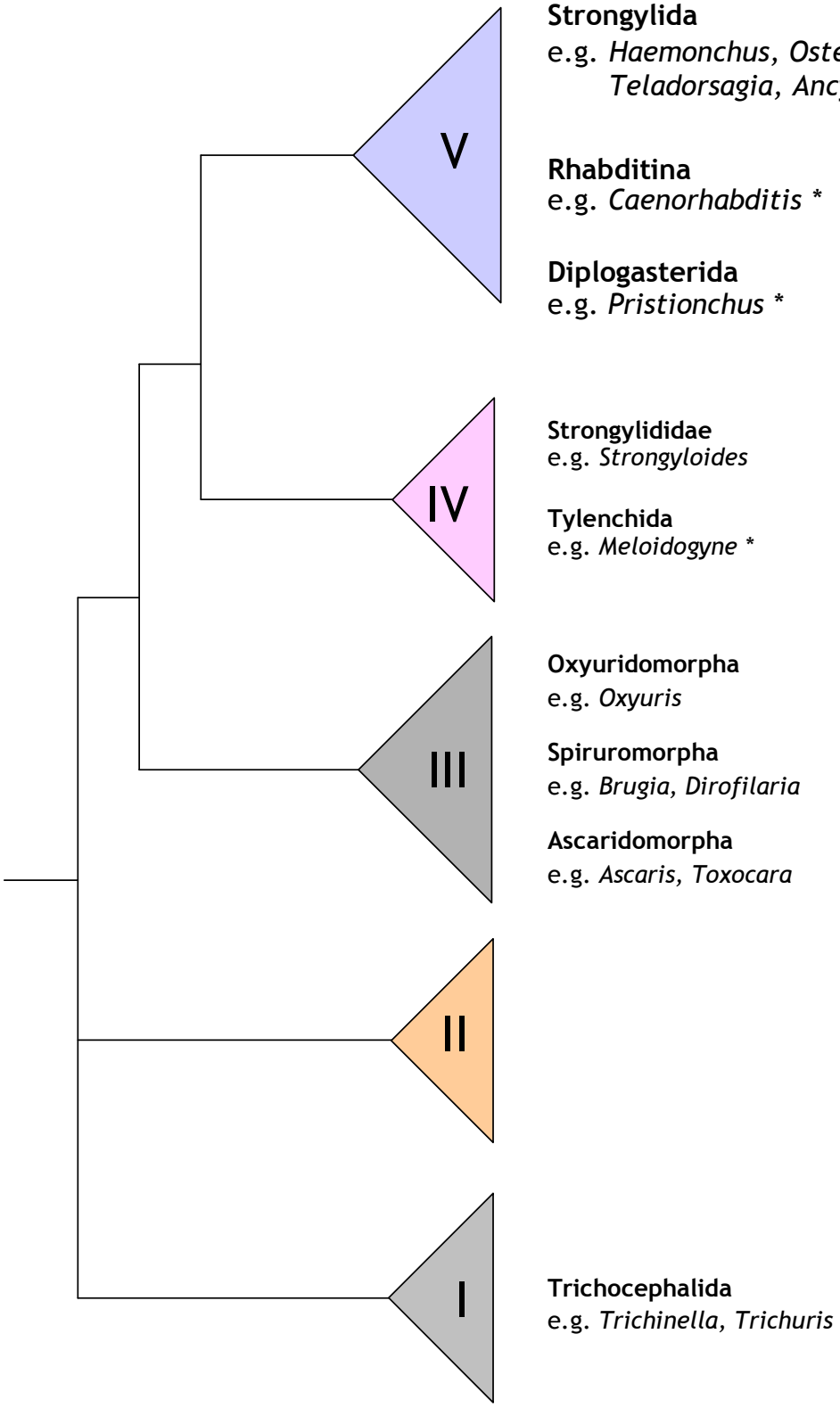


Figure 1.1: Nematode phylogenetic tree showing parasite species affecting animals and humans. Adapted from Blaxter et al 1998 and Coghlan 2005. Roman numerals relate to five clades of the phylum Nematoda deduced from analysis of small subunit ribosomal DNA sequence. Parasites that do not infect animals or humans but are discussed in this thesis are marked *.

contributed to an increasing economic importance of the parasite in these regions (Waller *et al* 2006).

H. contortus has a typical trichostrongyloid lifecycle (Urquhart *et al* 1996): adult worms reside in the abomasum and produce eggs which are released in the faeces. These hatch and develop into L1, L2 and then infective L3 larvae, which are ingested by the grazing ruminant. L3 larvae exsheath in the host abomasum then moult twice to become L4 larvae then adult worms (see Figure 1.2). Temperature and moisture are the most important factors in the development of eggs to the infective L3 stage, which can occur in as little as three days under optimum conditions (O'Connor *et al* 2006). The pre-patent period is 2-3 weeks in sheep and four weeks in cattle (Urquhart *et al* 1996). As mentioned, the L4 stage can undergo arrested development in adverse conditions e.g. host immunity, then spontaneously resume development when exposed to a trigger e.g. host immune suppression. This can result in clinical haemonchosis 4-6 months after infection.

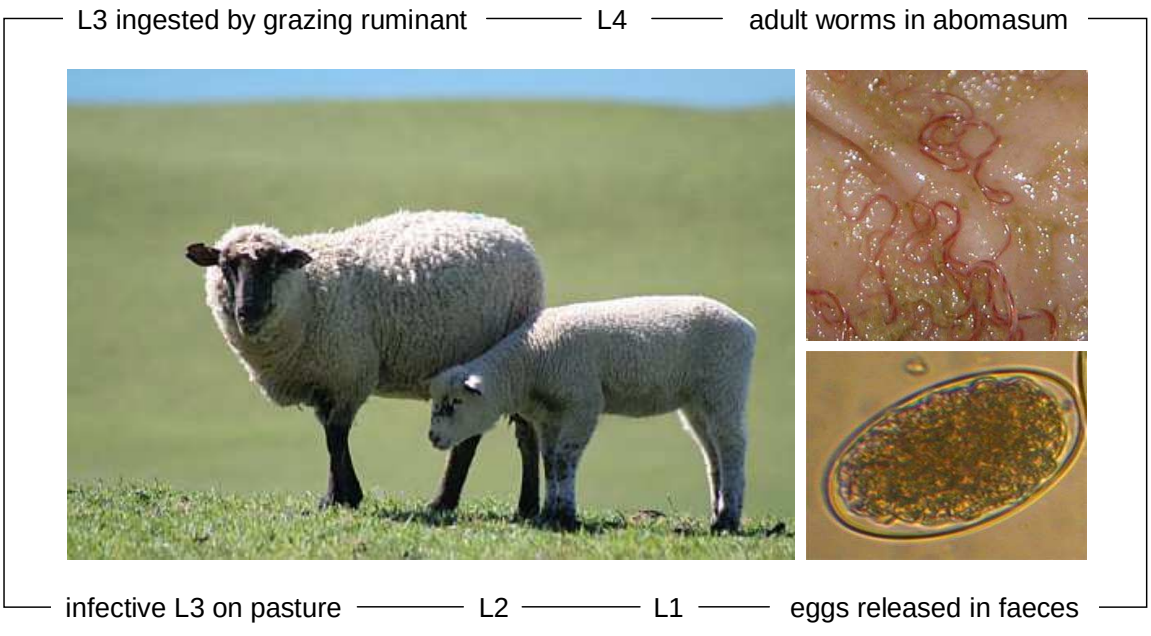


Figure 1.2: *H. contortus* lifecycle. Adult worms reside in the abomasum and produce eggs which are released in the faeces. Eggs hatch and develop on the pasture to the infective L3 stage, which is ingested by the grazing ruminant. This is a typical trichostrongyloid direct lifecycle with a 2-3 week pre-patent period in sheep.

H. contortus adult worms are prolific breeders: one female can produce up to 10,000 eggs per day (Prichard 2001) and each host can harbour thousands of

adult worms. Infection with *H. contortus* results in hyperacute, acute or chronic disease depending on the worm burden. Infection with up to 30,000 worms can cause hyperacute haemonchosis with severe haemorrhagic gastritis resulting in sudden death. More commonly, infections of 2000-20,000 (acute haemonchosis) result in severe anaemia, hypoproteinaemia, oedema and death. Chronic haemonchosis, from infections of several hundred worms, results in progressive weight loss and weakness, while subclinical infections may manifest only in reduced productivity. Diarrhoea is not normally observed (Urquhart *et al* 1996).

Anthelmintics are the main weapon in the treatment and prevention of haemonchosis but, as described above, *H. contortus* is frequently the first nematode to develop resistance to the drugs used in its control. The mechanisms by which anthelmintic resistance evolves are unclear. It may be a pre-adaptive phenomenon with the genes conferring resistance existing in the population before exposure to the drug (Jackson and Coop, 2000) in which case, the extremely high level of genetic diversity identified in different populations of *H. contortus* (see below) may explain the parasite's propensity to rapidly develop resistance. Alternative mechanisms include a new mutation occurring and spreading through the population under selection pressure, or multiple recurrent mutations arising in different populations under selection (Gilleard and Beech 2007, Skuce *et al* 2010). Again, the genetic diversity of *H. contortus* is suggestive of a highly plastic genome, which would be predicted to promote the development of resistance by these methods too.

1.4.1 Genetic diversity of *H. contortus*

Extremely high levels of genetic diversity have been reported in many species of nematodes (reviewed in: Blouin 1998, Gilleard and Beech 2007). In the case of *H. contortus*, this high genetic diversity is predicted to be a result of a large population size, rapid rate of reproduction and wide geographical spread (Prichard 2001). The naturally large population size of *H. contortus* is thought to be further increased by host movement; work by Blouin *et al* (1995) found the variation in mitochondrial DNA sequence in *H. contortus* populations across four North American states was higher within populations than between populations which would be consistent with a very high gene flow and thus a large effective

population size. With the limited range of dispersal of the free-living stages, this was predicted to be a result of extensive movement of sheep between farms/states. The worrying implication of this for the spread of resistance alleles was also highlighted.

Analysis of mitochondrial DNA sequence and amplified fragment length polymorphism (AFLP) has since identified high genetic differentiation between *H. contortus* populations from 14 different countries, but again genetic difference between different populations from the same continent was lower than that between individuals of the same population suggestive of a large effective population size (Troell *et al* 2006). More recently, microsatellite analysis of five different laboratory isolates showed a high degree of genetic differentiation, concurrent with the expected barrier to gene flow implicit in the maintenance of laboratory isolates (Redman *et al* 2008b).

1.4.2 *H. contortus* isolates

The nomenclature described by Redman *et al* (2008b) has been adopted for this project. The prefix 'MHco' identifies isolates which have been maintained by experimental infection at The Moredun Research Institute.

The MHco1 (MOSI) isolate, also known as SE, was originally isolated in East Africa in the 1950s. This isolate is fully susceptible to all classes of anthelmintic.

MHco3 (ISE) is an inbred laboratory isolate derived from progeny of a single adult female of the MHco1 (MOSI) isolate. This is the reference isolate for the *H. contortus* genome project (see Section 1.4.3).

MHco4 (WRS) is the White River Strain isolate from South Africa and is resistant to ivermectin, benzimidazoles, rafoxanide and closantel (van Wyk and Malan 1988).

MHco10 (CAVR) is the Chiswick Avermectin Resistant Strain, a field isolate from Australia, which is resistant to ivermectin. It also shows a low level of resistance to the benzimidazoles, but is sensitive to levamisole and closantel. Males of the

MHco10 (CAVR) isolate appear to be more sensitive to the effects of ivermectin than females (Le Jambre *et al*, 1995).

1.4.3 The *H. contortus* genome project

H. contortus has a haploid genome of five pairs of autosomal chromosomes with two X chromosomes in the female and one in the male (Redman *et al* 2008a). Leroy *et al* (2003) predicted the size of the genome to be ~53 Mb using flow cytometry, but current genome assembly data suggest it may be closer to 300 Mb (pers. com. Dr M Berriman).

The *H. contortus* genome project, an open resource provided by The Wellcome Trust Sanger Institute (WTSI), is sequencing and analysing the nuclear genome of the adult MHco3 (ISE) isolate (www.sanger.ac.uk/Projects/H_contortus). To date, ~1.25Gb reads of genomic sequence have been generated with a combination of capillary and 454 sequencing. This is currently available as a read database, a contig database (N50=2370) and a supercontig database (N50=9901), where N50 represents the size of the smallest contig in the minimum set of largest contigs whose sizes total 50% of all sequence in the database. An additional 2.24Mb sequence has been generated with capillary sequencing of 22 Bacterial Artificial Chromosome (BAC) inserts.

The concurrent WTSI *H. contortus* EST project has generated a database of 25,000 cDNAs. The ability to detect over 95% of the ~6000 single cluster ESTs in the *H. contortus* genomic databases suggest coverage is good (pers. com. Dr G Saunders), so the N50 values for both the contig and supercontig databases are unexpectedly small. This may reflect a high level of genetic variation present within the population, despite the use of the inbred MHco3 (ISE) isolate.

1.4.4 *C. elegans* as a model nematode

Genomic and technical resources for parasitic nematodes are greatly lacking (Geary and Thompson 2001, Gilleard 2004, Hashmi *et al* 2001). By and large, their development has been impeded by the lack of an *in vitro* culture system, so the availability of a free-living model nematode is a valuable alternative.

C. elegans has many innate qualities that make it suitable for use as a model organism. These include its ease of propagation, short generation cycle, large brood size, short lifespan, and transparency at all stages of development. It now has a fully sequenced genome, freely accessible online, and techniques for chemical mutagenesis and RNA interference (RNAi) have been developed to facilitate forward and reverse genetics respectively (Silverman *et al* 2009). *C. elegans* has also proven useful as a heterologous transformation system, expressing parasite promoter sequences and genes to study expression and function (reviewed in: Gilleard 2004, Hashmi *et al* 2001).

The major factors dictating the suitability of *C. elegans* as a model are the evolutionary distance between it and the species of interest, and the particular aspect of biology under investigation. Blaxter *et al* (1998) divided the phylum Nematoda into five clades based on phylogenetic analysis of 53 small subunit ribosomal DNA sequences, with *Caenorhabditis* residing in clade V (see Figure 1.1). It is likely that work with *C. elegans* will be highly applicable to other nematodes residing in clade V (Gilleard 2004) such as *Ancylostoma, Necator, Haemonchus, Ostertagia, Trichostrongylus* and *Nippostrongylus,* all of which are important parasites of humans or animals.

This however, does not preclude the use of *C. elegans* as a model for more distantly related species, as core biological pathways may be widely conserved. When utilised as a model for human pathology, homologues in *C. elegans* are identifiable for 40-75% human disease-related genes (reviewed in Silverman *et al* 2009).

For comparative studies in parasitic nematodes, *C. elegans* offers the advantage of sharing characteristics of interest such as a protective cuticle and development into a dauer stage (Hashmi *et al* 2001) and, as highlighted by Gilleard (2004), many biological characteristics that are shared between *C. elegans* and parasitic nematodes are not conserved in mammals, and as such are clear targets for control strategies. In contrast, the model worm is less likely to be useful for characteristics associated with a parasitic lifestyle, but the elucidation of these differences themselves may be of vital use in the identification of elements essential to parasitism.

Most of our understanding of anthelmintic mode of action was derived from investigations of drug resistance in *C. elegans*. The mechanisms of BZ, ML and levamisole function were all elucidated through work on the model worm. In the case of research into anthelmintic resistance, it may not be possible to directly extrapolate *C. elegans* data for parasitic nematodes since a species living in a controlled laboratory environment may tolerate resistance phenotypes that are incompatible with a parasitic lifestyle (Geary and Thompson, 2001). However, this approach can be used to identify candidate mutations for further investigation and the growing body of research based on comparative and transgenic studies with *C. elegans* continues to prove its value as a model organism. For closely related species such as *H. contortus*, the application of *C. elegans* genetic and genomic tools is expected to be particularly enlightening.

1.5 Drug transfer and metabolism in nematodes

1.5.1 Drug absorption

Anthelmintic drugs can enter parasitic nematodes by either oral ingestion or transcuticular diffusion, but the latter is thought to be most important (Alvarez *et al* 2007). This is evidenced by the high concentrations of the BZ albendazole (ABZ) measured in adult *H. contortus* recovered from infected treated sheep despite the lack of detectible ABZ in the peripheral circulation of the host and only low levels detectible in the portal blood (Alvarez *et al* 2000). Passive diffusion occurs across the lipid components of the cuticle and the rate of transfer is thought to be dependant on the concentration gradient and lipophilicity of the specific drug (Alvarez *et al* 2001, Alvarez *et al* 2007).

1.5.2 Drug metabolism and elimination

There are three stages to the detoxification of xenobiotics: phase I metabolism, phase II metabolism, and excretion. Phase I metabolism increases the solubility of the target drug, usually by adding or uncovering a hydrophilic group and phase II metabolism further increases solubility of the xenobiotic with the conjugation of an exogenous compound. The third phase is active transport of the metabolite from the cell.

Cytochrome P450 enzymes (CYPs) are the main enzymes involved in phase I metabolism and oxidation reactions are most common (Cvilink *et al* 2009, Guengerich 2006). Other enzymes with primarily oxidase functions such as peroxidases, flavin-containing mono-oxygenases (FMOs), monoamine oxidase and xanthine oxidase may also catalyse phase I metabolism, as may enzymes with primarily endogenous functions such as hydrolases, medium and short chain dehydrogenases and aldo-keto reductases.

In phase II metabolism, the solubility of the xenobiotic is increased further with the conjugation of glucuronic acid or glutathione catalysed by uridine dinucleotide phosphate glucuronosyl transferases (UGTs) and glutathione S transferases (GSTs) respectively. After the CYPs, the UGTs are the enzymes most commonly involved in xenobiotic metabolism (Cvilink *et al* 2009, Guengerich 2006). Alternative phase II reactions include acetylation by N-acetyltransferase (NAT) and sulphation by sulphotransferases (SULTs).

The resulting metabolite is then exported from the cell by ATP-binding cassette (ABC) transporters such as P-glycoproteins (PGPs). Excretion can occur after phase I metabolism and phase II metabolism can occur independently of phase I metabolism, but the full chronological succession of steps is thought to be the norm. The relative contribution of phase I and phase II reactions will vary depending on the structure of the individual compound (Glue and Clement 1999).

The CYPs, UGTs and GSTs are considered ubiquitous, as they have been identified in all organisms that have been tested for their presence (Cvilink *et al* 2009). In *C. elegans*, 80 genes encoding CYPs, 72 genes encoding UGTs, 48 genes encoding GSTs and 68 genes encoding short chain dehydrogenases genes have been identified (Gotoh 1998, Lindblom and Dodd 2006).

Enzymes involved in xenobiotic metabolism have their expression induced on exposure to their substrate (see Section 1.6.5). In *C. elegans*, this has facilitated investigation of pathways involved in drug metabolism through analysis of gene up-regulation in response to xenobiotic exposure. Menzel *et al* (2001) used a whole genome microarray approach to show induction of members of the *C. elegans* CYP, UGT and GST families in response to exposure to β-naphthoflavone.

In humans and animals, the enzymes and transporters involved with xenobiotic metabolism and their functions in biotransformation pathways have been well studied (de Groot 2006, Guengerich 2006). In nematodes, the identification of these proteins and/or orthologues of the genes that encode them and evidence of increased activity with exposure to a xenobiotic is suggestive that the same pathways may be conserved. However, there have only been a limited number of functional studies undertaken to confirm this. Further, the identification of a number of atypical metabolites (not seen in humans and animals) indicates that there may be nematode-specific and even nematode-species-specific pathways of biotransformation, warning against the assumption that knowledge of a pathway of xenobiotic metabolism in one species will be transferable to another (Cvilink *et al* 2009).

1.6 Cytochrome P450s (CYPs)

The CYPs are an ancient superfamily of enzymes, playing key roles in many biotransformation pathways. They act on both endogenous and exogenous substrates, catalysing steps in the biosynthesis and catabolism of steroids, retinoids, prostaglandins, bile acids and fatty acids and the detoxification of drugs, toxins and insecticides (Mansuy 1998, Nebert 1994, Nelson *et al* 1993, Thomas 2007). Their implication in the tolerance of and resistance to a range of xenobiotics in many organisms make them intriguing candidates for research into the mechanisms of parasite drug resistance.

CYPs are capable of catalysing a remarkable range of substrates. This is partly due to the large size of the family, but is also due to the wide substrate diversity of a number of individual enzymes. Although some CYPs have specific individual substrates, especially those acting on endogenous compounds, many of those associated with xenobiotic metabolism have a much broader range. The metabolism of around 75% of drugs in clinical use can be undertaken by just five of the 57 human CYPs (Wienkers and Heath 2005b,Williams *et al* 2004) with one enzyme, CYP3A4, catalysing more than 50% of these reactions (Fujita 2004, Smith and Jones 1992). Some individual substrates can be also catalysed by a number of different CYPs, providing alternative routes of metabolism if one enzyme is deficient. A negative side effect of this versatility is that reactions

involving different substrates or alternative CYP pathways can vary in their efficiency and in the metabolites produced. Indeed, a number of these reactions can convert exogenous compounds into active carcinogens and CYPs are infamous for their lead role in human drug-drug interactions and drug side effects (Amacher 2010, Guengerich 2006).

In addition to their wide substrate diversity, CYPs exhibit a remarkable catalytic versatility. Although they primarily act as mono-oxygenases, catalysing the transfer of one atom of molecular oxygen to a substrate and reducing the other to water, they also demonstrate oxidase, reductase, dehydrogenase, dehydrase and isomerase activity amongst others. In total, CYPs have been shown to catalyse at least 60 chemically distinct reactions (Gillam 2008, Mansuy 1998).

1.6.1 Location

While prokaryotic P450s are soluble proteins, most eukaryotic P450s are membrane-bound enzymes associated with the endoplasmic reticulum and mitochondria. In humans, they are synthesised predominantly in the liver, followed by the small intestine, kidneys and adrenal glands. The most frequently implicated CYP in human drug metabolism, CYP3A4, is most highly expressed in the small intestine (Fujita 2004).

In *C. elegans*, a number of CYPs are highly expressed in the intestine (Chakrapani *et al* 2008, Menzel *et al* 2001), which is thought to be the prime site of detoxification in nematodes (McGhee 2007). In *Drosophila melanogaster*, the main organs of detoxification are the midgut, Malpighian (renal) tubules and fat body and again, CYP expression is highest at these sites, with a small number of CYPs expressed most highly in the head, gonads and hindgut (Chung *et al* 2009).

1.6.2 Structure

Many CYP proteins are not amenable to structural analysis by experimental methods such as electron microscopy and x-ray crystallography, because they lack the pre-requisite solubility or ease of crystallisation necessary for such techniques. However, models have been generated for all human CYPs involved

in drug metabolism based on sequence comparison with bacterial and fungal CYPs of known crystal structure (de Groot 2006).

CYP polypeptide sequences are highly divergent but are typically 450-500 amino acids in length. They are characterised by a heme-binding domain towards the C-terminus, containing a conserved cysteine ligand for the heme iron, and conserved PERF and K-helix domains (Poulos 2005, Tijet *et al* 2001). In the case of microsomal proteins, there is also a conserved N-terminal region for attachment to the microsomal membrane. Six especially diverse substrate recognition sites have also been identified and these constitute ~16% of the residues in the polypeptide (Gotoh 1992). See Figure 1.3.

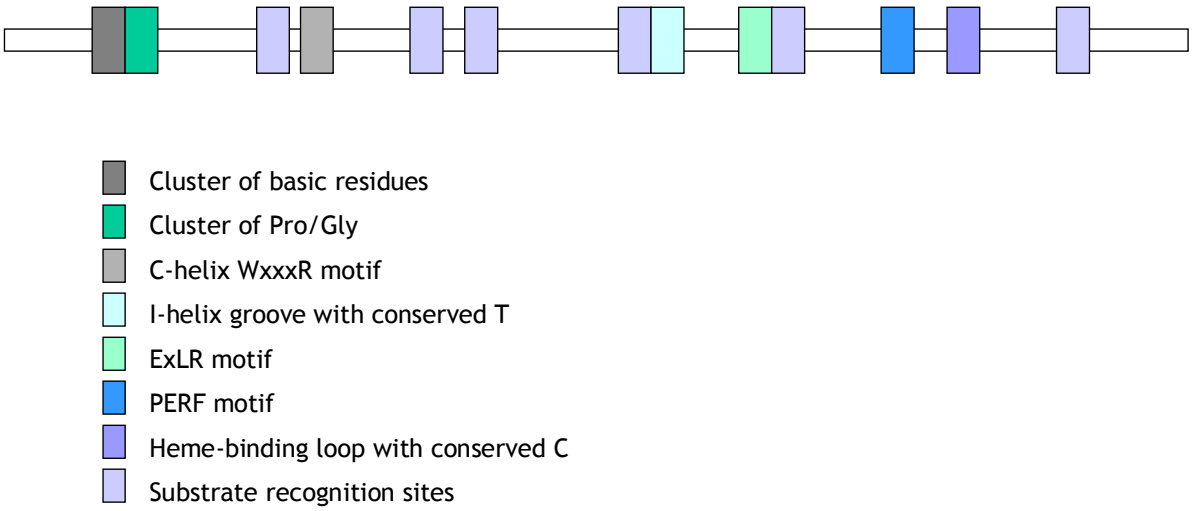


Figure 1.3: Conserved regions of a microsomal CYP. Adapted from Feyereisen 2005.

Despite the sequence diversity, all CYP proteins have a highly conserved three-dimensional structure. This is especially striking in light of the vast range of compounds the CYPs interact with, but a degree of flexibility in the secondary structure adopted for each compound is thought to facilitate this (Poulos 2005). The x-ray crystallography of *Pseudomonas putida* CYP450cam (bound to camphor) revealed that a number of dispersed amino acid sites throughout the protein interacted with the substrate. The resulting active binding site is well sequestered within the enzyme, requiring further dynamic fluctuations of the protein to expose this region for substrate-binding (Poulos *et al* 1987). It is hypothesised that individual CYPs acting on a range of compounds may be able

to change shape according to that of the substrate. The binding of a redox partner may also elicit structural change, triggering electron transfer and O_2 binding (reviewed in Poulos 2005).

1.6.3 Nomenclature

The standard nomenclature for CYPs is a number identifying the family, followed by a letter identifying the subfamily, then a number identifying each orthologous gene e.g. *cyp13a1*. Family members must share more than 40% amino acid identity, and members of a subfamily must share more than 55% amino acid identity, although for some the shared identity is as high as 99% (Nelson *et al* 1993). To constitute an individual member, the amino acid sequence of a new CYP must differ from the others by more than 3% (Gotoh 1998).

1.6.4 CYP family structure and evolution

The human genome encodes 57 CYPs, which can be classified into 18 families and 42 sub-families. Enzymes in CYP1, CYP2 and CYP3 families are involved in the metabolism of xenobiotics, while members of the remaining families have important endogenous roles. These include fatty acid metabolism (CYP4), bile acid biosynthesis (CYP7) and steroid metabolism (CYP11, CYP17) amongst others (Amacher 2010, Gonzalez 1992). CYP families can be divided into those showing variation in family or subfamily size (CYP2, CYP3 and CYP4) and those that do not vary (all others), which may relate to the expansion of xenobiotic metabolising CYP families and the conservation of CYPs with endogenous functions (Nelson 1999) and is discussed further below. The families of CYPs with endogenous functions also appear more highly conserved between species and tend to demonstrate greater substrate and product specificities (Gonzalez 1992, Nelson *et al* 2004).

In insects, CYPs are involved in the synthesis of the ecdysteroids and juvenile hormones, which regulate growth, development and reproduction. They also metabolise natural plant compounds, pesticides and insecticides, resulting in their detoxification (Feyereisen 2005). Insect CYPs can be divided into four clades: clade 2, clade 3, clade 4 and the mitochondrial clade (Feyereisen 2006). Clade 3 is the largest and encodes CYPs associated with xenobiotic metabolism

and insecticide resistance including the multi-drug resistance gene *cyp6g1* (see Section 1.6.8), although some individual CYPs in clade 2, clade 4 and the mitochondrial clade have also been associated with xenobiotic metabolism. The clade 3 CYPs are most related to the mammalian CYP3 and CYP5 families, but appear to have undergone multiple insect-specific expansions.

Nematodes were historically thought to lack CYPs (Barrett 1998, Pemberton and Barrett 1989, Precious and Barrett 1989a,Precious and Barrett 1989b) but the sequencing of the *C. elegans* genome revealed a large CYP family of 80 CYPs of 15 families and 24 subfamiles (www.wormbase.org). Gotoh (1998) classified *C. elegans* CYPs into three clades: clade A is most closely related to the mammalian CYP2 family, clade B to the mammalian CYP3 and CYP5 families and clade C to the mammalian CYP4 family. The functions of most *C. elegans* CYPs are unknown, but essential roles in signalling pathways have been identified for a small number. *cyp22a1* (*daf-9*) in clade A is associated with dauer formation, fat deposition and gonadal development (Gerisch *et al* 2001, Gerisch and Antebi 2004), *cyp33e2* in clade A and *cyp29a3* in clade C are associated with eicosapentaenoic acid metabolism, and *cyp31a2* and *cyp31a3* in clade C are associated with eggshell lipid production and embryo polarisation (Benenati *et al* 2009). Some members of clade A (*cyp35* family, *cyp14a3, cyp14a5, cyp33c2, cyp33e1, cyp34a9*), clade B (*13a7*) and clade C (*cyp29a2*) are inducible on exposure to xenobiotics (Chakrapani *et al* 2008, Menzel *et al* 2001, Menzel *et al* 2005).

Characteristics of the CYP superfamily suggest it is evolving rapidly. After 75-90 MY of divergence, only 34 putative CYP orthologues can be identified in the human and mouse genomes, relative to a whole genome comparison assigning 80% of all predicted proteins to orthologous pairs (Nelson *et al* 2004, Waterston *et al* 2002). In the *Drosophila* genus, divergence over 60 MY has resulted in only a third of the 53 CYPs studied having 1:1 orthologues in 11 species (Chung *et al* 2009) and in nematodes, a comparison of *C. elegans, C. briggsae* and *C. remanei* CYPs finds only 38% have 1:1 orthologues after 80-110 MY divergence (Thomas 2007).

CYP diversity is a result of successive gene duplications followed by sequence divergence and this has resulted in clusters of CYP genes in mammals, insects

and nematodes (Nelson *et al* 2004, Thomas 2006, Thomas 2007, Tijet *et al* 2001). These 'CYP blooms' affect different subfamilies in different species and are one of the most striking characteristics of CYP evolution (Feyereisen 2010). A comparison of human and mouse CYPs, shows a number of CYP subfamilies, such as *cyp2c*, have expanded in the mouse, resulting in a total of 102 mouse CYPs relative to 57 human CYPs (Nelson *et al* 2004).

There appears to be a distinct pattern of selection for expansion in some CYP subfamilies and selection against expansion in others. In an analysis of ten vertebrate genomes, Thomas (2007) found all CYPs with known endogenous substrates to be 'phylogeneticaly stable' and characterised by few or no gene duplications or losses, while CYPs with a role in xenobiotic detoxification were 'phylogenetically unstable', residing in a few dense gene clusters. Similarly in insects, 80% of *D. melanogaster* CYPs expressed in organs of detoxification were classified as unstable, while only 27% of CYPs expressed in other body tissues were unstable (Chung *et al* 2009). The same study used an RNAi screen to show that eight out of nine phylogenetically stable CYPs had essential endogenous functions.

In vertebrates, unstable CYP genes also appear to be affected by positive selection for amino acid substitutions, particularly in substrate-binding regions (Thomas 2007). This may also be true in nematodes as Gotoh (1998) observed the highest rate of nonsynonymous versus synonymous nucleotide changes in the central region encoding putative substrate recognition sites in closely related *C. elegans* CYPs.

The high frequency of gene duplication and divergence observed in xenobiotic-metabolising CYPs is consistent with a mechanism of adaptive evolution to cope with changing environmental conditions: the clusters of closely related CYPs could be considered as the 'raw materials' for selection to act upon.

1.6.5 Transcriptional regulation of CYPs

The expression of individual CYPs can be increased in response to exposure to their substrate. In vertebrates, this mechanism is primarily regulated by nuclear receptors (NRs), which reside in the cell cytoplasm. When activated, the NR

binds to a DNA response element in the promoter region of the target gene, inducing its mRNA expression (Amacher 2010). In humans, three of these transcription factors, aryl hydrocarbon receptor (AhR), pregnane X receptor (PXR) and constitutive androstane receptor (CAR), are most important in regulating the expression of xenobiotic-metabolising CYPs (Amacher 2010, Waxman 1999, Xu *et al* 2005). Activity of CYP3A4, the most abundantly expressed human CYP and the enzyme responsible for the metabolism of the majority of clinical drugs, is highly dependent on PXR, which in turn is activated by a particularly diverse range of xenobiotics and dietary compounds. CYP3A4 is also activated by a number of other NRs including CAR and it is thought that a degree of cross-talk between NRs determines final CYP expression (Urquhart *et al* 2007, Waxman 1999). In addition to inducing expression of CYPs, the xenosensor NRs can simultaneously up-regulate a number of phase II metabolising genes, including UGTs, GSTs and SULTs, and phase III drug transporters (Amacher 2010).

Three genes encoding NRs with homology to PXR and CAR have been identified in *C. elegans*; *daf-12, nhr-48* and *nhr-8* (Maglich *et al* 2001). DAF-12 is involved in the dauer pathway, the function of NHR-48 is currently unknown and NHR-8 is thought to play a role in xenobiotic metabolism. A study by Lindblom *et al* (2001) found *C. elegans nhr-8* deletion mutants had increased sensitivity to cholchicine and chloroquine, and this result was confirmed by *nhr-8* RNA-mediated interference. The same study used GFP reporter constructs to demonstrate *nhr-8* is exclusively expressed in the intestine, which is comparable to the human liver as the main organ of detoxification in nematodes (McGhee 2007). Human PXR and CAR are primarily expressed in the liver and intestine (Amacher 2010, Waxman 1999).

1.6.6 Factors affecting CYP activity

Genetic polymorphism is one of the most important factors in the variability of CYP expression; alleles causing quantitative increases or decreases in enzyme expression have been identified for many human CYPs, in addition to alleles resulting in increased or decreased activity. CYP polymorphism is common and can result in ‘pharmacogenetic heterogeneity’ between individuals, ranging from

poor metabolism phenotypes, with low enzyme activity towards a drug, to ultra-rapid metabolism phenotypes, with high activity towards a drug. The latter is the most widespread and most clinically important phenotype as it can result in drug failure or, in the case of pro-drugs, active drug excess or the formation of toxic metabolites (Guengerich 2006, Ingelman-Sundberg *et al* 2007).

In humans, the genetic basis for the polymorphism is usually copy number variation (CNV) due to one or more gene duplications, but single nucleotide polymorphisms (SNPs) and insertions and deletions (indels) have also been identified (Anzenbacher and Anzenbacherova 2001, Ingelman-Sundberg and Sim 2010, Nebert and Russell 2002). So far, 360 human CYP alleles have been identified and all but one of the CYPs expressed in the human liver are associated with genetic polymorphisms. Genetic biomarkers of CYP activity have been recognised as predictors of the outcome of drug treatment for a number of clinically important human drugs. Tamoxifen, an anti-oestrogen drug used in the treatment of breast cancer, requires activation by CYP2D6 and the presence of one or two null alleles results in reduced tamoxifen metabolism and a poorer therapeutic outcome (Higgins and Stearns 2010). Although tamoxifen is currently the only viable treatment option for younger women, for older women, genetic screening to determine CYP2D6 genotype may be useful in choosing the most effective chemotherapy regime.

Other innate factors affecting CYP activity in humans include age, gender and race in addition to environmental factors such as diet and lifestyle. For example, the activity of the drug-metabolising enzyme CYP1A2 can be increased by smoking, by the consumption of cruciferous vegetables and a diet high in protein but low in carbohydrates and decreased by the consumption of grapefruit juice and a diet low in protein and high in carbohydrates (Glue and Clement 1999).

1.6.7 Induction and inhibition of CYPs by xenobiotics

In humans, xenobiotics (including clinical drugs) act as ligands for AhR, PXR and CAR and activation of these NRs results in a cascade of gene up-regulation involving members of the CYP1, CYP2 and CYP3 families as well as phase II metabolising enzymes and phase III drug transporters (Amacher 2010). This can lead to drug tolerance due to repeated administration of a CYP-inducing drug

resulting in its own accelerated metabolism. CYP induction is also an important consideration for multi-drug prescriptions, as concurrent treatment with substrates for the same CYPs can result in drug failure due to increased clearance, drug excess in the case of a pro-drug, or toxicity due to the increased generation of reactive intermediates. Conversely, some widely used clinical drugs, such as the anti-histamine cimetidine and the anti-fungal ketoconazole, inhibit CYP activity and this response again influences the efficacy and safety of concurrent drug therapy.

In nematodes, CYP activity can also be induced or inhibited by xenobiotic exposure. In *H. contortus*, CYP activity was assessed by measuring microsomal activity towards the CYP substrates aldrin and ethoxycoumarin and was shown to increase up to 60-fold in response to exposure to phenobarbital (PB), a known inducer of multiple CYPs in humans (Kotze 1997). This is a similar magnitude to that observed in rats, where western blot analysis identified PB-induced increases in liver microsomal CYP proteins of up to 50-100-fold, which corresponded to an increase of mRNA levels of 20-50-fold (Waxman and Azaroff 1992). PB exposure has also been shown to up-regulate CYP mRNA levels in *C. elegans*, albeit at a lower magnitude, with expression of *cyp31a1* and *cyp31a3* shown to increase by around 2-fold using semi-quantitative RT-PCR (Menzel *et al* 2001).

Chakrapani *et al* (2008) used transgenic *C. elegans* expressing GFP under the control of the promotor elements of five CYPs, each of which was predicted to be the orthologue of a xenobiotic-inducible human CYP, to measure their expression after exposure to 17 xenobiotics. All five *C. elegans* CYPs were induced by 15 or more of the xenobiotics and showed similar expression profiles to their human orthologues. However, perhaps unsurprisingly, there are clear limitations for homology-based comparisons of CYPs in humans and nematodes. Menzel *et al* (2001) observed that almost all members of the CYP35 family were strongly inducible by a range of xenobiotics in *C. elegans*. A follow-up study demonstrated that *cyp35a1, cyp35a2, cyp35a5* and *cyp35c1* were induced by the xenobiotics atrazine, fluoranthene and lansoprasol in a concentration-dependant manner (Menzel *et al* 2005). The CYP35 genes showed most homology to human CYP2 family members. However, atrazine, fluoranthene and lansoprasol are more strongly linked to CYP1A induction in man. Further, the orthologue of the

NR known to activate CYP1A expression in humans, AhR, has been identified in *C. elegans* (*ahr-1*) and yet *ahr-1* null mutants still showed a strong induction on exposure to these compounds. These findings, combined with those of Section 1.6.4, suggest that the rapid evolution of the CYP family, especially affecting CYPs involved in xenobiotic metabolism, may prohibit inferences of shared function between species based solely on sequence homology.

Concurrent with this, piperonyl butoxide is a potent CYP inhibitor in most species, but it induces expression of two genes, *cyp6a1* and *cyp6b2,* in insects (Feyereisen 2005). In *H. contortus* L3, CYP activity was suppressed in a concentration-dependant manner with exposure to piperonyl butoxide (Kotze 1997) and in a later study, piperonyl butoxide was shown to have a synergistic action on the toxic effects of rotenone, an insecticide metabolised by CYPs, in *H. contortus* and *T. colubriformis* (Kotze *et al* 2006). Drug toxicity was assessed with a larval development assay and an adult motility assay, and the greatest piperonyl butoxide and rotenone synergy was observed in the larval stages of both species.

1.6.8 CYPs and insecticide resistance

The association of insect CYPs with insecticide metabolism and resistance is well established. CYP-mediated resistance has been reported in most classes of insecticide and in most species of pest insect (Berge *et al* 1998). Perhaps the most well-known example is the expression of a single gene, *cyp6g1,* which confers multi-drug resistance in 20 field strains of *D. melanogaster* (Daborn *et al* 2002). The mechanism responsible for the overexpression of this gene was identified as an *Accord* transposable element insertion in the 5' flanking sequence. In *D. simulans,* the insertion of a *Doc* transposable element in the 5' end flanking sequence of *cyp6g1* has also been shown to confer insecticide resistance (Schlenke and Begun 2004). Eight additional CYPs, *cyp4e2, cyp6a2, cyp6a8, cyp6a9, cyp6g2, cyp6w1, cyp12a4* and *cyp12d1/2,* have since been shown to confer insecticide resistance in *D. melanogaster* through either overexpression from mutations in the promoter regions or, in the case of *cyp6a2*, point mutations affecting enzyme activity in a manner that is yet to be elucidated (Amichot *et al* 2004, Bogwitz *et al* 2005, Daborn *et al* 2007, Maitra *et*

al 1996, Pedra *et al* 2004). CYP overexpression has also been identified as a mechanism of insecticide resistance in pyrethroid-resistant field populations of the important malaria vectors *Anopheles funestus* (*cyp6p9*) and *A. gambiae* (*cyp6z1*) (Amenya *et al* 2008, Nikou *et al* 2003).

1.6.9 CYPs and anthelmintic resistance

Anthelmintic drugs can be detoxified by CYPs. In mammals, ABZ is metabolised by a two-step oxidation reaction: CYP3A and FMO oxidise ABZ to the active sulphoxide metabolite (ABZSO) before CYP1A oxidises ABZSO to the inactive sulphone metabolite (ABSOO) (reviewed in Capece *et al* 2009). In the mouflon, a relative of domestic sheep, repeated ABZ administration has been shown to cause auto-induction of CYP1A resulting in an increased rate of anthelmintic deactivation (Velík *et al* 2005).

In nematodes, much less is known about ABZ metabolism. In *C. elegans*, ABZ has been shown to induce the expression of *cyp35a2, cyp35a5* and *cyp35c1* using a microarray approach confirmed by real-time PCR (Laing *et al*, 2010). The same project also identified ABZ-glucoside metabolites in whole worm homogenates incubated with ABZ using high performance liquid chromatography and mass spectrometry (HP-LC-MS). This concurs with the work of Cvilink *et al* (2008b), who used a similar approach to demonstrate that *H. contortus* metabolises ABZ to ABZSO and ABZ-glucoside and the related BZ, flubendazole (FLU), to reduced FLU and glucosides of both FLU and reduced FLU. These findings demonstrate that nematodes use both phase I (oxidation and reduction) and phase II (glucosidation) enzymes to metabolise BZ anthelmintics.

In the case of the MLs, again, knowledge of nematode anthelmintic metabolism lags far behind that of vertebrates. In humans, IVM is metabolised to at least ten metabolites and CYP3A4 is the major enzyme involved (Zeng *et al* 1998). Work by Alvinerie *et al* (2001) implicated CYPs in the metabolism of moxidectin (MXD) in *H. contortus* because the production of the single undefined metabolite identified with HP-LC was blocked by carbon monoxide, a known CYP inhibitor. Conversely, recent work with *C. elegans*, suggests that IVM is not metabolised by the nematode; in contrast to the HP-LC-MS analysis of whole worm incubations

with ABZ described above, no metabolites were detected for incubations with IVM (pers. com. Dr S Laing).

Despite the well-known importance of CYPs in xenobiotic resistance in insects, there appears to be little research regarding the putative role of CYP-mediated anthelmintic resistance in parasites. Drug resistant isolates of the liver fluke *Fasciola hepatica* have been shown to have a greater capacity to metabolise the BZ triclabendazole (TBZ), which is the main anthelmintic used to treat fluke infection (Alvarez *et al* 2005, Robinson *et al* 2004). TBZ is metabolised to the active TBZSO and inactive TBZSOO by phase I enzymes (FMO and CYPs) and phase II enzymes (GSTs) then excreted by phase III transporters (PGPs). Inhibition of the FMOs, CYPs and PGPs have been shown to reduce TBZ resistance, varying from a modest reduction in resistance with CYP inhibition to complete reversal to susceptibility with PGP inhibition (reviewed in Brennan *et al* 2007). Although work by Alvarez *et al* (2005) found no difference in the degree of inhibition of TBZ sulphoxidation with piperonyl butoxide between resistant and susceptible flukes, suggesting there was no difference in CYP activity between the isolates, a recent study by Devine *et al* (2010) investigating morphological changes in the fluke with TBZ treatment found piperonyl butoxide potentiated the effect of TBZ in resistant flukes to a greater extent than in susceptible flukes. This suggests that increased CYP activity may play a role in the resistance of flukes to anthelmintic treatment. Kotze (2000) compared CYP activity in ML-susceptible and –resistant *H. contortus*, but failed to detect any difference between the isolates. It is possible, however, that CYPs with low or no activity towards the substrates used in the study, aldrin and ethoxycoumarin, are more active in the resistant isolate.

1.7 Aims of this project

1. To identify all CYP sequences present in the WTSI *H. contortus* genome databases and assemble the CYP family in *H. contortus* for phylogenetic analysis with CYPs in other species

2. To develop a real-time PCR screen to measure constitutive CYP gene expression and to measure CYP gene induction in response to xenobiotic exposure

3. To use Illumina technology to sequence the *H. contortus* transcriptome to measure CYP gene expression and facilitate global analysis of gene expression

4. To use transcriptome data to investigate conservation between *H. contortus* and *C. elegans* at a gene and genome level and assess the impact of findings on the future completion of the *H. contortus* genome, an essential requirement for the definitive characterisation of the CYP family in the parasite

2 Materials and Methods

2.1 Standard reagents

Ethidium bromide: 10 mg/ml in sterile distilled water. Stored at room temperature.

Loading buffer (5X): 100 mM EDTA pH 7.5, 22% Ficoll (Sigma, F2637), 0.05% Bromophenol blue (Sigma, B0126). Stored at room temperature.

M9 buffer: 22 mM KH_2PO_4, 50 mM Na_2HPO_4, 85 mM NaCl, 1 mM $MgSO_4$. Stored at room temperature.

Saline for parasite work: 0.85% w/v NaCl in tap water. Stored at room temperature.

TAE (50X): 2 M Tris-base, 100 ml/L 0.5 M EDTA, 57.1 ml/L glacial acetic acid. Autoclaved and stored at room temperature.

TBE (5X): 0.45 M Tris-base, 0.45 M Boric acid, 100 ml/L 0.5 M EDTA. Autoclaved and stored at room temperature.

TE buffer: 10 mM Tris, 1 mM EDTA pH 8. Stored at room temperature.

2.2 Parasite

2.2.1 *H. contortus* maintenance and culturing

Experimental infections were performed at the Moredun Research Institute using orally administered infections of 5000 L3 into 4- to 9-month-old lambs that had been reared and maintained indoors under conditions designed to minimise the risk of infection with gastrointestinal nematodes.

2.2.1.1 Culture of L3

Faeces from an infected donor lamb were incubated in culture trays at 22°C for 7 days. Trays were flooded with tap water at 22°C, left to soak for 1 hour (to let the larvae migrate out of the faeces), and the contents were sieved through a 1 mm nylon mesh. The filtrate was sedimented at 4°C for 2 hours then cleaned with a Baermann filter as described in Jackson and Hoste (2010). L3 were stored in tap water in cell culture flasks with vented lids at 4-8°C for up to 3 months.

2.2.1.2 Exsheathing L3

The number of L3 in a suspension was roughly quantified by counting the number of larvae in a 10 μl aliquot under a microscope. ~ 0.5 million L3 were removed to a 50 ml falcon tube per treatment group. The L3 were stimulated to exsheath by the addition of 10 μl sodium hypochlorite (Sigma Aldrich, 425044) per 1 ml L3 suspension. A small aliquot of the solution was examined under a microscope after 10 minutes incubation to monitor exsheathment. The addition of sodium hypochlorite was repeated every 10 minutes until exsheathment was confirmed. The L3 suspension was centrifuged at 3000 rpm (955 g) at room temperature for 1 minute. The supernatant was removed and the L3 were washed three times with M9 buffer and centrifugation then either resuspended in 30 ml M9 buffer for immediate use in liquid culture experiments or snap-frozen in liquid nitrogen for storage at -80°C.

2.2.1.3 Harvest of adult *H. contortus*

The donor abomasum was removed at post-mortem and split longitudinally. Luminal contents were emptied into a bucket containing 5 L saline at 37°C and the mucosal surface was rinsed with a further 1 L saline. Any remaining clumps of adult worms were harvested from the mucosa with tweezers into falcon tubes containing saline at 37°C. Adult worms in the luminal contents were harvested using an agar/mesh floatation technique described in Jackson and Hoste (2010). Briefly, 9 g agar (Sigma Aldrich Ltd., 7002) was dissolved in 500 ml saline (0.85% w/v NaCl) at 59°C and mixed with 500 ml luminal contents. The resulting solution was poured over a 1mm nylon mesh stretched over a makeshift springform tray, created from the upturned top of a plastic bucket with a clip-on

lid, and left to set for 15 minutes. The agar slab was submerged in 5 L saline at 37°C in a 10 L funnel with a tap at the base. Worms migrate out of the agar into the warm saline (upwards or downwards) and can be collected after ~1 hour. Worms were rinsed twice in RPMI 1640 solution (Gibco, 21875-034) at 37°C and once in RPMI 1640 solution with 1% Penicillin-Streptomycin (Invitrogen, 15140-122) at 37°C. Worms were either resuspended in RPMI 1640 solution with 1% Penicillin-Streptomycin at 37°C for immediate use in liquid culture experiments or divided between 2 ml cryotubes, snap-frozen in liquid nitrogen and stored at -80°C.

2.2.2 Sexing adult *H. contortus*

Adult worms were placed in a 100 mm x 20 mm Petri dish containing RPMI 1640 solution with 1% Penicillin-Streptomycin at 37°C. The Petri dish was placed on a light box to aid viewing and maintain temperature. Adult males and females were separated based on the identification of the distinct male copulatory bursa and the white female ovaries wrapped around the blood-filled intestine. There was also a clear size difference between the smaller males and large females.

2.2.3 Xenobiotic exposure experiments

2.2.3.1 L3 liquid culture

Two batches of ~0.5 million exsheathed L3 in 30 ml M9 buffer (see above) were mixed and divided into six 75 ml cell culture flasks with vented caps to allow experiments to be performed in triplicate. Stock solutions of xenobiotic were added to the three flasks in the 'treatment' group and, where DMSO was used as a solvent, the corresponding volume of DMSO was added to the L3 suspensions in the three control flasks. The liquid cultures were maintained in a shaking incubator at 37°C and at 50 rpm for 4-24 hours depending on the xenobiotic and guided by published studies in *C. elegans* and *H. contortus* (see Chapter 3, Section 3.2.4). L3 suspensions were transferred to separate 50 ml falcon tubes and rinsed three times with M9 buffer and centrifugation at 3000 rpm (955 g), before being divided into appropriately labelled 2 ml cryotubes and snap-frozen with liquid nitrogen. Frozen L3 were stored at -80°C.

2.2.3.2 Adult liquid culture

Adult worms were divided between six 100 mm x 20 mm Petri dishes per experiment (the exact number of worms depended on the yield from the donor, but usually there was enough for at least 20 adult worms per dish) each containing 30 ml RPMI 1640 solution with 1% Penicillin-Streptomycin at 37°C. Stock solutions of xenobiotic were added to the three dishes in the 'treatment' group and, where DMSO was used as a solvent, the corresponding volume of DMSO was added to the three control Petri dishes. The liquid cultures were maintained in a 5% CO_2 incubator at 37°C for 5 hours (time was limited by the viability of adult worms outside their host) and gently stirred or swirled every hour. Worms from each Petri dish were transferred to separate 50ml falcon tubes and rinsed three times with RPMI 1640 at 37°C. Centrifugation was unnecessary as the adult worms rapidly sank to the bottom of the tubes. The worms were divided into appropriately labelled 2 ml cryotubes, snap-frozen with liquid nitrogen and stored at -80°C.

2.3 Template preparation

2.3.1 DNA lysates for microsatellite analysis

Microsatellite analysis was used to confirm the worms infecting each donor were of the expected isolate using a technique described in Redman (2008b). This is based on each strain showing different frequencies of particular alleles for a number of microsatellites (Redman 2008, unpublished data). Briefly, a pellet of ~100 L3 was reserved from every donor sheep. This pellet was added to 50 μl DNA 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). The lysate was kept at -80°C for 10 minutes before incubation at 60°C for 98 minutes followed by 20 minutes at 94°C to denature the proteinase K. Negative control lysates (without L3) were prepared simultaneously. Lysates were stored at -80°C. Dr L. Redman and Miss F. Whitelaw used 1 μl of a 1:50 dilution of lysate as template for PCR amplification of microsatellites to confirm the identity of the isolate.

2.3.2 RNA extraction

RNA was extracted using a standard Trizol procedure. Briefly, a 200 μl *H. contortus* pellet was frozen in liquid nitrogen then ground thoroughly in a mini glass homogeniser (Camlab, 1165622) with 1600 μl of Trizol reagent (Invitrogen, 15596-026). The sample was then vortexed and left at room temperature for 30 minutes. Insoluble debris was removed by centrifuging at 14000 rpm (20800 g) at 4°C for 10 minutes. The supernatant was removed to a fresh tube and 320 μl of chloroform added. The mixture was vortexed for 15 seconds and left at room temperature for 2-3 minutes. Following centrifuging at 14000 rpm (20800 g) at 4°C for 15 minutes, the aqueous layer was removed and 500 μl isopropanol added. The RNA was precipitated at -80°C for 1 hour. The RNA was pelleted by centrifuging at 14000 rpm (20800 g) at 4°C for 10 minutes. The RNA pellet was washed in 75% ethanol in DEPC-treated water, vortexed briefly and centrifuged at 7500 rpm (5974 g) at 4°C for 5 minutes. The supernatant was removed and the RNA pellet was air-dried for 10 minutes. The RNA pellet was then dissolved in RNase-free water and treated with DNase I (Qiagen, 79254) in solution for 10 minutes before purification and concentration using RNeasy columns (Qiagen, 74104) including the optional on-column DNase I treatment steps in the RNeasy protocol.

RNA samples were quantified by 260/280 absorption on a Gene Quant Pro spectrophotometer (Amersham Biosciences) and were analysed by gel electrophoresis (1.2% agarose, TBE gel, 100V, 1hr).

2.3.3 First strand cDNA Synthesis

First-strand cDNA was synthesized from 5 μg total RNA using random hexamer primers, following the manufacturer's protocol (Cloned AMV First-strand cDNA Synthesis Kit, Invitrogen, 12328-040). For each synthesis reaction, a concurrent no reverse transcriptase control was run, using 1 μl DEPC-treated water in place of 1 μl cloned AMV RT. Residual primers, nucleotides and enzymes were removed from the cDNA with a PCR purification kit (QIAquick PCR Purification Kit, QIAGEN, 28106) following the manufacturer's protocol.

2.4 PCR

PCR reactions were performed using a GeneAmp PCR system 9700 (Applied Biosystems) in a 20 μl volume. Routine PCR conditions were 95°C for 30 seconds, primer annealing at 58-60°C for 30 seconds and extension at 72°C for 1-2 minutes per 1Kb of target sequence. A total of 35-40 cycles were used. Final concentrations of 250-500 nM of forward and reverse primers and 250 μM of each dNTP were used. Oligonucleotide primers were purchased from Eurofins MWG Operon. The sequences of all primers used are presented in the Appendices. Amplitaq DNA polymerase (5 U/μl, Applied Biosystems, N808-0160) was used at a final concentration of 1 unit of enzyme per reaction.

2.5 Agarose gel electrophoresis

Nucleic acids were separated on 1-2% (w/v) agarose gels. Agarose (Invitrogen, 15510-027) was melted in 1XTAE, or 1XTBE for RNA separation, by heating until in solution. Ethidium bromide was then added to a final concentration of 0.1 μg/ml and gels cast. Gels were electrophoresed in 1XTAE or 1XTBE as appropriate using equipment from Amersham Pharmacia Biotech. Gels were imaged using a Flurochem 5500 UV transilluminator and image capture system (Alpha Inotech).

2.6 Real-time PCR screen

For experimental design, see Chapter 3, Section 3.2.2.

2.6.1 Normalising and control genes

A panel of control genes was developed based on genes used for real-time normalisation in *C. elegans* and constitutively expressed *H. contortus* genes described in the literature.

C. elegans ama-1 encodes the large subunit of RNA polymerase II. It is ubiquitously expressed and has been extensively used as a normalising gene in *C. elegans* RT-PCR experiments. The *H. contortus* homologue of this gene, *Hc-ama*,

was identified bioinformatically and annotated (see Chapter 3, Section 3.2.2, Figures 3-4 and 3-5). Primers were designed to amplify the predicted mRNA.

act-1 encodes a highly expressed and highly conserved actin isoform, commonly used as a normalising gene in RT-PCR experiments in *C. elegans*. An *H. contortus* homologue of this gene, *hc-act*, was identified bioinformatically, annotated, and primers were designed to amplify the predicted mRNA.

gtp-ch-1 encodes a GTP cyclohydrolase 1 precursor, which is highly expressed in *T. circumcincta* L3 larvae (Nisbet *et al* 2008). In an *H. contortus* comparative stage EST sequencing project it was detected in L3 databases only (Hoekstra *et al* 2000). Primers were designed against the published cDNA sequence (Accession no. AW670739).

H. contortus Hc27 is a homologue of the *C. elegans vit-6* gene, which encodes a vitellogenin, a precursor glycoprotein of egg yolk protein, expressed highly in the intestine of adult hermaphrodites. In *H. contortus,* it was shown by northern-blot analysis of PCR products to be expressed in adults only (see Figure 2-1) and by in situ hybridisation to be expressed abundantly in the intestine (Hartman *et al*, 2001). Primers were designed against the published mRNA sequence (Accession no. AF305957).

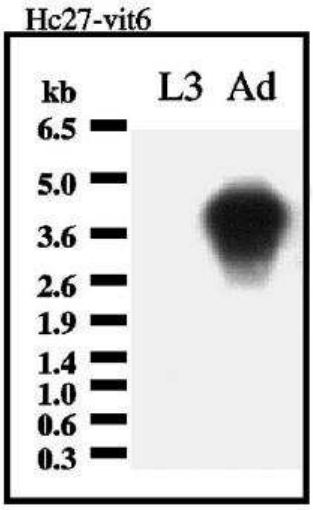


Figure 2.1: Northern blot of *Hc27* PCR products demonstrating high adult-only expression in *H. contortus*. From Hartman *et al* (2001).

2.6.2 Primer design

Primers were designed with predicted melting temperatures of 60+/-2°C, lengths of 18-20 nt, and GC contents of >45%. Where possible, amplicons of 160-200 bp (range 120-212 bp) were designed to span introns (see Appendix for primer sequences). Primer pairs were tested with PCR amplification of adult and L3 first strand cDNA. A single band of the expected size when run on an agarose gel was taken as a positive result. This was confirmed with melting curve analysis by the real-time PCR machine in each real-time experiment.

2.6.3 RT-QPCR reaction parameters

The screen was run on two 96-well plates. On both plates, the first 12 wells for each template were reserved for the panel of control genes (*Hc-ama* in triplicate), no reverse transcriptase and no template controls.

All reactions were carried out using Brilliant SYBR Green QPCR master mix (Stratagene, 600548). Experiments were carried out in a separate room from the RNA, mRNA and cDNA preparation and in a UV chamber to reduce the risk of contaminating nucleic acid. Samples were prepared on ice and the Brilliant SYBR Green QPCR master mix was protected from light at all times. The final concentration of primers was between 300 and 400 nM in a total reaction volume of 25 μl. A Stratagene Mx3000P QPCR system was used with the following parameters: 7 minutes and 30 seconds at 95°C, 40 cycles of 30 seconds at 95°C, 30 seconds at 60°C and 30 seconds at 72°C, and finally 1 minute at 95°C, followed by 30 seconds at 55°C and a gradient to 95°C. Fluorescence was measured at the end of the elongation phase (72°C) during each cycle for quantitation, and continuously during the final gradient from 55-95°C to assess dissociation curves. Data were captured and analysed using Stratagene MxPro software.

2.6.4 Analysis

The relative quantity of each gene on interest (GOI) within each biological replicate was calculated using the ΔΔCT method:

Relative quantity = $2^{-\Delta CT\ GOI}$ / $2^{-\Delta CT\ \textit{Hc-ama}}$

Where,

ΔCT = difference in threshold cycles between samples (unknown – control)

2.7 Illumina transcriptome analysis

2.7.1 RNA and cDNA preparation

Total RNA was isolated from the frozen pellet of 500 µl exsheathed L3 or mixed sex adult worms using the standard Trizol protocol described above. The quality and quantity of the total RNA yield was assessed with a Bioanalyzer 2100 (Agilent).

mRNA was isolated from 50 µg total RNA using an oligo(dT) magnetic bead protocol, which separates mRNA based on the poly(A)+ tail (FastTrack MAG mRNA Isolation Kit, Invitrogen, K1580-01). The mRNA was eluted in 35 µl RNase-free water and quantified with a NanoDrop 3300 Fluorospectrometer (Thermo Scientific).

Next, mRNA (0.5-1 µg) was fragmented with RNA Fragmentation Reagents (Ambion, AM8740); 31.5 µl mRNA was heated at 70^{o}C with 3.5 µl 10X Fragmentation Buffer for 5 minutes, before adding 3.5 µl Stop Buffer. The fragmented mRNA was precipitated with 3.5 µl NaOAC pH 5.2, 2 µl glycogen and 100 µl 100% ethanol and incubated at -80^{o}C for 30 minutes. The tube was centrifuged at 14000 rpm (20800 g) at room temperature for 15 minutes. The supernatant was removed and the pellet was washed with 1 ml 70% ethanol in DEPC-treated water, vortexed and centrifuged at 14000 rpm (20800 g) at room temperature for 10 minutes. The supernatant was removed and the pellet was air-dried for up to 30 minutes then re-suspended in 10.5 µl RNase-free water.

First-strand and second-strand cDNA were synthesized according to the manufacturer's protocol (Superscript Double-stranded cDNA Synthesis Kit,

Invitrogen) but with 3 µg/ul random hexamer primers (Invitrogen). The cDNA was cleaned using a QIAquick PCR Purification Kit (Qiagen).

2.7.2 Sequencing and analysis

The fragmented double-stranded cDNA was submitted to the Illumina Sequencing Team at The Sanger Institute for library preparation. Briefly, Illumina adapters were ligated to the cDNA fragments (after end repair and addition of a single A base). The fragments were gel purified with size selection at 200 bp (+/- 25 bp) then PCR amplified using Illumina primers designed against the adapter sequences. The cDNA fragments were sequenced on a Genome Analyser (Illumina).

Initially, 52 bp length reads were generated and mapped using the algorithms 'Mapping and Assembly with Qualities' (MAQ) and 'Sequence Search and Alignment by Hashing Algorithm' (SSAHA) (Li *et al* 2008, Ning *et al* 2001, Rutherford *et al* 2000). Later, with improvements in both sequencing technology and downstream analysis, 76 bp length reads were generated and mapped with 'Burrows-Wheeler Aligner' (BWA) algorithms and processed as SAM and BAM files (Carver *et al* 2010, Li and Durbin 2009).

The alignment of reads to genomic sequence with MAQ, SSAHA and BWA was performed by Martin Hunt at the Wellcome Trust Sanger Institute, Cambridge.

2.7.2.1 Generating coverage plots from mapped reference files

BAM files generated from mapping with BWA algorithms were re-formatted to contain 60 characters per line to make them compatible with SAMtools (http://samtools.sourceforge.net/). SAMtools was used to index the re-formatted BAM files and extract transcriptome read data for each supercontig or BAC insert sequence of interest to create individual pile-up files. These files were opened and viewed as coverage plots for each reference sequence in Artemis.

BAM files were also directly opened in Artemis after being sorted and indexed with SAMtools. This permitted visualisation of individual reads aligned to the reference sequence and mate pair reads could be identified.

2.7.2.2 Extracting read data for comparison of gene expression

SAMtools was used to extract and count all reads aligning to specified regions of a supercontig. This permitted comparison of the level of CYP gene expression between samples: reads aligned to annotated CYP genes were extracted and normalised to the number of reads aligned to the control gene *Hc-ama*.

2.7.2.3 Mapping reads containing spliced leader (SL) sequences

All reads containing published *H. contortus* SL1 or SL2 sequences (accession number Z69630 and AF215836) were extracted and the SL sequence was removed. There is a family of SL2 sequences in *C. elegans*, and the same may be true in *H. contortus*, so a single base pair mismatch from the published sequence was tolerated. The trimmed reads were aligned to the reference BAC insert sequences with BWA as two separate groups (depending on the SL sequence). The SL1 and SL2 BAM files were sorted and indexed with SAMtools and opened directly in Artemis to view the trimmed reads aligned to the annotated BAC insert sequences. This allowed the identification of genes trans-spliced to SL1 and SL2.

2.8 Bioinformatic identification of CYP sequence

Polypeptide sequences for all *C. elegans* CYPs listed in Wormbase were used to perform a tBLASTn (protein into translated DNA) search of the 93 Mb *H. contortus* assembled contigs 27/01/06 database. This search was repeated with CYPs of interest from other species: human CYP3A4 and *Drosophila melanogaster* CYP6G1 polypeptides. All contigs identified with $P<0.005$ were used for a BLASTx (translated DNA into protein) reciprocal search of the *C. elegans* Wormpep database to ensure the best hits were indeed CYPs. The closest matched protein for each was recorded.

This procedure was repeated for the revised 214 Mb *H. contortus* assembled contigs 12/11/07 database and 279 Mb *H. contortus* assembled contigs and supercontigs 21/08/08 databases as they became available. In addition, each *H. contortus* CYP sequence identified with a previous search was used for a BLASTn (DNA into DNA) search of each new database for completeness and to maintain continuity with the nomenclature.

2.9 Bioinformatics software

DNA sequence was viewed and annotated using both Vector NTI (Invitrogen) and Artemis (Sanger) software. The latter is free and available online (http://www.sanger.ac.uk/resources/software/artemis/).

Basic Local Alignment Search Tool (BLAST), provided by the National Centre for Biotechnology Information (NCBI) was used to search nucleotide and protein sequence databases. It was possible to BLAST search *C. elegans* sequence databases from within Wormbase (www.wormbase.org) and *H. contortus* sequence databases from within the *H. contortus* Sequencing Project website (www.sanger.ac.uk/Projects/H_contortus).

SAMtools was used to handle and process files of transcriptome reads aligned to reference sequence. This software is hosted by SourceForge and is available for free download (http://sourceforge.net/projects/samtools/files/).

3 Cytochrome P450 gene expression

3.1 Introduction

Cytochrome P450s (CYPs) are a large superfamily of enzymes found in almost all living organisms (Nelson *et al* 1993). Their ubiquity reflects both their ancient origin and their physiological importance. CYPs catalyse a wide range of reactions involving both endogenous and exogenous substrates. They are involved in the biosynthesis and catabolism of steroids, retinoids, prostaglandins and fatty acids and the detoxification of exogenous substrates including drugs and insecticides (Anzenbacher and Anzenbacherova 2001, Mansuy 1998, Nebert 1994, Nebert and Dieter 2000, Nebert and Russell 2002, Nelson *et al* 1993, Thomas 2007).

The impressive repertoire of substrates metabolised by CYPs is in part due to the large size of the family, but also due to the wide substrate diversity of a number of individual enzymes. Although a number of CYP proteins are specific for individual substrates, especially those acting on endogenous compounds, many of those associated with xenobiotic metabolism have a much broader range. This is well demonstrated by the five (of 57) human CYPs, which are responsible for the metabolism of around 75% of drugs in clinical use (Wienkers and Heath 2005a,Williams *et al* 2004) and the one enzyme, CYP3A4, which catalyses more than 50% of these reactions (Fujita 2004, Smith and Jones 1992). Incredibly, the total number of substrates these five enzymes can collectively metabolise is thought to reach into the millions (Guengerich 2009).

CYPs primarily function as mono-oxygenase enzymes, but this encompasses many distinct types of biotransformation; CYPs demonstrate oxidase, reductase, dehydrogenase, dehydrase and isomerase activity amongst others. In total, CYPs have been shown to catalyse at least 60 chemically distinct reactions (Gillam 2008, Mansuy 1998).

The expression of individual CYPs can be increased in response to exposure to their substrate. This mechanism is primarily regulated by nuclear receptors,

which bind to control elements upstream of the 5' end of the gene, inducing mRNA expression (reviewed in: Savas *et al* 1999, Xu *et al* 2005).

In insects, the association of overexpression of CYP genes with drug resistance is well established. In 2001, Daborn *et al*, found overexpression of a single gene, *cyp6g1*, was responsible for multi-insecticide resistance in 20 field strains of *D. melanogaster*. The mechanism responsible for the overexpression of *cyp6g1* was identified as an *Accord* transposable element insertion in the 5' flanking sequence. Work by Schlenke *et al* (2004) then identified a *Doc* transposable element insertion, also in the 5' flanking sequence of *cyp6g1*, capable of conferring insecticide resistance in *Drosophila simulans*. Eight additional CYPs, *cyp4e2, cyp6a2, cyp6a8, cyp6a9, cyp6g2, cyp6w1, cyp12a4* and *cyp12d1/2,* have since been shown to confer insecticide resistance in *D. melanogaster* through either overexpression from mutations in the promoter regions or, in the case of *cyp6a2*, point mutations affecting enzyme activity in a manner that is yet to be elucidated (Amichot *et al* 2004, Bogwitz *et al* 2005, Daborn *et al* 2007, Maitra *et al* 1996, Pedra *et al* 2004). Despite the wealth of information highlighting the importance of CYP-mediated drug metabolism in humans and xenobiotic resistance in insects, the CYP family in nematodes has received little attention.

The genome of the model nematode *C. elegans* encodes 80 CYPs. The function of most are unknown, although a number of CYPs have been shown to be induced with exposure to xenobiotics (Menzel *et al* 2001, Menzel *et al* 2005). The commonly used BZ anthelmintic, albendazole, has recently been shown to be metabolised by *C. elegans* and to induce the expression of *cyp35a2, cyp35a5* and *cyp35c1* (Laing *et al* 2010). In parasitic nematodes, CYP oxidase activity has been demonstrated in microsomal preparations from *H. contortus* L1 and L3 larvae, although only a low level was detected in the adult stage (Kotze 1997). However, the CYP inhibitor piperonyl butoxide has been shown to increase the toxicity of rotenone to both L3 and adult *H. contortus* and *T. colubriformis,* suggestive of active CYP-mediated metabolism in both stages (Kotze *et al* 2006).

The aim of this project was to identify the family of CYP genes in *H. contortus* and measure their expression levels in different life-stages, sexes, and tissues. Of particular interest were their expression levels in anthelmintic-exposed and anthelmintic-resistant worms relative to non-anthelmintic-exposed susceptible

worms, with a view to elucidating a putative role in xenobiotic metabolism and anthelmintic resistance.

3.2 Results

3.2.1 Identifying CYP gene sequences in an incomplete genome

3.2.1.1 Bioinformatics and comparative analysis

All CYPs are large single domain proteins, which facilitates their discovery using homology-based methods. A proven approach to the identification of CYP family members in a new genome is to perform multiple Basic Local Alignment Search Tool (BLAST) searches, using well-characterised CYP sequences from other species, aided where possible by EST sequences (Gotoh 1998, Tijet *et al* 2001).

The characterisation of the CYP family in *H. contortus* was facilitated by the availability of genomic, mRNA and amino acid sequence for the complete CYP family in *C. elegans* (www.wormbase.org). The main obstacle for annotating the family in the parasite was the lack of a completed genome (see Chapter 1). The unfavourable combination of small contigs and large genes complicated the identification of full length CYPs in the *H. contortus* genomic databases (www.sanger.ac.uk/Projects/H_contortus). An additional challenge was the high level of sequence identity and conservation of exon: intron organisation in CYPs of the same subfamily, making it difficult to distinguish contigs encoding regions of closely related but different genes from those containing unassembled polymorphic sequence from the same gene. The known propensity for CYP families to expand by gene duplication, potentially generating clusters of highly similar genes and pseudogenes, added to the complexity of confident annotation.

Due to this a ‘tag’ approach was developed; undertaking experimental work based on partial sequences (‘tags’) on each small contig, rather than full length genes. It was hoped that the predicted 1-3 fold coverage of the genome in the reference databases would permit identification of at least one tag for every CYP gene in the family. The resulting information would then be directly applicable to the full length CYPs identified when the genome is complete.

3.2.1.2 Identification of CYP gene tags

Polypeptide sequences for all *C. elegans* CYPs in Wormbase were used to perform a tBLASTn (protein into translated DNA) search of the 93 Mb *H. contortus* assembled contigs 27/01/06 database. This search was repeated with CYPs of interest from other species: human CYP3A4 and *Drosophila melanogaster* CYP6G1 polypeptides. 48 contigs were identified with P<0.005 and these were used for a BLASTx (translated DNA into protein) reciprocal search of the *C. elegans* Wormpep database to ensure the best hits were indeed CYPs. The closest matched protein for each was recorded.

This procedure was repeated for the revised 214 Mb *H. contortus* assembled contigs 12/11/07 database and 279 Mb *H. contortus* assembled contigs and supercontigs 21/08/08 databases as they became available. In addition, each *H. contortus* CYP sequence identified with a previous search was used for a BLASTn (DNA into DNA) search of each new database for completeness and to maintain continuity with the nomenclature.

In total, 97 partial CYP sequences ('tags') were identified on 61 contigs (see Figure 3-1).

3.2.1.3 Annotation of CYP tags

Each CYP tag was annotated based on BLASTx sequence alignment with the closest matched polypeptide in *C. elegans*. Methionine start codons (where the 5' end of a gene was present) and the GT:AG rule for intron boundaries were observed. The predicted polypeptide was then used to BLASTp search the NCBI protein databases and the closest matched protein was recorded (see Figure 3-2).

3.2.1.4 Nomenclature for *H. contortus* CYP tags

The standard CYP nomenclature system cannot be applied to the incomplete genome of *H. contortus* as the true amino acid identity between incomplete genes cannot be determined. In addition, the nomenclature *for H. contortus* genes based on their putative orthologues in *C. elegans* adopted by some

workers is especially risky for this family due to the likelihood of the rapidly evolving family generating a preponderance of paralogues. Further, a single amino acid substitution in a 500 residue CYP can dramatically change the catalytic activity, highlighting the importance of correct classification for this family (reviewed in Nelson *et al*, 1993).

Therefore, to avoid incorrect family classification or a misleading assumption of orthology, the CYP sequences were simply given the identifier 'Hc' for *H. contortus*, 'cyp' for cytochrome P450, 'tag' for partial sequence and a number e.g. Hc-cyp-tag1. CYP tags identified in previous searches retained their original designated number in each revised search.

3.2.2 Development of a PCR assay of CYP gene expression

Although the full sequence for every *H. contortus* CYP gene was not available, the identification of all partial CYP sequences present in the *H. contortus* genomic databases facilitated the design of a preliminary semi-quantitative RT-PCR screen. This was followed by the development of a real-time RT-PCR screen. When the semi-quantitative screen was designed, 76 CYP tags had been identified in the 12/11/07 contig database. By the time the real-time RT-PCR screen was developed, the 21/08/08 combined worm supercontig database had been released, and 97 CYP tags had been identified. By this stage, it was hoped that the predicted 1-3 fold coverage in the contig databases would be reflected in the presence of at least one tag for every CYP gene encoded in the *H. contortus* genome.

Initially, attempts were made to assay CYP gene expression semi-quantitatively by comparing the optical density of CYP tag RT-PCR products run on ethidium bromide agarose gels. At this stage, 76 CYP tags had been identified. Of these, 51 were both amenable to primer design and yielded amplicons of the expected size in either exsheathed L3 or adult cDNA. This preliminary screen was successful to a degree, as it was possible to detect large differences in CYP tag expression, such as those between L3 and adult stages of the parasite as shown in Figure 3.3 (some poorly expressed bands were more visible on screen than are reproduced in print). 27 tags were more highly expressed in the L3 stage, four

tags were more highly expressed in the adult and the remainder had similar expression levels between stages.

The differences that might be expected in CYP expression levels before and after induction with xenobiotics were unknown, but it was hoped that if present, they would be sufficiently large to be detected with the screen. However, no differences in expression were identified after exposing MHco3 (ISE) L3 larvae and adult worms to different xenobiotics at increasingly high concentrations (see Section 3.2.2.5). As it was unclear if this inability to detect a difference in CYP expression was due to a lack of induction or to the limitations of the semi-quantitative screen, it was decided that a more sensitive approach of measuring gene expression was required, so a real-time PCR screen was developed.

3.2.2.1 Primer design and reaction specificity

All primers were designed with predicted melting temperatures of 60+/-2°C, lengths of 18-20 nt, and GC contents of >45%. Where possible, amplicons of 160-200 bp (range 120-212 bp) were designed to span introns (see Appendix for primer sequences).

75 CYP tags were amenable to primer design. The remaining 22 partial CYP sequences were either too short or shared 100% nucleotide identity with the predicted coding sequence of another CYP tag for which primers had already been designed.

Primer pairs were tested with PCR amplification of adult and L3 first strand cDNA synthesised from total RNA after a double (in solution and on column) DNase digest to remove genomic contamination. A single band of the expected size when run on an ethidium bromide agarose gel was taken as a positive result. This was confirmed with melting curve analysis by the real-time PCR machine for each real-time experiment.

57 CYP tags were successfully amplified from the first pair of primers designed against their sequence and an additional 11 CYP tags were amplified from a second or third pair of primers designed after failure of the first set. Seven CYP tags could not be amplified from adult or L3 cDNA.

3.2.2.2 Normalisation

For real-time RT-QPCR, it is recommended that normalising genes have similar expression levels to those of the genes of interest. However, most CYPs have a very low constitutive expression and the use of a poorly expressed normalising gene was deemed more likely to introduce error and reduce the repeatability of the screen. For this reason, the *H. contortus* homologue of *C. elegans ama-1*, a gene encoding a subunit of RNA polymerase II was used (see Figure 3-4 and Figure 3-5). This gene has been extensively used as a reference transcript for expression analysis in *C. elegans* (Johnstone and Barry 1996). A panel of three control genes, *Hc-act, hc27* and *gtp-ch-1* (see Materials and Methods), were also included in the screen to monitor normalisation as well as acting as positive controls for the sample material.

3.2.2.3 Efficiency

The low expression level of the majority of the CYP tags prevented the use of standard curves as a measure of efficiency, as threshold was only reached in the most concentrated samples in a serial dilution. For this reason, the comparative CT method (assumes efficiency is 100% for both the normalising gene and the gene of interest) was used. Standard curves were run for the normalising gene *Hc-ama* to confirm its efficiency was close to 100% (see Figure 3-6).

In addition, linear regression on the fluorescence during the exponential phase of the PCR was used to estimate amplification efficiency (E) of all primer pairs using the computer software LinRegPCR (Ramakers *et al* 2003). For MHco3 (ISE) isolate adult cDNA, 59 of the 74 primer pairs had E>0.90, five had E=0.70-0.90 and ten had E<0.70 (although the latter group included six genes for which adult expression was only barely detected). All positive controls had E>0.98. For adults of both the MHco4 (WRS) isolate and the MHco10 (CAVR) isolate, only 35 of the CYP primer pairs had E>0.90 (23 of which were for the same tags in both isolates) confirming the limitations of applying a screen designed for MHco3 (ISE) cDNA to other isolates. The efficiencies of the control genes appeared relatively robust, with E>0.92 for the full panel of controls in both resistant isolates.

For reproducible experiments, run-to-run variability in primer efficiency is recommended to be less than 5% so the range of E values calculated for each primer pair by LinRegPCR was recorded for three biological replicates of MHco3 (ISE) isolate adult cDNA. However, only five primer pairs demonstrated a range of less than 5%, with significantly greater variability shown by the remainder.

3.2.2.4 Reproducibility

The technical reproducibility of the screen was assessed by running two replicate experiments using the same cDNA sample (intra-assay variation) and two replicate experiments using different batches of cDNA synthesised from the same RNA sample (inter-assay variation). *H. contortus* is known to be a highly polymorphic species, so two replicate experiments comparing cDNA from MHco3 (ISE) worm populations from two different donor sheep were also performed. In all experiments, reproducibility was best for the most highly expressed genes, with an increase in precision as CT value decreased (see Figure 3-7). Pearson's correlation co-efficient (r^2) was 0.956 for intra-assay variation, 0.953 for inter-assay variation, and 0.951 for MHco3 (ISE) worm populations between donors.

The screen was spread over two 96-well plates. As shown in graph A (Figure 3-7), the CT values for plate 1 and plate 2 replicates of the more highly expressed control genes were slightly staggered, suggesting a between-plate effect. Therefore, normalised CT values were also assessed, using three replicates of the *Hc-ama* gene per plate. This improved Pearson's correlation co-efficient for all data sets: 0.966 for intra-assay variation, 0.959 for inter-assay variation, and 0.952 for MHco3 (ISE) worm populations between donors.

3.2.2.5 Data analysis

Measurements of multiple genes in parallel are at risk of generating large numbers of false positives and statistical analysis must attempt to control this bias (Breitling 2006). Microarray studies are an extreme example of this, so the most common approaches used to analyse the large volume of data they generate were applied to the real-time screen.

Most published microarray studies consider a two-fold increase/decrease in measured level as potential differential expression, although there is no firm theoretical basis for selecting this level as significant (Mocellin and Rossi 2007). When this cut-off was applied to the preliminary results from the comparison of CYP expression between life stages, which, based on CYP expression data from *C. elegans,* were expected to show differences, 43 tags were identified as more highly expressed in one stage than the other. However, the Student's t-test found only 27 of these were significant at $P<0.05$.

Bonferroni's correction reduces the P-value threshold in proportion to the number of comparisons made, to correct for the increase in probability of identifying false positives inherent in larger scale studies. This simple correction was applied to the same adult versus L3 expression data for the student's t-test at $P<0.05$. However, at this level of stringency, no tags were identified as differentially expressed. The adjusted result was thought to be too strict for this real-time screen for a number of reasons. Firstly, knowledge of the CYP family in other species would suggest that a number of tags would show some differential expression, and the identification of a number of candidates using the student's t-test at both $P<0.05$ and $P<0.01$ but none after the Bonferroni correction, would suggest the adjustment was too severe. Secondly, the focus of this screen was on genes most likely to show differential expression rather than a non-discriminating genome-wide screen, so it could be argued that this would increase the probability that a tag identified as differentially expressed would be a true result. Thirdly, a number of the tags were expected to represent the same gene, which would theoretically decrease k, the number of comparisons in the dataset, and thus the proportional reduction in P inferred by Bonferroni's correction.

Although the results from the student's t-test at $P<0.05$ without adjustment would equate to a false discovery rate (FDR) of 13.8%, which would be on the high side for a strict microarray analysis (normally set between 1-10% (Breitling 2006)), the aim of the screen was to identify genes with possible higher expression for further analysis, so it was deemed best to tolerate an occasional false positive to improve the chances of finding such candidates.

3.2.3 Constitutive CYP gene expression in different life stages, sexes and tissues

The constitutive expression of the CYP tags was compared between samples, which, based on the CYP family in other species, would be expected to show marked differences in expression profiles. The reasons for this were threefold: firstly, to test the ability of the screen to detect differences in gene expression, secondly, to detect tags with similar expression profiles to help gene assembly and thirdly, to learn more about the CYP family in *H. contortus*.

Preliminary results had suggested that there were large differences in CYP expression between L3 larvae and adults in *H. contortus* as a stage specific constitutive expression had been clearly detected with the semi-quantitative screen (see Figure 3-3).

Therefore, comparisons between adult and two larval stages (L3 and L4), between male and female adult worms, and between intestine and soma (worm body with intestine removed) were undertaken with the real-time screen.

3.2.3.1 Real-time PCR screen comparison of CYP gene expression in L3 larvae and adult worms

Constitutive CYP expression in MHco3 (ISE) L3 larvae and 21-day adult worms was compared with the real-time screen (see Figure 3-8). Three biological replicates (material from different populations of worms from different donor sheep) were run. If an L3 or adult sample failed to reach threshold in one biological replicate for a particular CYP tag, data from the remaining two biological replicates were used (marked *). The expected profiles of adult-expressed *hc27* and L3-expressed *gtp-ch-1* genes were observed in the panel of controls (Hartman *et al* 2001, Hoekstra *et al* 2000) although low levels of expression of both *hc27* and *gtp-ch-1* were detected in the L3 and adult stages respectively, but not described in the literature.

The vast majority of CYP tags were more highly expressed in L3s. The exceptions were Hc-cyp-tag15, Hc-cyp-tag25, Hc-cyp-tag34, Hc-cyp-tag54 and Hc-cyp-tag60, which had higher constitutive expression in adult worms, although the

differences in Hc-cyp-tag34 and Hc-cyp-tag54 expression were not statistically significant at $P<0.05$.

Hc-cyp-tag9, Hc-cyp-tag17, Hc-cyp-tag23, Hc-cyp-tag28, Hc-cyp-tag29, Hc-cyp-tag30, Hc-cyp-tag33, Hc-cyp-tag36, Hc-cyp-tag37, Hc-cyp-tag44, Hc-cyp-tag51, Hc-cyp-tag58, Hc-cyp-tag64, Hc-cyp-tag65, Hc-cyp-tag69, Hc-cyp-tag76, Hc-cyp-tag77, Hc-cyp-tag86, Hc-cyp-tag89, Hc-cyp-tag94 and Hc-cyp-tag95 did not reach threshold in one or other stage in more than two experiments, so relative expression could not be determined for these genes. Theoretically, nominal relative expression levels could be calculated by assigning a value of 40 (maximum number of cycles) to tags that failed to cross the threshold in all three biological replicates for one or other sample. However, this practice was not used for the CYP screen because with the high CT values recorded for many of the expressed tags, there would be a risk that a tag expressed at a low level in one sample and not expressed in the other, but assigned a nominal value of 40, could appear more highly expressed in the latter due to normalisation. However, when absolute expression rather than relative expression of the above genes was considered (see Figure 3-10), Hc-cyp-tag29 was shown to be constitutively expressed in the L3 stage but not the adult, and Hc-cyp-tag77, Hc-cyp-tag94 and Hc-cyp-tag95 were constitutively expressed in the adult but not the L3 stage.

3.2.3.2 Illumina (RNA-seq) comparison of CYP gene expression in L3 larvae and adult worms

Results of CYP expression analysis from the real-time screen were compared to data generated by RNA-seq transcriptome analysis. Illumina technology was used to sequence the transcriptome of 21-day adult and L3 stages of the *H. contortus* MHco3 (ISE) isolate. 38 million and 19 million 76 bp reads were generated from sequencing runs of adult and L3 cDNA libraries respectively. The reads were mapped to reference genomic sequence in the 21/08/08 supercontig database (www.sanger.ac.uk/Projects/H_contortus) using Mapping and Assembly with Qualities (MAQ) algorithms. 48% of the adult reads and 43% of the L3 reads mapped uniquely to sequence in the database.

As expected from the findings of the real-time screen, adult expression of the majority of CYPs was low. This was reflected in few or no reads mapping to the predicted coding regions of most CYP tags. The exceptions were Hc-cyp-tag3, Hc-cyp-tag11, Hc-cyp-tag15, Hc-cyp-tag25, Hc-cyp-tag34, Hc-cyp-tag54 and Hc-cyp-tag60, which had relatively good coverage over all available predicted coding sequence. The predicted coding sequence of Hc-cyp-tag25 had the highest maximum read depth of 66.

The coverage of the CYPs mapped with L3 transcriptome reads was significantly better than for the adults, although still relatively low compared to many neighbouring genes on the contigs studied. Highest coverage was seen for Hc-cyp-tag23, with a depth of up to 136 reads mapping to the predicted coding sequence.

The number of reads mapping to each real-time PCR primer binding site was used to compare relative expression of the CYP tags between adult worms and L3 larvae. The number of reads mapping to the CYP tags were normalised for each sample with the number of reads mapping to *Hc-ama*. Encouragingly, the results closely matched those of the real-time screen (see Figure 3-8).

3.2.3.3 Real-time PCR screen comparison of CYP gene expression in L4 larvae and adult worms

A similar overall pattern of constitutive CYP expression in the larval stage was seen in three biological replicates of MHco3 (ISE) L4 larvae and 21-day adult worms compared with the real-time screen (see Figure 3-9). The L4 expression of the L3 (*gtp-ch-1*) and adult (*hc27*) positive control genes are not described in the literature, but the results of the screen suggested *hc27* was expressed more highly in the adult than the L4 and no difference was seen in expression of *gtp-ch-1* between the adult and L4, both of which are in-keeping with the known adult and L3 stage-specificity of the controls (Hartman *et al* 2001, Hoekstra *et al* 2000).

In common with the L3 and adult comparison, the large majority of CYP tags were more highly expressed in the larval stage with the exceptions of Hc-cyp-tag15 and Hc-cyp-tag25, which were again more highly expressed in the adult.

Unlike in the L3 and adult comparison, Hc-cyp-tag34, Hc-cyp-tag54 and Hc-cyp-tag60 were not more highly expressed in the adult stage than the L4.

A lack of constitutive expression of Hc-cyp-tag29 was again noted in the adult stage, although in contrast to the L3 stage, there appeared to be no L4 expression of this tag. No L4 expression was detected for Hc-cyp-tag23, Hc-cyp-tag28, Hc-cyp-tag33, Hc-cyp-tag36, Hc-cyp-tag44, Hc-cyp-tag58 or Hc-cyp-tag77 (see Figure 3-10).

3.2.3.4 Real-time PCR screen of CYP gene expression in soma and intestine of adult worms

The intestine is thought to be the main organ of detoxification in nematodes and a number of genes associated with xenobiotic metabolism such as CYPs, GSTs and PGPs show higher expression in the *C. elegans* intestine (An and Blackwell 2003, McGhee 2007, Menzel *et al* 2001). It was therefore of interest to investigate the level of CYP expression in the *H. contortus* intestine.

Constitutive CYP expression in the Beltsville isolate adult soma and intestine were compared with the real-time screen using oligo-dT-primed cDNA kindly provided by Professor D. Jasmer (see Figure 3-11). Soma consisted of worm bodies with the intestines removed, but could not be described as completely intestine free due to the nature of the extraction. Only one biological replicate was performed due to the limited amount of material available.

The majority of CYP tags were more highly expressed in the intestine. The exceptions were Hc-cyp-tag6, Hc-cyp-tag12, Hc-cyp-tag14, Hc-cyp-tag15, Hc-cyp-tag25, Hc-cyp-tag40, Hc-cyp-tag56, Hc-cyp-tag70, Hc-cyp-tag73, Hc-cyp-tag74, Hc-cyp-tag75, Hc-cyp-tag76, Hc-cyp-tag80, Hc-cyp-tag81 and Hc-cyp-tag95, which were more highly expressed in the soma.

3.2.3.5 Real-time PCR screen of CYP gene expression in adult male and female worms

A comparison of CYP expression in *H. contortus* males and females was of interest because a number of CYP genes in *C. elegans* have sex-specific

expression and are required for normal reproductive development and behaviour (Benenati *et al* 2009, Gerisch and Antebi 2004, Kleemann *et al* 2008), and may have functional orthologues in the parasite.

Constitutive CYP expression in adult MHco3 (ISE) male and female worms was compared with the real-time screen (see Figure 3-12). Three RNA replicates were performed using material isolated from a group of adult males and a group of adult females harvested from the same donor sheep. In cases where a single well failed to reach threshold, data from the remaining two replicates was used (marked *).

25 CYP tags were more highly expressed in the male sample at P<0.05 and *Hc-nhr* and *Hc-pgp* also appeared to be more highly expressed in males. Hc-cyp-tag15, Hc-cyp-tag25, Hc-cyp-tag73, Hc-cyp-tag74 and Hc-cyp-tag75 were more highly expressed in the female sample, although only the first two were statistically significant at P<0.05. Hc-cyp-tag15 and Hc-cyp-tag25 were also more highly expressed in the adult than larvae and more highly expressed in the soma than the intestine, while Hc-cyp-tag73, Hc-cyp-tag74 and Hc-cyp-tag75 were more highly expressed in the L3 than the adult and more highly expressed in the intestine than the soma. These tags are discussed further in Chapter 4.

3.2.4 CYP gene expression in xenobiotic-exposed *H. contortus*

CYPs are induced on exposure to their substrate (Xu *et al* 2005). In *C. elegans*, a number of CYPs have been shown to be induced on exposure to xenobiotics (Chakrapani *et al* 2008, Menzel *et al* 2001, Menzel *et al* 2005). It was of interest to investigate the induction of *H. contortus* CYPs on exposure to anthelmintics due to their possible role in drug metabolism and resistance.

Concentrations for *in vitro* drug exposures of *H. contortus* L3 larvae and adults were chosen based on those used in published studies of xenobiotic responses in *C. elegans* (Chakrapani *et al* 2008, Menzel *et al* 2001), *H. contortus* larvae (Boisvenue *et al* 1983, Kotze 1997), *H. contortus* adults (Cvilink *et al* 2008a,O'Grady and Kotze 2004) and clinically relevant doses. Acute drug exposures at very high concentrations over a short time period were also performed (see Table 3-1).

3.2.4.1 CYP gene expression in IVM-exposed L3

Initial attempts to detect differences in CYP expression after xenobiotic exposure were based on L3 material for three reasons: L3 larvae were more readily available and limited the number of donor sheep required for experiments; L3 were more amenable to *in vitro* work as they survived well in liquid culture; studies by Kotze (1997) had demonstrated greater oxidase activity towards model CYP substrates in *H. contortus* L3 than in adults and this was supported by the results of the real-time PCR screen, which indicated higher CYP expression in L3 than adults.

No CYP induction was identified with the semi-quantitative screen in L3 after exposure to an extremely high concentration (5 μg/ml) of IVM in 0.05% DMSO for four hours relative to control larvae exposed to 0.05% DMSO only. As it was unclear if the lack of a detectable increase in CYP expression was a true result or the product of an insensitive assay, the experiment was repeated after the real-time screen was developed. However, no tags with a higher expression were identified in the IVM-exposed L3 with this screen.

3.2.4.2 CYP gene expression in IVM-exposed adult worms

The expression of CYP tags in MHco3 (ISE) adult worms exposed to 5 μg/ml ivermectin dissolved in 0.05% DMSO for five hours in liquid culture were compared with adult worms exposed to 0.05% DMSO in liquid culture only. Worms harvested from one donor sheep were split into six batches, allowing the drug exposures to be run in triplicate. This generated material for three biological replicates.

Gene expression in IVM-exposed and DMSO-control worms had originally been compared with the semi-quantitative screen. However, the low constitutive adult CYP expression appeared largely unaffected by the drug exposures and it was not possible to detect any differences between the IVM-exposed and DMSO-control worms.

Conversely, the real-time screen did detect a number of differences in expression, but these were highly variable (see Figure 3-13). The only tag with

statistically significant higher expression in the IVM-exposed samples was Hc-cyp-tag9, but Hc-cyp-tag41, which is thought to represent the same gene (see Chapter 4), showed statistically significant lower expression. Hc-cyp-tag24 appeared more highly expressed in biological replicate one and three but not in biological replicate two, while Hc-cyp-tag62 appeared more highly expressed in biological replicates one and two, but not in biological replicate three.
The expression levels of the remainder of the tags were variable with no reproducible induction observed in the IVM-exposed worms. Indeed, a number of tags appeared to have slightly reduced expression in the IVM-exposed samples, but again the results were highly variable.

3.2.4.3 CYP gene expression in ABZ-exposed adult worms

The expression of CYP tags in MHco3 (ISE) adult worms exposed to 1 mg/ml albendazole for five hours in liquid culture was compared with adult worms maintained in RPMI medium only. Worms harvested from one donor sheep were split into six batches, allowing the drug exposures to be run in triplicate. This generated material for three biological replicates of the real-time screen.

Similarly to the IVM-exposure results, the expression levels were highly variable (see Figure 3-14). Unexpected higher expression of *Hc-act, gtp-ch-1* and *hc27* control genes were measured in the ABZ-exposed sample, complicating interpretation of the results. Four CYP tags show a statistically significant higher expression level in the ABZ-exposed samples: Hc-cyp-tag1, Hc-cyp-tag6, Hc-cyp-tag40 and Hc-cyp-tag95, and ten additional tags showed non-statistically significant higher expression in the ABZ-exposed sample in three biological replicates.

3.2.4.4 RNA-seq comparison of CYP gene expression in ABZ-exposed adult worms

Illumina technology was used to sequence the transcriptome of 21-day adults of the *H. contortus* MHco3 (ISE) isolate after exposure to 300 μg/ml albendazole (in 2% DMSO) or 2% DMSO for five hours in liquid culture. The resulting reads were mapped to MHco3 (ISE) reference genomic sequence in the 21/08/08 supercontig

database (www.sanger.ac.uk/Projects/H_contortus) using Mapping and Assembly with Qualities (MAQ) algorithms.

24 million and 28 million 76 bp reads were generated from sequencing runs of ABZ-exposed and DMSO-exposed worm libraries respectively.

Although the mapped read coverage for the panel of control genes were good for both libraries, few or no reads mapped to the CYP tags preventing the comparative analysis of this data for this particular gene family.

3.2.5 CYP gene expression in resistant isolates of *H. contortus*

3.2.5.1 Real-time PCR screen comparison of constitutive CYP gene expression in adult MHco3 (ISE) and MHco4 (WRS) isolates

The constitutive expression of CYP genes in adult MHco3 (ISE) and MHco4 (WRS) isolates was compared. The WRS isolate is a field-derived strain from South Africa. It is resistant to ivermectin, benzimidazoles, rafoxanide and closantel (van Wyk *et al* 1987, van Wyk and Malan 1988).

Adult worm material was harvested from three donor sheep for each isolate, allowing three biological replicates of the real-time screen to be run (see Figure 3-15). Primers were designed to amplify MHco3 (ISE) cDNA, so it was hypothesised that the cDNA sequence of a number of tags in the resistant strain would be too divergent to amplify (see Chapter 1). The results of the screen reflected this; primers for one tag would not amplify MHco4 (WRS) cDNA and 24 tags showed more than two-fold lower expression in the resistant isolate. This is expected to be an overestimate of the true figure due to polymorphism at the primer binding sites reducing PCR efficiency in the resistant isolate.

One of the control genes, *Hc-act*, showed higher expression in MHco4 (WRS), but the other control genes showed little difference in expression between isolates. Although not statistically significant at $P<0.05$, more than two-fold higher expression in the resistant isolate was detected for Hc-cyp-tag9, Hc-cyp-tag30, Hc-cyp-tag41 and Hc-cyp-tag77.

To investigate this further, individual RT-QPCR of RNA replicates of the three biological replicates used in the real-time PCR screen were performed to test for consistency and individual RT-QPCR experiments were run on two to three new biological replicates, depending on the availability of worm material. The results are shown in Figure 3-16 and reported below.

Hc-cyp-tag41 and Hc-cyp-tag9

Higher expression of Hc-cyp-tag41 was seen in the MHco4 (WRS) isolate for two of three real-time screen biological replicates, with the magnitude of difference ranging from 2.24 to 2.32 log fold change higher in the resistant isolate. The relative expression in the third biological replicate in the real-time PCR screen was unknown due to the MHco3 (ISE) well failing to reach threshold. However, little difference in expression was detected for two subsequent individual RT-QPCR experiments using worm material from the real-time screen third biological replicate (1.26 and 1.24 times higher in the resistant isolate). Three further individual RT-QPCR experiments, each with different worm material, found 2.67 to 20.1 fold higher expression in the MHco4 (WRS) isolate relative to MHco3 (ISE).

Hc-cyp-tag41, Hc-cyp-tag55, Hc-cyp-tag70 and Hc-cyp-tag9 are thought to amplify the same gene (see Chapter 4). Another tag, Hc-cyp-tag61, represents sequence encoding an identical putative 122 amino acid N-terminus to Hc-cyp-tag41. The coding sequence shares 98% nucleotide identity, but the genomic sequence is only 72% identical. Primers for Hc-cyp-tag41 could also amplify Hc-cyp-tag61 but the primers for Hc-cyp-tag61 should not amplify Hc-cyp-tag41 due to the specificity of the reverse primer.

Expression of Hc-cyp-tag9 was shown to be higher in biological replicates one and two of the real-time screen (2.04 and 2.66 log fold change), but again the MHco3 (ISE) isolate well failed to reach threshold in biological replicate three. Subsequent individual RT-QPCRs with RNA replicates of the three real-time screen biological replicates plus two new biological replicates detected no higher expression of Hc-cyp-tag9 in the resistant isolate (from 0.27 to 1.00 fold MHco3 (ISE) expression).

Higher expression of Hc-cyp-tag55, Hc-cyp-tag70 and Hc-cyp-tag61 were not detected in the MHco4 (WRS) isolate, even in samples showing higher expression of Hc-cyp-tag41.

Hc-cyp-tag2

With the real-time screen, markedly higher expression of Hc-cyp-tag2 (7.73 log fold difference) in the resistant isolate was detected in the first biological replicate but not in subsequent biological replicates. Individual real-time RNA replicates of the three real-time screen biological replicates failed to detect any higher expression in MHco4 (WRS) relative to MHco3 (ISE), although three subsequent biological replicates did generate a wide range of values from 1.0 to 25.1 fold higher expression in the resistant isolate.

Hc-cyp-tag77

Hc-cyp-tag77 showed a higher constitutive expression in the MHco4 (WRS) isolate in three real-time screen biological replicates of 1.01 to 3.49 log fold change. Individual real-time RNA replicates suggested a more modest difference of 1.88 to 2.73 fold higher expression in MHco4 (WRS). However, two subsequent individual real-time biological replicates detected no higher expression of Hc-cyp-tag77 in the resistant isolate.

Hc-cyp-tag4 is thought to represent the same CYP as Hc-cyp-tag77 (see Chapter 4). However results of the real-time screen for Hc-cyp-tag4 were variable, showing 1.54, 1.29 and 0.29 log fold changes for the three biological replicates.

Hc-cyp-tag94

Although only one biological replicate with the real-time screen showed higher expression of Hc-cyp-tag94 in MHco4 (WRS), it was included for individual real-time experiments due to its putative higher expression in resistant isolate MHco10 (CAVR) (see Section 3.2.5.2). The expression level of Hc-cyp-tag94 appeared similar in both MHco3 (ISE) and MHco4 (WRS) in three RNA replicates of the real-time screen material and in one of three new biological replicates.

However, higher expression of 3.21 to 13.9 fold that in the susceptible isolate was measured in two biological replicates.

Hc-cyp-tag30

Higher expression of Hc-cyp-tag30 in the resistant isolate was detected in two of three biological replicates with the real-time screen (3.83 and 1.93 log fold change) with the MHco4 (WRS) well failing to reach threshold in the remaining replicate. Experiments to confirm this with individual real-time RNA replicates were not undertaken, but three new biological replicates failed to detect a higher expression level in MHco4 (WRS).

3.2.5.2 Real-time screen comparison of constitutive CYP expression in adult MHco3 (ISE) and MHco10 (CAVR) isolates

The constitutive expression of CYP genes in adult MHco3 (ISE) and MHco10 (CAVR) isolates were compared. The CAVR isolate is a field-derived strain from Australia, which is resistant to ivermectin. It also shows moderate resistance to the benzimidazoles, but is sensitive to levamisole (Le Jambre *et al* 1995).

Worm material was harvested from three donor sheep for each isolate, allowing three biological replicates of the real-time screen to be run (see Figure 3-17). Individual quantitative real-time PCR experiments were then used to determine the comparative expression of any tags shown by the screen to be more highly expressed in the resistant isolate (see Figure 3-18).

As mentioned above, primers were designed to amplify MHco3 (ISE) cDNA, so it was hypothesised that the cDNA sequence of a number of tags in the resistant strain would be too divergent to amplify. The results of the screen reflected this; primers for nine tags would not amplify MHco10 (CAVR) cDNA and 21 tags showed more than two-fold lower expression in the resistant isolate.

Again, *Hc-act*, showed higher expression in the resistant isolate, but the other control genes showed little difference. A number of CYP tags showed a greater than two-fold higher expression in the MHco10 (CAVR) isolate and were followed

up with individual real-time experiments. These included Hc-cyp-tag61 and Hc-cyp-tag94 which were statistically significant at $P<0.05$.

Hc-cyp-tag61, Hc-cyp-tag41, Hc-cyp-tag55 and Hc-cyp-tag9

Hc-cyp-tag41, Hc-cyp-tag55, Hc-cyp-tag70 and Hc-cyp-tag9 are thought to represent the same gene (see Chapter 4). Another tag Hc-cyp-tag61 represents 386bp coding sequence sharing 98% nucleotide identity with Hc-cyp-tag41.

Higher expression of Hc-cyp-tag61 in the MHco10 (CAVR) isolate was detected for all three biological replicates with the real-time screen (2.125, 3.48 and 2.56 log fold change). The results of individual real-time RNA replicates of the real-time screen were however variable, with expression levels of 0.38 to 4.94 fold those of the MHco3 (ISE) isolate recorded. Two additional biological replicates failed to detect a higher expression in the resistant isolate.

Although higher expression of Hc-cyp-tag41 in MHco10 (CAVR) was only seen in two of three real-time screen biological replicates (3.486, 0.566 and -0.106 log fold change), two RNA replicates of the real time screen followed by two biological replicates showed a consistently higher expression of Hc-cyp-tag41 in MHco10 (CAVR) of 1.58 to 6.24 fold relative to the MHco3 (ISE) isolate.

Higher expression was also seen with the real-time screen for all biological replicates of Hc-cyp-tag55 and Hc-cyp-tag70 tags, although the magnitude was highly variable: 0.29 to 2.32 log fold change and 0.57 to 2.16 log fold change for Hc-cyp-tag55 and Hc-cyp-tag70 respectively. Four individual real-time experiments showed little or no higher expression in the resistant isolate for Hc-cyp-tag55, other than one RNA replicate of material used in the real-time screen, which showed a 2.0 fold higher expression.

Higher expression of Hc-cyp-tag9 was not detected in MHco10 (CAVR) with the real-time screen and four individual real-time experiments confirmed this result.

Hc-cyp-tag94

Higher constitutive expression of Hc-cyp-tag94 was seen for MHco10 (CAVR) in three real-time screen biological replicates. The magnitude of difference varied from 2.72 to 3.49 log fold change. This was confirmed with three individual gene RT-QPCR experiments, which showed 3.13 to 8.1 fold higher expression of Hc-cyp-tag94 in MHco10 (CAVR) relative to MHco3 (ISE).

Hc-cyp-tag77

Hc-cyp-tag4 and Hc-cyp-tag77 are thought to amplify the same gene (see Chapter 4). Although not statistically significant at $P<0.05$, both Hc-cyp-tag4 and Hc-cyp-tag77 showed a higher but variable expression (1.85, 0.69, 0.1 log fold changes for Hc-cyp-tag4 and 0.81, 1.54, 2.59 log fold changes for Hc-cyp-tag77) in MHco10 (CAVR) with three real-time screen biological replicates.

Individual real-time RNA replicates were undertaken for Hc-cyp-tag77 and showed 2.45 to 3.75 fold higher expression in the resistant isolate. However, two further biological replicates showed a variable expression of 0.39 and 7.69 fold difference.

3.2.5.3 RNA-seq comparison of CYP gene expression in adult worms of MHco3 (ISE), MHco4 (WRS), and MHco10 (CAVR) isolates

Illumina technology was used to sequence the transcriptome of 21-day adults of the *H. contortus* MHco3 (ISE), MHco4 (WRS) and MHco10 (CAVR) isolates. The resulting reads were mapped to MHco3 (ISE) reference genomic sequence in the 21/08/08 supercontig database (www.sanger.ac.uk/Projects/H_contortus) using Mapping and Assembly with Qualities (MAQ) algorithms.

21 million, 17 million and 20 million 54 bp reads were generated from sequencing runs of MHco3 (ISE), MHco4 (WRS) and MHco10 (CAVR) libraries respectively.

Although the mapped read coverage for the control genes *Hc-ama, Hc-act* and *Hc27* was excellent for all isolates, few or no reads mapped to the CYP tags in

the resistant isolates, preventing the comparative analysis of this data for this particular gene family.

3.3 Tables and Figures

SUPERCONTIG	SIZE (bp)	HC-CYP-TAG	REGION (amino acids)	CLOSEST *C. elegans* CYP
1	2193	54	2-87	14A3
2	1869	42	15-87	14A5
3	3074	63	2-148	14A3
4	1507	46	28-84	14A5
5	23637	50_94_97_62_36_35	425-462,115-191,293-421,293-423,374-461 ,1-95	33C7,14A5,33C2,33C9,14A2,14A5
6	15764	60_34_33_49_30	169-461, 115-190, 289-425	14A5, 14A4, 33C1
7	35550	32_68_69	1-90, 285-462	14A3, 33C6
8	11772	91_39_38_37	1-28, 3-187, 226-390	33E1, 33C9, 33C5
9	2455	95_56	360-425, 291-370	33C7, 33E2
10	1290	84	6-141	33C2
11	114720	16	4-470	33C9
12	1283	81	274-288	33C2
13	1908	45	89-143	33C11
14	1652	40	289-359	33C3
15	28419	22_14	11-42, 108-519	33C9, daf-9=22A1
16	11558	65_64	1-323	23A1
17	1623	82	75-177	23A1
18	1731	10	392-533	23A1
19	13166	66	68-107	13B1
20	1206	7	434-494	13B1
21	4342	90_19	1-62, 268-304	13B2, 13A12
22	1266	78	459-493	13A12
23	1411	8	303-358	13A12
24	1508	79	303-435	13A8
25	12604	2_6	17-439, 434-494	13A12, 13B1
26	23674	1_ 5	17-305, 433-486	13A12, 13B1
27	15981	51_92_93	207-359, 27-61, 15-107	32A1, 32A1, 13B1
28	7475	20	317-462	32B1
29	2551	44	239-341	32A1
30	7109	89	21-141	32A1
31	3426	86_87	234-325, 470-506	32A1, 37B1
32	9075	80	454-507	37B1
33	55737	53_67	463-507, 142-521	37B1, 32A1
34	32949	25_15	1-326, 364-493	31A3, 31A2
35	59191	12_13	22-68, 302-494	13A11, 13A10
36	34267	27	454-507	37B1
37	10155	17_18	1-145, 188-448	33A1, 33E1
38	2507	47_48	61-155, 293-329	34A4, 34A7
39	1522	57	150-194	34A7
40	1146	43	167-329	34A8
41	23723	71_72	28-196, 290-454	34A3, 35A2
42	46749	24_23_52	27-293, 295-429, 379-429	34A3, 34A4, 35A1
43	13088	28	24-147	34A4
44	1110	26	104-137	36A1
45	4635	59_58	140-500	36A1
46	1855	88	325-493	36A1
47	4182	29	38-136	36A1
48	53370	21_96	288-457, 419-506	37A1 , 37B1
49	1024	41	1-122	42A2
50	1400	61	1-122	42A1
51	42261	70_9	274-392, 395-508	42A1, 42A2
52	1054	55	123-210	42A1
53	12719	4	1-467	43A1
54	5373	77	42-142	43A1
55	58931	11_3	48-141, 292-467	43A1, 43A1
56	1501	85	440-481	44A1
57	1405	31	403-439	44A1
58	1268	75	413-482	44A1
59	1020	76	399-481	44A1
60	5818	73_74	96-405	44A1
61	2584	83	57-94	44A1

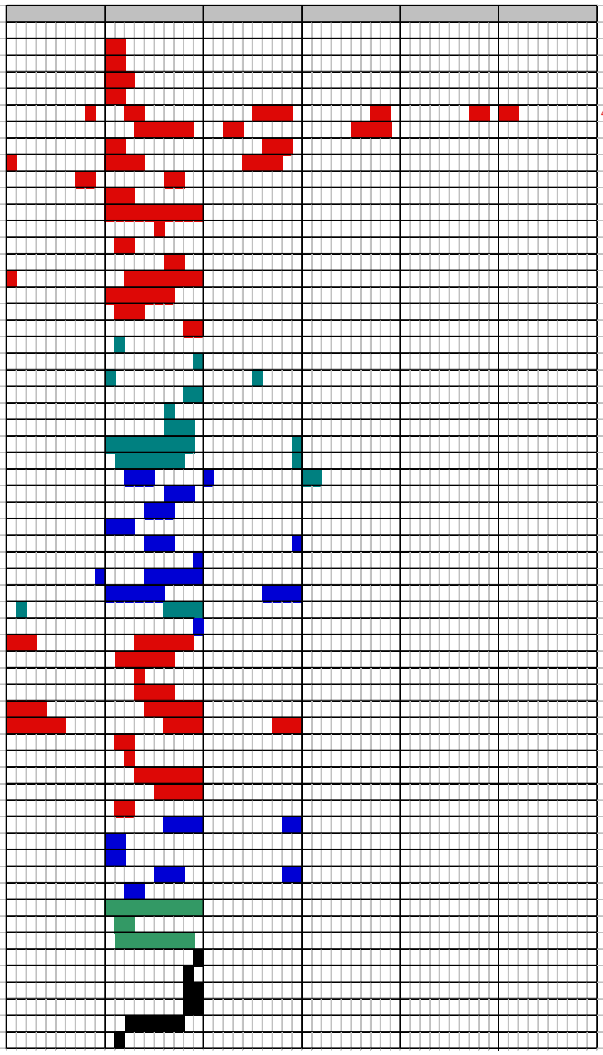
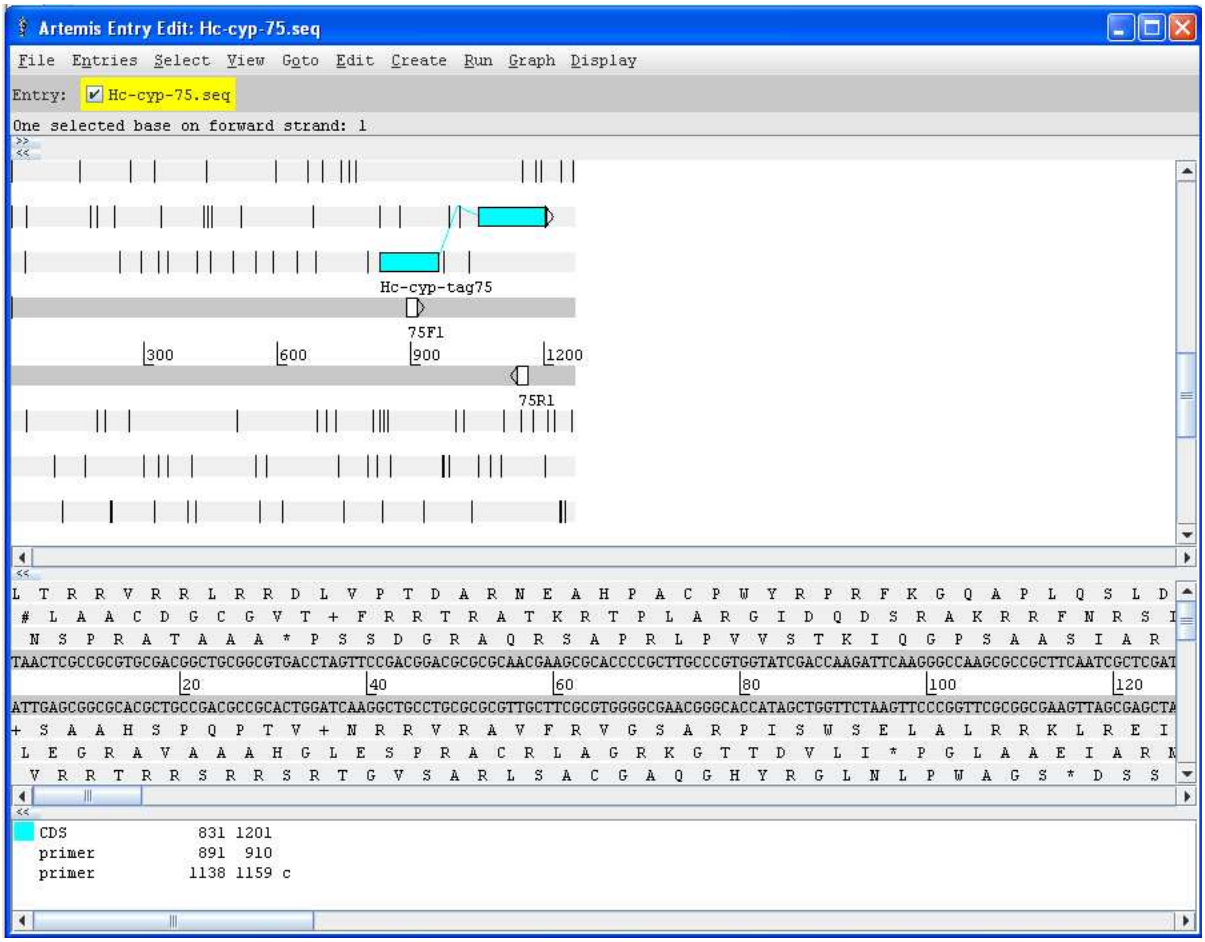


Figure 3.1: Schematic of 97 CYP tags on 61 supercontigs.
Each row represents a supercontig and each column represents a full-length CYP gene. Coloured regions show the proportion of the gene represented by each tag. Colours relate to the closest *C. elegans* CYP clade (see Chapter 4).

A:



B:

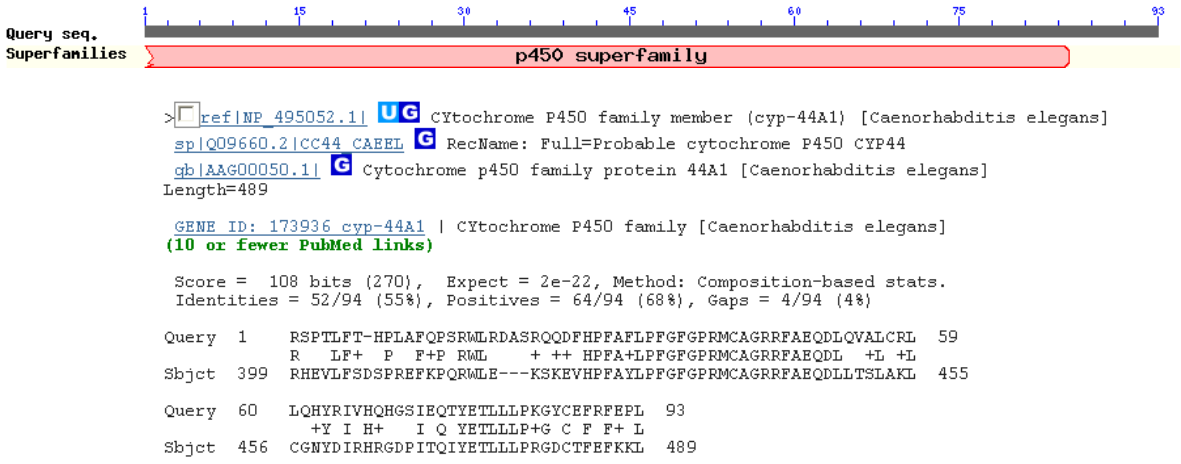


Figure 3.2: Example of a typical CYP tag sequence and typical BLAST result.
A: Coding sequence of Hc-cyp-tag75 is shown in blue and RT-PCR primers are in white
B: BLASTp search with conceptual translation of Hc-cyp-tag75 into NCBI database yields a P450 superfamily conserved domain (partial) and the top hit is a *C. elegans* CYP

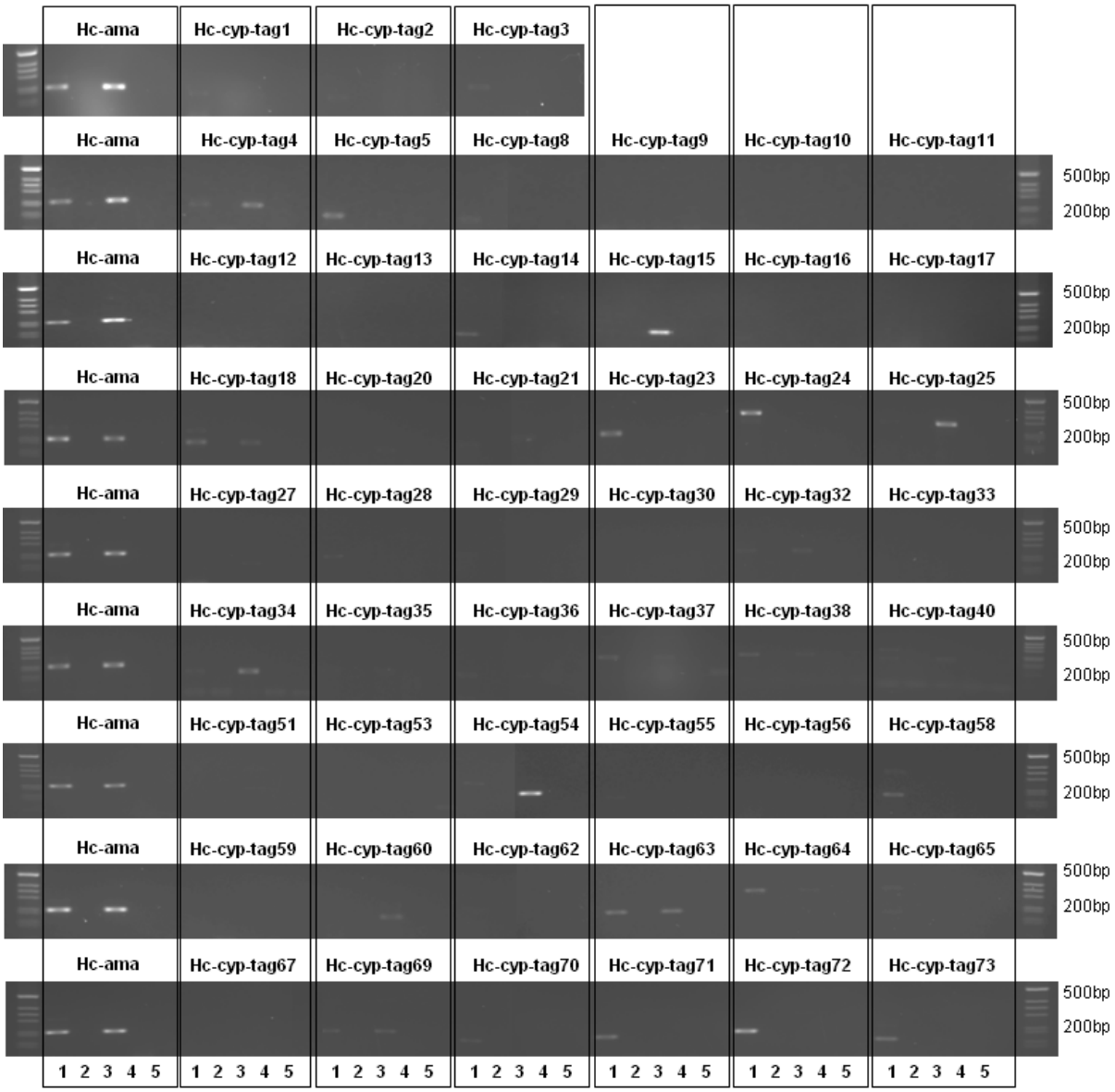


Figure 3.3: Preliminary RT-PCR screen of stage specific constitutive expression of 51 CYP tags in *H. contortus*.
Lanes: 1=L3 cDNA, 2=L3 no-RT control, 3=adult cDNA, 4=adult no-RT control, 5=no template control. *Hc-ama* was included on each gel to normalise CYP expression between stages. 35 PCR cycles were run.

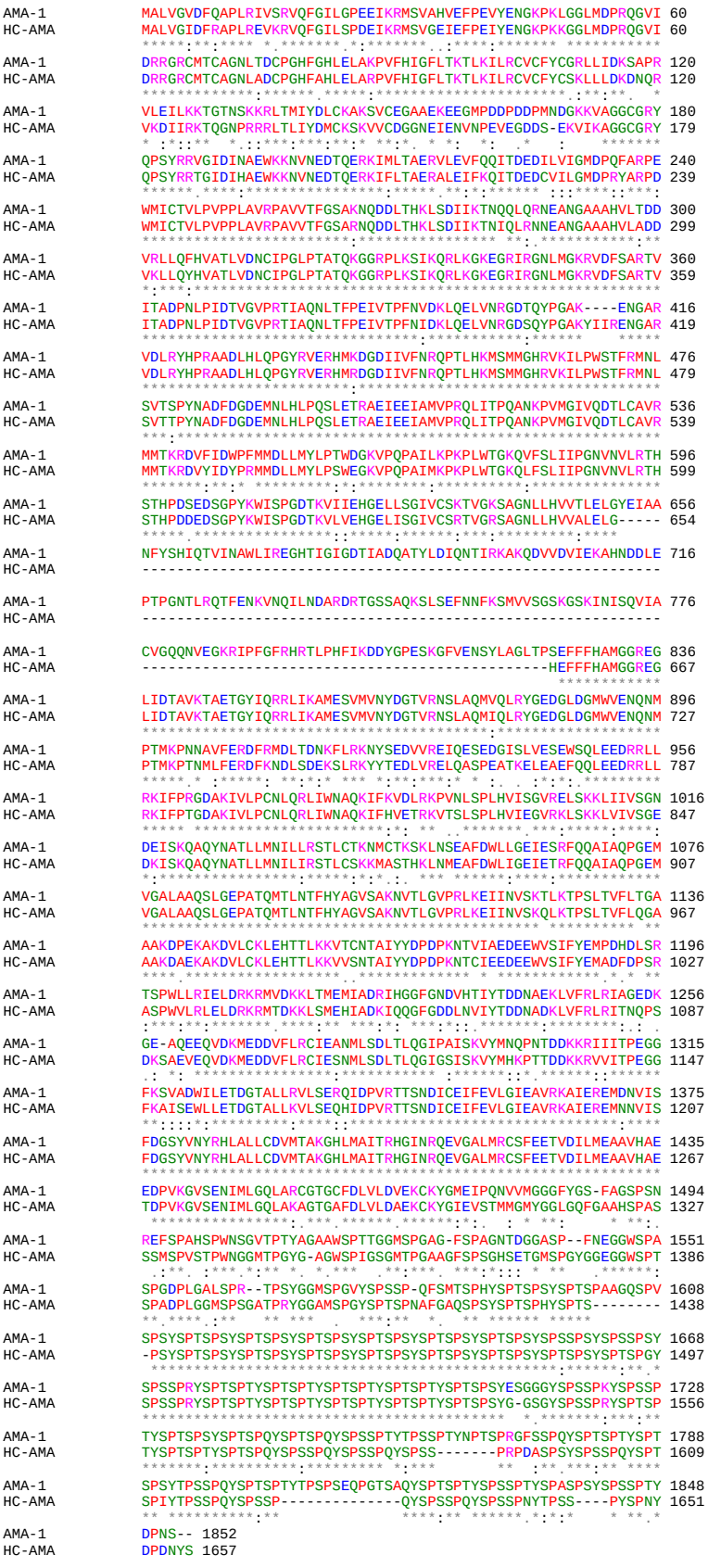


Figure 3.4: ClustalW alignment of *C. elegans* AMA-1 and *H. contortus* HC-AMA. 82% amino acid identity. ~172 amino acids are missing in HC-AMA due to a gap in the supercontig assembly.

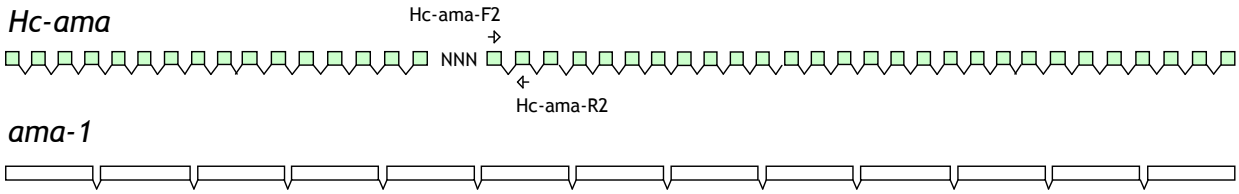


Figure 3.5: Diagram of *C. elegans ama-1* gene and *H. contortus Hc-ama* gene.

Not to scale. NNN indicates a gap in the supercontig assembly. White arrows indicate RT-PCR primer binding sites. The *C. elegans* gene spans 10.1 Kb and has 12 introns. The parasite gene spans at least 10.4 Kb and has at least 43 introns.

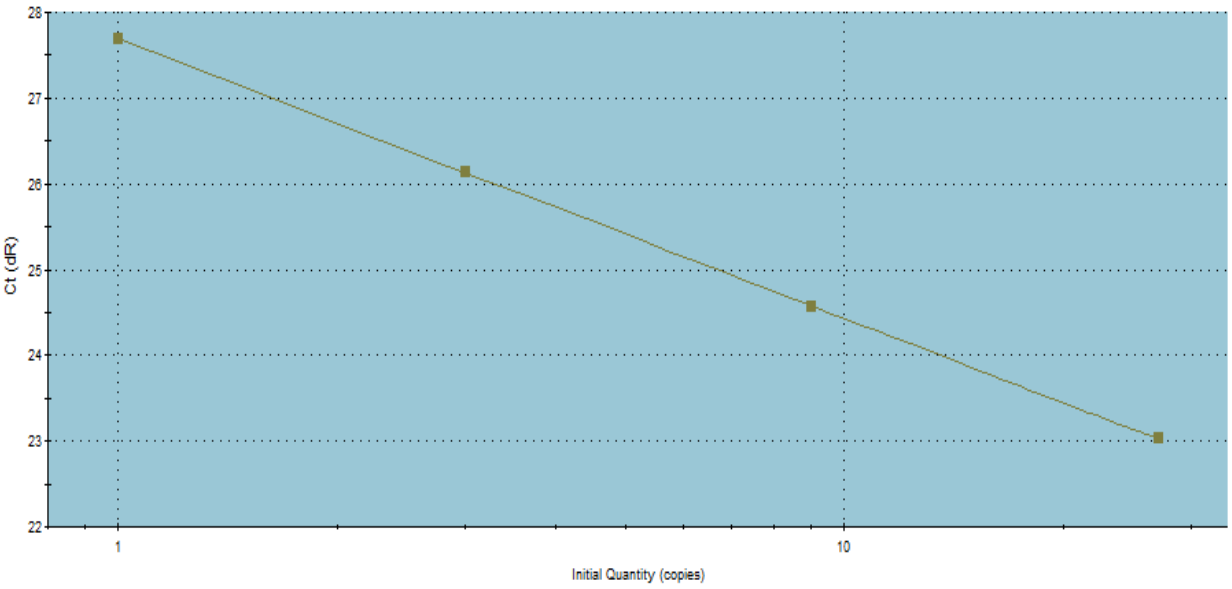
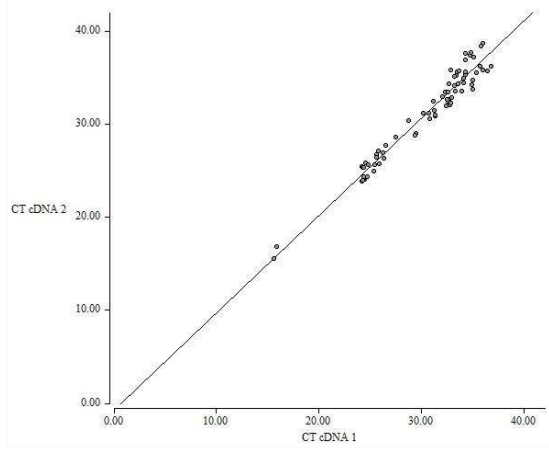


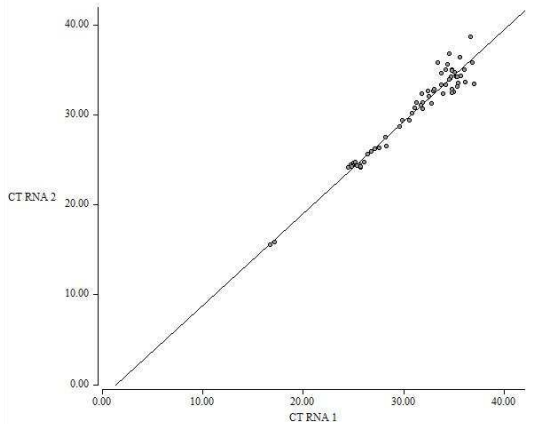
Figure 3.6: *Hc-ama* standard curve.

Adult *H. contortus* MHco3 (ISE) cDNA. 1 in 3 dilution over 4 orders of magnitude. RSq = 1.00, Efficiency = 102.7%.

A:



B:



C:

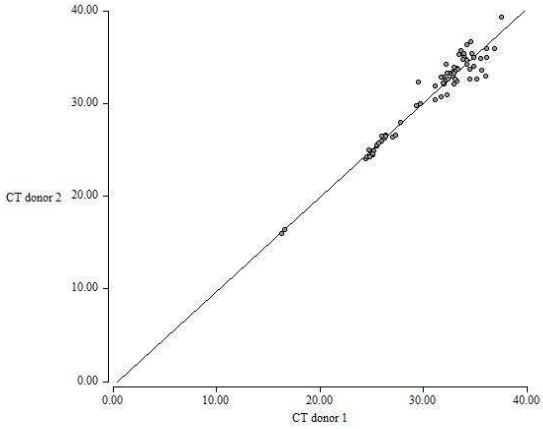


Figure 3.7: Reproducibility of real-time PCR screen.
Each real-time screen uses two 96-well plates. All genes (controls and CYPs) from both plates were included in each replicate. Reproducibility is reflected in the proximity of r2 to 1.
A: cDNA replicate (same RNA isolation): linear regression y = -0.71+1.05x, r2 = 0.956
B: RNA replicate (same worm donor): linear regression y = -1.37+1.02x, r2 = 0.953
C: Biological replicate (different worm donor): linear regression y = -0.44+1.02x, r2 = 0.951

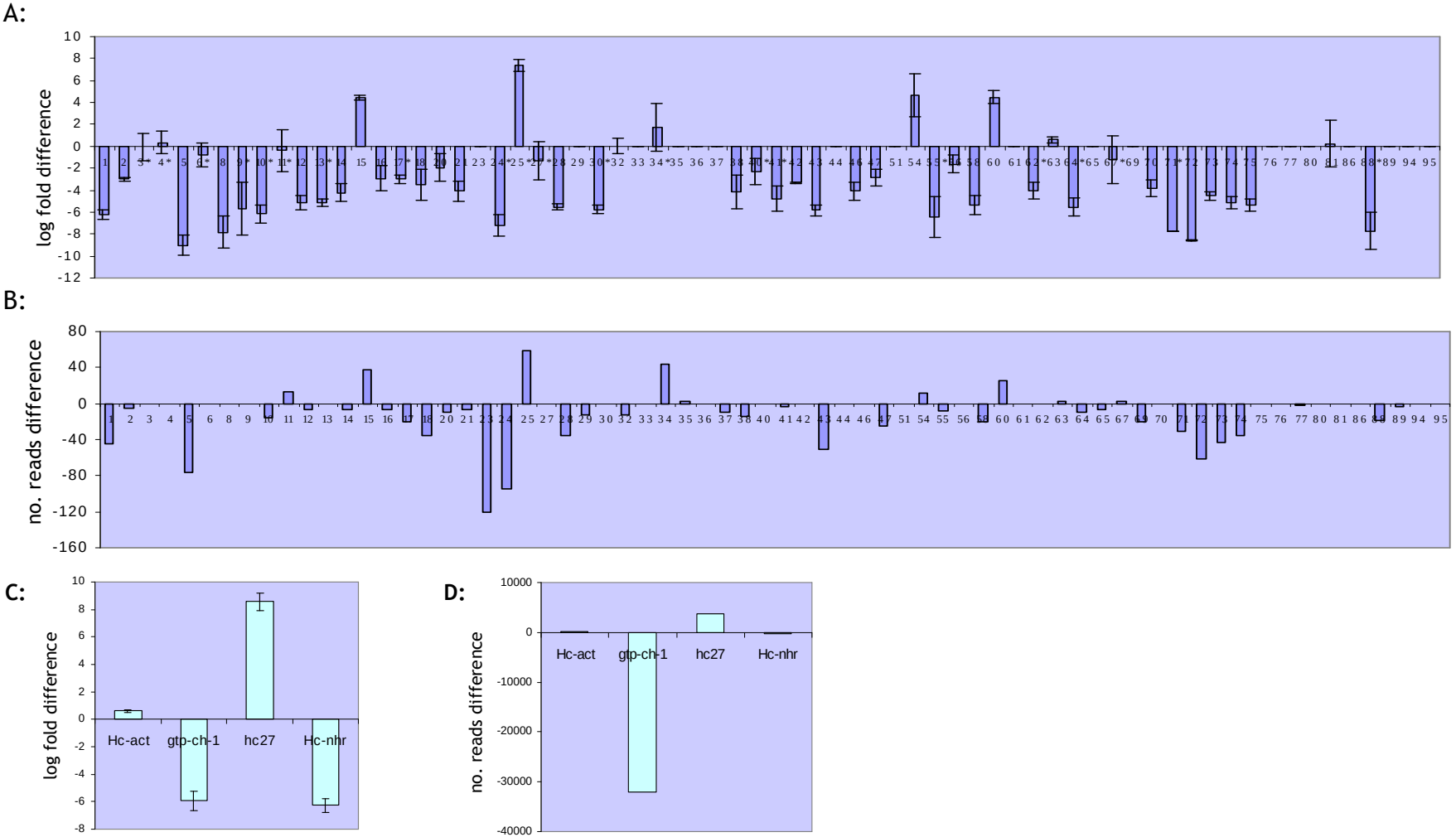
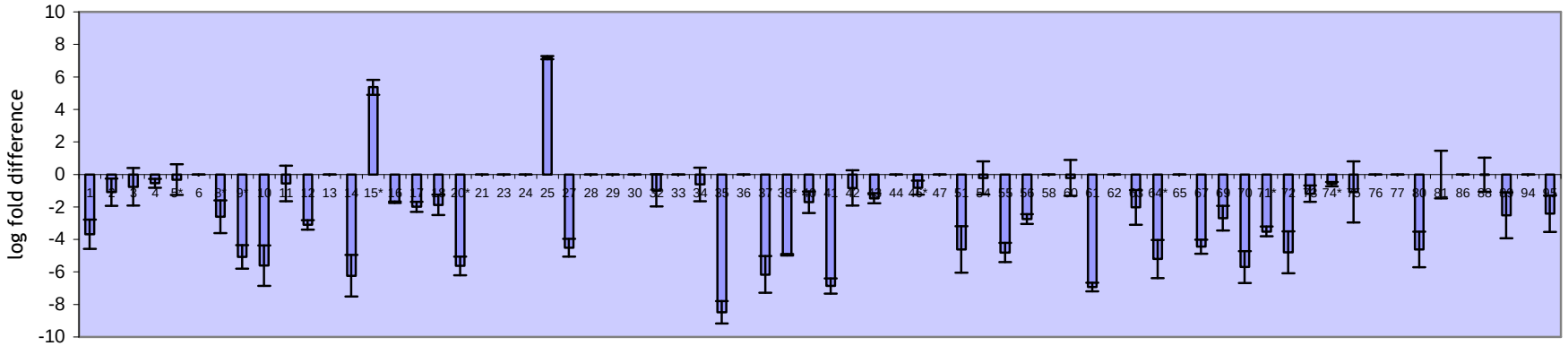


Figure 3.8: Relative expression of CYP tags in L3 and adult worms of MHco3 (ISE) isolate.
Figures on x-axis relate to Hc-cyp-tag numbers. Positive bars indicate higher expression in adults; negative bars indicate higher expression in L3 larvae. A: Three biological replicates of real-time screen. ΔΔCT normalised to *Hc-ama*. Panel of control genes (C) is shown below. B: Number of Illumina transcriptome reads mapped to binding sites of Hc-cyp-tag primers used in real-time screen. Normalised to number of transcriptome reads mapped to *Hc-ama* primer binding sites. Panel of control genes (D) is shown below.

STAGE	ISOLATE	XENOBIOTIC	FORMULATION	CONCENTRATION	SOLVENT/CONTROL	TIME	PHENOTYPE
L3	MHco3	Albendazole	Sigma	500 μg/ml	DMSO (2.5% v/v)	4 hours	normal
L3	MHco3	Albendazole	Albex	1 mg/ml	no solvent	5 hours	normal
L3	MHco4	Albendazole	Albex	1 mg/ml	no solvent	5 hours	normal
L3	MHco10	Albendazole	Albex	1 mg/ml	no solvent	5 hours	normal
L3	MHco4	Albendazole	Albex	10 mg/ml	no solvent	5 hours	normal
L3	MHco3	Ivermectin	Sigma	1 μg/ml	DMSO (0.5% v/v)	4 hours	-
L3	MHco3	Ivermectin	Sigma	2 μg/ml	DMSO (0.05% v/v)	5 hours	~90% paralysed (straight), remainder reduced motility
L3	MHco4	Ivermectin	Sigma	2 μg/ml	DMSO (0.05% v/v)	5 hours	~50% paralysed (straight), remainder reduced motility
L3	MHco10	Ivermectin	Sigma	4 μg/ml	DMSO (0.1% v/v)	5 hours	~80% paralysed (straight), remainder reduced motility
L3	MHco3	Ivermectin	Sigma	5 μg/ml	DMSO (2.5% v/v)	4 hours	100% paralysed (straight)
L3	MHco3	Phenobarbital	-	50 μg/ml	DMSO (0.5% v/v)	24 hours	normal
L3	MHco3	Phenobarbital	-	1 mg/ml	DMSO (1% v/v)	5 hours	normal
L3	MHco3	Phenobarbital	-	1 mg/ml	DMSO (1% v/v)	24 hours	normal
L3	MHco3	Phenobarbital	-	1 mg/ml	DMSO (1% v/v)	24 hours	normal
L3	MHco3	Caffeine	Sigma	10 mg/ml	no solvent	5 hours	~80% paralysed (curved), remainder reduced motility
L3	MHco3	Phenothiazine	Sigma	1mg/ml	no solvent	5 hours	~40% curling up, remainder reduced motility
L3	MHco3	Propanol	-	1% v/v	no solvent	5 hours	~60% paralysed (curved), remainder reduced motility
Adult	MHco3	Albendazole	Sigma	300 μg/ml	DMSO (2% v/v)	5 hours	normal
Adult	MHco3	Albendazole	Albex	1 mg/ml	no solvent	5 hours	normal
Adult	MHco4	Albendazole	Albex	1 mg/ml	no solvent	5 hours	normal
Adult	MHco10	Albendazole	Albex	1 mg/ml	no solvent	5 hours	normal
Adult	MHco3	Ivermectin	Virbamec	5 μg/ml	no solvent	5 hours	100% paralysed (straight)
Adult	MHco4	Ivermectin	Virbamec	5 μg/ml	no solvent	5 hours	100% paralysed (straight)
Adult	MHco10	Ivermectin	Virbamec	5 μg/ml	no solvent	5 hours	100% paralysed (straight)
Adult	MHco3	Ivermectin	Sigma	5 μg/ml	DMSO (0.05% v/v)	5 hours	100% paralysed (straight)
Adult	MHco10	Phenobarbital	-	500 μg/ml	DMSO (0.5% v/v)	4 hours	normal

Table 3.1: *In vitro* drug exposure experiments

A:



B:

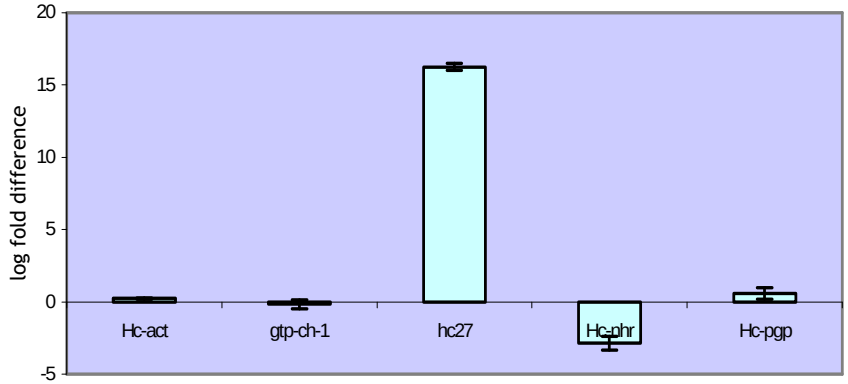


Figure 3.9: Relative expression of CYP tags in L4 larvae and adult worms of MHco3 (ISE) isolate assayed with real-time PCR screen. Three biological replicates. ΔΔCT normalised to *Hc-ama*. Figures on x-axis relate to Hc-cyp-tag numbers. Positive bars indicate higher expression in adults; negative bars indicate higher expression in L4 larvae. A: Hc-cyp-tags. B: Panel of control genes.

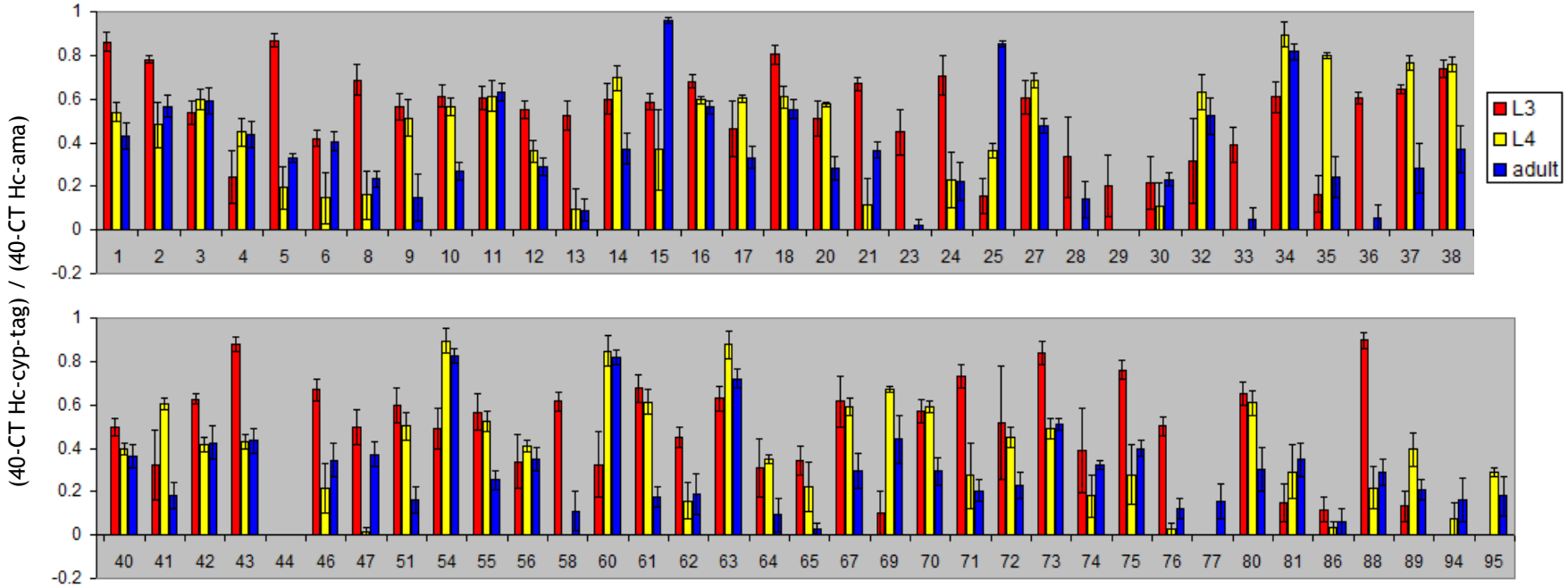
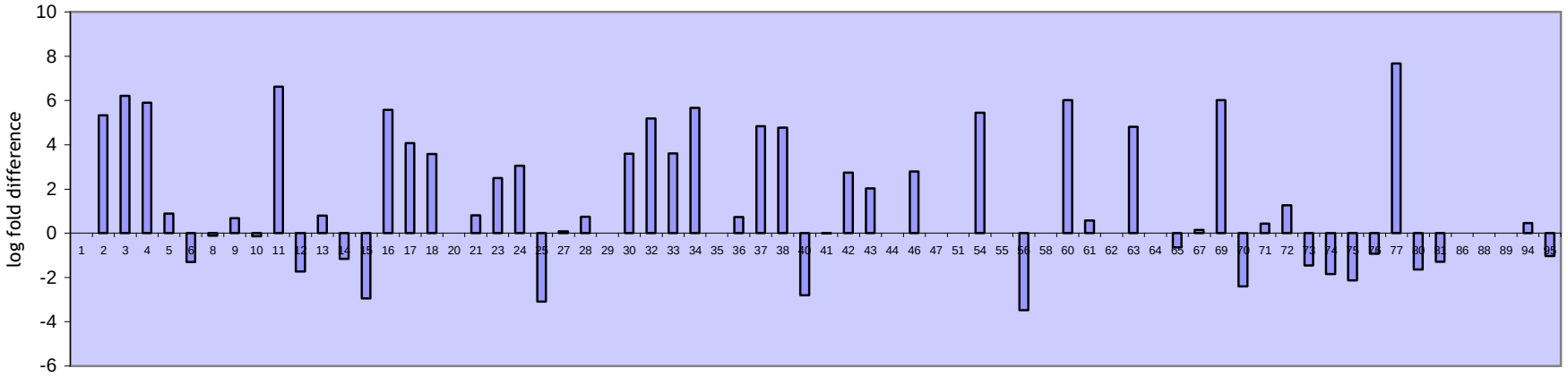


Figure 3.10: Histogram of CYP tag expression as proportion of *Hc-ama* expression in L3, L4 and adult stages assayed with real-time PCR screen

Figures on x-axis relate to Hc-cyp-tag numbers. Threshold values (CT) for CYP tags and *Hc-ama* were subtracted from 40 (maximum threshold value from 40 rounds of RT-QPCR amplification) then CYP tag value was divided by *Hc-ama* value. Adult CTs were pooled from the L3 versus adult real-time PCR screen (Figure 3-8) and the L4 versus adult real-time PCR screen (Figure 3-9).

A:



B:

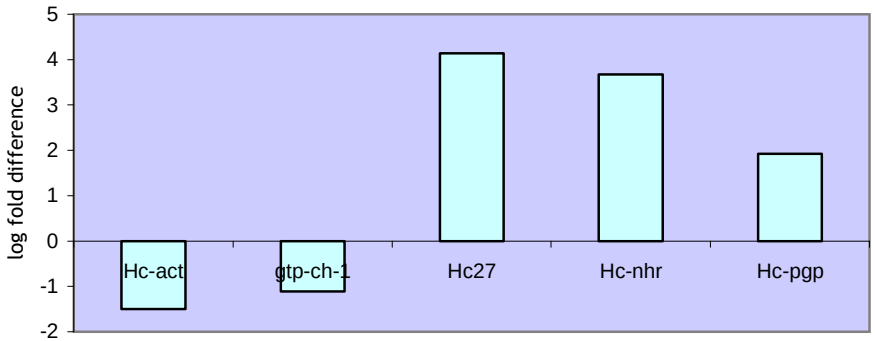
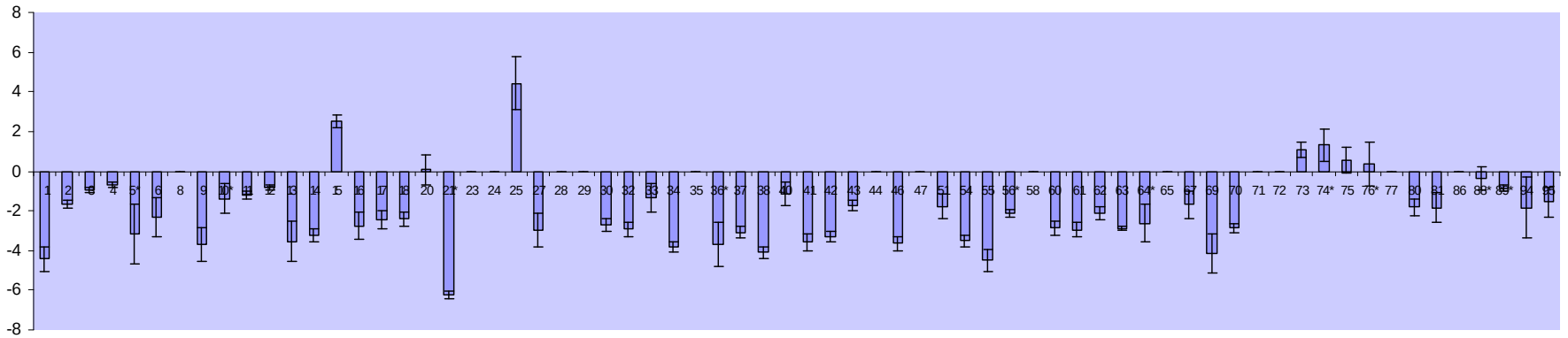


Figure 3.11: Relative expression of CYP tags in soma and intestine of adult worms of the Beltsville isolate assayed with real-time PCR screen.

ΔΔCT normalised to three *Hc-ama* replicates per plate. Figures on x-axis relate to Hc-cyp-tag numbers. Positive bars indicate higher expression in the intestine; negative bars indicate higher expression in the soma. A: Hc-cyp-tags. B: Panel of control genes.

A:



B:

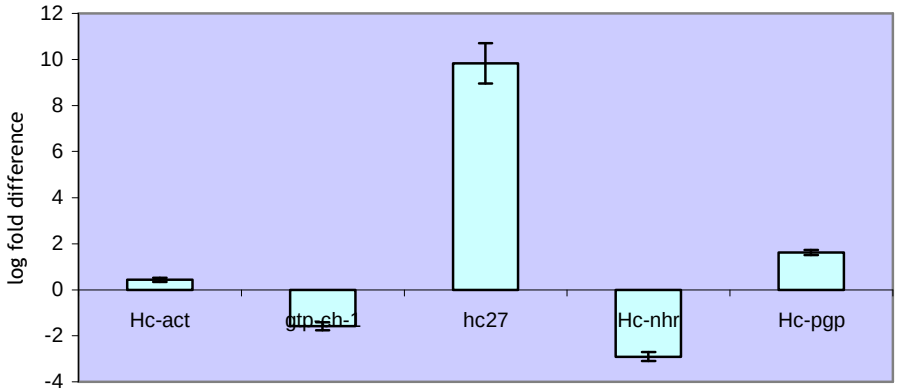
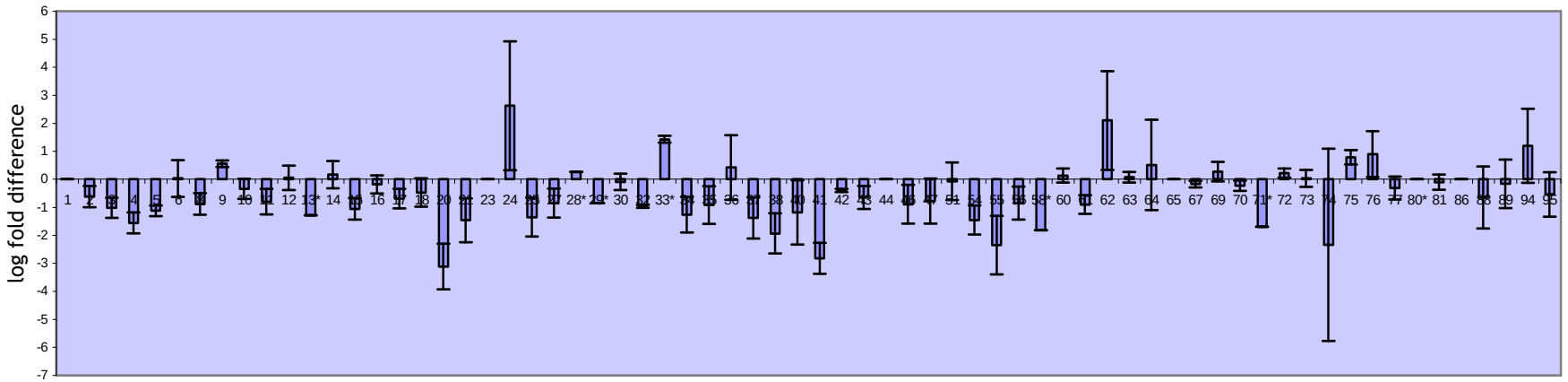


Figure 3.12: Relative expression of CYP tags in male and female adult worms of MHco3 (ISE) isolate assayed with real-time PCR screen.

Three RNA replicates. ΔΔCT normalised to three *Hc-ama* replicates per plate. Figures on x-axis relate to Hc-cyp-tag numbers. Positive bars indicate higher expression in female; negative bars indicate higher expression in male. A: Hc-cyp-tags. B: Panel of control genes.

A:



B:

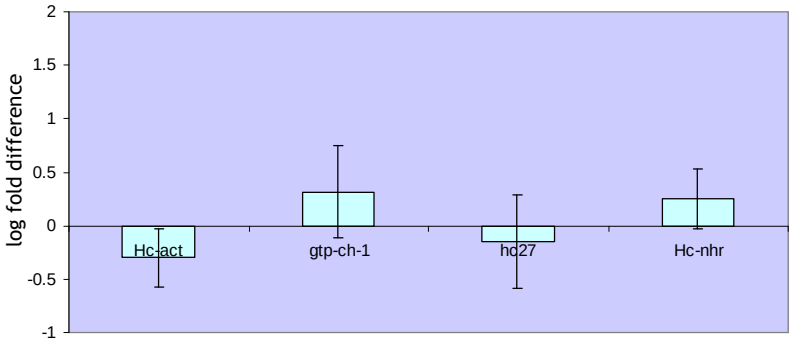
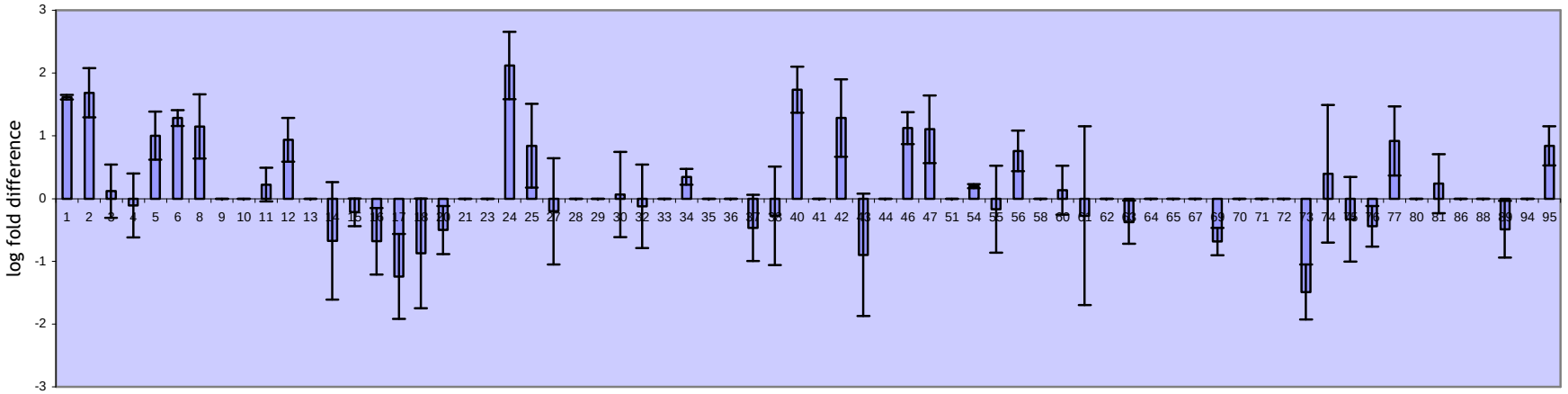


Figure 3.13: Relative expression of CYP tags in IVM-exposed and DMSO-exposed adult worms of MHco3 (ISE) isolate assayed with real-time PCR screen. Three biological replicates. ΔΔCT normalised to three *Hc-ama* replicates per plate. Figures on x-axis relate to Hc-cyp-tag numbers. Positive bars indicate higher expression in IVM-exposed worms; negative bars indicate lower expression in IVM-exposed worms. A: Hc-cyp-tags. B: Panel of control genes.

A:



B:

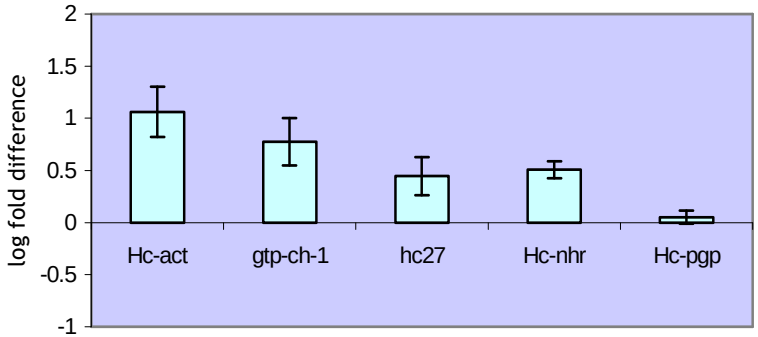
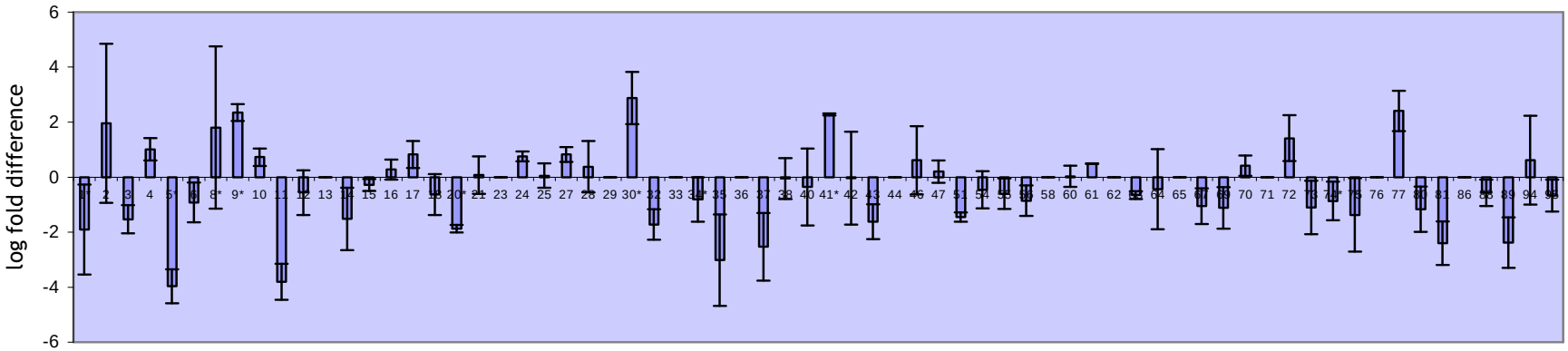


Figure 3.14: Relative expression of CYP tags in ABZ-exposed and non-ABZ-exposed adult worms of MHco3 isolate assayed with real-time PCR screen. Three biological replicates. ΔΔCT normalised to three *Hc-ama* replicates per plate. Figures on x-axis relate to Hc-cyp-tag numbers. Positive bars indicate higher expression in ABZ-exposed worms; negative bars indicate lower expression in ABZ-exposed worms. A: Hc-cyp-tags. B: Panel of control genes.

A:



B:

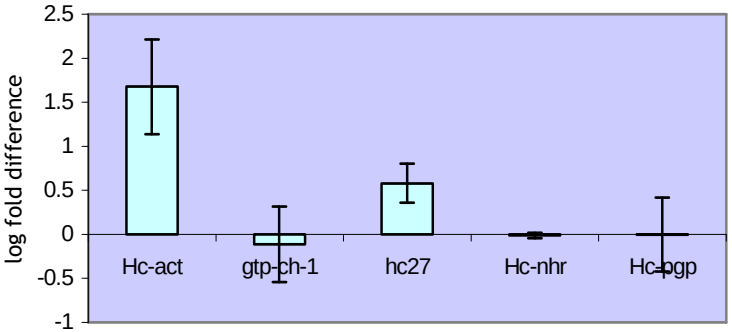


Figure 3.15: Relative expression of CYP tags in adult worms of MHco3 (ISE) and MHco4 (WRS) isolates assayed with real-time PCR screen. Three biological replicates. ΔΔCT normalised to three *Hc-ama* replicates per plate. Figures on x-axis relate to Hc-cyp-tag numbers. Positive bars indicate higher expression in MHco4 (WRS) isolate; negative bars indicate higher expression in MHco3 (ISE) isolate. A: Hc-cyp-tags. B: Panel of control genes.

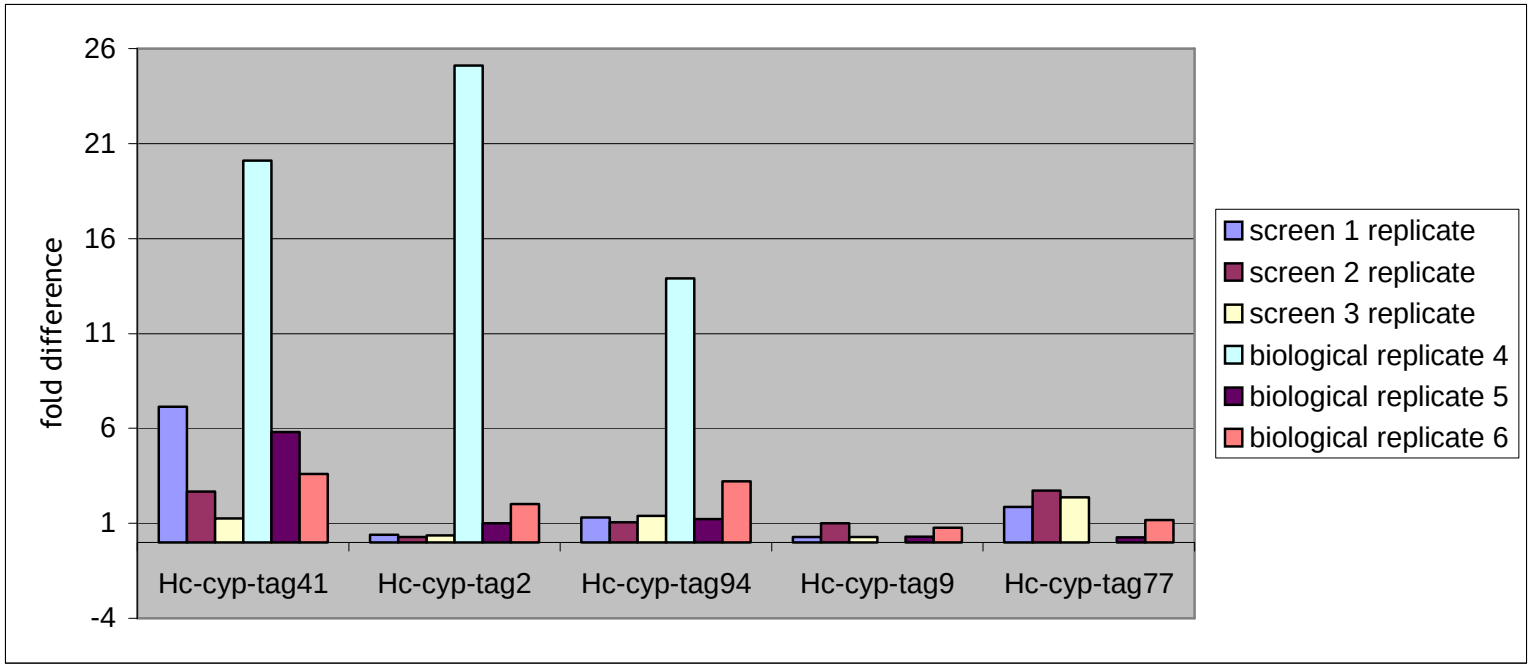
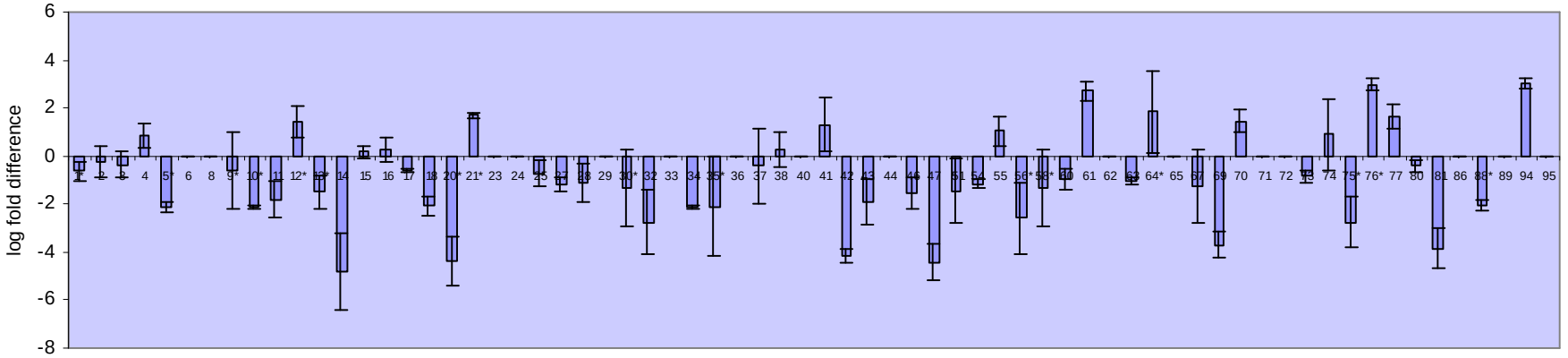


Figure 3.16: Individual RT-QPCR of CYP tag expression in adult MHco4 (WRS) worms relative to adult MHco3 (ISE) worms.
RNA replicates of three biological replicates used for each real-time screen plus three new biological replicates with worm material from different donor sheep. A value of one indicates no difference in expression between isolates; a value of greater than one indicates higher expression in the MHco4 (WRS) isolate. Scale is fold difference.
Worm material used for biological replicate 5 for Hc-cyp-tag9 and Hc-cyp-tag77 was from different donor sheep than that used for biological replicate 5 for other tags. Hc-cyp-tag9 and Hc-cyp-tag77 were not measured in biological replicate 4.

A:



B:

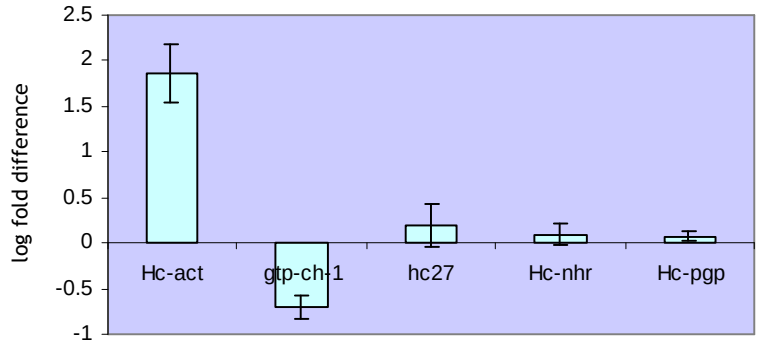


Figure 3.17: Relative expression of CYP tags in adult worms of MHco3 (ISE) and MHco10(CAVR) isolates assayed with real-time PCR screen. Three biological replicates. ΔΔCT normalised to three *Hc-ama* replicates per plate. Figures on x-axis relate to Hc-cyp-tag numbers. Positive bars indicate higher expression in MHco10 (CAVR) isolate; negative bars indicate higher expression in MHco3 (ISE) isolate. A: Hc-cyp-tags. B: Panel of control genes.

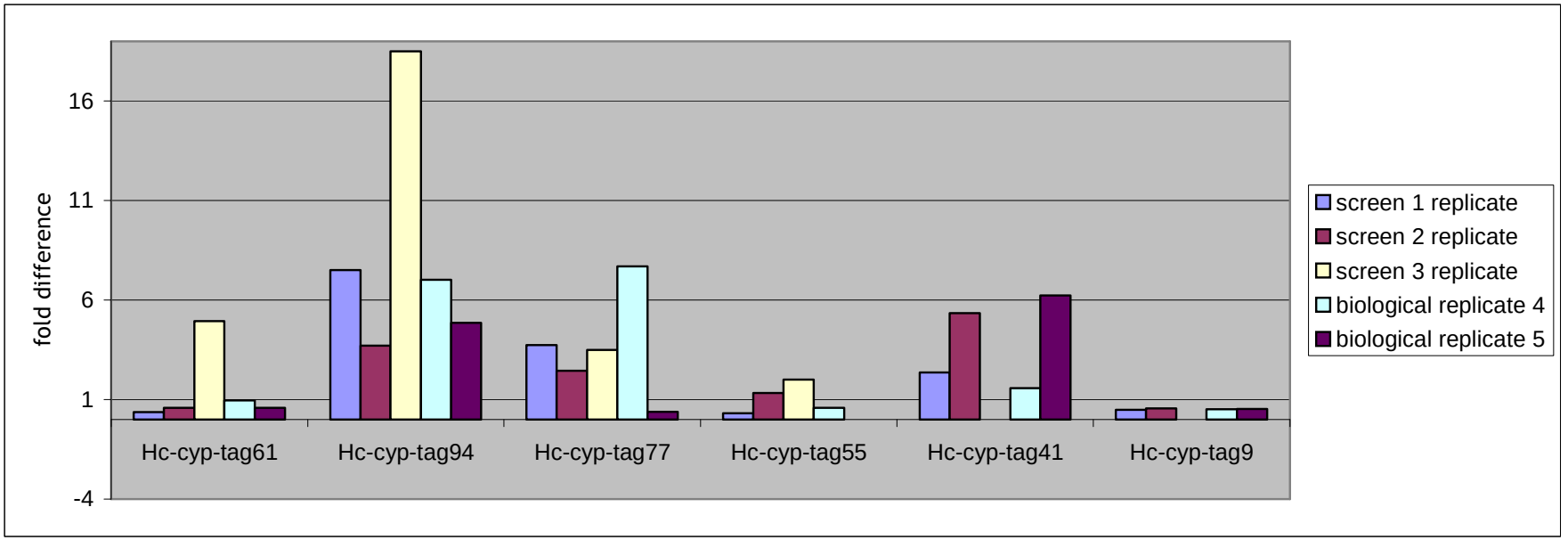


Figure 3.18: Individual RT-QPCR of CYP tag expression in adult MHco10 (CAVR) worms relative to adult MHco3 (ISE) worms
RNA replicates of three biological replicates used for each real-time screen plus three new biological replicates with worm material from different donor sheep. A value of one indicates no difference in expression between isolates; a value of greater than one indicates higher expression in the MHco10 (CAVR) isolate. Scale is fold difference.
A screen 3 replicate was not undertaken for Hc-cyp-tag41 and Hc-cyp-tag9.

3.4 Discussion

The low constitutive expression level of CYPs in adult *H. contortus* worms has, in the past, lead workers to question their existence (Barrett 1998, Pemberton and Barrett 1989, Precious and Barrett 1989a,Precious and Barrett 1989b). Despite compelling evidence of CYP-catalysed metabolism in *H. contortus* larvae (Kotze 1997) and the identification of a large family of CYPs in *C. elegans* (Gotoh 1998) the historic consensus that parasitic nematodes lack CYPs continues to prevail in today's literature (Barrett 2009, Yadav *et al* 2010).

3.4.1 Measuring expression of the CYP family

Due to the ongoing difficulties in assembling the *H. contortus* genome, a 'tag' approach for measuring CYP expression was developed. This allowed experimental work to be undertaken based on partial gene sequences only, with the resulting information directly applicable to the family of full length CYPs.

The results of a real-time PCR screen of CYP tag expression indicated that the constitutive expression of CYPs differ between life stages, tissues and sexes. CYP activity appeared to be highest in larval stages, with both L3 and L4 samples showing a higher constitutive expression than adults for the vast majority of CYP tags. The higher expression of CYPs in L3s than adults was confirmed with RNA-seq transcriptome data. This finding was consistent with work by Kotze (1997) where mono-oxygenase activity in microsomal preparations from three *H. contortus* life stages was measured, identifying high activity in L1 and L3 larvae, but minimal activity in adults.

Similarly, in *C. elegans*, CYP mRNA expression is up-regulated in the dauer stage, an arrested L3 form (Burnell *et al* 2005). This life stage is non-feeding and has a sealed intestine and protective cuticle, so its metabolic requirements may be comparable to those of the *H. contortus* L3. The dauer is thought to require increased CYP activity to metabolise toxic endogenous metabolites and an up-regulation of lipophilic hormones for the maintenance of and recovery from the dauer state. Free-living stages are also likely to be exposed to a wide range of

environmental toxins, so *H. contortus* L3 may require a higher CYP activity than adults to detoxify exogenous compounds.

For adult stages, the relevance of comparisons between *C. elegans* aerobic metabolism and parasite metabolism in the oxygen-limited environment of the host intestine may be more limited. *C. elegans* is capable of anaerobic respiration, although the efficiency of metabolism is depressed to 3-4% of aerobic levels, measured by ATP production (Foll *et al* 1999). Kotze (1997) conjectured that the decrease in CYP activity from L3 to adult could parallel a transition from aerobic metabolism outside the host to anaerobic metabolism in the host abomasum.

Interestingly, the results of the real-time screen indicated a high CYP expression in *H. contortus* L4, which like the adult worms, reside in the abomasum. This would suggest the abomasal environment does not necessarily prohibit CYP activity, as does the higher adult expression of a small subset of CYPs and the adult expression of NADPH-cytochrome reductase. One hypothesis would be that the parasite could derive an adequate supply of molecular oxygen from the host blood supply to facilitate CYP activity. An extreme example of this would be the high levels of CYP activity detected in microsomal preparations from adults of the blood-fluke *Schistosoma mansoni* (Saeed *et al* 2002). Adults of the hookworm *Nippostrongylus brasiliensis* reside in the small intestine, yet also maintain a functional aerobic respiratory chain, relying on oxygen as the terminal electron acceptor (Fry and Jenkins 1983, Fry and Jenkins 1984). This is facilitated by their attachment to the gut mucosa, where oxygen tensions are higher than in the lumen. *H. contortus* adults, which attach to the abomasal mucosa, were also shown to be capable of both aerobic and anaerobic respiration (Fry and Jenkins 1984).

Consistent with this, it was hypothesised that CYPs might catalyse both aerobic and anaerobic pathways of metabolism in *H. contortus*. In 1999, Kotze showed that *H. contortus* L3 were capable of peroxide-supported CYP activity independent of a supply of molecular oxygen *in vitro*. However, the results for adults were inconclusive: *in vitro* oxidase activity was demonstrated in microsomal preparations from adult worms, but was inhibited by both CYP and peroxidase inhibitors.

Highest expression for the majority of CYPs was detected in the parasite intestine. This would be consistent with a possible role in xenobiotic metabolism, as the nematode intestine is considered to be the prime site of detoxification (An and Blackwell 2003, McGhee 2007). In *C. elegans* a number of CYPs are known to be expressed in the intestine including *cyp13a7, cyp14a3, cyp33c2, cyp33e2* and *cyp35a2*, all of which have proven to be inducible on exposure to xenobiotics (Chakrapani *et al* 2008, Menzel *et al* 2001). In *D. melanogaster*, most CYPs are expressed in the midgut, Malpighian (renal) tubules and fat body, which constitute the main organs of detoxification in insects, with a smaller number expressed in the head, gonads and hindgut (Chung *et al* 2009). Expression of the insecticide resistance gene *cyp6g1* occurs in the organs of detoxification listed (Chung *et al* 2007).

Interestingly, a large number of CYP tags appeared to be more highly expressed in adult males than females. Although the smaller size of the males may have resulted in a greater proportion of intestinal material included in the male sample, the intestinal-expressed control gene *hc27* could not clarify this, as it is also more highly expressed in females (Hartman *et al* 2001). However, it is likely that a number of CYP tags would have male-specific roles. In *D. melanogaster*, CYPs have been associated with male behavioural phenotypes, with expression of *cyp6a20* linked to male aggressive behaviour in an inducible and reversible manner (Wang *et al* 2008) and *cyp4d21* facilitating mating in adult male flies (Fujii *et al* 2008). In *C. elegans, cyp22a1 (daf-9)* expression is thought to inhibit dauer formation and promote reproductive development (gonadal outgrowth) controlled via a positive feedback loop from the downstream NR, *daf-12* (Gerisch and Antebi 2004). The *daf-12* pathway is also thought to regulate mate searching behaviour in fully developed adult males (Kleemann *et al* 2008). In the real-time screen, putative *H. contortus* homologues of both these genes (*Hc-cyp-tag14* and *Hc-nhr*) showed highest expression in the larval stages and higher male than female expression in the adult.

Characteristic expression profiles were used to group and sequence between CYP tags representing the same genes (see Chapter 4). This facilitated the identification of putative orthologues of *C. elegans* CYPs with stage, tissue and sex specific expression. Hc-cyp-tag15 and Hc-cyp-tag25 were most highly expressed in the adult female body and represent a single gene homologue of

the closely related *C. elegans pod-7 (cyp31a2)* and *pod-8 (cyp31a3)* genes, which are expressed in gonads, oocytes and embryos and are involved in the production of eggshell lipids (Benenati *et al* 2009).

A small number of CYPs were more highly expressed in anthelmintic-resistant adult worms than in the susceptible isolate. Although the magnitude of difference was variable, in the context of such high variability in the expression of the other CYPs, their relative consistency is perhaps more noteworthy. Further, much of the variability could be attributed to the low level of expression in the susceptible isolate.

Hc-cyp-tag41 showed the most consistent higher expression in the ivermectin and benzimidazole-resistant MHco4 (WRS) isolate relative to the susceptible MHco3 (ISE) isolate. However, in one biological replicate, a higher expression level was not recorded, suggesting this finding was not universal. Hc-cyp-tag9, Hc-cyp-tag55 and Hc-cyp-tag70, which amplify the same gene as Hc-cyp-tag41, did not show a consistently higher expression. This may be due to polymorphism reducing the efficiency of primer binding for these tags, as primers were designed for the MHco3 (ISE) isolate, or alternatively the Hc-cyp-tag41 primers may also amplify a conserved region in a different gene with higher expression in the resistant isolate.

Hc-cyp-tag94 was more highly expressed in the ivermectin-resistant MHco10 (CAVR) isolate than the susceptible MHco3 (ISE) isolate in all five biological replicates. Interestingly, Hc-cyp-tag94 was not more highly expressed in the MHco4 (WRS) isolate, which would suggest the mechanisms of resistance vary between isolates, if up-regulation of this gene was involved in MHco10 (CAVR) resistance. Anthelmintic-resistant isolates have recently been generated from backcrosses of MHco3 (ISE) and MHco4 (WRS) worms and backcrosses of MHco3 (ISE) and MHco10 (CAVR) worms. These may provide valuable information regarding the mechanisms of resistance in the parent isolates, as it is hoped that only the resistant genes will differ between the backcross isolates and MHco3 (ISE).

Similarly to Hc-cyp-tag41 in MHco4 (WRS), Hc-cyp-tag77 showed higher expression in MHco10 (CAVR) in four biological replicates but no difference in

expression in biological replicate five. Although this underlined the importance of performing as many biological replicates as possible, a constant consideration was the requirement of donor sheep to generate the parasite material. In addition, the quantity of starting RNA required to measure a large number of genes expressed at low levels was not insignificant.

The importance of such a low level of adult CYP expression for xenobiotic metabolism and resistance has been questioned. Metabolism of CYP substrates ethoxycoumarin and aldrin in adult *H. contortus* microsomes was 500-fold to nearly 2000-fold lower than in mammalian liver microsomes for the former and 10 000-fold lower than in rodent liver microsomes for the latter (Kotze, 1997). However, as the real-time screen shows, adult worms do express a small number of CYPs relatively highly, which would be missed in a pooled measurement of CYP activity. Further, the choice of substrates used to measure overall activity could give significantly different results depending on the particular CYPs they target. In *D. melanogaster*, 10 to 100-fold higher expression of a single gene cyp6g1 is known to confer multi-drug resistance (Daborn *et al* 2002) and it would seem possible that a single highly expressed CYP could also confer resistance in adult nematodes in the context of a family of collectively low expression.

A number of CYPs showed no expression or significantly lower expression in the resistant isolates in all biological replicates. This was predicted to be a result of polymorphism at primer binding sites, as the screen was designed from and for MHco3 (ISE) sequence only. Since no reference sequence was available for the resistant isolates, no attempts were made to design primers for the divergent sequence, although it could be argued that these may be relevant genes for further research, as they could be xenobiotic responsive CYPs under selection pressure. An alternative interpretation would be that a number of these CYPs were truly down-regulated in the resistant isolates, which could benefit the parasite by decreasing the activation of a pro-drug or production of a toxic metabolite. For example, chlorpyrifos resistance in *D. melanogaster* has been shown to increase rather than decrease with administration of piperonyl butoxide (Ringo *et al* 1995). Interestingly, a comparison of oxidase activities between macrocyclic-susceptible (McMaster) and –resistant (CAVR) isolates of *H. contortus* found similar levels of 7-ethoxycoumarin *O*-deethylase and aldrin epoxidase activities in the L3 stages but lower aldrin epoxidase activities in the

resistant isolate in the adult stage (Kotze 2000). This was hypothesised to be a result of lower activity of soluble peroxidase enzymes in the CAVR adults rather than lower CYP activity, as the difference was only detected in whole worm samples (microsomal and soluble fractions) and not in exclusively microsomal samples.

3.4.2 Limitations of the real-time screen

One of the aims of this project was to design a large throughput screen to detect CYPs showing differential expression, which could then be confirmed by more sensitive means.

Linear regression analysis had suggested the overall reproducibility of the screen was good, but an inverse relationship between gene expression level and degree of variation measured between screen replicates was apparent. This was concerning for the detection of putative xenobiotic metabolising CYPs which would potentially have a low constitutive expression that would only be induced in response to the substrate or up-regulated in a resistant isolate. Consistent with this, Hc-cyp-tag41 which was more highly expressed in the anthelmintic-resistant MHco4 (WRS) isolate in five of six biological replicates, showed a low constitutive expression in the anthelmintic-susceptible MHco3 (ISE) isolate. Further, cDNA replicates clarified that much of the variation in magnitude of relative expression level was generated by fluctuations in measured CT values for Hc-cyp-tag41 in the MHco3 (ISE) isolate only. This may suggest a comparison of absolute expression levels for each isolate, rather than a relative expression level between isolates, would be valuable for genes with low expression.

Biological variation in gene expression between individuals can complicate gene expression studies. Studies into the expression of the CYP family may be especially prone to individual variation as enzymes are expressed differentially with factors such as life stage, sex and food intake. Adult worms were staged at 21 days, but were not sexed before being split into batches for biological replicates of drug exposure experiments, or before snap freezing for constitutive expression comparisons between isolates. In light of the large difference in CYP expression between sexes, it would be recommended to do this in future.

The drug-exposure experiments showed an especially high degree of variation. Although it is possible that the xenobiotic-treatment elicited a different response in individual worms, experimental error was deemed more likely. The iterative nature of this project meant that over time a number of CYP tags were identified as amplifying the same gene, but these were maintained in the real-time screen as internal controls. A number of these associated tags showed contradictory differences in expression in the drug-exposure experiments and in the case of the ABZ-exposure experiment, the expression of the control genes *Hc-act, gtp-ch-1* and *hc27* appeared to be higher in the drug-exposed samples. It is possible that this was due to anthelmintic residues affecting the PCR reactions, although all worm samples were rinsed after exposure, no abnormalities were noted in the RNA samples (analysed with light spectrophotometry and run on an Ethidium bromide agarose gel), and cDNA template was purified before use in the screen. Interestingly, the control gene *Hc-act*, also showed higher constitutive expression in both resistant isolates. This is unlikely to be an artefact of normalisation as the other control genes showed little difference in expression. In addition, the consistency of *Hc-act* expression in the real-time PCR screen comparisons of life stages, tissues (higher in soma as expected) and sexes suggest it is a reliable control gene for the MHco3 (ISE) isolate. It is possible that expression of this gene is higher in the resistant isolates, but it is perhaps more likely that the primers amplify more than one member of the actin family in MHco4 (WRS) and MHco10 (CAVR).

22 CYP tags identified in the genome were not amenable to primer design, so were not included in the screen. Seven tags could not be amplified from adult or L3 cDNA and may represent pseudogenes or CYPs with particularly low constitutive expression. It would be interesting to investigate these genes further and attempt to determine their expression patterns as it is possible this subset includes candidate xenobiotic metabolising genes, which have so far been missed.

3.4.3 Conclusion

The *H. contortus* genome encodes a large number of cytochrome P450 genes. CYP activity appears higher in the larval stages than in the adult, which may

reflect a role in metabolising endogenous toxins and compounds essential for development or the greater exposure of free-living stages to environmental toxins. The importance of CYPs in xenobiotic detoxification in the adult worm is unclear. However, the high expression of a small subset of CYP genes as well as NADPH-cytochrome P450 reductase, suggests that CYP-catalysed metabolism is important in the adult. Further, the large artillery of potentially inducible CYPs encoded in the genome would make a role in xenobiotic detoxification distinctly possible.

4 Assembly of the CYP family and characterisation of genes of interest

4.1 Introduction

Cytochrome P450s (CYPs) have been identified in virtually all living organisms including animals, plants, bacteria and fungi (Nelson *et al* 1996). The family size varies dramatically between species. At one extreme are the yeast genomes, encoding only two (*Schizosaccharomyces pombe)* or three (*Saccharomyces cerivisiae)* CYPs, while at the other is the rice genome, encoding 323 CYPs (Nelson *et al* 2004).

In insects, the honeybee has a dramatically reduced CYP family of only 46 genes (Claudianos *et al* 2006), relative to 90 in *D. melanogaster* (Tijet *et al* 2001) and 106 in *A. gambiae* (Feyereisen 2006). This corresponds to a similar decrease in the number of GSTs and carboxyl/cholinesterases encoded in the honeybee genome and is thought to play a role in its sensitivity to pesticides. A small subset of CYP family members have however undergone recent expansion in the honeybee and these may be involved in biosynthesis of pheromones to facilitate the high level of social organisation displayed by bees (Claudianos *et al* 2006).

The number of CYPs in nematode species also varies widely. In clade V, the genomes of closely related nematodes *C. elegans* and *C. briggsae* encode 80 and 73 CYPs respectively, but *P. pacificus* has 198 CYPs (Dieterich *et al* 2008, Gotoh 1998, Stein *et al* 2003). The CYP family expansion in *P. pacificus* is paralleled by an expansion in other families of detoxification genes such as the GSTs, SULTs and ABC-transporters (Dieterich *et al* 2008) and is thought to act as a defence against microbes and toxins inherent in the *P. pacificus* necromenic lifestyle.

These findings suggest the CYP family is evolving rapidly. Frequent gene duplications and losses have resulted in clusters of related CYP genes with a high number of pseudogenes and detritus exons arranged in tandem (Baldwin *et al* 2009, Nelson *et al* 2004, Thomas 2006, Thomas 2007, Tijet *et al* 2001). In the human and mouse CYP families, the number of CYP pseudogenes nearly matches the number of functional CYPs and as pseudogenes are not conserved by natural

selection, their presence suggests they are being generated continuously (Nelson *et al* 2004).

The mammalian CYP family is divided into families and subfamilies. Family members share more than 40% amino acid identity and subfamily members share more than 55% amino acid identity (Nelson *et al* 1993). Gotoh (1998) classified the *C. elegans* CYPs into three clades, A, B and C (see figure 4-1), which correspond closely to the mammalian CYP2, CYP3 and CYP4 families respectively. In many cases CYPs within families and subfamilies share significantly higher identity than the 40% and 55% amino acid identity cut-offs, but the divergence between CYPs of different families can be high (Gotoh 1998).

H. contortus was historically thought to lack CYPs (Barrett 1998, Pemberton and Barrett 1989, Precious and Barrett 1989a,Precious and Barrett 1989b) but the sequencing of the parasite's genome facilitated identification of 61 contigs encoding 97 partial CYP sequences (see chapter 3). The aim of the work presented in this chapter was to assemble the partial sequences into full length genes where possible and to group *H. contortus* CYPs into putative families for comparison with *C. elegans*. CYP genes shown by the real-time PCR screen to have characteristic expression patterns in different life stages, tissues and sexes or higher expression in anthelmintic-resistant isolates were investigated in more detail.

4.2 Nomenclature for *H. contortus* CYP genes

As described in Chapter 3, a nomenclature *for H. contortus* genes based on their putative orthologues in *C. elegans* was unsuitable for this rapidly evolving family and for the partial gene sequences, the standard CYP nomenclature system cannot be applied.

Therefore, to retain neutrality until the genome is complete, each full length CYP sequence was given the identifier 'Hc' for *H. contortus*, 'cyp' for cytochrome P450 and a number e.g. *Hc-cyp-1*. Incomplete CYP sequences retained their 'tag' name, but where multiple CYP tags were assembled on the same contig, the addition of "+" between tags thought to represent the same gene and "_" between tags representing different genes or where the

association could not be determined, was included in the nomenclature e.g. Hc-cyp-tag1_5.

4.3 Approaches to assemble full length CYP genes

4.3.1 Molecular biology approach

Firstly, attempts were made to amplify cDNA products by PCR between CYP tags located on the same supercontigs provided they were on the same strand and their assembly would result in the expected linear order of amino acids. This process was then extended to tags shown to have similar expression patterns in the real-time screen. For example, unlike the majority of CYP tags, Hc-cyp-tag15 and Hc-cyp-tag25 most were most highly expressed in adult, female and body samples, so PCR amplification and sequencing was used to confirm that they represented the same gene (see Figure 4-2A).

CYP tags with a high constitutive expression in MHco3 (ISE) adult worms were most amenable to this process, due to the availability of template and the relative ease of amplification. However, hemi-nested and nested PCR amplification between Hc-cyp-tag41, Hc-cyp-tag9, Hc-cyp-tag55 and Hc-cyp-tag70 facilitated sequencing of the single gene they represent (see Figure 4-2B), which was of interest due to the higher expression of Hc-cyp-tag41 identified in the MHco4 (WRS) isolate relative to MHco3 (ISE) isolate.

Analysis of expression profiles was a useful method of associating tags, but they were only considered to represent the same gene once a product had been PCR amplified between them, due to the risk that close family members could share similar expression patterns.

When RNA-seq data was generated and transcriptome reads were mapped to the reference sequence in the supercontig database, it revealed coding sequence that BLAST homology alone could not detect. This was especially useful for the most 5’ exons of *H. contortus* CYP genes, which had previously proven difficult to identify as the 5’ ends of CYPs encode the membrane-binding regions of the proteins and appear to be most divergent from those in *C. elegans*. Further annotation of the CYP tags guided by mapped MHco3 (ISE) adult and MHco3 (ISE)

L3 reads facilitated the assembly of a small number of tags and confirmed that a larger number of tags represented different genes by revealing divergent coding sequence.

Using these methods, the original 97 CYP tags were assembled into 73 larger CYP tags (see Figure 4-3).

4.3.2 Bioinformatic approach

When Gotoh (1998) assembled the *C. elegans* CYP family, he suggested close family members should differ by at least 1% at the amino acid level. For this distinction, at least 95% of the protein sequence was needed, which was lacking for many of the *H. contortus* CYP sequences identified.

16 CYP tags (eight pairs) encoded sequence with 100% amino acid identity to another tag encoded on a different contig, but alignment of the genomic sequence between pairs demonstrated significantly lower identity. For example, Hc-cyp-tag41 and Hc-cyp-tag61 shared 100% amino acid identity and 98% exonic nucleotide identity, but only 72% genomic nucleotide identity. For the purposes of the real-time PCR screen, these tags were considered to represent different genes to avoid incorrectly classifying close family members as single genes, but with the understanding there may be a degree of redundancy with some CYP tags. Although this conservative approach was appropriate for the screen, a more realistic model for assembling the family would include options for polymorphism and allelic variation.

The supercontig database was generated from genomic DNA isolated from multiple adult worms so allelic variation could generate variant CYP sequence. Differentiating this from the (often minimal) sequence divergence of two closely related but different CYP genes proved challenging. For example, in *C. elegans*, cyp25a4 and cyp25a5 are closely related yet distinct genes residing on chromosome III and chromosome IV respectively and their 2165 bp genomic sequences share 98% nucleotide identity.

Since close family member genes could potentially be more similar than allelic variants of the same gene, forcing assembly of consensus sequences was

predicted to be counter productive and more meaningful information could derived from the remaining tags in their unassembled state.

4.4 Classification into putative families

Despite falling short of the 95% sequence required to distinguish subfamily members (Gotoh 1998), it was possible to tentatively group the unassembled parasite CYP tags into families, based on those in *C. elegans*.

Table 4-1 shows the 73 CYP tags grouped into putative families, based on the closest matched CYP genes in *C. elegans*. Although there was a risk that the shorter tags would change groups when their full sequence was determined, the greater divergence between family groups ensured that longer tags could be more confidently classified. The 15 CYP tags encoding more than 75% of a polypeptide were aligned to the closest *C. elegans* polypeptide to determine the % identity. CYPs in the same family should share >40% amino acid identity and subfamily members should share >55% (Nelson *et al* 1993).

The greatest number of tags were grouped into the CYP33 family, which is the largest CYP family in *C. elegans* with 18 members. The CYP13, CYP14 and CYP34 families are also large in the model worm and the high number of CYP tags assigned to these groups in the parasite may reflect this. However, only one CYP tag was placed in the CYP35 family, which has ten members in *C. elegans*. Similarly, no tags were assigned to the CYP25 family, which has six members in *C. elegans*. Interestingly, all of the six CYP tags that grouped in the CYP37 family, which has two members in the model worm, encoded only the C termini of polypeptides, but in all other groups the entire lengths of polypeptides were more evenly represented.

4.5 Gene conservation and divergence

Figure 4-4 shows a neighbour-joining tree of conceptual translations of CYP tags aligned with the family of CYP polypeptides in *C. elegans*. Phylogenetic relationships could not be inferred between tags representing different regions of genes, so a number of short tags had to be excluded from the analysis.

In general, the positioning of the CYP tags within the phylogenetic tree, agreed with the putative family groupings in Table 4-1. However, a number of CYP tags which had been grouped in the CYP34 family, clustered equally with the CYP34 and CYP35 families in the tree, which may explain the lack of CYP35 homologues identified with BLAST searching alone. In general, bootstrap values were high, especially so for tags clustering with single-member CYPs in *C. elegans,* suggesting these may be more highly conserved between the species.

In a survey of ten vertebrates, (Thomas 2007) found that all CYP genes could be classified as either phylogenetically stable or phylogenetically unstable, and the divide was about even for each species studied. He found that CYPs with known endogenous functions were more likely to be stable, with no or few gene duplications or losses, while CYPs with a role in xenobiotic metabolism were most commonly found in unstable gene clusters, proposed to have arisen by local gene duplication.

4.5.1 Single family member genes

In *C. elegans cyp22a1, cyp23a1, cyp36a1, cyp42a1, cyp43a1* and *cyp44a1* are genes in single member families, so these were hypothesised to represent phylogenetically stable CYPs. As seen in Table 4-1 and Figure 4-4, these may also be stable in *H. contortus* as in general, fewer tags were grouped into these families, most represented contiguous sequence rather than numerous copies of the same region of a gene, and bootstrap values were 100 for all.

C. elegans cyp22a1 (daf-9) is a single-family member cytochrome P450 involved in promoting reproductive development and regulating the dauer pathway. Hc-cyp-tag22+14 was the only homologue identified.

C. elegans cyp23a1 is a single-family member CYP of unknown function sharing most identity to human CYP7B1, mutations of which lead to giant cell hepatitis. Both Hc-cyp-tag65+64 and Hc-cyp-tag82 encoded N-termini with high identity to CYP23A1. Although Hc-cyp-tag65+64+10 and Hc-cyp-tag82 shared 100% amino acid identity and 93% exonic nucleotide identity, the available intronic sequences only shared 74-85% identity due to a number of SNPs and two deletions (of 13 bp and 29 bp) in the Hc-cyp-tag82 genomic sequence.

Four CYP tags shared most identity with *cyp36a1,* a gene of unknown function in *C. elegans*. Hc-cyp-tag26 represented a single exon sharing 100% predicted amino acid identity with Hc-cyp-tag29 sequence, 92% coding nucleotide identity but less than 40% genomic nucleotide identity. *In C. elegans*, there are three known alleles of *cyp36a1,* due to transposon insertions in the corresponding region of the gene represented by Hc-cyp-tag26 and Hc-cyp-tag29, but there was no transposon-like or repetitive sequence on either contig in *H. contortus*. Hc-cyp-tag59+58 and Hc-cyp-tag88 both represented 3' ends of a homologous gene to *cyp36a1*, and shared 98% amino acid identity, 96% coding nucleotide identity and 94% genomic nucleotide identity with each other. Their intron: exon boundaries were conserved.

Hc-cyp-tag41+55+70+9 represented a *C. elegans cyp42a1* homologue (see Section 4.7.2.1). An additional contig, Hc-cyp-tag61, encoded an identical 122 amino acid N-terminus to Hc-cyp-tag41. The coding sequence shared 98% nucleotide identity with Hc-cyp-tag41, but the genomic sequence was only 72% identical.

Two *H. contortus* contigs, Hc-cyp-tag4+77 and Hc-cyp-tag11+3, encoded sequence with most identity to *C. elegans* single-family member *cyp43a1*. The conceptual translations of the parasite CYP tags differed by 13 amino acid substitutions. The gene is of unknown function in *C. elegans*.

cyp44a1 is the only mitochondrial CYP in *C. elegans*. Hc-cyp-tag31, Hc-cyp-tag75 and Hc-cyp-tag76 encoded 83, 92 and 92 amino acids of the C-terminus respectively. The conceptual translations of Hc-cyp-tag31 and Hc-cyp-tag76 shared 100% amino acid identity and Hc-cyp-tag75 differed by one amino acid. They all shared 92-98% coding nucleotide identity, 94-97% genomic nucleotide identity and conserved intron: exon boundaries. Only one tag, Hc-cyp-tag83, encoded a homologous N-terminus to CYP44A1 (see Section 4.7.1.2).

4.5.2 Clustered genes

Expansion by gene duplication is common in CYP family evolution (Baldwin *et al* 2009, Nelson *et al* 1993, Thomas 2006, Tijet *et al* 2001) and as described by Thomas (2006) may play a role in adaptive evolution in the face of xenobiotic challenge. The extent of CYP family clustering in the *C. elegans* genome is

shown in Figure 4-5. As can be seen, most clusters consist of close family members and probably arose by multiple gene duplications.

A minimum of three genes is required to fit the definition of a cluster. As shown in Figure 4-3, in *H. contortus* the biggest putative cluster was six genes on supercontig 5, consisting of Hc-cyp-tag50, Hc-cyp-tag94, Hc-cyp-tag97, Hc-cyp-tag62, Hc-cyp-tag36 and Hc-cyp-tag35 and there was a putative cluster of three CYPs on supercontig 27: Hc-cyp-tag51, Hc-cyp-tag92 and Hc-cyp-tag93. Two additional supercontigs, 6 and 8, each encoded three consecutive CYP tags, but for both supercontigs, two of the tags could potentially assemble into one gene as they encoded amino acids in the correct linear order. All 12 of the tags on supercontigs 5, 6 and 8 shared most homology with either CYP33 or CYP14 family members, which cluster in large arrays in *C. elegans*. As shown in Figure 4-5, the CYP33 genes are clustered on Chromosome IV and Chromosome V in *C. elegans*, while the CYP14 genes cluster on Chromosome X.

Interestingly, Hc-cyp-tag94, which lies within the biggest putative cluster of CYPs identified in the parasite (supercontig 5), was shown to have a higher constitutive expression in anthelmintic-resistant MHco10 (CAVR) isolate adults relative to the anthelmintic-susceptible MHco3 (ISE) isolate.

4.6 Genes associated with CYP pathways

4.6.1 Redox partners

C. elegans emb-8 is thought to encode the orthologue of NADPH-cytochrome P450 reductase, a protein that supplies electrons to cytochrome P450 enzymes. *H. contortus* supercontig_0047488 encodes a putative *emb-8* orthologue (see Figure 4-6). The parasite gene is predicted to encode a product with 69% amino acid identity to the *C. elegans* protein (BLASTp, E=0), although sequence potentially encoding ~55 amino acids is missing due to a gap in the supercontig assembly.

C. elegans emb-8 is a 3002bp, 6 exon gene, encoding a 1989bp mRNA. The *H. contortus emb-8* homologue is at least 6119bp, with at least 15 exons, based on the partial 1824bp mRNA that can be annotated from available sequence.

In both nematodes, the predicted EMB-8 polypeptide encodes a conserved NADP cytochrome P450 reductase (CYPOR) domain, containing all 40 expected residues for the associated NADP binding pocket, FAD binding pocket, FAD binding motif, phosphate binding motif, beta-alpha-beta structure motif and the four expected catalytic residues.

Based on Illumina transcriptome data, the parasite *emb-8* homologue appears to be highly expressed in both L3 and adult stages.

4.6.2 Nuclear Receptors

The *C. elegans* genome encodes 284 nuclear receptor (NR) genes, of which three have been predicted to reside in the same subfamily as the mammalian PXR and CAR proteins known to regulate the expression of xenobiotic-metabolising CYPs: *daf-12, nhr-8* and *nhr-48* (reviewed in Xu *et al,* 2005).

When the real-time screen was first designed, the closest matched sequence in the *H. contortus* 12/11/07 contig database to all three *C. elegans* genes was encoded on contig_0015966. A reciprocal BLASTx search of wormpep proved it to be most similar to DAF-12 (E=3e-45), as did alignment of the predicted mRNA and polypeptide, and primers were designed to amplify this gene in the real-time screen (*Hc-nhr*).

When the 21/08/08 supercontig database was released the following year, another tBLASTn search with the three *C. elegans* NRs of interest revealed 12 supercontigs encoding sequence with high identity (P<0.005), which facilitated identification of the full-length sequence of *Hc-nhr,* in addition to putative *C. elegans nhr-8* and *nhr-48* homologues.

Results of the real-time screen showed expression of the *daf-12* homologue *Hc-nhr* to be higher in the L3 and L4 stages than in the adult parasite, higher in the gut than in the soma, and higher in males than females. Similarly, in *C. elegans*,

studies have shown DAF12::GFP to be expressed from embryo to adult, but most highly in the larval stages, with peak expression in the L2 stage. Although lower in adults, DAF12::GFP expression was widespread, but strongest in the intestine, spermatheca and uterus.

In *C. elegans*, *daf-12* has been shown to be involved in many pathways and plays key roles in regulating development and lifespan. Analysis of loss-of-function and gain-of-function *daf-12* mutants relative to N2 worms showed an increased sensitivity in the former and increased resistance in the later to oxidative stress, based on survival following exposure to paraquat (Fisher and Lithgow 2006). The same study used microarray analysis of the *daf-12* mutants to identify a relatively small set of differentially expressed genes, which included members of the CYP and GST families, although both appeared to be down-regulated in the gain-of-function mutant.

4.7 CYP genes of interest

4.7.1 CYPs with high constitutive expression in the susceptible isolate

4.7.1.1 Hc-cyp-1 (Hc-cyp-tag15+25)

Hc-cyp-1 was the most highly expressed CYP identified in *H. contortus* adult worms. Its characteristic expression profile was unlike the majority of CYP tags in showing higher expression in the body than the intestine and higher expression in the female than the male (see Chapter 3).

Missing *Hc-cyp-1* sequence from 3708 bp to 3807 bp was amplified and sequenced from adult worm cDNA (see Figure 4-2A). The conceptual translation of *Hc-cyp-1* shared 57% amino acid identity with CYP31A2 (POD-7) and CYP31A3 (POD-8) in *C. elegans* (see Figure 4-7) and 57% amino acid identity with CBR-CYP-31A3 in *C. briggsae*. Hc-cyp-tag15 and Hc-cyp-tag25 were the only tags with high identity to the *C. elegans* CYP31A subfamily identified in the *H. contortus* reference sequence databases, so it is likely that this is a single family member CYP in the parasite.

pod-7 and *pod-8* are expressed in gonads, oocytes and embryos and are involved in the formation of lipids required for eggshell production (Benenati *et al* 2009). The encoded proteins have been shown to function with the NADPH reductase EMB-8, a homologue of which was identified *H. contortus* and described in Section 4.6.1.

4.7.1.2 Hc-cyp-2 (Hc-cyp-tag83+73+74+76)

Hc-cyp-2 was more highly expressed in L3 and L4 larvae than in adults. In the latter, it was most highly expressed in females and in the intestine. The closest *C. elegans* CYP was *cyp44a1*, the only mitochondrial CYP in the model worm. The conceptual translation of *Hc-cyp-2* shared 54% amino acid identity with the *C. elegans* polypeptide (see Figure 4-8) and with CBR-CYP-44A1 in *C. briggsae*. Most divergence was seen at the N terminus, with only the MRRS motif of the first four amino acids conserved.

Two additional tags encoded 83 amino acids and 92 amino acids of a homologous C-terminus. They all shared 92-98% coding nucleotide identity, 94-97% genomic nucleotide identity and conserved intron: exon boundaries, but the product encoded by Hc-cyp-tag31 shared 100% amino acid identity with HC-CYP-2 while the product encoded by Hc-cyp-tag75 differed by one amino acid. These tags may represent alleles or paralogues.

Vertebrate mitochondrial CYPs catalyse steroid, bile and vitamin D synthesis, but no function has been ascribed to *cyp44a1* in *C. elegans* or the eight mitochondrial CYPs in *Drosophila melanogaster* (Tijet *et al* 2001). RNAi of the mitochondrial CYP in *C. elegans* causes no obvious abnormalities (Kamath *et al* 2003).

4.7.2 CYP genes with higher expression in resistant isolates

4.7.2.1 Hc-cyp-3 (Hc-cyp-tag41+55+70+9)

Hc-cyp-tag41, Hc-cyp-tag55, Hc-cyp-tag70 and Hc-cyp-tag9 represent contiguous regions of the same gene, which is homologous with *cyp42a1* in *C. elegans* (BLASTx, E=0). Hc-cyp-tag41 showed the most consistent higher expression in the ivermectin and benzimidazole-resistant MHco4 (WRS) isolate relative to the susceptible MHco3 (ISE) isolate. Hc-cyp-tag55, Hc-cyp-tag70 and Hc-cyp-tag9 did not show a consistently higher expression, but this may reflect between-isolate polymorphism.

Missing cDNA sequence between the predicted coding sequences the on Hc-cyp-tag55 contig and the Hc-cyp-tag70+9 contig was determined by hemi-nested PCR amplification followed by direct sequencing (see Figure 4-2B). This additional 211 bp region of cDNA sequence was completely absent from all *H. contortus* sequence databases, despite clearly representing a region homologous to *C. elegans* CYP42A1 (BLASTx, E=3e-11). An additional tag, Hc-cyp-tag61, encoded an identical 122 amino acid N-terminus to Hc-cyp-tag41. The coding sequence shared 98% nucleotide identity with Hc-cyp-tag41, but the genomic sequence was only 72% identical.

The full-length 1536 bp Hc-cyp-3 cDNA shared 68% identity to the 1539 nt *C. elegans cyp42a1* mRNA. The 512 amino acid conceptual translation was 73% identical to CYP42A1 in *C. elegans* (see Figure 4-9), 72% identical to CBR-CYP42A1 in *C. briggsae* and 64% identical to XP_001901707.1 in *B. malayi* (BLASTp, E=0 for all). These genes are of unknown function.

4.7.2.2 Hc-cyp-tag94

3 to 8 fold higher expression of Hc-cyp-tag94 was recorded in the anthelmintic-resistant MHco10 (CAVR) isolate relative to the MHco3 (ISE) isolate.

Hc-cyp-tag94 shares most identity with CYP14A1 in *C. elegans* (BLASTp, E=2e-10), which is a gene of unknown function. Nine other *H. contortus* CYP tags share most identity with *C. elegans* CYP14A family members, but only Hc-cyp-tag94 showed higher expression in the real-time screen.

Hc-cyp-tag94 lies within the biggest putative cluster of CYPs identified in the parasite (see Figure 4-3).

4.8 Tables and Figures

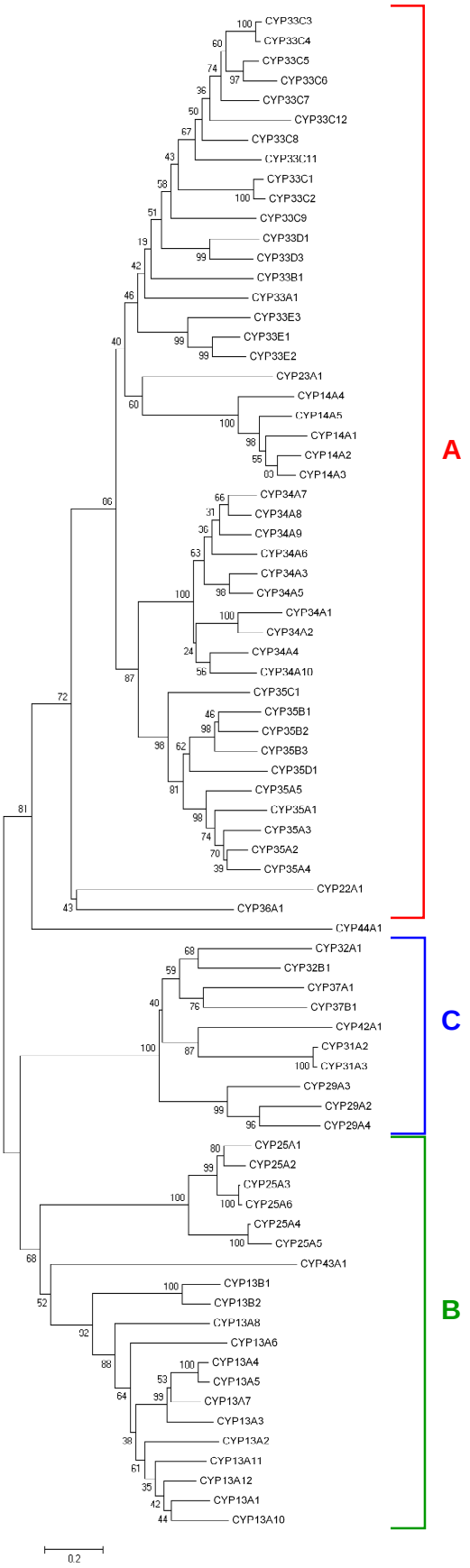
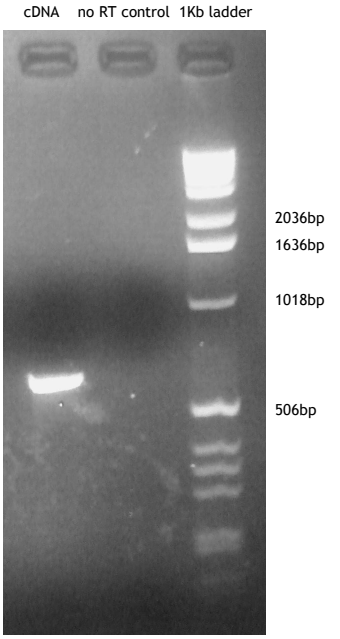


Figure 4-1: Neighbour-joining tree of *C. elegans* CYP polypeptides. The evolutionary history was inferred using the Neighbor-Joining method (Saitou *et al.*, 1987). The optimal tree with the sum of branch length = 21.52513544 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (2000 replicates) are shown next to the branches (Felsenstein, 1985). The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Poisson correction method (Zuckerkandl and Pauling, 1965) and are in the units of the number of amino acid substitutions per site. All positions containing gaps and missing data were eliminated from the dataset (Complete deletion option). There were a total of 143 positions in the final dataset. Phylogenetic analyses were conducted in MEGA4 (Tamura *et al.*, 2007). A, B and C relate to clades described by Gotoh (1998). CYP44A1 is the only mitochondrial CYP in *C. elegans*.

A:



B:

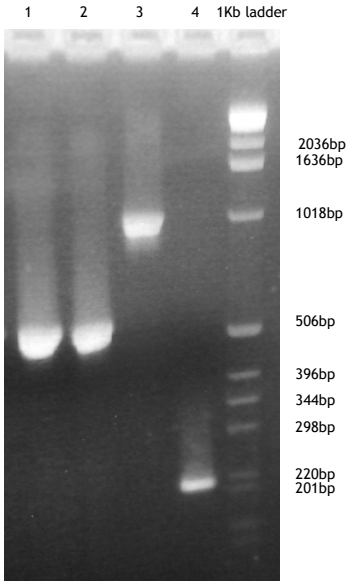


Figure 4.2: RT-PCR amplification between CYP tags
A: RT-PCR product amplified from adult worm cDNA with Hc-cyp-tag25F and Hc-cyp-tag15R primers. This product was sequenced to reveal 99 bp coding sequence missing due to a gap in the supercontig assembly, showing both tags represented the same gene, *Hc-cyp-1*.
B: Hemi-nested RT-PCR amplifications from adult worm cDNA. No products were visualized on a gel after one round of RT-PCR amplification with Hc-cyp-tag41F + Hc-cyp-tag9R or Hc-cyp-tag70R primers (data not shown). A second round of hemi-nested RT-PCR amplified the expected products (shown above), demonstrating Hc-cyp-tag41, Hc-cyp-tag55, Hc-cyp-tag9 and Hc-cyp-tag70 represent the same gene, *Hc-cyp-3*. The product of lane 3 was sequenced to reveal 211 bp of coding sequence missing from reference databases.
1= Hc-cyp-tag41F + Hc-cyp-tag9R product as template for Hc-cyp-tag41F + Hc-cyp-tag55R
2= Hc-cyp-tag41F + Hc-cyp-tag70R product as template for Hc-cyp-tag41F + Hc-cyp-tag55R
3= Hc-cyp-tag41F + Hc-cyp-tag70R product as template for Hc-cyp-tag55F + Hc-cyp-tag9R
4= Hc-cyp-tag41F + Hc-cyp-tag70R product as template for Hc-cyp-tag41F + Hc-cyp-tag41R

SUPERCONTIG	HC-CYP-TAG	REGION (amino acids)	CLOSEST *C. elegans* CYP
1	54	2-87	14A3
2	42	15-87	14A5
3	63	2-148	14A3
4	46	28-84	14A5
5	50_94_97_62_36_35	425- 462115-191,293-421,293-423,374-461,1-95	33C7,14A5,33C2,33C9,14A2,14A5
6	60+34_33_49+30	169-461, 115-190, 289-425	14A5, 14A4, 33C1
7	32_68+69	1-90, 285-462	14A3, 33C6
8	91_39+38_37	1-28, 3-187, 226-390	33E1, 33C9, 33C5
9	95_56	360-425, 291-370	33C7, 33E2
10	84	6-141	33C2
11	16	4-470	33C9
12	81	274-288	33C2
13	45	89-143	33C11
14	40	289-359	33C3
15	22+14	11-519	22A1
16+18	65+64+10	1-323	23A1
17	82	75-177	23A1
19	66	68-107	13B1
20	7	434-494	13B1
21	90_19	1-62, 268-304	13B2, 13A12
22	78	459-493	13A12
23	8	303-358	13A12
24	79	303-435	13A8
25	2+6	17-439, 434-494	13A12, 13B1
26	1_ 5	17-305, 433-486	13A12, 13B1
27	51_92_93	207-359, 27-61, 15-107	32A1, 32A1, 13B1
28	20	317-462	32B1
29	44	239-341	32A1
30	89	21-141	32A1
31	86_87	234-325, 470-506	32A1, 37B1
32	80	454-507	37B1
33	53_67	463-507, 142-521	37B1, 32A1
34	25+15	1-508	31A2
35	12+13	22-68, 302-494	13A11, 13A10
36	27	454-507	37B1
37	17+18	1-448	33E2.
38	47+48	61-329	34A4
39	57	150-194	34A7
40	43	167-329	34A8
41	71+72	28-453	34A3
42	24+23+52	27-469	34A9
43	28	24-147	34A4
44	26	104-137	36A1
45	59+58	140-500	36A1
46	88	325-493	36A1
47	29	38-136	36A1
48	21_96	288-457, 419-506	37A1
49+51+52	41+55+9+70	1-512	42A1
50	61	1-122	42A1
53+54	4+77	1-467	43A1
55	11+3	48-467	43A1
57	31	403-439	44A1
59	76	399-481	44A1
56+58+60+61	83+73+74+76+85	1-480	44A1

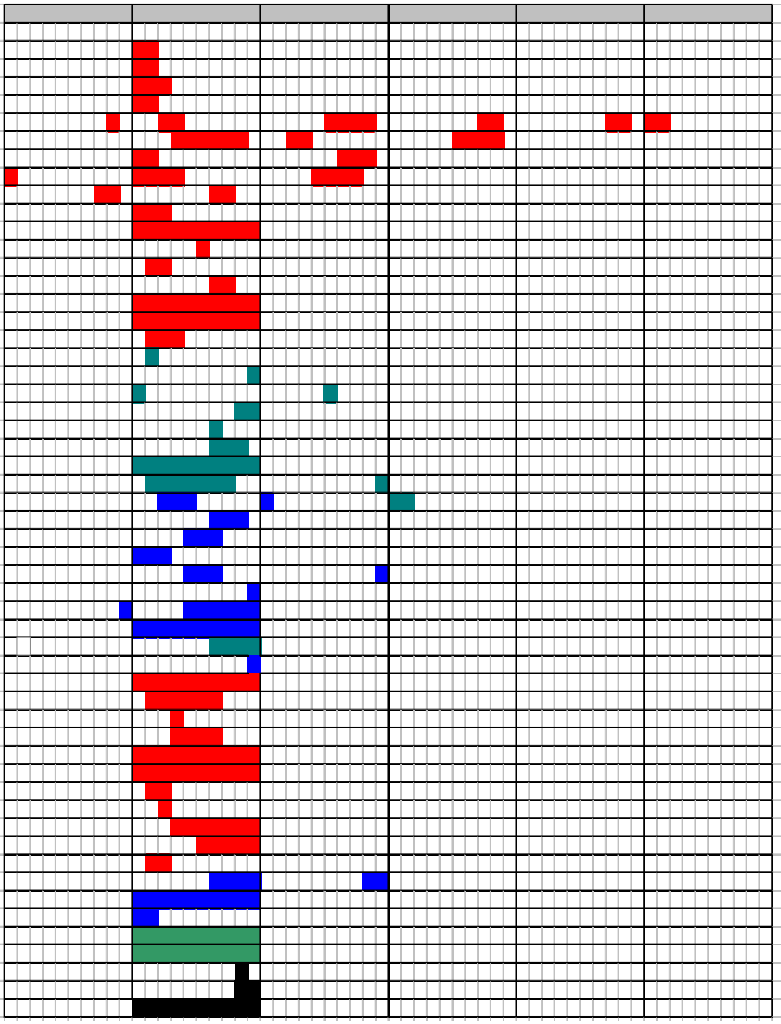


Figure 4.3: Schematic of 73 partial CYP sequences assembled from 97 CYP tags. Rows represent supercontigs and columns represent full-length genes. Coloured regions show the proportion of the gene represented by each partial CYP, colour relates to the clade of the closest *C. elegans* CYP (see Figure 4-1): red=clade A, green=clade B, blue=clade C. Tags numbers with a "+" between them represent the same gene.

Putative family	HC-CYP-TAG	Region of polypeptide (amino acids)	*C. elegans* homologue	Identity with homologue
CYP13	13	302-494	13A10	
	12	22-68	13A11	
	1	1-305	13A12	44%
	2+6	1-494	13A12	43%
	19	268-304	13A12	
	8	303-358	13A12	
	78	459-493	13A12	
	79	303-435	13A8	
	93	15-107	13B1	
	5	433-486	13B1	
	7	434-494	13B1	
	90	.1-62	13B2	
CYP14	94	115-191	14A1	
	62	287-461	14A3	
	32	.1-90	14A3	
	63	1-148	14A3	
	54	1-87	14A3	
	33	115-190	14A4	
	35	.1-95	14A5	
	42	15-87	14A5	
	60+34	169-461	14A5	34%
	46	28-84	14A5	
CYP22	22+14	1-519	22A1 (daf-9)	37%
CYP23	65+64+10	1-533	23A1	52%
	82	75-177	23A1	
CYP31	25+15	1-508	31A2	57%
CYP32	67	142-521	32A1	
	51	207-359	32A1	
	89	21-141	32A1	
	86	234-325	32A1	
	44	239-341	32A1	
	92	27-61	32A2	
	20	317-462	32B1	
CYP33	66	70-106	33B1	
	49+30	289-425	33C1	
	45	89-143	33C11	
	97	293-421	33C2	
	84	1-141	33C2	
	81	274-288	33C2	
	40	289-359	33C3	
	37	226-390	33C5	
	68+69	285-462	33C6	
	95	360-425	33C7	
	50	425-426	33C7	
	91+19	.1-390	33C9	51%
	39+38	1-187	33C9	
	16	1-470	33C9	48%
	36	355-461	33C9	
	17+18	1-448	33E1.	41%
	56	291-370	33E2.	
CYP34	28	1-147	34A4	
	47+48	1-329	34A4	
	57	150-194	34A7	
	43	167-329	34A8	
	24+23+52	1-469	34A9	34%
CYP35	71+72	1-453	35A4	35%
CYP36	26	104-137	36A1	
	29	1-160	36A1	
	59+58	140-500	36A1	
	88	325-493	36A1	
CYP37	21	288-457	37A1	
	96	419-506	37B1	
	27	454-507	37B1	
	80	454-507	37B1	
	53	463-507	37B1	
	87	470-506	37B1	
CYP42	61	1-121	42A1	
	41+55+9+70	1-508	42A2	73%
CYP43	11+3	1-467	43A1	35%
	4+77	1-497	43A1	42%
CYP44	31	403-439	44A1	
	75	413-482	44A1	
	83+73+74+76+85	1-481	44A1	54%

Table 4.1: Grouping of Hc-cyp-tags into putative families based on *C. elegans* CYP families. Identity with homologue shows % amino acid identity between conceptual translations of CYP tags encoding >75% gene and the closest *C. elegans* CYP polypeptide.

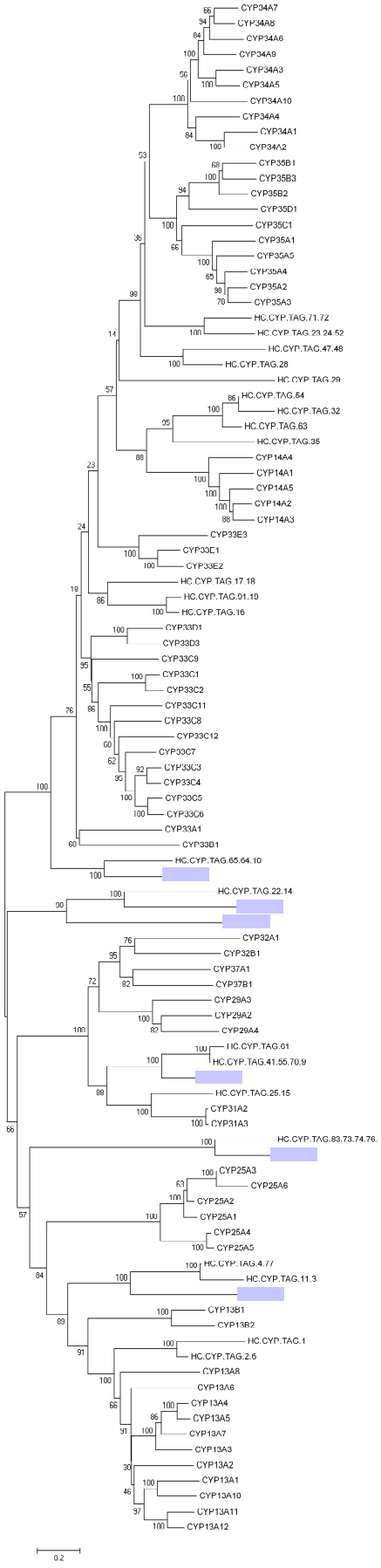


Figure 4-4: Neighbour-joining tree of conceptual translations of *H. contortus* CYP tags and *C. elegans* CYP polypeptides. The evolutionary history was inferred using the Neighbor-Joining method (Saitou *et al.*, 1987). The optimal tree with the sum of branch length = 30.99985086 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (2000 replicates) are shown next to the branches (Felsenstein, 1985). The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Poisson correction method (Zuckerkandl and Pauling, 1965) and are in the units of the number of amino acid substitutions per site. All positions containing alignment gaps and missing data were eliminated only in pairwise sequence comparisons (Pairwise deletion option). There were a total of 710 positions in the final dataset. Phylogenetic analyses were conducted in MEGA4 (Tamura *et al.*, 2007). *C. elegans* single family member CYPs are highlighted.

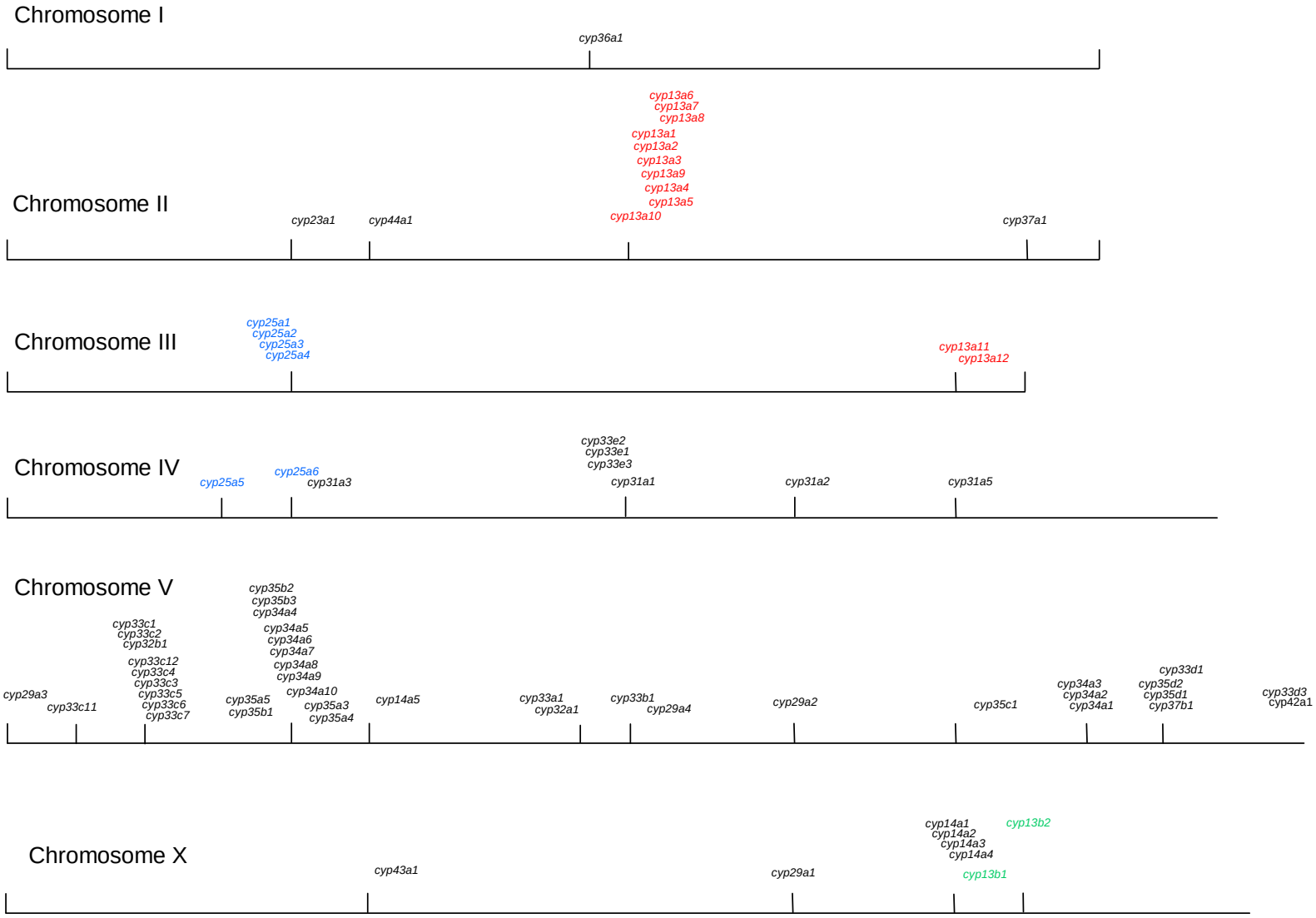


Figure 4.5: Clustering of CYP genes in *C. elegans* genome

A:

```
EMB-8          MLAWIVSGLDTSDLVVLTLLAGGAIIFLFMKVFNQQP---SSSRYSPTVASVTTSAAASK 57
HC-0047488     MWDWLVSQLDVTDFIILSALMLGGMYFLVKYRLGGTSRWVSLFRYKPTMTPITPTAKPK- 59

EMB-8          SNQSFIDRMKNENRQVLIMYGSQTGTAEEMSGRLAKDLTRYTKKAVVVDPEDIECEDLNR 117
HC-0047488     -DMSFISRMKNENRQVLIMYGSQTGTAEELSGRLAKD---------------------- 95

EMB-8          LSEVEDALLVLCIATYGEGDPTDNAVTLVEYLNAGDCDLSGVRFAVFGLGNKTYEHFNEI 177
HC-0047488     ---------------------------------------------VFGLGNKTYEHFNAV 110

EMB-8          GIQMDKQLEKLGAKRIFHLGLGDDDANLEEDFMIWREAFLPKVAEEFGWELNTEAETMRQ 237
HC-0047488     GKLFDKKLEELGAERAFPLGLGDDDANLEEDFMRWREAFLPTVAQRFGWELNTDAETLRQ 170

EMB-8          YQLEPVEEGK--ALFKGEFGRLGAYERPRPPFDVKNPYLATVAINDELHTEHSDRSCRHI 295
HC-0047488     YRLELVDHNANVTLFKGEYGRLGAFERLRPPFDQKNPFPATIAVNRELHTEKSERSCRHI 230

EMB-8          EFSVEGSRIRYEAGDHLAVFPTNDPVLVDRLINMLQFDPDHAFRLVNVDEDASKRHPFPC 355
HC-0047488     EFAVEGSRIRYEAGDHLAVFPTNDPELVDAVISLLDFDPEQAFRLINVDEESSKRNPFPC 290

EMB-8          PTTFRTALSHYVDICAPVKSHVLKAISEYCTDDTEKEFLNKLSTANEEGLKEYARYIVKE 415
HC-0047488     PCTYRTALTHYVDICAPLKSHVLKAISEYCSDEKEKAYLQLLSTATEEGLVQYSSYIVKE 350

EMB-8          RRSIVDVLTDQKTCKPPIEYLLELLPRLQARYYSIASSPRLNEEKIAICAVVTKYSIGDR 475
HC-0047488     RRSIIDVLRAHPSCKPPIEYLLELLPRLQARYYSIASSPKHQENRIAICCIVTKYTIGDR 410

EMB-8          DINGVCTRYLTTKDAGSKSPVFVRKSTMRLPHRTTTQVIMIGPGTGFAPFRGFLQDRQFH 535
HC-0047488     LIKGVCTNYLAGKDVNDRTPVFVRKSQMRLPHRTNTPVIMIGPGTGFAPFRAFLQERKFQ 470

EMB-8          KNAGKEIGAMHLYYGCRHPDHDYIYKDELAKFQEDEVLTHLVCAFSRAQEHKIYVQDRLW 595
HC-0047488     KNQGKEIGPIMLYYGCRHPEHDYIYQDEIEEMVKDGVITDLYCAFSRAQEHKIYVQNRLW 530

EMB-8          ETRDRIWDAINVGAHVYICGDARNMARDVQATLQKIFREIGGKSETEAVAYFKDMEKTKR 655
HC-0047488     ESRDKVWSAIEAGAHIYVCGDARNMARDVQNVLLRILQEVGGKSEEESATLFKNLERQRR 590

EMB-8          YQADVWS 662
HC-0047488     YQADVWS 597
```

B:

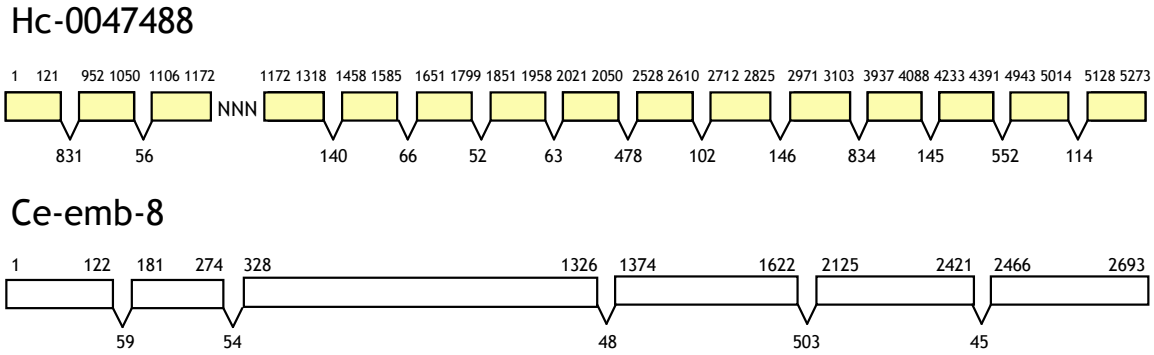


Figure 4.6: *C. elegans* EMB-8 and putative *H. contortus* orthologue HC-0047488
A: Amino acid alignment of EMB-8 and HC-0047488 polypeptides. B: Schematic of *emb-8* and *Hc-0047488* genes

A:

```
CYP31A2          MGVIIPAVLLAMATVIAWLLYKHLRMRQVLKHLNQPRSYPIVGHGLITKPDPEGFMNQVI 60
CYP31A3          MGVIIPAVLLASATIIAWLLYKHLRMRQALKHLNQPRSYPIVGHGLVTKPDPEGFMNQVI 60
HC-CYP-1         MGLLSAFLFVTVVTSIVYFIAKHLQLRNQLKGINCPRSYPLIGHGLLKKPDMEGFINQVM 60
                 **:: . :::: .* *.::: ***::*: ** :* *****::*****:.*** ***:***:

CYP31A2          GMGYLYPD-PRMCLLWIGPFPCLMLYSADLVEPIFSSTKHLNKGFAYVLLEPWLGISILT 119
CYP31A3          GMGYLYPD-PRMCLLWIGPFPCLMLYSADLVEPIFSSTKHLNKGFAYVLLEPWLGISILT 119
HC-CYP-1         GMAQMYPDSPRMVLFWLGPVPVVMLYSARLVEKILNCSQHLNKGIAYYFFESWLGQGIIT 120
                 **. :*** *** *:*:**.* :***** *** *:..::*****:** ::*.*** .*:*

CYP31A2          SQKEQWRPKRKLLTPTFHYDILKDFLPIFNEQSKILVQKLCCLGADEEVDVLSVITLCTL 179
CYP31A3          SQKEQWRPKRKLLTPTFHYDILKDFLPIFNEQSKILVQKMCSLGAEEEVDVLSVITLCTL 179
HC-CYP-1         SNVDNWRPKRKLLTPTFHYDILKDFVPIFNDQAQILVKKFASLPEGYPVELMSYITLCAL 180
                 *: ::********************:****:*::***:*:..*     *:::* ****:*

CYP31A2          DIICETSMGKAIGAQLAENNEYVWAVHTINKLISKRTNNPLMWNSFIYNLTEDGRTHEKC 239
CYP31A3          DIICETSMGKAIGAQLAENNEYVWAVHTINKLISKRTNNPLMWNSFIYNLTEDGRTHEKC 239
HC-CYP-1         DIICETSMGKSLNAQLDKESEYVKAVHTVNDLVQKRTKSPLYWNDYFYNKFGEGETEKKC 240
                 **********::.*** ::.*** **** :*.*:.***:.** **.::**   :*.*.:**

CYP31A2          LRILHDFTKKVIVERKEALQENDYKMEGRLAFLDLLLEMVKSGQMDETDVQAEVDTFMFE 299
CYP31A3          LRILHDFTKKVIVERKEALQENDYKMEGRLAFLDLLLEMVKSGQMDETDVQAEVDTFMFE 299
HC-CYP-1         IDILHSFTNKVIAERRKELEDRQWRFEGRRAFLDLLLDMANSGQLEASEIQEQVDTLMFA 300
                 : ***.**:***.**:: *::.::::*** *******:*.:***:: :::* :***:**

CYP31A2          GHDTTSTGLMWAIHLLGNHPEVQRKVQAELDEVMGDDEDVTIEHLSRMKYLECALKEALR 359
CYP31A3          GHDTTSTGLMWAIHLLGNHPEVQRKVQAELDEVMGDDEDVTIEHLSRMKYLECALKEALR 359
HC-CYP-1         GHDTTSTGSSWALFLFGCYPEIQRKVQEEIDEVLEDSDYILPEHLPRLKYLECCLKESLR 360
                 ********  **:.*:* :**:***** *:***: *.: :  ***.*:*****.***:**

CYP31A2          LFPSVPIITRELSDDQVIGGVNIPKGVTFLLNLYLVHRDPAQWKDPDVFDPDRFLPENSI 419
CYP31A3          LFPSVPIITRELSDDQVIGGVNIPKGVTFLLNLYLVHRDPSQWKDPDVFDPDRFLPENSI 419
HC-CYP-1         LCTPVPMIMRKLGADQELEGVTLPKGTQVVLNQYMVHRDPMYWPDPEKFDPDRFLPENCI 420
                 * ..**:* *:*. ** : **.:***. .:** *:***** * **: **********.*

CYP31A2          GRKSFAFIPFSAGSRNCIGQRFALMEEKVIMAHLLRNFNIKAVELMHEVRPKMEIIVRPV 479
CYP31A3          ARKSFAFIPFSAGSRNCIGQRFALMEEKVIMAHLLRNFNVKAVELMHEVRPKMEIIVRPV 479
HC-CYP-1         GRHPFAFIPFSAGSRNCIGQRFGLMEIKVVVSWMLRHFNVTAVQRRCDLKSKIEIILRPQ 480
                 .*:.***************.*** **::: :**:**:.**:   :::.*:***:**

CYP31A2          TPIHMKLTRRRPIVSP------- 495
CYP31A3          TPIHMKLTRRRPIVSP------- 495
HC-CYP-1         DGIHVFLEKRRAIADGFRSSLIA 503
                   **: * :**.*..
```

B:

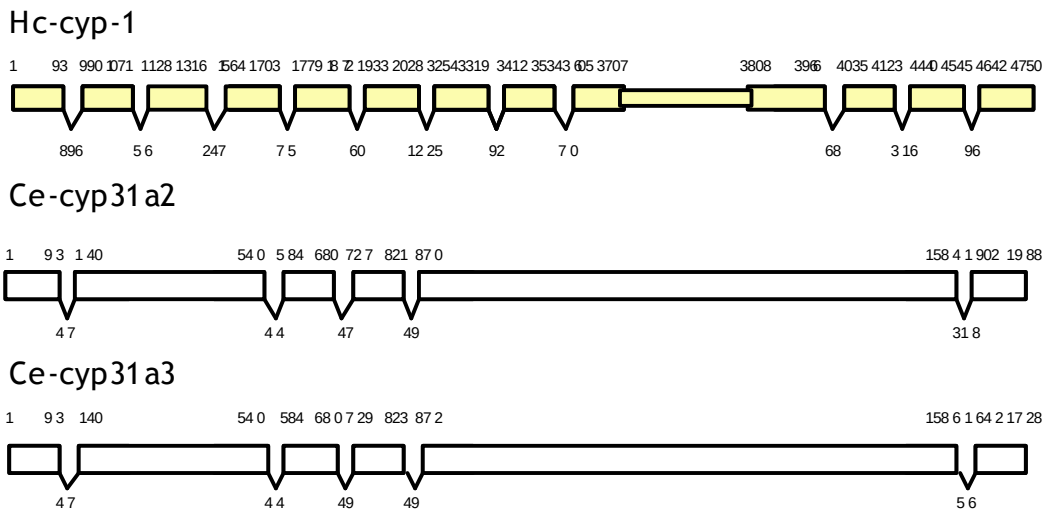


Figure 4.7: *H. contortus* HC-CYP-1 and *C. elegans* CYP31A2 and CYP31A3
A: Amino acid alignment of HC-CYP-1, CYP31A2 and CYP31A3 polypeptides. B: Schematic of *Hc-cyp-1*, *cyp31a2* and *cyp31a3* genes.

A:

```
CYP44A1         MRRSIR---NLAENVEKCPYSPTSS---PNTPPRTFSEIPGPREIPVIGNIGYFKYAVKS 54
HC-CYP-2        MRRSVVVAASSAKSMLNCPISASTATVEQEIEARPFEEIPGPN--VFERHFGRNRTVFRS 58
                ****:    . *:.: :** *.:::    :  .*.*.*****.   .  ::*  : ..:*

CYP44A1         DAKTIENYNQHLEEMYKKYGKIVKENLGFGRKYVVHIFDPADVQTVLAADGKTPFIVPLQ 114
HC-CYP-2        -KRSIANYFQWLVDLHKRYGPIVRVEQGFGRGAVVHVFDPEDARRVFASDGRQPFIVPLQ 117
                  ::* ** * * :::*:** **: : ****  ***:*** *.: *:*:**: *******

CYP44A1         ETTQKYREMKGMNPGLGNLNGPEWYRLRSSVQHAMMRPQSVQTYLPFSQIVSNDLVCHVA 174
HC-CYP-2        ETTQRYRQMKGMNPGLGNLNGDEWYRLRSSVQQVMMRPQAVQKYLPYTNEVAQELVDHVR 177
                ****:**:************ **********:.*****:**.***::: *:::** **

CYP44A1         DQQKR--FGLVDMQKVAGRWSLESAGQILFEKSLGSLGNRSEWADGLIELNKKIFQLSAK 232
HC-CYP-2        RESETSTSGEVDVSKIAGRWALESSALTVFEKRIGALTDRIEWADGLVNLNKAIFRLSAV 237
                 :.:    * **:.*:****:***:.  :*** :*:* :* ******::*** **:***

CYP44A1         MRLGLPIFRLFSTPSWRKMVDLEDQFYSEVDRLMDDALDKLKVNDSDSKDMRFASYLINR 292
HC-CYP-2        LKFAFPLYQYFPTPKWKKMVELEDRFY--------------RFN-----EMKFASLLINR 278
                :::.:*::: *.**.*:***:***:**              :.*     :*:*** ****

CYP44A1         KELNRRDVKVILLSMFSDGLSTTAPMLIYNLYNLATHPEALKEIQKEIKEDPASS---KL 349
HC-CYP-2        KELNVNDVKIILLSMFSDGLSTTAPMLVYNLFNIATHPDVQAEIRDEVNAAVQRNEGWSL 338
                **** .***:****************:***:*:****:.  **:.*::     .   .*

CYP44A1         TFLRACIKETFRMFPIGTEVSRVTQKNLILSGYEVPAGTAVDINTNVLMRHEVLFSDSPR 409
HC-CYP-2        PLLRACIKETFRLFPIGTEISRIPQKDIVLSNYHIPAGTPVDINTNVLMRSPTLFND-PL 397
                .:**********:******:**:.**:::**.*.:****.**********  .**.* *

CYP44A1         EFKPQRWLE---KSKEVHPFAYLPFGFGPRMCAGRRFAEQDLLTSLAKLCGNYDIRHRGD 466
HC-CYP-2        AFQPSRWLRDVSRQQDFHPFAFLPFGFGPRMCAGRRFAEQDLQVALCRLLQHYRIVHQHG 457
                 *:*.***.   :.::.****:******************** .:*.:*  :* * *: .

CYP44A1         PITQIYETLLLPRGDCTFEFKKL 489
HC-CYP-2        SIEQTYETLLLPKGYCEFRFEPL 480
                .* * *******:* * *.*: *
```

B:

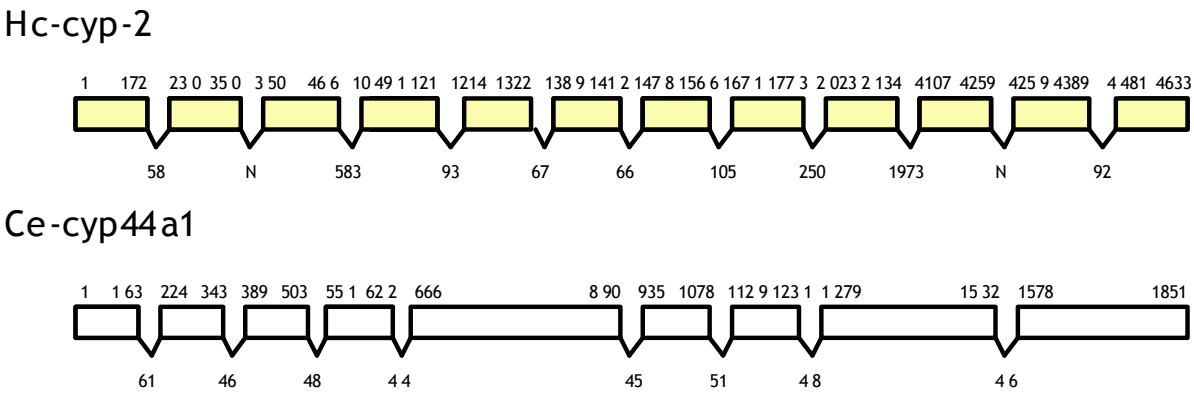


Figure 4.8: *H. contortus* HC-CYP-2 and *C. elegans* CYP44A1
A: Amino acid alignment of HC-CYP-2 and CYP44A1 polypeptides. B: Schematic of *Hc-cyp-2* and *cyp44a1* genes

A:

```
CYP42A1         MGIITASLIVLTITWIIHFAFRKAKFIYNKLTVFQGPAALPLIGNFHQFHFSPEEFFEQS 60
HC-CYP-3        MGLVAIALAISVFTWITNFIYKKAKYIFDRIAVFQGPVALPFIGNLHQFHFTPEEFFEQA 60
                **::: :* : .:*** :* ::***:*:::::*****.***:***:*****:*******:

CYP42A1         QGIAYMMRKGDERITRVWLGGLPFVLLYGAHEVEAILGSPKMLNKPFLYGFLSAWIGDGL 120
HC-CYP-3        QGLAYMLRKNRDRMTRVWFGPLPYVLIYGPEECEAVLGSSKMLNKTMQYSFLSAWIGEGL 120
                **:***:**. :*:****:* **:**:**..* **:***.*****.: *.*******:**

CYP42A1         LISKPDKWRPRRKLLTPTFHYDILKDFVEVYNRHGRTLLSKFEAQAGTGEYSDVFHTITL 180
HC-CYP-3        LISKPDKWRPRRKLLTPTFHYEILKDFVEVYNRHGRTLLGKFLKHAEDGQYENIFHTVTL 180
                **********************:*****************.**  :*  *:*.::***:**

CYP42A1         CTLDVICEAALGTSINAQKDPHSPYLDAVFKMKDIVFQRLLRPHYFSDTIFNLIGPGKEH 240
HC-CYP-3        CTLDVICEAALGICLDVQKNPHSPYLDSVFKMKVLIQKRLVKPQYYPEFLFNLFGAGREQ 240
                ************ .::.**:*******:***** :: :**::*:*:.: :***:*.*:*:

CYP42A1         DECVKILHEFTSKAIYARKAKVDAAGGVEQLLAQETAEGRRRMAFLDLMLDMNSKGELPM 300
HC-CYP-3        ARCVKILHEFTGNVIRARKAKADAAGGVEKLLAQESAEGRRRMAFLDLMLDMNAKGELPM 300
                 .*********.:.* *****.*******:*****:*****************:******

CYP42A1         EGICEEVDTFTFEGHDTTSAAMNWFLHLMGANPEIQSKVQKEIDEVLGEADRPVSYEDLG 360
HC-CYP-3        DGVQEEVDTFTFEGHDTTSASINWFLHLMGANPDIQEKVQREVDEVLGEVDRPVTYEDLG 360
                :*: ****************::***********:**.***:*:******.****:*****

CYP42A1         KLKYLEACFKETLRLYPSVPLIARQCVEDIQVRGHTLPSGTAVVMVPSMVHKDPRYWDDP 420
HC-CYP-3        GLKYLEACIKETLRLYPSVPLIARQTVEDIKIKDHVLPSGTGVVVVPSMVHRDPNYWDDP 420
                 *******:**************** ****:::.*.*****.**:******:**.*****

CYP42A1         EIFNPERFITGELKHPYAYIPFSAGSRNCIGMRFAMMEEKCILAIILKNLKVKAKLRTDE 480
HC-CYP-3        EVFRPERFIDGELKHPYAYIPFSAGSRNCIGQRFAMMEEKCILALLMRHLRVRSLLRTDE 480
                *:*.***** ******************** ************:::::*:*:: *****

CYP42A1         MRVAAELIIRPLYGNELKFEKREFGDYTSIY- 511
HC-CYP-3        MRVAAELIIRPLHGNRIKFEKRAYGDYTHCSP 512
```

B:

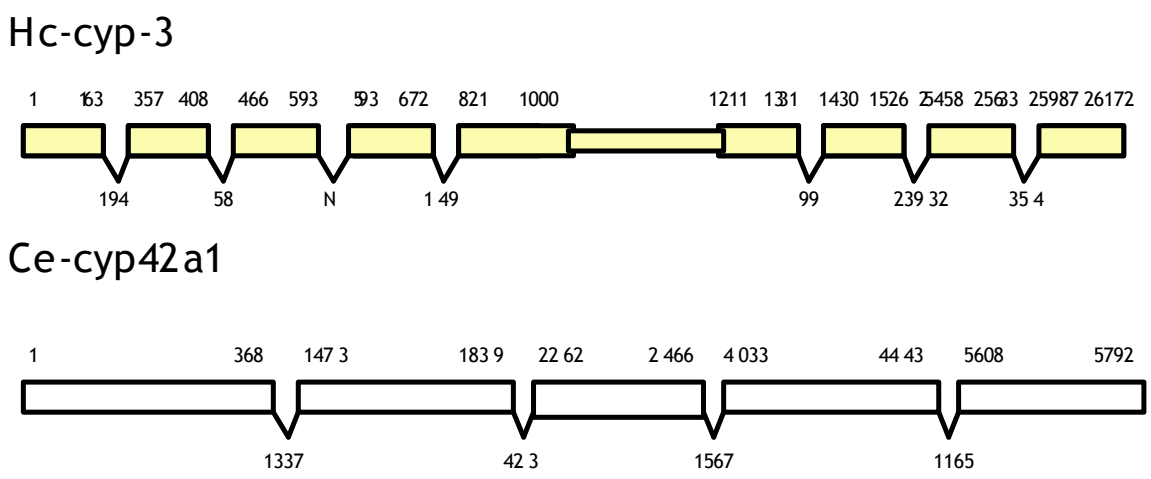


Figure 4.9: *H. contortus* HC-CYP-3 and *C. elegans* CYP42A1
A: Amino acid alignment of HC-CYP-3 and CYP42A1 polypeptides. B: Schematic of *Hc-cyp-3* and *cyp42a1* genes

4.9 Discussion

This project identified 73 partial CYP genes in the incomplete *H. contortus* genome. A number of these sequences may assemble into a smaller number of full-length genes and conversely, if a significant proportion of genome sequence is missing from the reference databases, the number of CYPs could increase. However, the findings that over 95% of the available *H. contortus* ~6000 single cluster ESTs can be identified in the genomic databases suggest coverage is good (pers. com. Dr G Saunders).

These findings are supported by the identification of at least one homologous *H. contortus* CYP tag for all single family member CYPs in *C. elegans*, which are predicted to be conserved between species. One caveat would be the possibility that current genome assembly algorithms might assemble reads from single family member CYP genes more easily than multi-member CYP families containing related genes with high nucleotide identity, which would remain in the unassembled reads databases. This project was based only on sequences from the contig and supercontig databases because they were more amenable to study than the read databases, which at 600-800 bp were prohibitively short for further analysis.

In most cases, even the *H. contortus* supercontigs were too short to encode full-length CYP genes, with the majority encoding less than 20% of the mRNA. One of the main challenges for the assembly of these partial genes was distinguishing alleles from subfamily members. CYP families in other species are known to expand via multiple gene duplications, generating new members with conserved intron: exon boundaries and with high shared sequence identity even at the genomic level (Gotoh 1998, Thomas 2007, Tijet *et al* 2001). In the unassembled *H. contortus* genome this created a jigsaw with extra pieces, missing pieces, and pieces that would fit in multiple combinations (guided by a picture from a different puzzle). However, 75-100% of the sequences of 15 CYP genes were assembled.

Pseudogenes and detritus exons provide an additional challenge to assembly of the family. As pseudogenes are predominantly duplicated genes rather than retrotransposed genes (Nelson *et al* 2004), pseudogenes and detritus exons sharing high identity with their relatives may be present within CYP clusters. This is especially problematic for assembling the *H. contortus* CYP family as the majority of CYPs within a cluster are currently only partial sequences, due to gaps in the supercontig assembly.

Expression data generated with this project was found to be of limited value for the detection of pseudogenes. The low constitutive expression of a number of functional CYP genes and their stage specificity (see Chapter 3) suggests the lack of ability to detect expression of a CYP with RT-PCR or RNA-seq does not necessarily prove a lack of expression. Also, the high sequence identity of tags representing closely related family members means that hypothetically, primers designed against the genomic sequence of a pseudogene could amplify the transcript of an expressed CYP and conversely, transcriptome reads generated by an expressed CYP could map to a pseudogene.

The clustering of CYPs identified at the Hc-cyp-tag-94 locus is of interest due to the higher constitutive expression of the gene in the anthelmintic-resistant MHco10 (CAVR) isolate. In a study of ten vertebrate genomes, Thomas (2007) found nearly all CYPs with known xenobiotic substrates were found in gene clusters, so clustering at the Hc-cyp-94 locus would be consistent with the presence of xenobiotic response genes. In *C. elegans*, the closest CYPs to all of the clustered parasite genes are CYP14 and CYP33 family members, which also cluster in large arrays in *C. elegans*. Of these, *cyp14a3, cyp33c2* and *cyp33e1* have been predicted in other studies to be orthologues of human xenobiotic inducible genes *cyp1a2, cyp2e1* and *cyp2c19*, and are also induced in *C. elegans* on exposure to a range of xenobiotics (Chakrapani *et al* 2008, Menzel *et al* 2001).

This work also identified putative orthologues of the *C. elegans* nuclear hormone receptor *daf-12* and the cytochrome P450 reductase in *H. contortus*, suggesting key pathways in CYP transcription and electron transfer may be conserved between the species. This in turn highlights the likely possibility that *H.*

contortus may utilise the full range of biotransformation pathways and xenobiotic responses known in insects and mammals.

4.10 Conclusion

The *H. contortus* genome encodes a large number of CYP genes. The 97 partial CYP genes identified in the contig and supercontig databases have been assembled into 73 larger sequences, of which 15 encode 75-100% of a CYP gene.

A number of full length *H. contortus* CYP genes share high identity with CYPs in *C. elegans* (up to 73% amino acid identity of conceptual translations) and life stage, tissue and sex specific expression data supports their classification as putative orthologues.

It is hoped that improvements in sequencing technology and assembly algorithms will allow completion of the *H. contortus* genome in the near future. This will allow identification of the full complement of *H. contortus* CYP genes, and facilitate more powerful comparative analysis with CYPs in other species. Homologues of xenobiotic metabolising CYPs in humans and *C. elegans* will be of particular interest for research into parasite drug metabolism and resistance, as will any CYP families that have undergone recent expansion in the parasite. It is hoped, however, that this work provides a valuable platform for further study of the *H. contortus* CYP family.

5 A transcriptomic approach to gene annotation in *H. contortus*

5.1 Introduction

The completed *C. elegans* genome has proven to be a powerful resource for nematode gene annotation. The degree of conservation between *H. contortus* and *C. elegans* at a gene, contig and genome level was investigated and the impact of findings on the future assembly and annotation of the *H. contortus* genome, and thus CYP family, was investigated. To date, comparative genetics with *C. elegans* has been successfully applied to the global annotation of *C. briggsae, P. pacificus* and *B. malayi* genomes (Stein *et al* 2003, Dieterich *et al* 2008, Ghedin *et al* 2003) in addition to a plethora of studies investigating individual genes or gene families in a multitude of species.

Analysis of expressed sequence tag (EST) data by Parkinson *et al* (2004) found ~65% of *H. contortus* genes had significant similarity to proteins in *C. elegans*, indicating a substantial degree of conservation at the polypeptide level. However, the degree of conservation at gene level e.g. intron size, intron number, is unknown, and such factors could greatly affect the ability of gene prediction software to accurately detect parasite genes, if based solely on model worm parameters. The same study also found that ~19% of *H. contortus* genes lacked similarity to known genes in any species, highlighting the need for species-specific data.

Until recently, EST analysis and tiling microarrays were the most viable species-specific whole genome approaches for gene discovery. However, both are expensive, time-consuming and require a large amount of starting RNA. They also have their own approach-specific limitations such as cloning bias in EST data and cross-hybridisation in microarrays (Marguerat and Bahler 2010, Mortazavi *et al* 2008, Wang *et al* 2009).

An alternative approach became available with the advent of next generation sequencing (NGS) technologies. Illumina, Roche 454 and AB SOLiD platforms can be used to directly sequence expressed transcripts, with the resulting high depth of short reads aligned against reference sequence to guide annotation, in a technique known as 'RNA-seq' (Nagalakshmi *et al* 2010). Similarly to EST and tiling microarray analysis, RNA-seq can also be used to measure gene expression.

All NGS platforms are based upon the sequencing of DNA fragments in parallel, producing large numbers of short reads (Marguerat and Bahler 2010). Illumina sequencing involves the attachment of short fragments of cDNA to a glass surface, the flow cell. This binding is facilitated by the hybridisation of oligo-adapters attached to the DNA fragments with complementary oligos covalently bonded to the interior surface of the cell. PCR amplification of the DNA fragments then generates high-density clusters to be sequenced by synthesis. Polymerase is added with four differentially labelled dNTPs with reversible 3' terminators, to ensure only a single base is incorporated per cycle. After each cycle, the fluorescing reaction is imaged to identify the newly incorporated base, before the 3' terminators are chemically removed allowing the next dNTP to be added to the sequence (Fox *et al* 2009, Mardis 2008).

The resulting large raw image files are then processed to extract numerical signals for every synthesis event, which are used for base calling. The resulting data files consist of lists of sequences, with their base call qualities, which can be filtered and directly mapped to a reference genome or assembled into contigs (Marguerat and Bahler 2010, Wang *et al* 2009).

Although RNA-seq avoids many of the issues affecting the former approaches it does have its own limitations. One issue is read coverage for rare transcripts. In general, larger genomes have more complex transcriptomes that require greater sequencing depth, but as coverage tends to reach a plateau (despite increasing depth), this becomes a law of diminishing returns (Wang *et al* 2009). The use of normalised cDNA samples or libraries differing in their expression profiles, may prove more effective than doubling the depth of sequencing.

Another consideration is that reads may map equally well to a number of loci, which can be especially problematic for genomes containing large families of paralogous genes or for internally repeated domains between genes (Mortazavi *et al* 2008). However, the availability of paired-end sequencing, increased read length and better mapping algorithms has already improved the situation.

This technology is evolving rapidly. The Illumina Genome Analyser can now produce reads of more than 100 bp length. One run currently takes 2-9.5 days (depending on read length; from 35-100 bp) and produces 1.5-2.1 Gb sequence per day (http://www.illumina.com). This in itself creates new informatic challenges for the storage, retrieval and processing of large amounts of data (Wang *et al* 2009), although compressed data formats and new processing software are being developed apace to address this problem (Li *et al* 2009).

The aim of this project was to use Illumina technology to sequence the *H. contortus* transcriptome and use RNA-seq to annotate the largest region of contiguous genomic sequence currently available. Conservation between *H. contortus* and *C. elegans* at a gene, contig and genome level was investigated and the impact of findings on the future assembly and annotation of the *H. contortus* genome were assessed.

5.2 Results

Illumina technology was used to sequence the transcriptome of the adult *H. contortus* MHco3 (ISE) isolate. 21 million 52 bp reads were generated and mapped onto 409 Kb of contiguous genomic sequence, assembled as the consensus of five overlapping BAC insert sequences: haemapobac13c1, haemapobac7n11, haembac15g16, haembac18h7 and haembac18g2, derived from the X chromosome (Redman *et al* 2008a), and a database of 22 BAC insert sequences. Reads were mapped using the algorithms ‘Mapping and Assembly with Qualities’ (MAQ) and ‘Sequence Search and Alignment by Hashing Algorithm’ (SSAHA) and viewed as read plots over genomic sequence in Artemis (Li *et al* 2008, Ning *et al* 2001, Rutherford *et al* 2000).

Later, with improvements in both sequencing technology and downstream analysis, 38 million and 19 million 76 bp reads were generated from Illumina sequencing runs of adult and L3 MHco3 (ISE) cDNA libraries respectively. These were mapped with 'Burrows-Wheeler Aligner' (BWA) algorithms, opened as BAM files directly in Artemis, and viewed as stacked paired reads over genomic sequence as shown in Figure 5.1 (Carver *et al* 2010, Li and Durbin 2009).

The 409 Kb contig and a 181 Kb BAC insert sequence, BH4E20, were annotated guided by the alignment of reads over transcribed regions. Gene prediction software and BLASTx homology were used to guide annotation of loci mapped by too few reads to confidently predict splice sites with transcriptome data alone. Conserved synteny was assessed by aligning the *H. contortus* 409 Kb contig and BAC insert with *C. elegans* chromosome X and chromosome I respectively, with the Artemis Comparison Tool (ACT) (Abbott *et al* 2005, Carver *et al* 2005).

Genes were named '*hc*' for *H. contortus*, followed by an identifier for the BAC insert they were located on, followed by a number e.g. *hc-bh4e20-1*.

5.2.1 Gene Annotation

53 transcripts were identified. 37 were coding sequences mapped with transcriptome data: 16 transcripts on the X chromosome fragment and 21 transcripts on BH4E20. An additional 14 coding sequences with a low coverage of mapped reads were predicted on the X chromosome fragment with the gene prediction software Genefinder. One ß-tubulin gene, *hc-18h7-1*, was annotated from sequenced cDNA (pers. com. Dr G Saunders) and gene *hc-18h7-3* was retrospectively identified within a region of conserved microsynteny. 12 of the 53 transcripts were identified as mobile elements, so were excluded from the analysis, resulting in a total of 41 putative protein-coding genes.

All *H. contortus* genes annotated in this study had an ATG start codon, 5′-GT/AG-3′ intron boundaries, and a TAG, TAA or TGA terminal stop codon. In total, the 590 Kb genomic sequence studied had a GC content of 43%, with a 47.6% and 46.3% GC content for the identified coding sequence in the X chromosome fragment and BH4E20 respectively. These figures were slightly higher than the *C.*

elegans genome GC content of 35.4% and exon GC content of 42.7% (The *C. elegans* Sequencing Consortium 1998, Stein *et al* 2003).

5.2.2 Gene Density

41 putative genes were identified in a total of 590 Kb genomic sequence, which is a density of one gene per 14.39 Kb. The genome average for *C. elegans* is one gene per 5 Kb (The *C. elegans* Sequencing Consortium, 1998).

23 of these transcripts were identified on the 409 Kb contig, which is a density of one gene per 17.78 Kb. The average gene density on the X chromosome in *C. elegans* is one gene per 6.54 Kb (The *C. elegans* Sequencing Consortium, 1998).

18 transcripts were identified on the 181 Kb BAC insert, which is a density of one gene per 10.05 Kb, relative to a *C. elegans* average of one gene per 4.77-5.06 Kb on chromosome I, the range reflecting a higher density in the central cluster region than in the arms (The *C. elegans* Sequencing Consortium, 1998).

5.2.3 Similarities to Known Proteins

The conceptual translations of 22 transcripts shared highest identity with *C. elegans* proteins, eleven transcripts with *C. briggsae* predicted proteins and three with *Brugia malayi* proteins, based on current NCBI databases. The conceptual translations of three transcripts shared most identity to proteins outside *Nematoda*. One transcript, *hc-bh4e20-1*, predicted to encode a P-glycoprotein (PGP) shared most identity with a published *H. contortus* PGP polypeptide, but was also highly similar to *C. elegans PGP-2*.

Homologous proteins (BLASTp, E>1e-10) were identifiable in *C. elegans* for all three predicted polypeptides with most identity to *B. malayi* proteins. For the three genes with a closest match outside *Nematoda*, two encoded proteins that were highly conserved in many species (*hc-13c1-3* encoded a putative RNA-binding protein and *hc-bh4e20-15* encoded a putative high mobility group protein) and both had homology (BLASTp, E>1e-10) to proteins in *C. elegans*. The third gene, *hc-13c1-2*, encoded a conserved F-box domain yet shared little

identity with any sequence in NCBI databases, the best BLASTp hit (only E=0.023) being an S-phase kinase in *Xenopus tropicalis*.

5.2.4 Conserved domain analysis

The predicted polypeptides encoded by 27 of the 41 predicted genes had conserved domains identifiable in NCBI and Pfam online databases. Seven of the predicted polypeptides contained at least one incomplete conserved domain, although the closest matching proteins in *C. elegans* for two of these, one of which has been confirmed by cDNA sequencing, also had incomplete conserved domains.

5.2.5 Comparison of *H. contortus* and *C. elegans* putative orthologues

To assess the extent of conservation of gene structure between *H. contortus* and *C. elegans*, a subset of putative orthologues of *C. elegans* genes were identified in the annotated parasite sequence.

The inherent risk in inferring an orthologous or paralogous relationship between *C. elegans* genes and those from the incomplete *H. contortus* genome is that closer relatives may be identified when the parasite genome is fully sequenced. With this caveat in mind, a conservative subset of 24 genes was identified in the parasite with significant homology to *C. elegans* genes; the predicted polypeptide encoded by each putative orthologue had greater than 45% amino acid identity to a *C. elegans* protein, over greater than 80% of its length, and no sequence with higher identity to the *C. elegans* protein was present in the *H. contortus* contig databases. Two neighbouring genes on the X chromosome fragment, *hc-13c1-5* and *hc-13c1-4*, were both homologous with *C. elegans folt-1*, so for the purposes of this study were considered as an orthologue (50% amino acid identity) and a paralogue (48% amino acid identity) respectively. *hc-13c1-4* was excluded from the subset of orthologues and this proposed gene duplication is discussed in more detail later. Three genes were included in the subset of orthologues due to conserved microsynteny, although their amino acid identity was below the arbitrary cut-off. Gene *hc-18h7-3* lies within the second intron of

gene *hc-18h7-2* on the complementary strand, an identical relationship to their closest matched genes, *zk154.1* and *zk154.4*, in *C. elegans* (see Figure 5.2). Gene *hc-bh4e20-6* lies on the same strand and directly upstream of the orthologue of *C. elegans ath-1*, and shares most homology with *k04g2.11*, the corresponding upstream gene in *C. elegans* (see Figure 5.3).

This conservative subset of 24 genes was used in the following *H. contortus* to *C. elegans* gene comparison (see Table 5.1).

5.2.6 Gene size

The parasite genes had a similar spliced transcript size to their homologues in *C. elegans*, but unspliced transcript size was invariably larger, which was a function of both a greater number of introns and a larger intron size in *H. contortus* (see Table 5.1) Average unspliced transcript length was 5.72 Kb (median 4.84 Kb) compared to an average of 2.74 Kb (median 1.85 Kb) for the homologous gene set in *C. elegans*, with an average of 2.5 Kb (median 1.91 Kb) in the *C. elegans* genome (Duret and Mouchiroud 1999, Stein *et al* 2003). Predicted UTRs were excluded in the above calculations and any transcribed sequence (e.g. putative transposable elements) located within an intron of an *H. contortus* gene was subtracted from the unspliced transcript size.

The average number of introns was 8.83 per gene (median 7.5) in *H. contortus* compared to an average of 5.25 introns per gene (median 4) in the *C. elegans* homologues, with a genome average of 4 per gene (median 5) in the model worm (Deutsch and Long 1999, Stein *et al* 2003).

The average intron size in this *H. contortus* subset was 517 bp compared to an average intron size of 360 bp in the *C. elegans* homologues. The genome average intron size for *C. elegans* is 466.6 bp, although this is skewed by a small number of very large introns, giving a more representative median of 65 bp (Deutsch and Long 1999, Spieth and Lawson 2006). The average for *C. elegans* in this subset was inflated by a large first intron in gene *m60.4*, which is conserved in the orthologous gene, *hc-18g2-4*, in *H. contortus* (see Figure 5.2). As might be expected, the second intron in this gene was significantly larger in *H. contortus*

(799 bp compared to 57 bp), although the third intron was similarly sized in both nematodes. The *H. contortus* gene has gained an additional 4th intron.

5.2.7 Synteny/Colinearity

A previous study showed the 409 Kb contig was syntenic with the *C. elegans* X chromosome based on inheritance of a panel of six microsatellites along the fragment (homozygous in all males) and the comparison of predicted *H. contortus* gene loci with *C. elegans* proteins using a BLASTx search (Redman *et al* 2008a).

This work supports these findings. 12 of the subset of 24 *C. elegans* orthologues were on the 409Kb contig; ten of these shared most identity to *C. elegans* proteins encoded by genes on the X chromosome, and the remaining two were homologous to genes on chromosome V (see Figure 5.2). A region of conserved microsynteny was apparent between genes *hc-18h7-1, hc-18h7-2* and *hc-18h7-3* in *H. contortus* and *mec-7, zk154.1* and *zk154.4* on the X chromosome in *C. elegans*.

On contig BH4E20, 11 of the predicted polypeptides were homologous to *C. elegans* proteins encoded on chromosome I and one, HC-BH4E20-11, was homologous to a protein encoded on chromosome III (see Figure 5.3). Again, regions of microsynteny were apparent: *H. contortus* genes *hc-bh4e20-8* and *hc-bh4e20-9* had a conserved relationship relative to the orthologous *C. elegans* genes *y54e10br.1* and *arx-7;* as did *H. contortus* genes *hc-bh4e20-12, hc-bh4e20-13* and *hc-bh4e20-14* to *C. elegans* genes *dylt-1, ttx-7* and *f13g3.6*; and *H. contortus* genes *hc-bh4e20-5* and *hc-bh4e20-6* to *C. elegans* genes *ath-1 and k04g2.11*.

5.2.8 BAC end analysis

To investigate whole genome synteny between *H. contortus* and *C. elegans*, a survey of BAC ends was undertaken. The BAC end database contains 20,828 sequences of an average of 760 bp, corresponding to each end of 10,414 BAC inserts. This dataset was used for a BLASTx search of *C. elegans* Wormpep. The

locus of the best-matched *C. elegans* gene for every hit with P<0.01 was recorded, before matching up each BAC end with its mate pair and comparing putative homologue loci.

233 BAC inserts had matches to *C. elegans* proteins at both ends using these criteria. 118 pairs (50.64%) of BAC ends hit *C. elegans* genes on the same chromosome. *C. elegans* has six chromosomes, so if linkage had been random, 1 in 6 pairs (16.67%) would be expected to have hit the same chromosome. Although the BLASTx matches cannot be claimed to represent orthologous pairs, a random selection of genes would not be expected to yield a higher linkage estimate. This suggests there is a high degree of chromosomal synteny between *H. contortus* and *C. elegans* genomes.

5.2.9 Trans-splicing

5.2.9.1 Alternative splicing

Gene *hc-bh4e20-1a* encodes a full-length P-glycoprotein, with homology to *C. elegans* PGP-2. A truncated gene encoding a 794 amino acid polypeptide, with 99% amino acid identity to the C terminus of HC-BH4E20-1A, lies 4.05 Kb downstream of the 3' end of the full-length P-glycoprotein. This truncated gene, *hc-bh4e20-1b*, is thought to represent an alternate splice variant of *hc-bh4e20-1a*. Transcriptome reads mapped to sequence with homology to a mariner transposase within the 9th intron of *hc-bh4e20-1b* and seven copies of a repetitive element, HcRep (Callaghan and Beh 1994, Grillo *et al* 2006, Hoekstra *et al* 1997), were identifiable at the break point in *hc-bh4e20-1a,* which may be associated with the gene duplication event (see Figure 5.7 and Section 5.2.11).

5.2.9.2 Spliced leaders

A 22 nucleotide transcribed sequence, sharing 100% identity with the spliced leader SL1 in other species, was identified in the *H. contortus* X chromosome fragment. In *C. elegans*, the 22 nt sequence is donated by a 100 nt small-nuclear ribonucleoprotein particle (snRNP), which is encoded by 110 tandem 1 Kb repeats on chromosome V only (The *C. elegans* Sequencing Consortium, 1998, Blumenthal 2005, Nelson and Honda 1985). In the *H. contortus* X chromosome

fragment there was only one copy of SL1, which was flanked by three upstream repeats of 400 bp and three downstream repeats of 106-146 bp. A BLASTn search of the *H. contortus* supercontig database identified contigs containing up to 71 copies of the SL1 sequence at approximately 300 bp intervals, suggesting the tandem array pattern conserved in both *C. elegans* and *C. briggsae* (The *C. elegans* Sequencing Consortium, 1998, Nelson and Honda 1985, Nelson and Honda 1989) may be present elsewhere in the *H. contortus* genome (see Figure 5.4). Up to five copies of the three downstream repeats (together at 400 bp intervals), and four contiguous copies of the 400 bp repeat, were present on numerous other contigs, but none were associated with the SL1 sequence.

Trans-splicing with SL2-like sequences has previously been described in *H. contortus* (Redmond and Knox 2001, Rufener *et al* 2009). A search of the supercontig database identified only four supercontigs encoding an SL2 with 100% identity to the published sequence (accession no. AF215836), with an additional two SL2-like sequences differing from the published sequence by one nucleotide and seven SL2-like sequences differing from the published sequence by three nucleotides. The array pattern associated with the SL1 sequence was not observed; one supercontig encoded five copies of SL2-like sequences but these were not 100% identical and they were not evenly spaced (see Figure 5.5). In *C. elegans*, there are 18 SL2 genes and they are dispersed throughout the genome (Blumenthal 2005, Stein *et al* 2003).

Transcriptome reads containing SL1 and SL2 sequences were extracted from the raw Illumina files and trimmed to remove the spliced leader sequence. These trimmed reads were mapped against annotated sequence of the X chromosome fragment and BH4E20 to identify the spliced leader sequence used to trans-splice each transcript (SL RNA-seq). The results are shown in Table 5.3. No SL RNA-seq reads were available for 26 genes. Of the remainder, most transcripts were trans-spliced to SL1, with the transcripts of four genes being trans-spliced to SL2 sequences.

Two SL2-spliced genes, *hc-15g16-2* and *hc-bh4e20-3*, had no upstream genes on the same strand within >35Kb and >47Kb respectively. However, SL2-spliced *hc-bh4e20-5* is 1403bp downstream from *hc-bh4e20-6* on the same strand, and SL2-

spliced *hc-bh4e20-14* is 403bp downstream from *hc-bh4e20-13* on the same strand. The *C. elegans* orthologues of *hc-bh4e20-6* and *hc-bh4e20-5* are co-expressed in an operon as are the *C. elegans* orthologues of *hc-bh4e20-13* and *hc-bh4e20-14* (see Section 5.2.9.3 and Figures 5.6 and 5.7).
SL RNA-seq data identified alternative start codons for six genes and the same spliced leader was used for different transcripts of the same gene. Significantly fewer reads mapped to the more upstream start codon of alternately spliced gene *hc-bh4e20-15*, suggesting it is the start methionine of a rarer transcript. A similar number of reads mapped to both start codons for the other five genes.

5.2.9.3 Evidence of Operons

Around 15% of *C. elegans* genes are in operons, but their occurrence in *H. contortus* is unknown. Once gained, selection is thought to conserve operonic structure, as any subsequent loss would leave downstream genes without promoter sequences (Guiliano and Blaxter 2006). This is reflected in the conservation of 96% operons between *C. elegans* and *C. briggsae* (Blumenthal and Gleason 2003, Stein *et al* 2003). Operons are predicted to have evolved before SL2 trans-splicing (Guiliano and Blaxter 2006), so the identification of SL2-spliced genes in *H. contortus* is suggestive of the presence of operons.

Functional constraints are thought to conserve intergenic regions within operons to approximately 100 bp in *C. elegans* although increased intergenic distances have been reported in a small number of downstream genes trans-spliced with SL1 (Blumenthal and Gleason 2003, Graber *et al* 2007, Stein *et al* 2003). Recent work by Rufener *et al*, 2009 may have identified an operon encoding two genes *Hco-des-2H* and *Hco-deg-3H,* the latter being SL2 trans-spliced. In this example, the intergenic distance is reported to be ten times larger than that between the orthologous pair of genes in a *C. elegans* operon.

Studies have also identified two operons with intergenic distances of 336 and 482 bp respectively in *B. malayi* (Liu *et al* 2010), although the filarial nematode trans-splices all downstream genes in putative operons with SL1, as it lacks any SL2-like sequences, in-keeping with its clade III phylogeny (Blumenthal and Gleason 2003).

Of the 12 *C. elegans* X chromosome orthologues, only *gob-1* is in an operon (CEOPX136), perhaps reflecting the lower frequency of operons on the X chromosome (Blumenthal and Gleason 2003), and is transcribed with the downstream gene *h13n06.4*. The putative *gob-1* orthologue in *H. contortus*, *hc-13c1-1*, had no identifiable gene within the 10 Kb available downstream sequence based on BLASTx homology, RNA-seq or Genefinder predictions. A reciprocal BLAST search with the H13N06.4 polypeptide hit no sequence within the BAC contig database as expected, but did identify homologous sequence in the supercontig database (N terminus on supercontig_0011013: tBLASTn E=8.1e-27, C terminus on supercontig_0033730: tBLASTn E=2.3e-30), suggesting the orthologous gene exists elsewhere in the *H. contortus* genome. How far from the available 10 Kb downstream of *hc-13c1-1* was not determined, as no upstream sequence was available. The *h13n06.4* homologue also appears to be expressed as RNA-seq reads map to its coding sequence on both supercontigs. *hc-13c1-1* is SL1 trans-spliced.

Of the 12 *H. contortus* genes on BH4E20 that have *C. elegans* orthologues on chromosome I, six of the *C. elegans* genes are known to be in operons: *coq-4* in CEOP1124, *k04g2.11* and *ath-1* in CEOP1449 and *f13g3.6, ttx-7*, *dylt-1* in CEOP1388.

coq-4 is transcribed in an operon with upstream gene *pgk-1*, but an orthologue for the latter was not identified in the sequence studied. However, as for the example above, homologous sequence was identified in the supercontig database (N terminus on supercontig_0059510: tBLASTn E=4e30, C terminus on supercontig_0055090: tBLASTn E=4.5e-35), suggesting the orthologue exists elsewhere in the *H. contortus* genome, but no downstream gene was present in the available 600bp. Both the putative *coq-4* orthologue, *hc-bh4e20-10*, and the putative *pgk-1* orthologue identified in the supercontig database are expressed, as evidenced by a large number of RNA-seq reads mapping to their loci. A 396 bp transcript (*hc-h4320-11*) overlaps the predicted 5'end of the putative *coq-4* orthologue *hc-bh4320.10* by eight nucleotides. The hypothetical polypeptide HC-BH4E20-11 encodes a conserved zinc-finger domain and shares homology with *C. elegans* F53A3.7 (but little/none with PGK-1). The majority of zinc-finger

transcription factors, including *f53a3.7,* are not encoded in operons in *C. elegans* (Blumenthal and Gleason 2003). It is unclear if these genes are SL1 or SL2 trans-spliced in *H. contortus* as no SL-trimmed transcriptome reads mapped to their loci.

ath-1 is transcribed in the middle of a three gene operon in *C. elegans* with *k04g2.11* upstream and *cdc-48.3* downstream. An *ath-1* orthologue, *hc-bh4e20-5*, was identified in *H. contortus* on BH4E20. Although orthologous genes for neither *k04g2.11* nor *cdc-48.3* were annotated in the original draft, knowledge of this operon allowed the retrospective annotation of the *k04g2.11* orthologue, *hc-bh4e20-6*, where RNA-seq data and BLASTx homology were too weak to previously identify the gene (see Figure 5.6). The intergenic region between the parasite genes was 1538 bp. An orthologue of the downstream gene *cdc-48.3* was not identified on the sequence studied, but was on supercontig_0059449 (tBLASTn E=7.5e-84). RNA-seq data suggested it was highly expressed and no upstream gene was identified in the available 18 Kb. No SL-trimmed reads mapped to *hc-bh4e20-6*, but both SL1 and SL2 reads mapped to *hc-bh4e20-5*, which is concurrent with its putative expression as a downstream gene in an operon. The *C. elegans* orthologue of this gene, *ath-1*, is also trans-spliced by both SL1 and SL2.

f13g3.6, ttx-7 and *dylt-1* are transcribed in a four gene operon with downstream gene *f13g3.3.* (see Figure 5.7). The orthologues of *f13g3.6, ttx-7* and *dylt-1* were collinear on BH4E20, but an orthologue for *f13g3.3* was not identified. However, as shown in Figure 5.7, sequence with homology to the F13G3.3 polypeptide was present in the *H. contortus* supercontig database (supercontig_0004013 tBLASTn E=6.7e-23, supercontig_0050792 tBLASTn E=4.3e-17, supercontig_0055990 tBLASTn E=6.2e-12), although no upstream sequence was available for analysis. No RNA-seq reads mapped to the predicted coding sequence with homology to *f13g3.3* on the three supercontigs, so the gene may not be expressed in *H. contortus*. No SL-trimmed reads mapped to *hc-bh4e20-12* or *hc-bh4e20-13,* but both SL1 and SL2 reads mapped to *hc-bh4e20-14.* This is concurrent with its putative expression as a downstream gene in an operon and the *C. elegans* orthologue of this gene, *dylt-1*, is also trans-spliced by both SL1 and SL2.

In summary, there seems to be partial conservation of operons in *C. elegans* and *H. contortus*. Genes that are 'missing' from putative *H. contortus* orthologues of *C. elegans* operons can be identified elsewhere in the genome. Some of these genes appear to be expressed, even in the case of orthologues of downstream genes in *C. elegans* operons, which would be predicted to lack their own promoter sequences.

5.2.10 Mobile elements

Of the 53 annotated transcripts, 12 encoded putative transposable elements (TEs), eight of which contained transposon-associated conserved domains. Four putative polypeptides, sharing homology with the retrotransposon *rte-1* in *C. elegans*, had exonuclease endonuclease phosphatase, reverse transcriptase-like superfamily and non-long-terminal repeat retrotransposon and non-long terminal repeat retrovirus reverse transcriptase domains. Three putative polypeptides had conserved transposase-1 domains and one polypeptide had conserved pao retrotransposon peptidase and reverse transcriptase-like superfamily domains. Four transcripts shared most identity with TEs in other species but did not encode conserved domains (see Figure 5.2, Figure 5.3 and Table 5.2).

RNA-seq reads mapped to nine of the 12 TEs. Of the remaining three, TE6 and TE7 differed by only one nucleotide, so if reads were to map to these loci, they could be subject to filtering, and TE2 has no start methionine, so may represent a pseudogene. However, the extent to which any TE loci are transcribed is unclear; inherent in the process of aligning whole transcriptome reads to only a portion of the genome, is the possibility that reads from a transcribed element elsewhere in the genome could be mapped to a non-transcribed loci e.g. reads from a functional mobile element mapping to non-functional daughter progeny or to remnant sequence at ancient loci.

Interestingly, transcriptome reads mapped to mobile element sequence within introns of the downstream genes involved in both gene duplications previously described in this chapter. *hc-13c1-5* has a retrotransposon insertion in intron 6, and *hc-bh4e20-1b* has a mariner transposase insertion in intron 9 (see Figure

5.8). As mentioned, the break point in *hc-bh4e20-1a* contains seven copies of HcRep.

5.2.11 Repetitive Sequence

Tandem Repeats Finder (Benson 1999) was used to identify repetitive sequence on the X fragment and BH4E20 using the default programme parameters ('basic search'). 33 tandem repeat loci were identified on the X fragment and 24 were identified on BH4E20. On the X fragment, tandem repeats were slightly more A (30.8%) and T (30.2%) rich, but codon usage was almost equal on BH4E20. The average period size was 34.4 nucleotides and 20.4 nucleotides and the average copy number was 3.4 repeats and 4.4 repeats on the x fragment and BH4E20 respectively.

Characteristic repeat elements 'HcRep' and 'TcRep' have previously been described in the genomic sequences of *H. contortus* and *T. circumcincta* respectively (Callaghan and Beh 1994, Grillo *et al* 2006). Seven copies of HcRep are present on BH4E20, as an array in intron 15 of *hc-bh4e20-1a*. As shown in Figure 5.9, the first three copies (A-C) are most divergent, copies D-F are highly similar to HcRep, and copy G is similar to the TcRep in *T. circumcinta*. The repeat elements are associated with an upstream $(GTCT)_{14}$ tandem repeat.

Previous studies have estimated over 0.1% of the *H. contortus* genome is related to HcRep1 based on Southern blot hybridisation intensity using the SE isolate (Hoekstra *et al* 1997). A BLASTn search of the *H. contortus* supercontig database identified over 3000 matches (P<1e-05) to the published consensus HcRep (accession U86701). RNA-seq reads mapped to all seven copies of the repeat element, indicating HcRep sequences are expressed in *H. contortus*. However the reads may be mapping from expressed HcRep elements elsewhere in the genome, so expression of these particular copies on BH4E20 is not known.

The HcRep locus identified in this project is in intron 15 of *hc-bh4e20-1a* and corresponds to the exact break point relating to the upstream partial gene, hc-bh4e20-1b as described above.

5.3 Tables and Figures

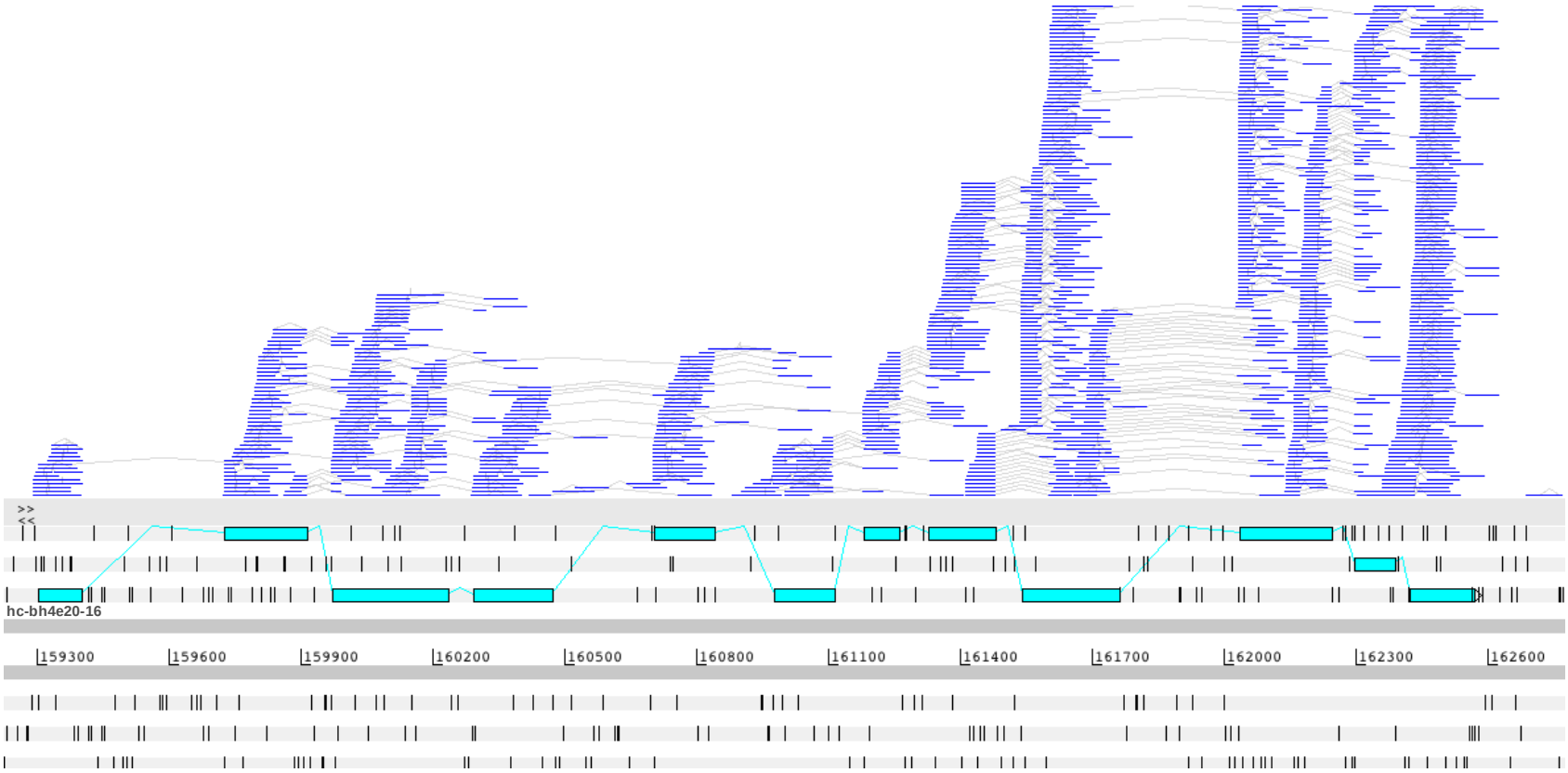


Figure 5.1: Artemis screenshot of RNA-seq reads aligned to genomic sequence to guide annotation of gene *hc-bh4e20-16*
Paired stack view is selected to assess positions of reads relative to their mate pairs (connected by grey line).

H. contortus						*C. elegans*							
GENE	UNSPLICED	SPLICED	PROTEIN	EXONS	INTRONS	GENE	UNSPLICED	SPLICED	PROTEIN	EXONS	INTRONS	CHROMOSOME	BLASTP
hc-13c1-1	4695	1437	478	12	11	*gob-1*	3408	1407	468	9	8	X	8.00E-136
hc-13c1-5	4392	1200	399	11	10	*folt-1*	1566	1233	410	5	4	V	5.00E-105
hc-7n11-3	8243	1596	531	16	15	*klc-2*	2402	1623	540	5	4	V	0
hc-15g16-2	6782	622	207	5	4	*e01g6.2*	1898	615	204	4	3	X	3.00E-41
hc-18h7-1	10663	1326	441	13	12	*mec-7*	1596	1326	441	5	4	X	0
hc-18h7-2	7731	642	213	6	5	*zk154.4*	5132	612	203	6	5	X	5.00E-17
hc-18h7-3	3341	912	303	6	5	*zk154.1*	895	630	209	5	4	X	1.00E-78
hc-18h7-4	8954	1059	352	8	7	*sms-2*	3616	1008	335	6	5	X	6.00E-149
hc-18h7-7	6811	1443	480	12	11	*arr-1*	3230	1308	435	10	9	X	8.00E-180
hc-18g2-2	9409	2667	888	17	16	*f25f6.1*	5991	2340	779	13	12	X	1.00E-135
hc-18g2-3	3556	879	292	9	8	*dhhc-1*	2218	888	295	6	5	X	1.00E-92
hc-18g2-4	5360	486	162	5	4	*m60.4*	3871	489	162	4	3	X	1.00E-59
hc-bh4e20-1a	18110	3915	1304	34	33	*pgp-2*	9012	3819	1272	14	13	I	0
hc-bh4e20-4	4766	1035	344	9	8	*gly-18*	2048	1323	440	8	7	I	7.00E-78
hc-bh4e20-5	3295	660	220	6	5	*ath-1*	1656	672	223	4	3	I	7.00E-83
hc-bh4e20-6	1371	357	118	4	3	*k04g2.11*	352	258	85	3	2	I	3.00E-08
hc-bh4e20-7	2799	846	282	9	8	*air-2*	1206	918	305	5	4	I	4.00E-87
hc-bh4e20-8	4921	2646	881	18	17	*y54e10br.1*	9937	2739	912	13	12	I	0
hc-bh4e20-9	2633	459	152	5	4	*arx-7*	645	459	152	3	2	I	8.00E-56
hc-bh4e20-10	1558	708	235	5	4	*coq-4*	844	696	231	4	3	I	1.00E-87
hc-bh4e20-11	1258	396	131	4	3	*f53a3.7*	485	360	119	3	2	III	7.00E-26
hc-bh4e20-12	6998	1173	390	10	9	*f13g3.6*	1343	1074	357	6	5	I	4.00E-92
hc-bh4e20-13	6517	858	285	8	7	*ttx-7*	1805	858	285	6	5	I	3.00E-118
hc-bh4e20-14	3107	348	115	4	3	*dylt-1*	543	321	106	3	2	I	1.00E-31
AVERAGE	5719.58	1152.92	383.46	9.83	8.83		2737.46	1124.00	373.67	6.25	5.25		
MEDIAN	4843.5	895.5	297.5	8.5	7.5		1851.5	903	300	5	4		

Table 5.1: Subset of 24 *C. elegans* and *H. contortus* putative orthologues
Sizes in nucleotides or amino acids as appropriate. Unspliced length is genomic sequence from ATG to terminal stop codon (does not include 5' or 3' UTR). Spliced length is predicted protein coding sequence only.

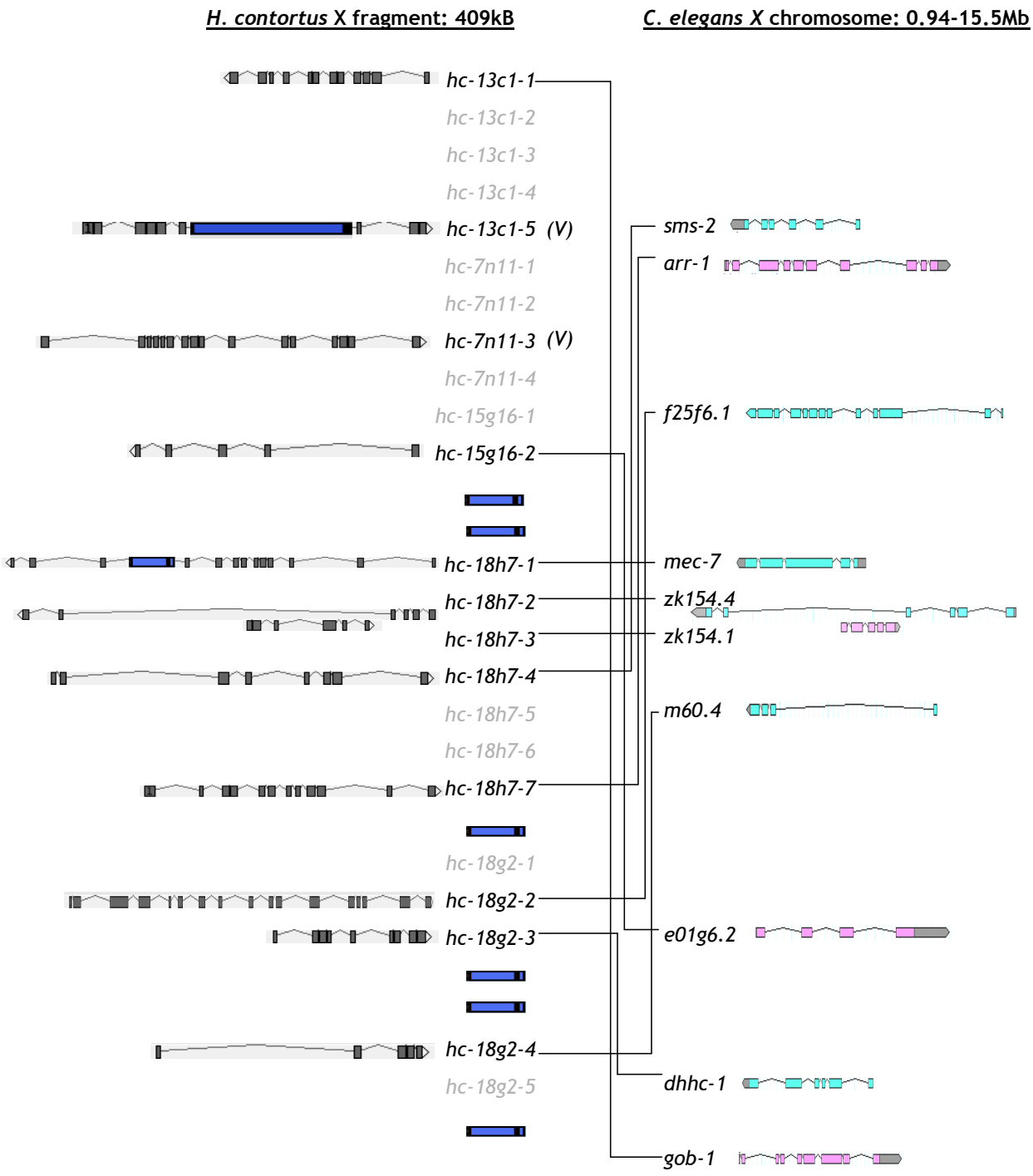


Figure 5.2: Conserved microsynteny between *H. contortus* X fragment and *C. elegans* X chromosome
Putative orthologues are in black type and transposon insertions are shown as dark blue ORFs. Colinearity is maintained in *H. contortus* orthologues of *C. elegans mec-7*, *zk154.4* and *zk154.1*. *hc-13c1-5* and *hc-7n11-3* are orthologous with genes on *C. elegans* chromosome V.

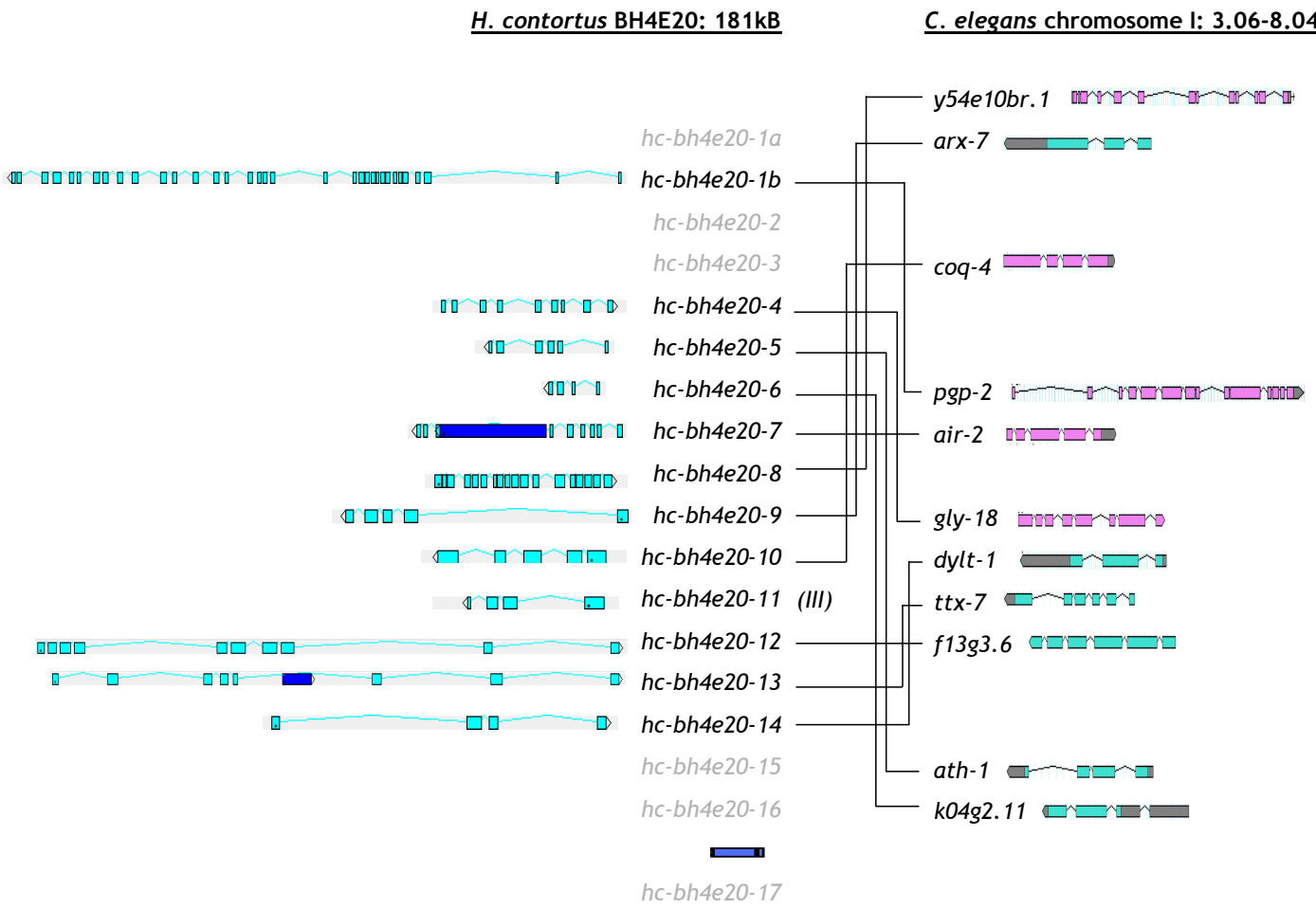
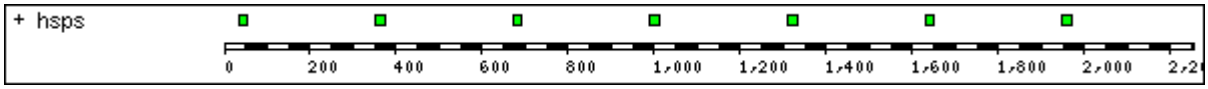


Figure 5.3: Conserved microsynteny between *H. contortus* BH4E20 and *C. elegans* chromosome I:
Putative orthologues are in black type and transposon insertions are shown as dark blue ORFs. Colinearity is maintained in *H. contortus* orthologues of *C. elegans y54e10br.1* and *arx-7*, orthologues of *C. elegans dylt-1*, *ttx-7* and *f13g3.6* and orthologues of *C. elegans ath-1* and *k04g2.11*. The last two gene sets are in operons in *C. elegans*. *hc-bh4e20-11* is orthologous with *f53a3.7* on *C. elegans* chromosome III.

Transposable element	Locus	Conserved Domains
TE1	intron 10 of *hc-18h7-1*	transposase-1
TE2	-	-
TE3	-	Pao retrotransposon peptidase and reverse transcriptase-like
TE4	within 3kB of 5' end *hc-18g2-4*	reverse transcriptase, retrotransposon, DYN1 and exo/endonuclease phosphatase
TE5	-	transposase-1
TE6	-	-
TE7	-	-
TE8	intron 6 of *hc-13c1-5*	reverse transcriptase, retrotransposon and exo/endonuclease phosphatase
TE9	intron 9 of *hc-bh4e20-1b*	transposase-1
TE10	intron 6 of *hc-bh4e20-7*	exo/endonuclease phosphatase superfamily, non-LTR retrotransposon and retrovirus reverse transcriptase
TE11	intron 5 of *hc-bh4e20-13*	-
TE12	within 4kB of 3' end *hc-bh4e20-16*	reverse transcriptase

Table 5.1: 12 transposable elements identified in 490 Kb genomic sequence
Locus describes TE insertion relative to nearest gene as shown in Figure 5.2 and Figure 5.3

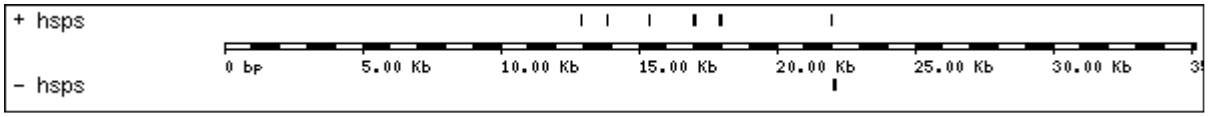
```
>Supercontig_0013605
 Length = 2257
```




```
Plus Strand HSPs:

 Score = 110 (22.6 bits), Expect = 1.8e+02, P = 1.00000
 Identities = 22/22 (100%), Positives = 22/22 (100%)

Query:     1 GGTTTAATTACCCAAGTTTGAG 22
             ||||||||||||||||||||||
Sbjct:  1948 GGTTTAATTACCCAAGTTTGAG 1969

 Score = 110 (22.6 bits), Expect = 1.8e+02, P = 1.00000
 Identities = 22/22 (100%), Positives = 22/22 (100%)

Query:     1 GGTTTAATTACCCAAGTTTGAG 22
             ||||||||||||||||||||||
Sbjct:  1629 GGTTTAATTACCCAAGTTTGAG 1650

 Score = 110 (22.6 bits), Expect = 1.8e+02, P = 1.00000
 Identities = 22/22 (100%), Positives = 22/22 (100%)

Query:     1 GGTTTAATTACCCAAGTTTGAG 22
             ||||||||||||||||||||||
Sbjct:  1310 GGTTTAATTACCCAAGTTTGAG 1331

 Score = 110 (22.6 bits), Expect = 1.8e+02, P = 1.00000
 Identities = 22/22 (100%), Positives = 22/22 (100%)

Query:     1 GGTTTAATTACCCAAGTTTGAG 22
             ||||||||||||||||||||||
Sbjct:   991 GGTTTAATTACCCAAGTTTGAG 1012

 Score = 110 (22.6 bits), Expect = 1.8e+02, P = 1.00000
 Identities = 22/22 (100%), Positives = 22/22 (100%)

Query:     1 GGTTTAATTACCCAAGTTTGAG 22
             ||||||||||||||||||||||
Sbjct:   672 GGTTTAATTACCCAAGTTTGAG 693

 Score = 110 (22.6 bits), Expect = 1.8e+02, P = 1.00000
 Identities = 22/22 (100%), Positives = 22/22 (100%)

Query:     1 GGTTTAATTACCCAAGTTTGAG 22
             ||||||||||||||||||||||
Sbjct:   353 GGTTTAATTACCCAAGTTTGAG 374

 Score = 110 (22.6 bits), Expect = 1.8e+02, P = 1.00000
 Identities = 22/22 (100%), Positives = 22/22 (100%)

Query:     1 GGTTTAATTACCCAAGTTTGAG 22
             ||||||||||||||||||||||
Sbjct:    34 GGTTTAATTACCCAAGTTTGAG 55
```

Figure 5.4: Typical example of a BLASTn search of the supercontig database with published *H. contortus* SL1 nucleotide sequence. This supercontig encodes an array of seven copies of identical SL1 sequences with ~300 bp separating each one. 144 supercontigs were identified as encoding at least one copy of SL1 with 100% nucleotide identity to the published sequence (accession no. Z69630).

```
>Supercontig_0057982
 Length = 35,136
```




```
Plus Strand HSPs:

 Score = 110 (22.6 bits), Expect = 1.9e+02, P = 1.00000
 Identities = 22/22 (100%), Positives = 22/22 (100%)

Query:     1 GGTTTTAACCCAGTATCTCAAG 22
             ||||||||||||||||||||||
Sbjct: 12944 GGTTTTAACCCAGTATCTCAAG 12965

 Score = 101 (21.2 bits), Expect = 4.9e+02, P = 1.00000
 Identities = 21/22 (95%), Positives = 21/22 (95%)

Query:     1 GGTTTTAACCCAGTATCTCAAG 22
             |||||||||||||||| ||||
Sbjct: 17950 GGTTTTAACCCAGTATCACAAG 17971

 Score = 83 (18.5 bits), Expect = 3.1e+03, P = 1.00000
 Identities = 19/22 (86%), Positives = 19/22 (86%)

Query:     1 GGTTTTAACCCAGTATCTCAAG 22
             |||| ||||||||||||  |||
Sbjct: 17019 GGTTATAACCCAGTATCAAAAG 17040

 Score = 83 (18.5 bits), Expect = 3.1e+03, P = 1.00000
 Identities = 19/22 (86%), Positives = 19/22 (86%)

Query:     1 GGTTTTAACCCAGTATCTCAAG 22
             |||| ||||||||||||  |||
Sbjct: 15375 GGTTATAACCCAGTATCAAAAG 15396

 Score = 83 (18.5 bits), Expect = 3.1e+03, P = 1.00000
 Identities = 19/22 (86%), Positives = 19/22 (86%)

Query:     1 GGTTTTAACCCAGTATCTCAAG 22
             |||| ||||||||||||  |||
Sbjct: 13838 GGTTATAACCCAGTATCAAAAG 13859

 Score = 60 (15.1 bits), Expect = 2.2e+04, P = 1.00000
 Identities = 12/12 (100%), Positives = 12/12 (100%)

Query:     7 AACCCAGTATCT 18
             ||||||||||||
Sbjct: 21988 AACCCAGTATCT 21999

Minus Strand HSPs:

 Score = 110 (22.6 bits), Expect = 1.9e+02, P = 1.00000
 Identities = 22/22 (100%), Positives = 22/22 (100%)

Query:    22 CTTGAGATACTGGGTTAAAACC 1
             ||||||||||||||||||||||
Sbjct: 22086 CTTGAGATACTGGGTTAAAACC 22107
```

Figure 5.5: BLASTn search of the supercontig database with a published *H. contortus* SL2 nucleotide sequence. Unlike the SL1 BLASTn result (see Figure 5.4) a number of non-identical SL2-like sequences were identified, and only one supercontig encoded multiple copies of SL2 (shown above). The copies encoded on this supercontig are not identical and the distance between them varies. Published SL2 sequence was accession no. AF215836.

GENE	SPLICED LEADER	ALTERNATE START CODON	INTERGENIC DISTANCE IF SL2 TRANS-SPLICED
hc-13c1-1	SL1	first codon exon 2 (SL1)	
hc-13c1-4			
hc-13c1-5			
hc-18g2-2			
hc-18g2-3	SL1	mid exon 2 (SL1)	
hc-18g2-4	SL1		
hc-18h7-1			
hc-18h7-2			
hc-18h7-3			
hc-18h7-4			
hc-18h7-7	SL1	mid exon 2 (SL1)	
hc-7n11-3			
hc-18h7-6			
hc-13c1-2			
hc-13c1-3	SL1		
hc-15g16-1	SL1		
hc-15g16-2	SL2		no upstream gene for >35Kb
hc-18h7-5			
hc-7n11-1			
hc-7n11-2	SL1		
hc-7n11-4			
hc-18g2-1			
hc-18g2-5			
hc-bh4e20-1a			
hc-bh4e20-2			
hc-bh4e20-3	SL2		all upstream genes on opposite strand for >47Kb
hc-bh4e20-4			
hc-bh4e20-5	SL1/SL2	first codon exon 2 (SL1/SL2)	1403bp downstream from *hc-bh4e20-6* on same strand
hc-bh4e20-6			
hc-bh4e20-7			
hc-bh4e20-8	SL1		
hc-bh4e20-9			
hc-bh4e20-10			
hc-bh4e20-11			
hc-bh4e20-12			
hc-bh4e20-13			
hc-bh4e20-14	SL1/SL2		403bp downstream from *hc-bh4e20-13* on same strand
hc-bh4e20-15	SL1	mid exon 2 (SL1)	
hc-bh4e20-16	SL1		
hc-bh4e20-17			

Table 5.3: Trans-splicing
The spliced leader trans-spliced to each transcript was identified using SL-trimmed RNA-seq data where available. Alternative start codons were identified for six transcripts and the same spliced leader was used for alternate transcripts of the same gene. Intergenic distance of the nearest upstream gene is shown for the four sequences trans-spliced to SL2.

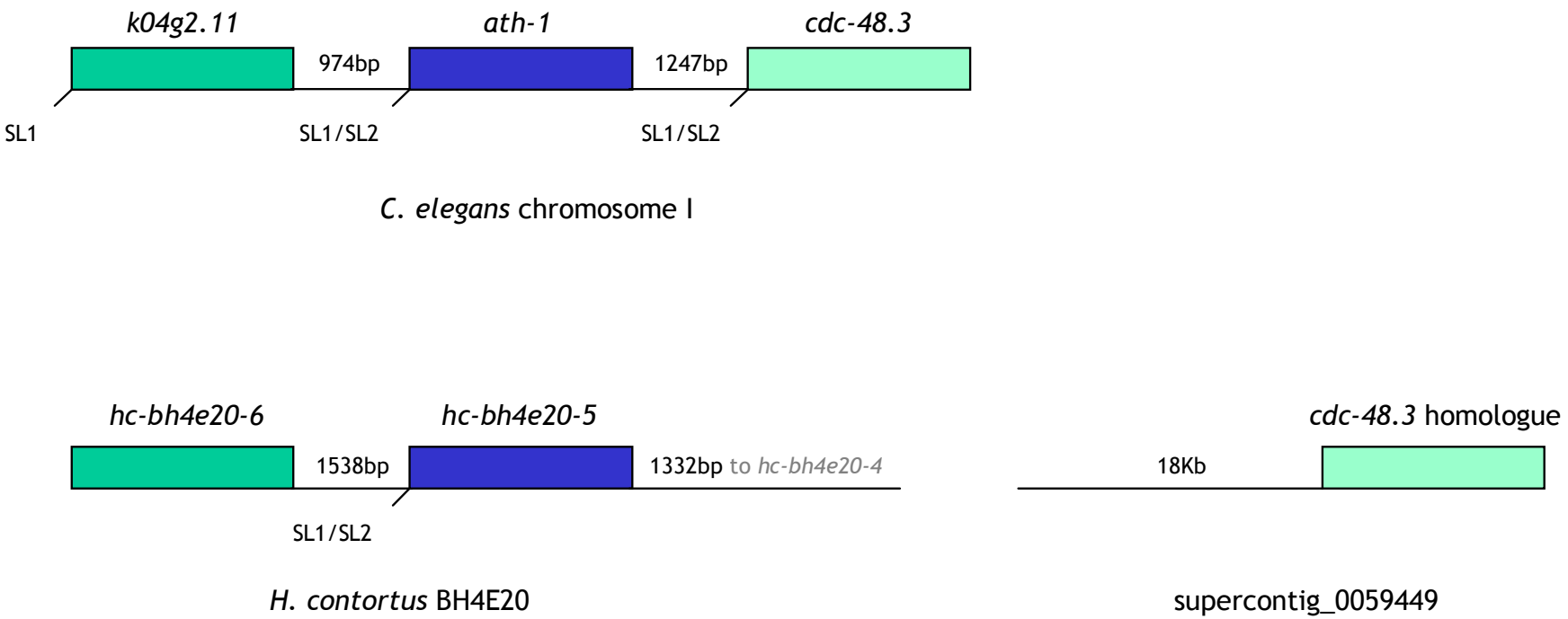


Figure 5.6: *C. elegans* operon CEOP1449 and orthologues on *H. contortus* BH4E20 of two of the three constituent genes.
An orthologue of the third gene, *cdc-48.3*, is not encoded with the two upstream genes on BH4E20 but appears to be present elsewhere in the genome: homologous sequence can be identified on supercontig_0059449 (top hit from tBLASTn search of supercontig database with CDC-48.3 polypeptide; reciprocal BLASTx search with supercontig_0059449 into Wormpep confirms best hit CDC-48.3). No coding sequence was identified in the 18 Kb upstream sequence of the *H. contortus cdc-48.3* homologue. Intergenic distance was 974 bp for *k04g2.11* and *ath-1* in *C. elegans* and 1538 bp for the orthologous genes in *H. contortus*. The next *H. contortus* gene on BH4E20, *hc-bh4e20-4*, is only 1332 bp downstream but is encoded on the opposite strand. SL-trimmed RNA-seq data was only available for *hc-bh4e20-5* but showed the transcript was trans-spliced to both SL1 and SL2 spliced leader sequences.

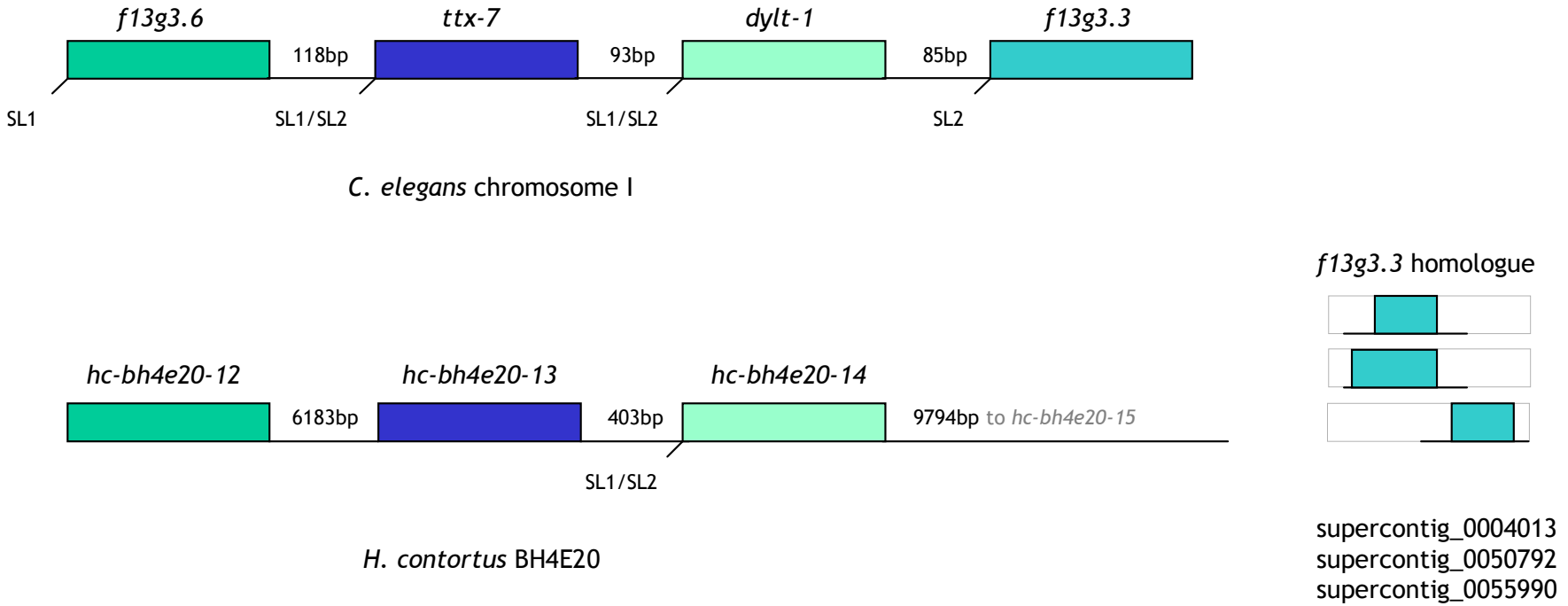


Figure 5.7: *C. elegans* operon CEOP1388 and orthologues on *H. contortus* BH4E20 of three of the four constituent genes
An orthologue of the fourth gene, *f13g3.3*, is not encoded with the three upstream genes on BH4E20 but appears to be present elsewhere in the genome: homologous sequence can be identified on three supercontigs (top hits from tBLASTn search of supercontig database with F13G3.3 polypeptide; reciprocal BLASTx searches with same three supercontigs into Wormpep confirm best hit F13G3.3). No upstream sequence of the *H. contortus f13g3.3* homologue is available. Intergenic distance ranges from 85-118 bp for the four *C. elegans* genes in CEOP1388 and from 403-6183 bp for the three orthologous genes in *H. contortus*. The next *H. contortus* gene on BH4E20, *hc-bh4e20-15*, is 9794 bp downstream and is encoded on the reverse strand. SL-trimmed RNA-seq data was only available for *hc-bh4e20-14* but showed the transcript was trans-spliced to both SL1 and SL2 spliced leader sequences.

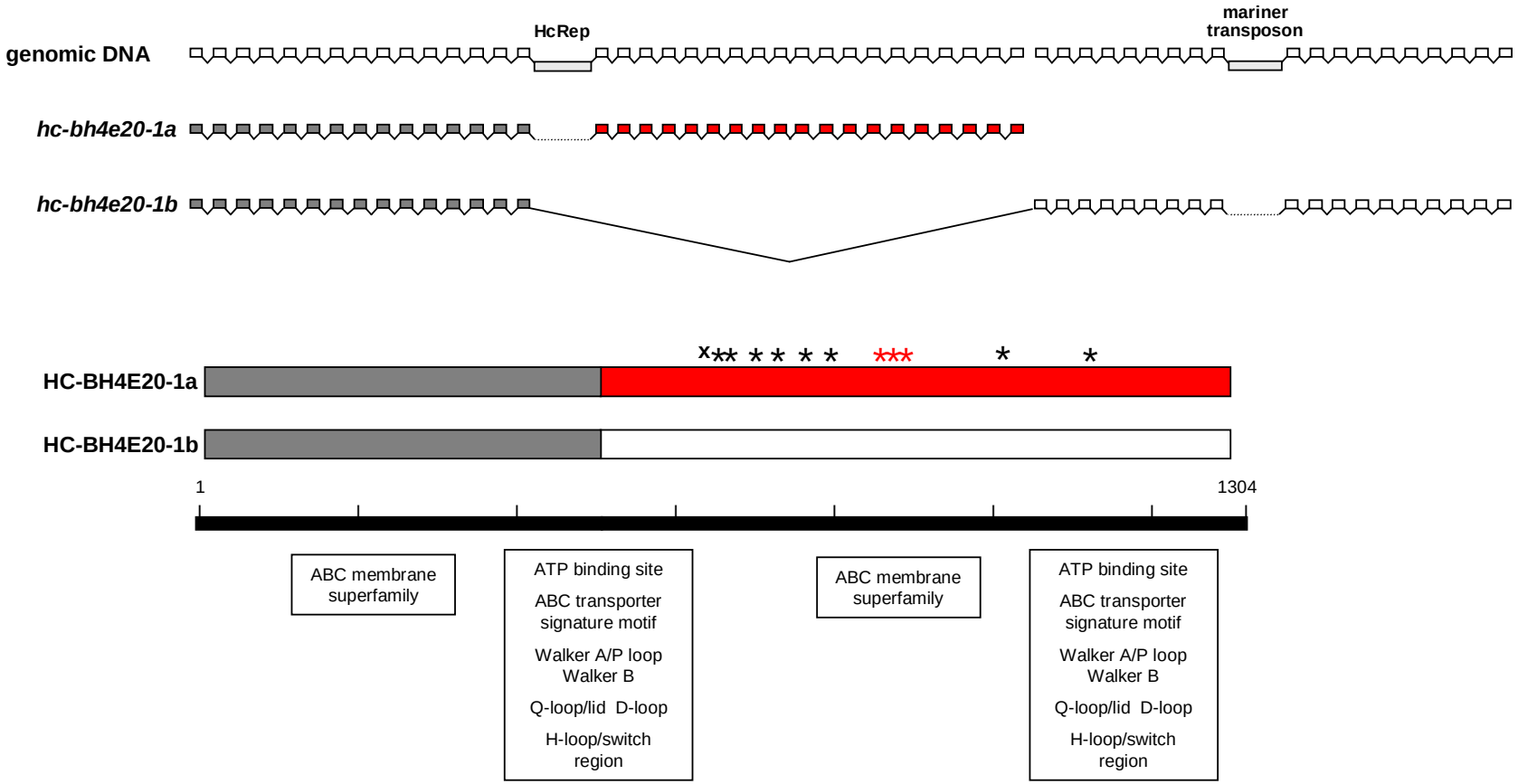


Figure 5.8: Alternate splicing of gene *hc-bh4e20-1*, a putative p-glycoprotein.

The alternate 3' ends of gene *hc-bh4e20-1* share 97% nucleotide identity, and the C-termini of the predicted polypeptides share 99% amino acid identity. Seven of the nine amino acid substitutions (each marked *) lie outside conserved domains and all but one (x) are synonymous. HC-BH4E20-1b has a three amino acid deletion (marked*)**

The break-point in *hc-bh4e20-1a* is associated with seven copies of the repetitive element HcRep and sequence sharing identity with a mariner transposon is present within intron 8 of *hc-bh4e20-1b*.

Not to scale.

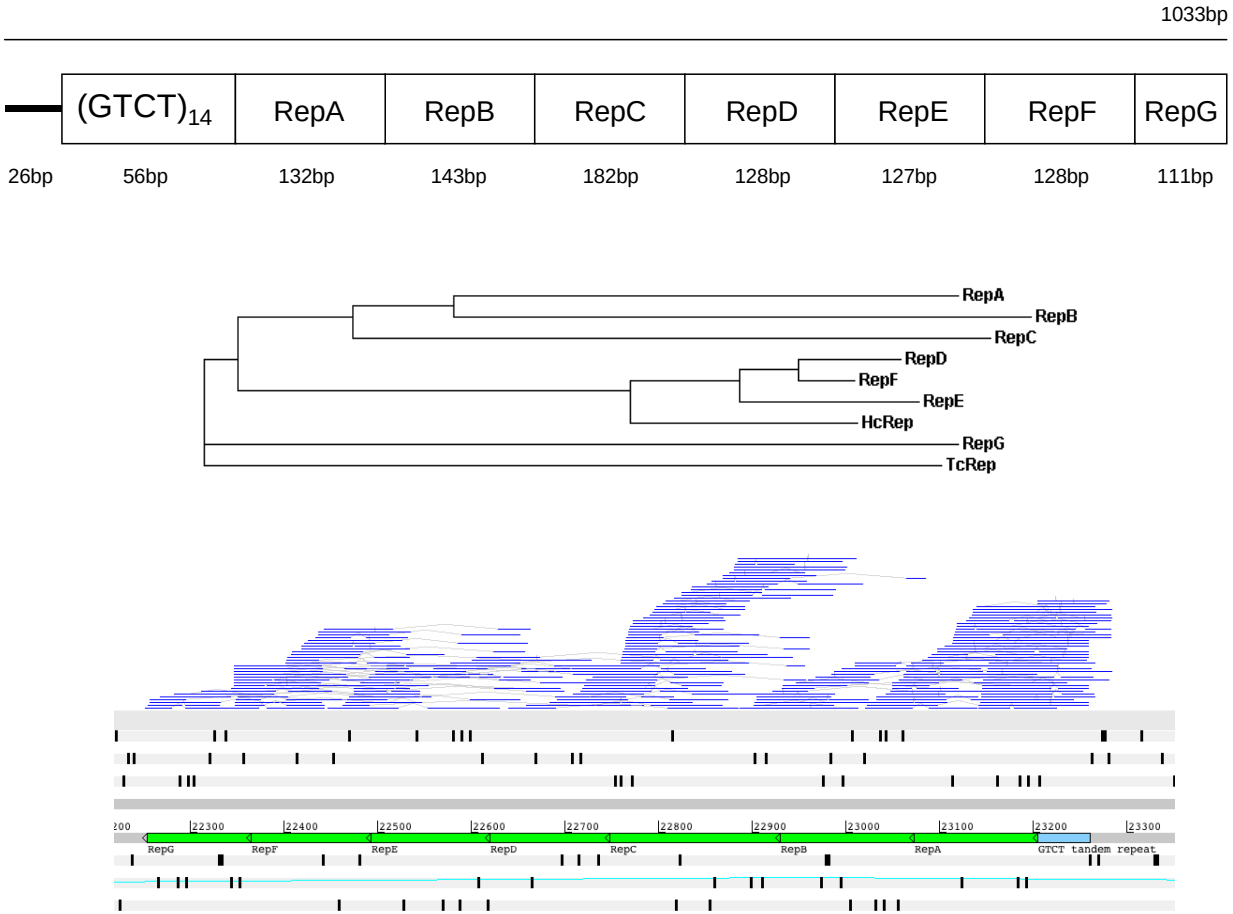


Figure 5.9: Copies of HcRep sequence identified on BAC BH4E20
Seven copies of the repeat element were identified on BH4E20 and RNA-seq suggests they are expressed. A ClustalW guide tree based on multiple sequence nucleotide alignment shows adjacent HcRep copies are most similar to each other. RepD, RepE and RepF share most homology with the published *H. contortus* HcRep consensus sequence (accession no. U86701), while RepG is more similar to *T. circumcincta* repeat element TcRep (accession no. M84610).

	Best BLASTx hit in *C. elegans*			Second best BLASTx hit in *C. elegans*		
***H. contortus* BAC insert**	Gene	Chromosome	Expect value	Gene	Chromosome	Expect value
haembac8e5	*avr-1*	I	3e-46	*gbr-2b*	I	1e-28
	f39b2.3	I	1e-20	-	-	-
	TE			*TE*		
	TE			*TE*		
	dnj-12	I	8e-20	*dnj-19*	V	7e-13
haembac18a19	*trxr-1*	IV	1e-15	*trxr-2*	III	6e-14
	tpk-1	III	1e-17	-	-	-
	unc-32	III	3e-29	*vha-6*	II	9e-22
	TE		5e-57	*TE*		1e-50
	lin-9	III	4e-31	-	-	-
	cyp43a1	X	5e-31	*cyp13a1*	III	1e-20
bha11l19	*TE*		4e-61	*TE*		7e-47
	zc132.4	V	6e-27	-	-	-
	TE		1e-63	*TE*		1e-63
	vap-1	X	9e-25	*t05a10*	X	1e-18
	c47f8.2	X	1e-07	-	-	-
	mek-2	I	2e-06	*jkk-1*	X	3e-05
	TE		5e-23	*TE*		7e-23
	TE		1e-19	*TE*		2e-19
	ztf-2	I	5e-26	*syd-9*	X	1e-04
	col-89	III	7e-14	*col-10*	V	2e-13
	TE		2e-05	*TE*		2e-05

Table 5.4: BLAST survey of three additional *H. contortus* BAC insert sequences
Each sequence was divided into 50 Kb sections, which were used for a BLASTx search of *C. elegans* Wormpep. The best and second best hit for each locus was recorded. TEs represent transposable elements.

5.4 Discussion

5.4.1 Genomic sequence comparison of *H. contortus* and *C. elegans*

5.4.1.1 *H. contortus* gene structure and implications for genome size

The gene density in the annotated *H. contortus* sequence was 2-3 times lower than that in *C. elegans*. It is possible that this is an overestimate of the difference as there could be more parasite genes to be identified; genes with stage-specific (embryo, L1 or L4) or low expression would not be detected with RNA-seq and genes that are *H. contortus*-specific or very divergent from other species might not be detected with sequence homology.

Further, the density of genes on the *C. elegans* X chromosome is lower than that on the autosomes (The *C. elegans* Sequencing Consortium, 1998). Difference in gene density between the X fragment and BH4E20 suggests this might also be the case in the parasite, which would again make the estimated density in this study an underestimate for the genome.

The main reason for the lower gene density in *H. contortus* was gene size. The average and median gene sizes were more than 2-fold greater in the parasite due to a larger intron number and length. The true difference in gene size may be greater yet, as sequences with homology to TEs were subtracted from calculations of parasite intron size, but these may represent ancient remnants, so arguably could be incorporated into the intron size statistics. The benefit of this intronic expansion is unclear. It would be expected to reduce the efficiency of transcription (Castillo-Davis *et al* 2002) but perhaps this cost is offset by an advantage conferred by functional elements encoded in the introns or by facilitating TE insertion to promote gene evolution.

5.4.1.2 Homology

All putative protein-coding genes other than *hc-13c1-2* had putative homologues (BLASTp, E>1e-10) in *C. elegans*. This represents 97.6% gene conservation, a significantly higher percentage than the ~65% estimated by Parkinson *et al* (2004). The lowest BLASTp score of E=3e-11 recorded in this study corresponded to a BLASTx score of E=8e-11, while Parkinson *et al* (2004) used a BLASTx cut-off of E=10e-5 to signify homology. The higher figure for gene conservation in this study is therefore not explained by a lower stringency level for the identification of homologues. However, the use of RNA-seq may have biased this project's results for more conserved genes, as constitutively expressed housekeeping genes are more likely to be highly expressed and easily detected with transcriptome data. The other possibility is that homology is more likely to be detected for a survey of full-length genes than for ESTs, as a number of the latter will represent only the divergent regions of otherwise homologous genes.

5.4.1.3 Synteny

Studies comparing the *C. elegans* genome with *C. briggsae* and *B. malayi* genomes have shown high rates of rearrangement (Coghlan and Wolfe 2002, Ghedin *et al* 2007, Guiliano *et al* 2002, Hillier *et al* 2007, Stein *et al* 2003). The rearrangement rate of *C. briggsae* has been estimated at 0.4-1 chromosomal breakages per Mb per million years (Coghlan and Wolfe 2002), which is at least four times that of *D. melanogaster* (Ranz *et al* 2001).

Intra-chromosomal rearrangements are more common than inter-chromosomal rearrangements (Coghlan and Wolfe 2002, Ghedin *et al* 2007, Guiliano *et al* 2002, Hillier *et al* 2007, Stein *et al* 2003). This has been suggested to occur because intra-chromosomal rearrangements require fewer DNA breaks and because the conformation of the nuclear scaffold may maintain the association of local regions (Guiliano *et al* 2002). The results of this study suggest the same may be true in *H. contortus*: overall synteny was not conserved, yet ten of 12 putative orthologues on the X fragment were on the *C. elegans* X chromosome and 11 of 12 putative orthologues on BH4E20 were on *C. elegans* chromosome I.

Interestingly, the X chromosome appears to have a lower rate of structural evolution than the autosomes in *C. elegans*. In a comparison with *C. briggsae*, the X chromosome contains two of the three largest conserved segments and has undergone around half as many rearrangements as the autosomes since the worms diverged (Stein *et al* 2003). It has been suggested that its involvement in sex determination selects against structural rearrangements (Coghlan *et al* 2008) as in *B. malayi*, which has a Y chromosome, the X chromosome appears to have undergone a higher rate of inter-chromosomal rearrangement than the autosomes (Whitton *et al* 2004).

Although this pilot annotation covered only 590 Kb of sequence, the emerging pattern of selection for intra-chromosomal rearrangements is expected to be a true reflection of the whole genome. 50% of *H. contortus* BAC ends had matches on the same *C. elegans* chromosomes, which is significantly higher than the value expected had inter-chromosomal rearrangements been as likely as intra-chromosomal rearrangements. It is lower, however, than might be expected from the 590 Kb annotation. In a similar survey of *B. malayi* BAC ends, 60% had matches on the same *C. elegans* chromosomes, yet there was very little conserved microsynteny (Whitton *et al* 2004). This suggests either a greater degree of conservation in the *H. contortus* X chromosome and chromosome I sequences annotated, or that the failure to ensure homologues identified in the BAC end analysis were true orthologous pairs has led to the incorrect association of a number of genes, artificially lowering the predicted level of chromosomal synteny. A combination of both these factors is perhaps most likely.

This was further examined with survey of an additional three randomly selected BAC insert sequences. Redman *et al* (2008) had previously shown the *H. contortus* 409Kb contig was syntenic with the *C. elegans* X chromosome by performing a BLASTx search of *C. elegans* Wormpep using 50 Kb sections of the contig. A BLAST hit with an expect value of ≥1e-08, where the next best hit had an e value of >0.01, was considered to represent an orthologue. When this method was applied to the three BAC sequences however, none of the second best BLAST hits had an e value of >0.01, with the majority of loci yielding multiple hits with e values of ≥ 1e-08 (see Table 5.4). These findings are, again,

suggestive of a more highly conserved X chromosome and highlight the dangers of inferring orthology from BLAST scores alone.

In this 590 Kb pilot annotation, regions of conserved microsynteny between *H. contortus* and *C. elegans* were apparent. Four regions with collinear genes were identified, two of which contained orthologues of genes in operons in *C. elegans*.

Once formed, operons are thought to be difficult to break, as downstream genes would be left without promoters (Blumenthal and Gleason 2003). However, none of the *H. contortus* regions sharing synteny with *C. elegans* operons encoded the full complement of genes. Putative orthologues for all 'missing' genes could be identified elsewhere in the genome and interestingly, RNA-seq reads mapped to all but one of these genes, suggesting they are expressed.

96% of *C. elegans* operons are conserved in *C. briggsae* (Stein *et al* 2003) but it has been shown that the 4% of operons that are not conserved are a result of not only operon gains in *C. elegans* but also losses in *C. briggsae* (Qian and Zhang 2008). The same study identified a four-gene operon in *C. elegans* in which the first three genes were translocated to a different chromosome in *C. briggsae*. Despite this, all four genes were expressed in *C. briggsae* and the authors suggested the separated downstream gene had formed an operon with its new upstream gene. A similar mechanism may have facilitated expression of *hc-bh4e20-10* in *H. contortus*. This is a putative orthologue of *C. elegans coq-4*, a gene expressed downstream of *pgk-1* in two-gene operon CEOP1124. Despite *hc-bh4e20-10* lacking an upstream orthologue of *pgk-1,* RNA-seq data shows it is expressed. However, a different upstream gene, *hc-bh4e20-11*, overlaps the predicted 5' end of *hc-bh4e20-10* by eight nucleotides. *hc-bh4e20-11* shares homology with *C. elegans f53a3.7,* which is not in an operon. It is possible that these genes have formed a new operon in *H. contortus*. No SL-trimmed RNA-seq reads mapped to this gene, so it is unclear if it is SL2 trans-spliced.

5.4.1.4 Mobile elements and repetitive DNA

Approximately 12% of the *C. elegans* genome is comprised of TEs, although most of these are thought to be no longer mobile (The *C. elegans* Sequencing Consortium, 1998). The 12 *H. contortus* TEs identified in this study represent

just over 3% of the 590 Kb genomic sequence analysed but the significant number of transcriptome reads mapping to TE loci suggests these elements may be highly active within the *H. contortus* genome.

Along with mutations generated by DNA-polymerase errors, TE insertions are the main internal drivers of genetic change, and the association of mobile elements with chromosome rearrangements in *Caenorhabditis* and *Drosophila* are well known (Bessereau 2006, Caceres *et al* 1999, Duret *et al* 2000). An association of repetitive sequence with synteny break points in *Caenorhabditis* has also been identified (Coghlan and Wolfe 2002, Stein *et al* 2003). This may be an indirect association, if repetitive sequences represent derivatives of TEs or are generated by TE insertions, or a direct association if repetitive sequences induce ectopic recombination between repeats (Coghlan and Wolfe 2002).

Both putative gene duplications identified in this study were associated with intronic transposon insertions, and one was also associated with the repeat element HcRep. Studies have suggested that TEs might be enriched within or flanking environmental response genes, such as the cytochrome P450 family (Chen and Li 2007). Genomic plasticity would be predicted to be advantageous at these loci, as it would facilitate adaptive response to changes in the environment. Consistent with this hypothesis, a gene duplication associated with an intronic mariner transposon insertion and the repeat element HcRep was identified at a P-glycoprotein locus in this study. A TE insertion was observed in intron 5 of *hc-bh4e20-13*, which may also be an environmental response gene: the putative orthologue in *C. elegans, ttx-7,* encodes a myo-inositol monophosphatase required for normal thermotaxis and chemotaxis to sodium (Tanizawa *et al* 2006). However, a retrotransposon-associated gene duplication was also identified involving the orthologue of *C. elegans folt-1*, a folate transporter, and intronic TE insertions were identified in two *H. contortus* genes more likely to have constitutive functions: orthologues of the *C. elegans mec-7* (a β-tubulin) and *air-2* (a serine/threonine protein kinase).

Although the number of TE insertions might be expected to be lower in the X chromosome because recessive deleterious insertions would be selected against since males are XO, Duret *et al* (2000) found a slightly higher overall TE density

in the X chromosome. However, within this, certain families were over-represented and certain families appeared to be selected against. In contrast, in this study, the X fragment did have a slighter lower frequency of one TE per 51Kb, relative to one TE per 45Kb in BH4E20, but it was unclear if there was selection bias on any specific TE family in this small subset.

5.4.2 Implications for the genome project

Nematode genomes are evolving rapidly. Comparisons of *C. elegans* and *C. briggsae* genomes relative to mouse and human genomes (the former pair diverged ~80-110 million years ago, the latter ~75 million years ago), showed the nematodes have fewer 1:1 orthologue pairs, more genes lacking matches in both species, a nearly 3-fold higher nucleotide substitution rate and a dramatically higher chromosomal rearrangement rate (Stein *et al* 2003). The same authors highlighted the incongruity between the almost indistinguishable morphological differences between *C. elegans* and *C. briggsae* and the difference in physical appearance between the more genetically similar mouse and human.

C. elegans and *H. contortus* are estimated to have diverged 400 million years ago (Vanfleteren *et al,* 1994) but the degree of conservation between the two genomes would be expected to underpin the success or otherwise of the *H. contortus* genome project. The impact of the *H. contortus* adaptation to parasitism on the (potentially highly plastic) nematode genome is unknown, but it was predicted that the retention of a conserved genetic core relating to essential biological processes would facilitate a comparative bioinformatic approach for gene discovery and annotation (Gilleard 2004).

In this study, gene discovery and annotation based on RNA-seq data and sequence homology with *C. elegans* have been found to correspond closely, and a number of putative orthologues have been identified. It is unclear if positional information of genes in *C. elegans* will be of use for finding orthologues in the parasite due to intra-chromosomal rearrangements, but homologous genes appear to reside on the same chromosomes frequently, and where present, local regions of microsynteny can be used as supporting evidence of orthologous relationships.

The generation of databases of *H. contortus* transcriptome reads for different isolates and life stages has provided a valuable resource. They can be aligned to genomic sequence to guide annotation of the genome and have proven especially useful for mapping splice variants and divergent sequence that escapes identification by comparative approaches. RNA-seq can also be used for the identification of transcribed regulatory sequences and non-coding RNAs.

The technology and software for RNA-seq is developing rapidly. Over the course of this project alone, Illumina sequencing lengths have increased from 35bp to 100bp, and improved aligners are faster, allow more reads to be mapped and clip ends for quality to better demarcate intron: exon boundaries. In addition, it is simple to format aligner output files to open directly into Artemis, allowing reads to be filtered for quality, mate-pairs to be traced, and alignments to be viewed at an individual read or individual base level of clarity.

de novo assembly of the *H. contortus* transcriptome is also underway. This has the benefit that it does not rely on a reference genome, so can be undertaken independently until the genome is assembled. It is hoped, through comparison with the *C. elegans* transcriptome, that this will this aid research into parasite drug resistance by facilitating identification and analysis of known drug resistance genes in other species and parasite-specific drug targets.

This work also indicates that the previous prediction of the genome size of *H. contortus* at 53 Mb may be an underestimate (Leroy *et al* 2003). The *C. elegans* genome is 100 Mb and if the parasite genome has a similar gene complement, a gene density and size differing from the model worm by a 2-3 fold magnitude would be suggestive of a ~200-300 Mb genome. This has obvious repercussions for the genome project, but may in part explain the current difficulties with assembly. If genome size does prove to be significantly larger, the availability of NGS should make the problem eminently solvable. In addition, the increased depth of sequencing facilitated by this technology would also be expected to resolve ongoing issues of polymorphism.

6 General Discussion

This project had initially focused on the partial sequences of four CYP genes, shown by phylogenetic analysis to be most similar to the multi-drug resistance gene *cyp6g1* in insects and the multi-drug/anthelmintic-metabolising gene *cyp3a4* in humans (data not shown). However, during the course of this study, the genomes of other nematodes (*C. briggsae, P. pacificus, M. hapla, M. incognita, B. malayi*) and trematodes (*S. mansoni*) were completed in addition to a number of animal and insect genomes, allowing more extensive comparisons of the CYP family both within and between phyla. From this, it became apparent that the superfamily was evolving so rapidly that orthologues between even closely related species might be difficult to identify. Further, progress with the *H. contortus* genome project revealed a CYP family that grew incrementally larger with each new assembly, so a wider and more inclusive approach was deemed possible. Consequently, a semi-quantitative PCR screen of all identified CYPs was developed but due to a lack of sensitivity was succeeded by a real-time PCR screen. The *H. contortus* transcriptome was also sequenced to facilitate comparison of gene expression on a global scale and to guide genome annotation.

Attempts to assemble the *H. contortus* CYP family in this project were only partially successful, but it was possible to provide an overview of the gene family. Reassuringly, many of the challenges faced were common to the assembly of the CYP family in incomplete genomes of other species e.g. clustering, recent duplications, pseudogenes and detritus exons arranged in tandem. In insect genomes, the high identity of adjacent CYPs has been known to result in problems with assembly resulting in gaps at these loci (Feyereisen 2010) and this may have been a confounding factor in the struggle with the *H. contortus* genome assembly. An additional consideration for the CYP family in all species is the occurrence of within-species polymorphism in CYP gene copy number i.e. the number of CYP genes may vary between individuals (Feyereisen 2010). This is seen in a comparison of the two reference human genomes, where, for example, Craig Venter has two copies of *cyp1a2*. Gene copy polymorphism might be especially likely in an innately polymorphic species such

as *H. contortus* and if present will further complicate the assembly of the reference genome from a population of worms.

Due to the common challenges of assembling the CYP family, databases of large numbers of partial CYP sequences are accumulating for many species. Fogleman and Danielson (2000) developed algorithms to apply to partial sequences of human and insect CYPs in an attempt to infer the phylogeny of the full-length genes. Although this was successful to a degree, the optimum regions to use for analysis varied between CYP families, so this technique is likely to be more applicable to degenerate primer PCR generated fragments (amplifying the same region) than the bioinformatically identified fragments in this project. Other workers maintain that attempts to classify and name CYPs based on partial sequences remain ill advised (Feyereisen 2010).

In spite of the ongoing difficulties in assembling the *H. contortus* CYP family, the work outlined in Chapters 3 and 4 developed an assay to examine the expression of a large number of CYPs and revealed differences in constitutive expression between life stages, sexes and tissues. As shown in Chapter 4, the characteristic expression patterns of *H. contortus* CYP tags allowed a number to be assembled as single genes. It was also apparent that the expression patterns observed in *H. contortus* CYPs may relate closely to the expression patterns of their putative orthologues in *C. elegans*. Conserved CYPs are likely to fit the 'phylogenetically stable' gene classification, residing in small families and with roles in essential pathways.

Results regarding the putative role of CYPs in the resistance of *H. contortus* to anthelmintic treatment remain unclear, but the high intestinal expression would be supportive of a role in xenobiotic metabolism. In contrast to the findings of Kotze (2000), where no increase in CYP activity was detected in a resistant isolate, this project found that a small number of CYPs appear to more highly expressed in anthelmintic resistant isolates. The functional relevance of this up-regulation has not been tested. The inclusion of the CYP inhibitor piperonyl butoxide might be expected to increase anthelmintic efficacy in resistant isolates if CYP up-regulation is a mode of resistance. However, previous studies of this nature in different species have yielded mixed results. In TBZ-resistant *Fasciola hepatica*, piperonyl butoxide has been shown to decrease TBZ

resistance in TBZ-resistant isolates but to have no effect on the efficacy of the drug in susceptible isolates (Devine *et al* 2010). Conversely, Sargison *et al* (2010) found that inhibition of CYP activity with piperonyl butoxide increased IVM resistance in IVM-resistant and MXD-resistant isolates of *T. circumcinta* but not in an IVM-susceptible isolate. Despite the apparent incongruity of the effect of piperonyl butoxide on resistant parasites, the consistent lack of effect on susceptible parasites is suggestive of a difference in CYP expression associated with resistance. In addition to the CYPs with higher expression in resistant isolates, this project found a number of CYPs with lower expression, although in some cases this may be due to between-isolate polymorphism at the primer binding sites. Therefore, the mixed results from piperonyl butoxide studies might reflect roles of both CYP up-regulation and CYP down-regulation in anthelmintic resistance or may reflect differing effects of piperonyl butoxide itself, as although it is generally accepted as a CYP inhibitor, it has been shown to induce specific CYPs in insects (Feyereisen 2005).

A more directed future approach might include over-expressing individual CYPs shown to have higher constitutive expression in resistant isolates of *H. contortus* in transgenic *C. elegans*. If a CYP construct conferred anthelmintic resistance to *C. elegans*, it would be supportive of a role in *H. contortus* anthelmintic resistance. The most likely candidate gene from this project would be Hc-cyp-tag94 due to the relative consistency of its up-regulation in the anthelmintic-resistant MHco10 (CAVR) isolate between biological replicates, its low constitutive expression in the anthelmintic-susceptible MHco3 (ISE) isolate and the clustering of other CYP genes at its locus. Interestingly, it was not shown to be more highly expressed in the anthelmintic-resistant MHco4 (WRS) isolate, suggesting (if this gene confers resistance) that the different isolates have evolved different mechanisms of resistance. The application of other standard methods to test the functional importance of candidate CYP genes may currently be more limited: analysis of the rescue of *C. elegans* loss-of-function mutants or the rescue of RNAi-treated *C. elegans*, would require the identification of *C. elegans* orthologues, which based on knowledge of xenobiotic-metabolising CYP families in other species may be difficult; and although a mechanism for RNAi also exists in *H. contortus*, which would theoretically allow direct analysis of gene suppression in the parasite, it is currently unreliable (Geldhof *et al* 2007).

RNA-seq transcriptome analysis was also investigated as a method to assay CYP expression and as a global approach to identify genes conferring resistance in *H. contortus*. Although the comparison of CYP gene expression by RNA-seq was limited by poor coverage in this study, it is hoped that this could be overcome with future work. Attempts were made to normalise the cDNA libraries using a hybridisation protocol (Trimmer, Evrogen, NK001) to select for rare transcripts before Illumina sequencing. Briefly, the double stranded cDNA was denatured and allowed to re-associate. Hybridisation kinetics result in the abundant transcripts hybridising, while the rare transcripts remain as single strands. A double-strand specific nuclease was then added to degrade the hybridised (abundant) DNA, leaving the rare single-strand transcripts intact for PCR amplification. Preliminary attempts to use this method in the present study were unsuccessful, but if the protocol is optimised for Illumina libraries and for *H. contortus* cDNA, it should theoretically generate a greater depth of reads to map to the CYP family.

Another problem for CYP RNA-seq was the risk of reads mapping to more than one locus, an inherent problem in transcriptome analysis of a gene family sharing high sequence identity. However, over the course of this project, the RNA-seq mapping algorithms were improved significantly, allowing a greater number of previously ambiguous reads to be mapped based on mate pair information. Further progress with the genome project will also facilitate RNA-seq mapping. Reads in this project were mapped to the supercontig and BAC insert databases. The latter is expected to accurately reflect the genomic sequence but the former includes gaps and allelic sequence, both of which will reduce the number of reads that can be mapped. When the reference sequence assembly is improved, the coverage and accuracy of RNA-seq will improve. The transcriptome read databases generated in this project remain a valuable resource for when that time comes.

It is possible that, if genomic sequence was generated for the resistant isolates MHco4 (WRS) and MHco10 (CAVR), RNA-seq would reveal a high number of reads mapping to a CYP that is divergent from the genomic CYP sequence of the reference isolate used to map all reads in this project and hence missed in this study. Again, should such resources become available, the transcriptome databases generated with this work will be immediately applicable.

Work is currently underway at The Wellcome Trust Sanger Institute in the *de novo* assembly of the *H. contortus* transcriptome read databases. This should facilitate the comparison of gene expression between the adult and L3 stage and between anthelmintic-susceptible and anthelmintic-resistant isolates and will be independent of a fully assembled reference genome. The applicability to the CYP family may remain limited, due to low expression levels and complexities of CYP gene assembly, but again it is possible that a highly expressed divergent CYP could be identified in a resistant isolate *de novo* assembly.

Other workers in the Gilleard group in collaboration with the Moredun Research Institute (Dr Neil Sargison and Dr Frank Jackson) have recently generated two anthelmintic-resistant *H. contortus* isolates from backcrosses of anthelmintic-susceptible and –resistant isolates. These were created from a backcross of MHco3 (ISE) and MHco4 (WRS) adult worms and a backcross of MHco3 (ISE) and MHco10 (CAVR) adult worms. Transcriptome analysis will be undertaken to compare gene expression between the susceptible parent isolate (the transcriptome generated in this project) and the resistant backcross isolates. Theoretically, this should allow the susceptible genetic background to be subtracted, revealing the resistance-associated genes. It is hoped that work with the backcross isolates should also overcome or at least significantly reduce the challenge of between-isolate polymorphism experienced in this project, allowing a larger majority of reads from the backcross isolates to be mapped to genomic sequence of the reference (susceptible) isolate. Indeed a lack of reads mapping to a site might be indicative of a resistance-associated polymorphism. Again, the status of the reference genome will impact on the success of RNA-seq read mapping and *de novo* transcriptome assembly might be a worthwhile approach. Transcriptome analysis of the backcross isolates will begin next month.

Should the CYP family be implicated in anthelmintic resistance in *H. contortus*, potential mechanisms would include: a change in the CYP promoter or associated NR resulting in gene up-regulation, one or more gene duplications resulting in higher CYP expression, or a change in structure affecting CYP activity. As described by Feyereisen (2010), polymorphisms provide the variation upon which selection for resistance to xenobiotics can act and the CYP family in *H. contortus* certainly appears to be highly polymorphic (numerous alleles of the same genes), rapidly evolving (numerous recent gene duplications) or perhaps

most likely a combination of both. The true extent of both of these factors will remain unknown until there is a fully assembled genome.

For this parasite in particular, the significance of polymorphisms identified in candidate genes such as the CYPs must also be considered in context: the *H. contortus* genome shows a remarkably high degree of polymorphism globally. This almost certainly facilitates the rapid selection for drug resistance on exposure to new anthelmintics and thus ensures the success of *H. contortus* as a parasite, but is also likely to be observed at many loci that do not directly confer resistance. For instance, IVM-resistance in *H. contortus* has been associated with polymorphisms in numerous genes associated with drug resistance or metabolism in other species, such β-tubulin, the GSTs, PGPs, GABA and GluCl channels (Blackhall *et al* 1998b,Blackhall *et al* 1998a,Blackhall *et al* 2003, Eng *et al* 2006, Xu *et al* 1998). While it is possible and likely that IVM-resistance is multi-genic, the plasticity of the whole genome may complicate the detection of the most significant mutations. Again, functional studies with transgenic *C. elegans* are likely to be useful in confirming the degree of resistance conferred by each.

A high level of polymorphism also presents a significant bioinformatic challenge, as demonstrated by the current difficulties in assembling the genome (which perhaps further ensures the success of *H. contortus* as a parasite). It is hoped however, that new sequencing technology will evolve more rapidly than the parasite and will overcome this. The speed and relatively low cost of NGS should allow a high depth of short reads to be generated, which could be forced into a consensus genome assembly. Alternatively, new 'third generation' sequencing technology has been recently released, capable of producing very long single reads of up to 10 Kb (McCarthy 2010). This would be ideal for repetitive sequence and, for instance, loci of clustered CYP genes in the *H. contortus* genome and would significantly reduce the burden of assembly. It is likely that a combination of both of these approaches would be useful for this genome. The results of this project also suggest the genome size may be larger than previously estimated, which may be a confounding factor in the difficulties with the current genome assembly, but again is imminently solvable with further sequencing.

Nonetheless, as demonstrated by this project, the current WTSI *H. contortus* reference genome databases are a valuable resource. They permit bioinformatic analysis and facilitate molecular biological investigation, supported by the complete *C. elegans* genome. It is hoped that the *H. contortus* transcriptome databases generated with this work will provide a useful addition to this public resource and contribute to a better understanding of this intriguing parasitic genome.

Appendices

6.1 RT-QPCR primers and typical reaction efficiencies for MHco3 (ISE) adult cDNA

An asterisk indicates the efficiency was calculated from MHco3 (ISE) L3 data due to lack of adult expression

Primer	Sequence (5'-3')	Tm (°C)	cDNA Product (bp)	Efficiency (%)
Hc-ama				
Hc-ama-F2	tatgggaggtcgtgaaggtc	60	214	101.1
Hc-ama-R2	gtgggcttcatagtgggcata	61	214	
Hc-act				
Hc-actF2	gggtttgctggagatgacg	60	180	100.1
Hc-actR2	ccagttggtgacgattccg	60	180	
Hc27				
Hc27F1	gtctcggttgacttgcacag	60	199	112.2
Hc27R1	cggacaacaaggtgctccat	60	199	
gtp-ch-1				
gtp-ch-1F1	ggctgcgaaagcgatgttg	60	180	102.1
gtp-ch-1R1	cagatacccgatgtgcacc	60	180	
Hc-nhr				
Hc-nhrF2	ctgaaggagagctgtgaggt	60	185	103.1
Hc-nhrR2	tccgttatacagtctcttggc	60	185	
Hc-pgp				
Hc-pgpF3	tccaagaagcgtgccgtgt	60	181	101.9
Hc-pgpR3	gcactggtggcttcatcga	60	181	
Hc-cyp-tag1				
Hc-cyp-tag1F1	actcgattactggaagcggc	60	178	98.9
Hc-cyp-tag1R1	tcgctgattacaatcgtccgt	60	178	
Hc-cyp-tag2				
Hc-cyp-tag2F1	cgtgtcgtcacttctagcct	60	204	100.9
Hc-cyp-tag2R1	gcgaatcatatcaacatcgctg	60	204	
Hc-cyp-tag3				
Hc-cyp-tag3F1	tgagatttctccactgctgtg	60	184	93.5
Hc-cyp-tag3R1	cttctacgtcgtctccccat	60	184	
Hc-cyp-tag4				
Hc-cyp-tag4F2	ggaacctgcattgtcatcga	60	181	99.8
Hc-cyp-tag4R2	atctccataagtgcgaatcgc	60	181	

Hc-cyp-tag5				
Hc-cyp-tag5F1	ctgaaacgacggtgaggtct	60	135	98.0
Hc-cyp-tag5R1	gcaatgatgtcataccgcctga	60	135	
Hc-cyp-tag6				
Hc-cyp-tag6F1	tggactgaatcgacggagag	60	167	97.6
Hc-cyp-tag6R1	agcggaatcacctcggtatc	60	167	
Hc-cyp-tag8				
Hc-cyp-tag8F1	gtcacaaagcatttaacgcgg	60	132	71.2
Hc-cyp-tag8R1	tctgctgtactcgaggattct	60	132	
Hc-cyp-tag9				
Hc-cyp-tag9F1	gagagctgaaacatccctatg	60	196	98.2
Hc-cyp-tag9R2	aacgcacacggagatgtcg	60	235	
Hc-cyp-tag10				
Hc-cyp-tag10F2	cctgccgatcaacctgcta	60	167	31.1
Hc-cyp-tag10R2	ccgtcgtccgtaaggaatc	60	167	
Hc-cyp-tag11				
Hc-cyp-tag11F1	tctccttctgcacattccaca	60	195	95.9
Hc-cyp-tag11R1	acgtgctcatccaagaccg	60	195	
Hc-cyp-tag12				
Hc-cyp-tag12F1	aaggcgttcatgttcgttctg	60	194	97.3
Hc-cyp-tag12R1	tttccgatatgcactgctcca	60	194	
Hc-cyp-tag13				
Hc-cyp-tag13F1	cttactgggtgaaacgaggc	60	126	101.0
Hc-cyp-tag13R1	ccgtaggtatttccatattctgc	58	126	
Hc-cyp-tag14				
Hc-cyp-tag14F1	cgcttacttgctgtatcatcc	60	166	94.0
Hc-cyp-tag14R1	ggaatagcccatggtagaaca	60	166	
Hc-cyp-tag15				
Hc-cyp-tag15F2	tgcgtaaattgggcgctgatc	61	159	110.0
Hc-cyp-tag15R2	gacgaccgatacagttctcag	61	159	
Hc-cyp-tag16				
Hc-cyp-tag16F3	tgctgggatggagactacct	60	162	107.5
Hc-cyp-tag16R3	gcattggtgtaaggcagacg	60	162	
Hc-cyp-tag17				
Hc-cyp-tag17F3	ggaaacgaaggaatctaccac	60	209	96.5
Hc-cyp-tag17R3	agcaccatccttgactaacgt	60	209	
Hc-cyp-tag18				
Hc-cyp-tag18F1	cgtcgatacatgaggaaatgga	60	190	98.3
Hc-cyp-tag18R1	tgcctgacgggatatggtag	60	190	
Hc-cyp-tag20				
Hc-cyp-tag20F1	gctacggctcactccatca	60	179	94.7
Hc-cyp-tag20R1	tggatagtattcttcgggacg	60	179	

Hc-cyp-tag21				
Hc-cyp-tag21F1	ggttcgtcagatcgcagttg	60	191	106.5
Hc-cyp-tag21R1	cgtgaatcatggcaggcgt	60	191	
Hc-cyp-tag23				
Hc-cyp-tag23F1	cgaccaggaccaagccc	60	201	83.6*
Hc-cyp-tag23R1	tcgccttgttagcttcttgaaa	58	201	
Hc-cyp-tag24				
Hc-cyp-tag24F1	cgtaatcgttggcagcgtga	60	186	94.2
Hc-cyp-tag24R1	cttctctccggtctttctcc	60	186	
Hc-cyp-tag25				
Hc-cyp-tag25F1	gctgtgcatactgtcaacgat	60	170	104.1
Hc-cyp-tag25R1	cctgctcctggatctcgc	61	170	
Hc-cyp-tag27				
Hc-cyp-tag27F1	gaggttcgctgtgatggaag	60	150	100.8
Hc-cyp-tag27R1	tcgcctttctataatcagtggg	60	150	
Hc-cyp-tag28				
Hc-cyp-tag28F1	aatacggtcccgtccatactt	60	181	70.3
Hc-cyp-tag28R1	gcaaagcgacgctgttctac	60	181	
Hc-cyp-tag29				
Hc-cyp-tag29F1	gctgtggctaccgtatccta	60	187	98.7*
Hc-cyp-tag29R1	cagcgtatggagtgtgaatc	58	187	
Hc-cyp-tag30				
Hc-cyp-tag30F1	ggctggtatggagacgact	60	165	95.5
Hc-cyp-tag30R1	tcacccctctaaccgaagtat	60	165	
Hc-cyp-tag32				
Hc-cyp-tag32F1	cctgccctgattctgttcct	60	195	84.3
Hc-cyp-tag32R1	ggaagccagacagtgaagac	60	195	
Hc-cyp-tag33				
Hc-cyp-tag33F2	tattacggatttcgatggtgtc	58	120	91.5*
Hc-cyp-tag33R2	ccatcaagttcttccccattc	60	120	
Hc-cyp-tag34				
Hc-cyp-tag34F1	cgatagaaatgacaaggcagac	60	163	102.0
Hc-cyp-tag34R1	aaggtgtcagctgggatagg	60	163	
Hc-cyp-tag35				
Hc-cyp-tag35F1	tggtttggctgtactatgagaa	60	178	98.0
Hc-cyp-tag35R1	cagctataaccacatgcggc	60	178	
Hc-cyp-tag36				
Hc-cyp-tag36F1	tactcggcgagatccacca	60	169	105.2*
Hc-cyp-tag36R1	ccaacgcctctcctgcacat	60	169	
Hc-cyp-tag37				
Hc-cyp-tag37F1	ccttgatttatggtttgctggg	60	176	91.6
Hc-cyp-tag37R1	agcattggtgtaaggcagact	60	176	

Hc-cyp-tag38				
Hc-cyp-tag38F1	atatggcgtcgtggaaaccg	60	136	103.6
Hc-cyp-tag38R1	ataactgatccaacggccact	60	136	
Hc-cyp-tag40				
Hc-cyp-tag40F1	gaatgtctgcctggacctgt	60	173	69.9
Hc-cyp-tag40R1	atggaggtgattcttgtcagc	60	173	
Hc-cyp-tag41				
Hc-cyp-tag41	gcccttgccatatctgttttc	60	212	99.3
Hc-cyp-tag41	aaactcttgtcatcctgtctcg	60	212	
Hc-cyp-tag42				
Hc-cyp-tag42F1	acactaagcgctatcccaaag	60	182	95.9
Hc-cyp-tag42R1	aatgcctctttgacaccatcg	60	182	
Hc-cyp-tag43				
Hc-cyp-tag43F1	gccatagctaacggaactcac	61	171	106.4
Hc-cyp-tag43R1	gtgatagtagtttcttgaccagc	61	171	
Hc-cyp-tag44				
Hc-cyp-tag44F2	ggtctaattgaaggaaacggg	60	184	81.4*
Hc-cyp-tag44R2	ctagccaccacagtaaaac	60	184	
Hc-cyp-tag46				
Hc-cyp-tag46F1	ggaccacctccttacccatt	60	165	103.1
Hc-cyp-tag46R1	acgaatgcctctttgactcca	60	165	
Hc-cyp-tag47				
Hc-cyp-tag47F1	cccgttgccacggtacag	61	160	97.7
Hc-cyp-tag47R1	tcgtcgctgctccgtcca	61	160	
Hc-cyp-tag51				
Hc-cyp-tag51F2	gtgggatcgtgagcggc	60	171	80.5
Hc-cyp-tag51R2	ggtttcctatgcgggtcc	60	171	
Hc-cyp-tag54				
Hc-cyp-tag54F1	gtttgctgcccttattctgttc	60	199	101.9
Hc-cyp-tag54R1	ggaagccagacagtgaagac	60	199	
Hc-cyp-tag55				
Hc-cyp-tag55F1	aactgttgacgcccactttc	58	169	114.0
Hc-cyp-tag55R1	ccgcttcgcagatgacatct	60	169	
Hc-cyp-tag56				
Hc-cyp-tag56F1	gacctgtggatcgctggaa	60	158	104.5
Hc-cyp-tag56R1	ggaggtaagtcttgtcagcc	60	158	
Hc-cyp-tag58				
Hc-cyp-tag58F1	gctacaaactgcgaaagggt	58	176	101.2
Hc-cyp-tag58R1	ctcccaagcacgctcgtt	58	176	
Hc-cyp-tag60				
Hc-cyp-tag60F1	ccgtcagagaggatgtgga	60	153	100.1
Hc-cyp-tag60R1	catcgaagagttgtggctgta	60	153	

Hc-cyp-tag61				
Hc-cyp-tag61F2	gtcgcgctccctttcatc	58	173	97.9
Hc-cyp-tag61R2	gcttcgcattcttctggtcc	60	173	
Hc-cyp-tag62				
Hc-cyp-tag62F1	cttctacttggctggtatgga	60	158	96.4
Hc-cyp-tag62R1	aggtaatttctgccgatctgac	60	158	
Hc-cyp-tag63				
Hc-cyp-tag63F1	cgtcttcactgtctggcttc	60	188	102.0
Hc-cyp-tag63R1	cgccgttgttccttccaga	60	188	
Hc-cyp-tag64				
Hc-cyp-tag64F1	cttcacggcttactcgatcaa	60	182	97.3*
Hc-cyp-tag64R1	gacttctcgcgatgctcct	60	182	
Hc-cyp-tag65				
Hc-cyp-tag65F1	cggcgacgacttccacc	60	173	79.2*
Hc-cyp-tag65R1	cgacggtaaatctccaaccat	60	173	
Hc-cyp-tag67				
Hc-cyp-tag67F1	cagcaggaaaatcgtcttacag	60	166	106.6
Hc-cyp-tag67R1	accgattacttccttgacctct	60	166	
Hc-cyp-tag69				
Hc-cyp-tag69F1	atcggttcagttcctgttcca	60	187	108.1
Hc-cyp-tag69R1	cattccttctccgacgcac	60	187	
Hc-cyp-tag70				
Hc-cyp-tag70F1	ggatatgaacgcaaaaggagag	60	166	112.3
Hc-cyp-tag70R1	ctcatcgacttcacgttgtac	60	166	
Hc-cyp-tag71				
Hc-cyp-tag71F1	cactttacggacattacggga	60	188	70.5
Hc-cyp-tag71R1	gattcgattgataacactgccg	60	188	
Hc-cyp-tag72				
Hc-cyp-tag72F1	tcgaaccaaggaaggcaaac	60	201	110.2
Hc-cyp-tag72R1	caccgagacatgaccgtttc	60	201	
Hc-cyp-tag73				
Hc-cyp-tag73F1	aggtggctcaggaattggtg	60	161	109.5
Hc-cyp-tag73R1	agcccattcaatgcgatcagt	60	161	
Hc-cyp-tag74				
Hc-cyp-tag74F1	cccctatgcttgtctacaac	58	201	93.9
Hc-cyp-tag74R1	atatctactggtgtgccagc	58	201	
Hc-cyp-tag75				
Hc-cyp-tag75F1	gcttctcgccagcaggattt	60	176	110.3
Hc-cyp-tag75R1	aagtgtctcataggtctgttcg	60	176	
Hc-cyp-tag76				
Hc-cyp-tag76F1	ctcgcattccaaccaagtagat	60	206	83.3*
Hc-cyp-tag76R1	gtgtctcataggtctgttcgat	60	206	

Hc-cyp-tag77				
Hc-cyp-tag77F1	tctctccttctgcacattctac	60	195	94.5
Hc-cyp-tag77R1	tgctcatccgaaaccggg	58	195	
Hc-cyp-tag80				
Hc-cyp-tag80F1	ggccagaggtttgctgtgat	60	162	101.0
Hc-cyp-tag80R1	ctcagttcgcttttcaataatcag	60	162	
Hc-cyp-tag81				
Hc-cyp-tag81F1	gcgacagattacgtggcag	60	154	107.5
Hc-cyp-tag81R1	ctccccatgaaagcgtgttg	60	154	
Hc-cyp-tag86				
Hc-cyp-tag86F1	tgagtggatacggtttcgagt	60	186	105.7
Hc-cyp-tag86R1	cctcatcactcaactggttag	60	186	
Hc-cyp-tag88				
Hc-cyp-tag88F1	caagaagtctatcgtgctggc	58	167	107.6
Hc-cyp-tag88R1	atgattagttccggtcgagg	58	167	
Hc-cyp-tag89				
Hc-cyp-tag89F1	ggaggaatggggacgacaat	60	147	102.8
Hc-cyp-tag89R1	ccggttcctatccaatctga	58	147	
Hc-cyp-tag94				
Hc-cyp-tag94F1	gaactggagagaacaacgtc	58	184	103.2
Hc-cyp-tag94R1	ctcgtttgcagcgtccatca	60	184	
Hc-cyp-tag95				
Hc-cyp-tag95F1	gaatctactccacgaaacactt	58	202	104.2
Hc-cyp-tag95R1	gcccaagcactgacgtttc	60	202	

6.2 Wellcome Trust Sanger Institute references for CYP genomic sequence described in thesis

SUPERCONTIG	CONTIG(S)	WTSI DATABASE REFERENCE
1	54	Supercontig_0029042_cw_200808
2	42	Supercontig_0029255_cw_200808
3	63	Supercontig_0037463_cw_200808
4	46	Supercontig_0006244_cw_200808
5	50_94_62_36_35	Supercontig_0055835_cw_200808
6	60+34_33_49+30	Supercontig_0053015_cw_200808
7	32_68+69	Supercontig_0059022_cw_200808
8	91_39+38_37	Supercontig_0047793_cw_200808
9	95_56	Supercontig_0024087_cw_200808
10	84	Supercontig_0036329_cw_200808
11	16	Supercontig_0000158_cw_200808
12	81	Supercontig_0022776_cw_200808
13	45	Supercontig_0004472_cw_200808
14	40	Supercontig_0000588_cw_200808
15	22_14	Supercontig_0053000_cw_200808
16	65+64	Supercontig_0058211_cw_200808
17	82	Supercontig_0024845_cw_200808
18	10	Contig_0080247 (12/11/07)
19	66	Supercontig_0055713_cw_200808
20	7	Supercontig_0008400_cw_200808
21	90_19	Supercontig_0029120_cw_200808
22	78	Supercontig_0009042_cw_200808
23	8	Supercontig_0041241_cw_200808
24	79	Supercontig_0018164_cw_200808
25	2_6	Supercontig_0047659_cw_200808
26	1_ 5	Supercontig_0047698_cw_200808
27	51_92_93	Supercontig_0057680_cw_200808
28	20	Supercontig_0016742_cw_200808
29	44	Supercontig_0012299_cw_200808
30	89	Supercontig_0050076_cw_200808
31	86_87	Supercontig_0044030_cw_200808
32	80	Supercontig_0022729_cw_200808
33	53_67	Supercontig_0059036_cw_200808
34	25+15	Supercontig_0057952_cw_200808
35	12_13	Supercontig_0058664_cw_200808
36	27	Supercontig_0059583_cw_200808
37	17_18	Supercontig_0036840_cw_200808
38	47+48	Supercontig_0006640_cw_200808
39	57	Supercontig_0046553_cw_200808
40	43	Supercontig_0023588_cw_200808
41	71+72	Supercontig_0057401_cw_200808
42	24+23+52	Supercontig_0059253_cw_200808
43	28	Supercontig_0055649_cw_200808
44	26	Supercontig_0004356_cw_200808
45	59+58	Supercontig_0016803_cw_200808
46	88	Supercontig_0045832_cw_200808
47	29	Supercontig_0039964_cw_200808
48	21_96	Supercontig_0057945_cw_200808
49	41	Supercontig_0004463_cw_200808

50	61	Supercontig_0046836_cw_200808
51	9_70	Supercontig_0058643_cw_200808
52	55	Supercontig_0042738_cw_200808
53	4	Supercontig_0001791_cw_200808
54	77	Supercontig_0000221_cw_200808
55	11+3	Supercontig_0052968_cw_200808
56	85	Supercontig_0038404_cw_200808
57	76	Supercontig_0035762_cw_200808
58	75	Contig_0022272 (12/11/07)
59	76	Contig_0031347 (12/11/07)
60	73+74	Supercontig_0036826_cw_200808
61	83	Supercontig_0027055_cw_200808

6.3 References

Abbott JC, Aanensen DM, Rutherford K, Butcher S, Spratt BG (2005), WebACT--an online companion for the Artemis Comparison Tool, *Bioinformatics* 21: 3665-3666

Alvarez LI, Imperiale FA, Sanchez SF, Murno GA, Lanusse CE (2000), Uptake of albendazole and albendazole sulphoxide by *Haemonchus contortus* and *Fasciola hepatica* in sheep, *Vet.Parasitol.* 94: 75-89

Alvarez LI, Mottier ML, Lanusse CE (2007), Drug transfer into target helminth parasites, *Trends Parasitol.* 23: 97-104

Alvarez LI, Mottier ML, Sanchez SF, Lanusse CE (2001), Ex vivo diffusion of albendazole and its sulfoxide metabolite into *Ascaris suum* and *Fasciola hepatica*, *Parasitol.Res.* 87: 929-934

Alvarez LI, Solana HD, Mottier ML, Virkel GL, Fairweather I, Lanusse CE (2005), Altered drug influx/efflux and enhanced metabolic activity in triclabendazole-resistant liver flukes, *Parasitology* 131: 501-510

Alvinerie M, Dupuy J, Eeckhoutte C, Sutra JF, Kerboeuf D (2001), In vitro metabolism of moxidectin in *Haemonchus contortus* adult stages, *Parasitol.Res.* 87: 702-704

Amacher DE (2010), The effects of cytochrome P450 induction by xenobiotics on endobiotic metabolism in pre-clinical safety studies, *Toxicol.Mech.Methods* 20: 159-166

Amenya DA, Naguran R, Lo TC, Ranson H, Spillings BL, Wood OR, Brooke BD, Coetzee M, Koekemoer LL (2008), Over expression of a cytochrome P450 (CYP6P9) in a major African malaria vector, *Anopheles funestus*, resistant to pyrethroids, *Insect Mol.Biol.* 17: 19-25

Amichot M, Tarès S, Brun-Barale A, Arthaud L, Bride JM, Bergé JB (2004), Point mutations associated with insecticide resistance in the *Drosophila* cytochrome P450 Cyp6a2 enable DDT metabolism, *Eur.J.Biochem.* 271: 1250-1257

An JH, Blackwell TK (2003), SKN-1 links *C. elegans* mesendodermal specification to a conserved oxidative stress response, *Genes Dev.* 17: 1882-1893

Anzenbacher P, Anzenbacherova E (2001), Cytochromes P450 and metabolism of xenobiotics, *Cell Mol.Life Sci.* 58: 737-747

Baldwin WS, Marko PB, Nelson DR (2009), The cytochrome P450 (CYP) gene superfamily in *Daphnia pulex*, *BMC.Genomics* 10: 169

Barrett J (1998), Cytochrome P450 in parasitic protozoa and helminths, *Comp. Biochem.Physiol C.Pharmacol.Toxicol.Endocrinol.* 121: 181-183

Barrett J (2009), Forty years of helminth biochemistry, *Parasitology* 136: 1633-1642

Bartley DJ, Jackson F, Jackson E, Sargison N (2004), Characterisation of two triple resistant field isolates of *Teladorsagia* from Scottish lowland sheep farms, *Vet.Parasitol.* 123: 189-199

Benenati G, Penkov S, Muller-Reichert T, Entchev EV, Kurzchalia TV (2009), Two cytochrome P450s in *Caenorhabditis elegans* are essential for the organization of eggshell, correct execution of meiosis and the polarization of embryo, *Mech.Dev.* 126: 382-393

Benson G (1999), Tandem repeats finder: a program to analyze DNA sequences, *Nucleic Acids Res.* 27: 573-580

Berge JB, Feyereisen R, Amichot M (1998), Cytochrome P450 monooxygenases and insecticide resistance in insects, *Philos.Trans.R.Soc.Lond B Biol.Sci.* 353: 1701-1705

Bessereau JL (2006), Transposons in *C. elegans*, *WormBook.* 1-13

Blackhall WJ, Liu HY, Xu M, Prichard RK, Beech RN (1998a), Selection at a P-glycoprotein gene in ivermectin- and moxidectin-selected strains of *Haemonchus contortus*, *Mol.Biochem.Parasitol.* 95: 193-201

Blackhall WJ, Pouliot JF, Prichard RK, Beech RN (1998b), *Haemonchus contortus*: selection at a glutamate-gated chloride channel gene in ivermectin- and moxidectin-selected strains, *Exp.Parasitol.* 90: 42-48

Blackhall WJ, Prichard RK, Beech RN (2003), Selection at a gamma-aminobutyric acid receptor gene in *Haemonchus contortus* resistant to avermectins/milbemycins, *Mol.Biochem.Parasitol.* 131: 137-145

Blaxter ML, De Ley P, Garey JR, Liu LX, Scheldeman P, Vierstraete A, Vanfleteren JR, Mackey LY, Dorris M, Frisse LM, Vida JT, Thomas WK (1998), A molecular evolutionary framework for the phylum Nematoda, *Nature* 392: 71-75

Blouin MS (1998), Mitochondrial DNA diversity in nematodes, *J.Helminthol.* 72: 285-289

Blouin MS, Yowell CA, Courtney CH, Dame JB (1995), Host movement and the genetic structure of populations of parasitic nematodes, *Genetics* 141: 1007-1014

Blumenthal T (2005), Trans-splicing and operons, *WormBook.* 1-9

Blumenthal T, Gleason KS (2003), *Caenorhabditis elegans* operons: form and function, *Nat.Rev.Genet.* 4: 112-120

Bogwitz MR, Chung H, Magoc L, Rigby S, Wong W, O'Keefe M, McKenzie JA, Batterham P, Daborn PJ (2005), Cyp12a4 confers lufenuron resistance in a natural population of *Drosophila melanogaster*, *Proc.Natl.Acad.Sci.U.S.A* 102: 12807-12812

Boisvenue RJ, Brandt MC, Galloway RB, Hendrix JC (1983), In vitro activity of various anthelmintic compounds against *Haemonchus contortus* larvae, *Vet.Parasitol.* 13: 341-347

Boulin T, Gielen M, Richmond JE, Williams DC, Paoletti P, Bessereau JL (2008), Eight genes are required for functional reconstitution of the *Caenorhabditis elegans* levamisole-sensitive acetylcholine receptor, *Proc.Natl.Acad.Sci.U.S.A* 105: 18590-18595

Breitling R (2006), Biological microarray interpretation: the rules of engagement, *Biochim.Biophys.Acta* 1759: 319-327

Brennan GP, Fairweather I, Trudgett A, Hoey E, McCoy, McConville M, Meaney M, Robinson M, McFerran N, Ryan L, Lanusse C, Mottier L, Alvarez L, Solana H, Virkel G, Brophy PM (2007), Understanding triclabendazole resistance, *Exp.Mol.Pathol.* 82: 104-109

Burnell AM, Houthoofd K, O'Hanlon K, Vanfleteren JR (2005), Alternate metabolism during the dauer stage of the nematode *Caenorhabditis elegans*, *Exp.Gerontol.* 40: 850-856

Caceres M, Ranz JM, Barbadilla A, Long M, Ruiz A (1999), Generation of a widespread *Drosophila* inversion by a transposable element, *Science* 285: 415-418

Callaghan MJ, Beh KJ (1994), Characterization of a tandemly repetitive DNA sequence from *Haemonchus contortus*, *Int.J.Parasitol.* 24: 137-141

Capece BP, Virkel GL, Lanusse CE (2009), Enantiomeric behaviour of albendazole and fenbendazole sulfoxides in domestic animals: pharmacological implications, *Vet.J.* 181: 241-250

Carver T, Bohme U, Otto TD, Parkhill J, Berriman M (2010), BamView: viewing mapped read alignment data in the context of the reference sequence, *Bioinformatics.* 26: 676-677

Carver TJ, Rutherford KM, Berriman M, Rajandream MA, Barrell BG, Parkhill J (2005), ACT: the Artemis Comparison Tool, *Bioinformatics* 21: 3422-3423

Castillo-Davis CI, Mekhedov SL, Hartl DL, Koonin EV, Kondrashov FA (2002), Selection for short introns in highly expressed genes, *Nat.Genet.* 31: 415-418

Chakrapani BP, Kumar S, Subramaniam JR (2008), Development and evaluation of an in vivo assay in *Caenorhabditis elegans* for screening of compounds for their effect on cytochrome P450 expression, *J.Biosci.* 33: 269-277

Chen S, Li X (2007), Transposable elements are enriched within or in close proximity to xenobiotic-metabolizing cytochrome P450 genes, *BMC.Evol.Biol.* 7: 46

Chung H, Bogwitz MR, McCart C, Andrianopoulos A, Ffrench-Constant RH, Batterham P, Daborn PJ (2007), *Cis*-regulatory elements in the *Accord* retrotransposon result in tissue-specific expression of the *Drosophila melanogaster* insecticide resistance gene *Cyp6g1*, *Genetics* 175: 1071-1077

Chung H, Sztal T, Pasricha S, Sridhar M, Batterham P, Daborn PJ (2009), Characterization of *Drosophila melanogaster* cytochrome P450 genes, *Proc.Natl.Acad.Sci.U.S.A* 106: 5731-5736

Claudianos C, Ranson H, Johnson RM, Biswas S, Schuler MA, Berenbaum MR, Feyereisen R, Oakeshott JG (2006), A deficit of detoxification enzymes: pesticide sensitivity and environmental response in the honeybee, *Insect Mol.Biol.* 15: 615-636

Coghlan A (2005), Nematode genome evolution, *WormBook*.1-15.

Coghlan A, Fiedler TJ, McKay SJ, Flicek P, Harris TW, Blasiar D, Stein LD (2008), nGASP--the nematode genome annotation assessment project, *BMC.Bioinformatics.* 9: 549

Coghlan A, Wolfe KH (2002), Fourfold faster rate of genome rearrangement in nematodes than in *Drosophila*, *Genome Res.* 12: 857-867

Coles GC, Jackson F, Pomroy WE, Prichard RK, Samson-Himmelstjerna G, Silvestre A, Taylor MA, Vercruysse J (2006), The detection of anthelmintic resistance in nematodes of veterinary importance, *Vet.Parasitol.* 136: 167-185

Conway DP (1964), Variance in the effectiveness of thiabendazole against *Haemonchus contortus* in sheep, *Am.J.Vet.Res.* 25: 844-846

Cully DF, Vassilatis DK, Liu KK, Paress PS, Van der Ploeg LH, Schaeffer JM, Arena JP (1994), Cloning of an avermectin-sensitive glutamate-gated chloride channel from *Caenorhabditis elegans*, *Nature* 371: 707-711

Cvilink V, Kubicek V, Nobilis M, Krizova V, Szotakova B, Lamka J, Varady M, Kubenova M, Novotna R, Gavelova M, Skalova L (2008a), Biotransformation of flubendazole and selected model xenobiotics in *Haemonchus contortus*, *Vet.Parasitol.* 151: 242-248

Cvilink V, Lamka J, Skalova L (2009), Xenobiotic metabolizing enzymes and metabolism of anthelminthics in helminths, *Drug Metab. Rev.* 41: 8-26

Cvilink V, Skalova L, Szotakova B, Lamka J, Kostiainen R, Ketola RA (2008b), LC-MS-MS identification of albendazole and flubendazole metabolites formed ex vivo by *Haemonchus contortus*, *Anal.Bioanal.Chem.* 391: 337-343

Daborn PJ, Lumb C, Boey A, Wong W, Ffrench-Constant RH, Batterham P (2007), Evaluating the insecticide resistance potential of eight *Drosophila melanogaster* cytochrome P450 genes by transgenic over-expression, *Insect Biochem.Mol.Biol.* 37: 512-519

Daborn PJ, Yen JL, Bogwitz MR, Le Goff G, Feil E, Jeffers S, Tijet N, Perry T, Heckel D, Batterham P, Feyereisen R, Wilson TG, Ffrench-Constant RH (2002), A single p450 allele associated with insecticide resistance in *Drosophila*, *Science* 297: 2253-2256

de Groot MJ (2006), Designing better drugs: predicting cytochrome P450 metabolism, *Drug Discov.Today* 11: 601-606

Demeler J, Kuttler U, El Abdellati A, Stafford K, Rydzik A, Varady M, Kenyon F, Coles G, Hoglund J, Jackson F, Vercruysse J, Samson-Himmelstjerna G (2010), Standardization of the larval migration inhibition test for the detection of resistance to ivermectin in gastro intestinal nematodes of ruminants, *Vet.Parasitol.* 174(1-2):58-64.

Dent JA, Davis MW, Avery L (1997), avr-15 encodes a chloride channel subunit that mediates inhibitory glutamatergic neurotransmission and ivermectin sensitivity in *Caenorhabditis elegans*, *EMBO J.* 16: 5867-5879

Dent JA, Smith MM, Vassilatis DK, Avery L (2000), The genetics of ivermectin resistance in *Caenorhabditis elegans*, *Proc.Natl.Acad.Sci.U.S.A* 97: 2674-2679

Deutsch M, Long M (1999), Intron-exon structures of eukaryotic model organisms, *Nucleic Acids Res.* 27: 3219-3228

Devine C, Brennan GP, Lanusse CE, Alvarez LI, Trudgett A, Hoey E, Fairweather I (2010), Inhibition of cytochrome P450-mediated metabolism enhances ex vivo susceptibility of *Fasciola hepatica* to triclabendazole, *Parasitology* 137: 871-880

Dieterich C, Clifton SW, Schuster LN, Chinwalla A, Delehaunty K, Dinkelacker I, Fulton L, Fulton R, Godfrey J, Minx P, Mitreva M, Roeseler W, Tian H, Witte H, Yang SP, Wilson RK, Sommer RJ (2008), The *Pristionchus pacificus* genome provides a unique perspective on nematode lifestyle and parasitism, *Nat.Genet.* 40: 1193-1198

Drudge JH, Leland SE, Jr., Wyant ZN (1957), Strain variation in the response of sheep nematodes to the action of phenothiazine. II. Studies on pure infections of *Haemonchus contortus*, *Am.J.Vet.Res.* 18: 317-325

Drudge JH, Szanto J, Wyant ZN, Elam G (1964), Field studies on parasite control in sheep: comparison of thiabendazole, ruelene, and phenothiazine, *Am.J.Vet.Res.* 25: 1512-1518

Duret L, Marais G, Biemont C (2000), Transposons but not retrotransposons are located preferentially in regions of high recombination rate in *Caenorhabditis elegans*, *Genetics* 156: 1661-1669

Duret L, Mouchiroud D (1999), Expression pattern and, surprisingly, gene length shape codon usage in *Caenorhabditis, Drosophila*, and *Arabidopsis, Proc.Natl.Acad.Sci.U.S.A* 96: 4482-4487

Eng JK, Blackhall WJ, Osei-Atweneboana MY, Bourguinat C, Galazzo D, Beech RN, Unnasch TR, Awadzi K, Lubega GW, Prichard RK (2006), Ivermectin selection on beta-tubulin: evidence in *Onchocerca volvulus* and *Haemonchus contortus*, *Mol.Biochem.Parasitol.* 150: 229-235

Felsenstein J (1985), Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* 39:783-791.

Feyereisen R (2005), Insect Cytochrome P450, Elsevier, Oxford, UK

Feyereisen R (2006), Evolution of insect P450, *Biochem.Soc.Trans.* 34: 1252-1255

Feyereisen R (2010), Arthropod CYPomes illustrate the tempo and mode in P450 evolution, *Biochim.Biophys.Acta*. 1814(1):19-28.

Fisher AL, Lithgow GJ (2006), The nuclear hormone receptor DAF-12 has opposing effects on *Caenorhabditis elegans* lifespan and regulates genes repressed in multiple long-lived worms, *Aging Cell* 5: 127-138

Fogleman JC, Danielson PB (2000), Analysis of fragment homology among DNA sequences from cytochrome P450 families 4 and 6, *Genetica* 110: 257-265

Foll RL, Pleyers A, Lewandovski GJ, Wermter C, Hegemann V, Paul RJ (1999), Anaerobiosis in the nematode *Caenorhabditis elegans*, *Comp. Biochem.Physiol B Biochem.Mol.Biol*. 124: 269-280

Fox S, Filichkin S, Mockler TC (2009), Applications of ultra-high-throughput sequencing, *Methods Mol.Biol*. 553: 79-108

Fry M, Jenkins DC (1983), *Nippostrongylus brasiliensis* and *Ascaridia galli*: mitochondrial respiration in free-living and parasitic stages, *Exp.Parasitol*. 56: 101-106

Fry M, Jenkins DC (1984), Nematoda: aerobic respiratory pathways of adult parasitic species, *Exp.Parasitol*. 57: 86-92

Fujii S, Toyama A, Amrein H (2008), A male-specific fatty acid omega-hydroxylase, SXE1, is necessary for efficient male mating in *Drosophila melanogaster*, *Genetics* 180: 179-190

Fujita K (2004), Food-drug interactions via human cytochrome P450 3A (CYP3A), *Drug Metabol.Drug Interact*. 20: 195-217

Geary TG, Sims SM, Thomas EM, Vanover L, Davis JP, Winterrowd CA, Klein RD, Ho NF, Thompson DP (1993), *Haemonchus contortus*: ivermectin-induced paralysis of the pharynx, *Exp.Parasitol*. 77: 88-96

Geary TG, Thompson DP (2001), *Caenorhabditis elegans*: how good a model for veterinary parasites?, *Vet.Parasitol*. 101: 371-386

Geerts S, Gryseels B (2000), Drug resistance in human helminths: current situation and lessons from livestock, *Clin.Microbiol.Rev*. 13: 207-222

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* 134: 609-619

Gerisch B, Antebi A (2004), Hormonal signals produced by DAF-9/cytochrome P450 regulate *C. elegans* dauer diapause in response to environmental cues, *Development* 131: 1765-1776

Gerisch B, Weitzel C, Kober-Eisermann C, Rottiers V, Antebi A (2001), A hormonal signaling pathway influencing *C. elegans* metabolism, reproductive development, and life span, *Dev.Cell* 1: 841-851

Ghedin E, Wang S, Spiro D, Caler E, Zhao Q, Crabtree J, Allen JE, Delcher AL, Guiliano DB, Miranda-Saavedra D, Angiuoli SV, Creasy T, Amedeo P, Haas B, El

Sayed NM, Wortman JR, Feldblyum T, Tallon L, Schatz M, Shumway M, Koo H, Salzberg SL, Schobel S, Pertea M, Pop M, White O, Barton GJ, Carlow CK, Crawford MJ, Daub J, Dimmic MW, Estes CF, Foster JM, Ganatra M, Gregory WF, Johnson NM, Jin J, Komuniecki R, Korf I, Kumar S, Laney S, Li BW, Li W, Lindblom TH, Lustigman S, Ma D, Maina CV, Martin DM, McCarter JP, McReynolds L, Mitreva M, Nutman TB, Parkinson J, Peregrin-Alvarez JM, Poole C, Ren Q, Saunders L, Sluder AE, Smith K, Stanke M, Unnasch TR, Ware J, Wei AD, Weil G, Williams DJ, Zhang Y, Williams SA, Fraser-Liggett C, Slatko B, Blaxter ML, Scott AL (2007), Draft genome of the filarial nematode parasite *Brugia malayi, Science* 317: 1756-1760

Gillam EM (2008), Engineering cytochrome p450 enzymes, *Chem.Res.Toxicol.* 21: 220-231

Gilleard JS (2004), The use of *Caenorhabditis elegans* in parasitic nematode research, *Parasitology* 128 Suppl 1: S49-S70

Gilleard JS (2006), Understanding anthelmintic resistance: the need for genomics and genetics, *Int.J.Parasitol.* 36: 1227-1239

Gilleard JS, Beech RN (2007), Population genetics of anthelmintic resistance in parasitic nematodes, *Parasitology* 134: 1133-1147

Glue P, Clement RP (1999), Cytochrome P450 enzymes and drug metabolism--basic concepts and methods of assessment, *Cell Mol.Neurobiol.* 19: 309-323

Gonzalez FJ (1992), Human cytochromes P450: problems and prospects, *Trends Pharmacol.Sci.* 13: 346-352

Gotoh O (1992), Substrate recognition sites in cytochrome P450 family 2 (CYP2) proteins inferred from comparative analyses of amino acid and coding nucleotide sequences, *J.Biol.Chem.* 267: 83-90

Gotoh O (1998), Divergent structures of *Caenorhabditis elegans* cytochrome P450 genes suggest the frequent loss and gain of introns during the evolution of nematodes, *Mol.Biol.Evol.* 15: 1447-1459

Graber JH, Salisbury J, Hutchins LN, Blumenthal T (2007), *C. elegans* sequences that control trans-splicing and operon pre-mRNA processing, *RNA.* 13: 1409-1426

Grillo V, Jackson F, Gilleard JS (2006), Characterisation of *Teladorsagia circumcincta* microsatellites and their development as population genetic markers, *Mol.Biochem.Parasitol.* 148: 181-189

Guengerich FP (2006), Cytochrome P450s and other enzymes in drug metabolism and toxicity, *AAPS.J.* 8: E101-E111

Guengerich FP (2009), Cataloging the Repertoire of Nature's Blowtorch, P450, *Chem.Biol.* 16: 1215-1216

Guiliano DB, Blaxter ML (2006), Operon conservation and the evolution of trans-splicing in the phylum Nematoda, *PLoS.Genet.* 2: e198

Guiliano DB, Hall N, Jones SJ, Clark LN, Corton CH, Barrell BG, Blaxter ML (2002), Conservation of long-range synteny and microsynteny between the genomes of two distantly related nematodes, *Genome Biol.* 3: RESEARCH0057

Hartman D, Donald DR, Nikolaou S, Savin KW, Hasse D, Presidente PJ, Newton SE (2001), Analysis of developmentally regulated genes of the parasite *Haemonchus contortus, Int.J.Parasitol.* 31: 1236-1245

Hashmi S, Tawe W, Lustigman S (2001), *Caenorhabditis elegans* and the study of gene function in parasites, *Trends Parasitol.* 17: 387-393

Higgins MJ, Stearns V (2010), CYP2D6 polymorphisms and tamoxifen metabolism: clinical relevance, *Curr.Oncol.Rep.* 12: 7-15

Hillier LW, Miller RD, Baird SE, Chinwalla A, Fulton LA, Koboldt DC, Waterston RH (2007), Comparison of *C. elegans* and *C. briggsae* genome sequences reveals extensive conservation of chromosome organization and synteny, *PLoS.Biol.* 5: e167

Hoekstra R, Criado-Fornelio A, Fakkeldij J, Bergman J, Roos MH (1997), Microsatellites of the parasitic nematode *Haemonchus contortus*: polymorphism and linkage with a direct repeat, *Mol.Biochem.Parasitol.* 89: 97-107

Hoekstra R, Visser A, Otsen M, Tibben J, Lenstra JA, Roos MH (2000), EST sequencing of the parasitic nematode *Haemonchus contortus* suggests a shift in gene expression during transition to the parasitic stages, *Mol.Biochem.Parasitol.* 110: 53-68

Holden-Dye L, Walker RJ (1990), Avermectin and avermectin derivatives are antagonists at the 4-aminobutyric acid (GABA) receptor on the somatic muscle cells of *Ascaris;* is this the site of anthelmintic action?, *Parasitology* 101 Pt 2: 265-271

Holmes PH (1985), Pathogenesis of trichostrongylosis, *Vet.Parasitol.* 18: 89-101

Ingelman-Sundberg M, Sim SC (2010), Pharmacogenetic biomarkers as tools for improved drug therapy; emphasis on the cytochrome P450 system, *Biochem.Biophys.Res.Commun.* 396: 90-94

Ingelman-Sundberg M, Sim SC, Gomez A, Rodriguez-Antona C (2007), Influence of cytochrome P450 polymorphisms on drug therapies: pharmacogenetic, pharmacoepigenetic and clinical aspects, *Pharmacol.Ther.* 116: 496-526

Jackson F, Coop RL (2000), The development of anthelmintic resistance in sheep nematodes, *Parasitology* 120 Suppl: S95-107

Jackson F, Hoste H (2010), *In vitro* screening of plant resources for extra-nutritional attributes in ruminants: nuclear and related methodologies, *In Vitro Methods for the Primary Screening of Plant Products for Direct Activity against Ruminant Gastrointestinal Nematodes*, Springer, Netherlands.

Johnstone IL, Barry JD (1996), Temporal reiteration of a precise gene expression pattern during nematode development, *EMBO J.* 15: 3633-3639

Kamath RS, Fraser AG, Dong Y, Poulin G, Durbin R, Gotta M, Kanapin A, Le Bot N, Moreno S, Sohrmann M, Welchman DP, Zipperlen P, Ahringer J (2003), Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi, *Nature* 421: 231-237

Kaminsky R (2003), Drug resistance in nematodes: a paper tiger or a real problem?, *Curr.Opin.Infect.Dis.* 16: 559-564

Kaminsky R, Ducray P, Jung M, Clover R, Rufener L, Bouvier J, Weber SS, Wenger A, Wieland-Berghausen S, Goebel T, Gauvry N, Pautrat F, Skripsky T, Froelich O, Komoin-Oka C, Westlund B, Sluder A, Maser P (2008), A new class of anthelmintics effective against drug-resistant nematodes, *Nature* 452: 176-180

Kaplan RM (2004), Drug resistance in nematodes of veterinary importance: a status report, *Trends Parasitol.* 20: 477-481

Kleemann G, Jia L, Emmons SW (2008), Regulation of *Caenorhabditis elegans* male mate searching behavior by the nuclear receptor DAF-12, *Genetics* 180: 2111-2122

Knox DP, Redmond DL, Newlands GF, Skuce PJ, Pettit D, Smith WD (2003), The nature and prospects for gut membrane proteins as vaccine candidates for *Haemonchus contortus* and other ruminant trichostrongyloids, *Int.J.Parasitol.* 33: 1129-1137

Kotze AC (1997), Cytochrome P450 monooxygenase activity in *Haemonchus contortus* (Nematoda), *Int.J.Parasitol.* 27: 33-40

Kotze AC (2000), Oxidase activities in macrocyclic-resistant and -susceptible *Haemonchus contortus*, *J.Parasitol.* 86: 873-876

Kotze AC, Dobson RJ, Chandler D (2006), Synergism of rotenone by piperonyl butoxide in *Haemonchus contortus* and *Trichostrongylus colubriformis* in vitro: potential for drug-synergism through inhibition of nematode oxidative detoxification pathways, *Vet.Parasitol.* 136: 275-282

Krecek RC, Waller PJ (2006), Towards the implementation of the "basket of options" approach to helminth parasite control of livestock: emphasis on the tropics/subtropics, *Vet.Parasitol.* 139: 270-282

Kwa MS, Kooyman FN, Boersema JH, Roos MH (1993), Effect of selection for benzimidazole resistance in *Haemonchus contortus* on beta-tubulin isotype 1 and isotype 2 genes, *Biochem.Biophys.Res.Commun.* 191: 413-419

Kwa MS, Veenstra JG, Roos MH (1994), Benzimidazole resistance in *Haemonchus contortus* is correlated with a conserved mutation at amino acid 200 in beta-tubulin isotype 1, *Mol.Biochem.Parasitol.* 63: 299-303

Laing S, Ivens A, Laing R, Ravikumar S, Butler V, Woods DJ, Gilleard JS. Characterisation of the xenobiotic response of *C. elegans* to albendazole and the identification of novel drug glucoside metabolites. Biochemical Journal . 2010. Ref Type: In Press

Larsen M (2006), Biological control of nematode parasites in sheep, *J.Anim Sci.* 84 Suppl: E133-E139

Le Jambre LF, Gill JH, Lenane IJ, Lacey E (1995), Characterisation of an avermectin resistant strain of Australian *Haemonchus contortus*, *Int.J.Parasitol.* 25: 691-698

Leignel V, Silvestre A, Humbert JF, Cabaret J (2010), Alternation of anthelmintic treatments: a molecular evaluation for benzimidazole resistance in nematodes, *Vet.Parasitol.* 172: 80-88

Leroy S, Duperray C, Morand S (2003), Flow cytometry for parasite nematode genome size measurement, *Mol.Biochem.Parasitol.* 128: 91-93

Li H, Durbin R (2009), Fast and accurate short read alignment with Burrows-Wheeler transform, *Bioinformatics.* 25: 1754-1760

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R (2009), The Sequence Alignment/Map format and SAMtools, *Bioinformatics.* 25: 2078-2079

Li H, Ruan J, Durbin R (2008), Mapping short DNA sequencing reads and calling variants using mapping quality scores, *Genome Res.* 18: 1851-1858

Lindblom TH, Dodd AK (2006), Xenobiotic detoxification in the nematode *Caenorhabditis elegans*, *J.Exp.Zool.A Comp Exp.Biol.* 305: 720-730

Lindblom TH, Pierce GJ, Sluder AE (2001), A *C. elegans* orphan nuclear receptor contributes to xenobiotic resistance, *Curr.Biol.* 11: 864-868

Little PR, Hodges A, Watson TG, Seed JA, Maeder SJ (2010), Field efficacy and safety of an oral formulation of the novel combination anthelmintic, derquantel-abamectin, in sheep in New Zealand, *N.Z.Vet.J.* 58: 121-129

Liu C, Oliveira A, Chauhan C, Ghedin E, Unnasch TR (2010), Functional analysis of putative operons in Brugia malayi, *Int.J.Parasitol.* 40: 63-71

Maglich JM, Sluder A, Guan X, Shi Y, McKee DD, Carrick K, Kamdar K, Willson TM, Moore JT (2001), Comparison of complete nuclear receptor sets from the human, *Caenorhabditis elegans* and *Drosophila* genomes, *Genome Biol.* 2: RESEARCH0029

Maitra S, Dombrowski SM, Waters LC, Ganguly R (1996), Three second chromosome-linked clustered *Cyp6* genes show differential constitutive and barbital-induced expression in DDT-resistant and susceptible strains of *Drosophila melanogaster*, *Gene* 180: 165-171

Mansuy D (1998), The great diversity of reactions catalyzed by cytochromes P450, *Comp. Biochem.Physiol C.Pharmacol.Toxicol.Endocrinol.* 121: 5-14

Mardis ER (2008), The impact of next-generation sequencing technology on genetics, *Trends Genet.* 24: 133-141

Marguerat S, Bahler J (2010), RNA-seq: from technology to biology, *Cell Mol.Life Sci.* 67: 569-579

Martin PJ, Anderson N, Jarrett RG (1989), Detecting benzimidazole resistance with faecal egg count reduction tests and in vitro assays, *Aust.Vet.J.* 66: 236-240

McCarthy A (2010), Third generation DNA sequencing: pacific biosciences' single molecule real time technology, *Chem.Biol.* 17: 675-676

McGhee JD (2007), The *C. elegans* intestine, *WormBook.* 1-36

Menzel R, Bogaert T, Achazi R (2001), A systematic gene expression screen of *Caenorhabditis elegans* cytochrome P450 genes reveals CYP35 as strongly xenobiotic inducible, *Arch.Biochem.Biophys.* 395: 158-168

Menzel R, Rodel M, Kulas J, Steinberg CE (2005), CYP35: xenobiotically induced gene expression in the nematode *Caenorhabditis elegans*, *Arch.Biochem.Biophys.* 438: 93-102

Mocellin S, Rossi CR (2007), Principles of gene microarray data analysis, *Adv.Exp.Med.Biol.* 593: 19-30

Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B (2008), Mapping and quantifying mammalian transcriptomes by RNA-Seq, *Nat.Methods* 5: 621-628

Nagalakshmi U, Waern K, Snyder M (2010), RNA-Seq: a method for comprehensive transcriptome analysis, *Curr.Protoc.Mol.Biol.* Chapter 4: Unit-13

Nebert DW (1994), Drug-metabolizing enzymes in ligand-modulated transcription, *Biochem.Pharmacol.* 47: 25-37

Nebert DW, Dieter MZ (2000), The evolution of drug metabolism, *Pharmacology* 61: 124-135

Nebert DW, Russell DW (2002), Clinical importance of the cytochromes P450, *Lancet* 360: 1155-1162

Nelson DR (1999), Cytochrome P450 and the individuality of species, *Arch.Biochem.Biophys.* 369: 1-10

Nelson DW, Honda BM (1985), Genes coding for 5S ribosomal RNA of the nematode *Caenorhabditis elegans. Gene* 38: 245–251.

Nelson DR, Kamataki T, Waxman DJ, Guengerich FP, Estabrook RW, Feyereisen R, Gonzalez FJ, Coon MJ, Gunsalus IC, Gotoh O, (1993), The P450 superfamily: update on new sequences, gene mapping, accession numbers, early trivial names of enzymes, and nomenclature, *DNA Cell Biol.* 12: 1-51

Nelson DR, Koymans L, Kamataki T, Stegeman JJ, Feyereisen R, Waxman DJ, Waterman MR, Gotoh O, Coon MJ, Estabrook RW, Gunsalus IC, Nebert DW (1996), P450 superfamily: update on new sequences, gene mapping, accession numbers and nomenclature, *Pharmacogenetics* 6: 1-42

Nelson DR, Zeldin DC, Hoffman SM, Maltais LJ, Wain HM, Nebert DW (2004), Comparison of cytochrome P450 (CYP) genes from the mouse and human genomes, including nomenclature recommendations for genes, pseudogenes and alternative-splice variants, *Pharmacogenetics* 14: 1-18

Neveu C, Charvet CL, Fauvin A, Cortet J, Beech RN, Cabaret J (2010), Genetic diversity of levamisole receptor subunits in parasitic nematode species and abbreviated transcripts associated with resistance, *Pharmacogenet.Genomics* 20: 414-425

Nikou D, Ranson H, Hemingway J (2003), An adult-specific CYP6 P450 gene is overexpressed in a pyrethroid-resistant strain of the malaria vector, *Anopheles gambiae, Gene* 318: 91-102

Ning Z, Cox AJ, Mullikin JC (2001), SSAHA: a fast search method for large DNA databases, *Genome Res.* 11: 1725-1729

Nisbet AJ, Redmond DL, Matthews JB, Watkins C, Yaga R, Jones JT, Nath M, Knox DP (2008), Stage-specific gene expression in *Teladorsagia circumcincta* (Nematoda: Strongylida) infective larvae and early parasitic stages, *Int.J.Parasitol.* 38: 829-838

O'Connor LJ, Walkden-Brown SW, Kahn LP (2006), Ecology of the free-living stages of major trichostrongylid parasites of sheep, *Vet.Parasitol.* 142: 1-15

O'Grady J, Kotze AC (2004), *Haemonchus contortus*: in vitro drug screening assays with the adult life stage, *Exp.Parasitol.* 106: 164-172

Parkinson J, Mitreva M, Whitton C, Thomson M, Daub J, Martin J, Schmid R, Hall N, Barrell B, Waterston RH, McCarter JP, Blaxter ML (2004), A transcriptomic analysis of the phylum Nematoda, *Nat.Genet.* 36: 1259-1267

Pedra JH, McIntyre LM, Scharf ME, Pittendrigh BR (2004), Genome-wide transcription profile of field- and laboratory-selected dichlorodiphenyltrichloroethane (DDT)-resistant *Drosophila, Proc.Natl.Acad.Sci.U.S.A* 101: 7034-7039

Pemberton KD, Barrett J (1989), The detoxification of xenobiotic compounds by *Onchocerca gutturosa* (Nematoda: *Filarioidea*), *Int.J.Parasitol.* 19: 875-878

Poulos TL (2005), Structural biology of heme monooxygenases, *Biochem.Biophys.Res.Commun.* 338: 337-345

Poulos TL, Finzel BC, Howard AJ (1987), High-resolution crystal structure of cytochrome P450cam, *J.Mol.Biol.* 195: 687-700

Precious WY, Barrett J (1989a), The possible absence of cytochrome P-450 linked xenobiotic metabolism in helminths, *Biochim.Biophys.Acta* 992: 215-222

Precious WY, Barrett J (1989b), Xenobiotic metabolism in helminths, *Parasitol.Today* 5: 156-160

Prichard R (2001), Genetic variability following selection of *Haemonchus contortus* with anthelmintics, *Trends Parasitol.* 17: 445-453

Prichard RK (1990), Anthelmintic resistance in nematodes: extent, recent understanding and future directions for control and research, *Int.J.Parasitol.* 20: 515-523

Qian H, Martin RJ, Robertson AP (2006), Pharmacology of N-, L-, and B-subtypes of nematode nAChR resolved at the single-channel level in *Ascaris suum*, *FASEB J.* 20: 2606-2608

Qian W, Zhang J (2008), Evolutionary dynamics of nematode operons: easy come, slow go, *Genome Res.* 18: 412-421

Ramakers C, Ruijter JM, Deprez RH, Moorman AF (2003), Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data, *Neurosci.Lett.* 339: 62-66

Ranz JM, Casals F, Ruiz A (2001), How malleable is the eukaryotic genome? Extreme rate of chromosomal rearrangement in the genus *Drosophila*, *Genome Res.* 11: 230-239

Redman E, Grillo V, Saunders G, Packard E, Jackson F, Berriman M, Gilleard JS (2008a), Genetics of mating and sex determination in the parasitic nematode *Haemonchus contortus*, *Genetics* 180: 1877-1887

Redman E, Packard E, Grillo V, Smith J, Jackson F, Gilleard JS (2008b), Microsatellite analysis reveals marked genetic differentiation between *Haemonchus contortus* laboratory isolates and provides a rapid system of genetic fingerprinting, *Int.J.Parasitol.* 38: 111-122

Redmond DL, Knox DP (2001), *Haemonchus contortus* SL2 trans-spliced RNA leader sequence, *Mol.Biochem.Parasitol.* 117: 107-110

Ringo J, Jona G, Rockwell R, Segal D, Cohen E (1995), Genetic variation for resistance to chlorpyrifos in *Drosophila melanogaster* (Diptera: Drosophilidae) infesting grapes in Israel, *J.Econ.Entomol.* 88: 1158-1163

Robinson MW, Lawson J, Trudgett A, Hoey EM, Fairweather I (2004), The comparative metabolism of triclabendazole sulphoxide by triclabendazole-susceptible and triclabendazole-resistant *Fasciola hepatica*, *Parasitol.Res.* 92: 205-210

Rufener L, Maser P, Roditi I, Kaminsky R (2009), *Haemonchus contortus* acetylcholine receptors of the DEG-3 subfamily and their role in sensitivity to monepantel, *PLoS.Pathog.* 5: e1000380

Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, Barrell B (2000), Artemis: sequence visualization and annotation, *Bioinformatics.* 16: 944-945

Saeed HM, Mostafa MH, O'Connor PJ, Rafferty JA, Doenhoff MJ (2002), Evidence for the presence of active cytochrome P450 systems in *Schistosoma mansoni* and *Schistosoma haematobium* adult worms, *FEBS Lett.* 519: 205-209

Saitou N, Nei M (1987), The neighbour-joining method: a new method for reconstructing phylogenetic trees, *Mol Biol Evol.* 4:406-25.

Sangster N, Batterham P, Chapman HD, Duraisingh M, Le Jambre L, Shirley M, Upcroft J, Upcroft P (2002), Resistance to antiparasitic drugs: the role of molecular diagnosis, *Int.J.Parasitol.* 32: 637-653

Sangster NC (1999), Anthelmintic resistance: past, present and future, *Int.J.Parasitol.* 29: 115-124

Sangster NC, Riley FL, Wiley LJ (1998), Binding of [^{3}H]*m*-aminolevamisole to receptors in levamisole-susceptible and -resistant *Haemonchus contortus*, *Int.J.Parasitol.* 28: 707-717

Sargison N, Scott P, Jackson F (2001), Multiple anthelmintic resistance in sheep, *Vet.Rec.* 149: 778-779

Sargison ND, Jackson F, Bartley DJ, Moir AC (2005), Failure of moxidectin to control benzimidazole-, levamisole- and ivermectin-resistant *Teladorsagia circumcinda* in a sheep flock, *Vet.Rec.* 156: 105-109

Sargison ND, Jackson F, Bartley DJ, Wilson DJ, Stenhouse LJ, Penny CD (2007), Observations on the emergence of multiple anthelmintic resistance in sheep flocks in the south-east of Scotland, *Vet.Parasitol.* 145: 65-76

Sargison ND, Jackson F, Wilson DJ, Bartley DJ, Penny CD, Gilleard JS (2010), Characterisation of milbemycin-, avermectin-, imidazothiazole- and benzimidazole-resistant *Teladorsagia circumcincta* from a sheep flock, *Vet.Rec.* 166: 681-686

Savas U, Griffin KJ, Johnson EF (1999), Molecular mechanisms of cytochrome P-450 induction by xenobiotics: An expanded role for nuclear hormone receptors, *Mol.Pharmacol.* 56: 851-857

Sayers G, Sweeney T (2005), Gastrointestinal nematode infection in sheep--a review of the alternatives to anthelmintics in parasite control, *Anim. Health Res.Rev.* 6: 159-171

Schlenke TA, Begun DJ (2004), Strong selective sweep associated with a transposon insertion in *Drosophila simulans, Proc.Natl.Acad.Sci.U.S.A* 101: 1626-1631

Silverman GA, Luke CJ, Bhatia SR, Long OS, Vetica AC, Perlmutter DH, Pak SC (2009), Modeling molecular and cellular aspects of human disease using the nematode *Caenorhabditis elegans*, *Pediatr.Res.* 65: 10-18

Skuce P, Stenhouse L, Jackson F, Hypsa V, Gilleard J (2010), Benzimidazole resistance allele haplotype diversity in United Kingdom isolates of *Teladorsagia circumcincta* supports a hypothesis of multiple origins of resistance by recurrent mutation, *Int.J.Parasitol.* 40: 1247-1255

Smith DA, Jones BC (1992), Speculations on the substrate structure-activity relationship (SSAR) of cytochrome P450 enzymes, *Biochem.Pharmacol.* 44: 2089-2098

Spieth J, Lawson D (2006), Overview of gene structure, *WormBook.* 1-10

Stein LD, Bao Z, Blasiar D, Blumenthal T, Brent MR, Chen N, Chinwalla A, Clarke L, Clee C, Coghlan A, Coulson A, D'Eustachio P, Fitch DH, Fulton LA, Fulton RE, Griffiths-Jones S, Harris TW, Hillier LW, Kamath R, Kuwabara PE, Mardis ER, Marra MA, Miner TL, Minx P, Mullikin JC, Plumb RW, Rogers J, Schein JE,

Sohrmann M, Spieth J, Stajich JE, Wei C, Willey D, Wilson RK, Durbin R, Waterston RH (2003), The genome sequence of *Caenorhabditis briggsae:* a platform for comparative genomics, *PLoS.Biol*. 1: E45

Tamura K, Dudley J, Nei M, Kumar S (2007), MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Molecular Biology and Evolution* 24:1596-1599.

Tanizawa Y, Kuhara A, Inada H, Kodama E, Mizuno T, Mori I (2006), Inositol monophosphatase regulates localization of synaptic components and behavior in the mature nervous system of *C. elegans*, *Genes Dev*. 20: 3296-3310

Taylor HR, Greene BM (1989), The status of ivermectin in the treatment of human onchocerciasis, *Am.J.Trop.Med.Hyg*. 41: 460-466

Taylor MA, Hunt KR, Goodyear KL (2002), Anthelmintic resistance detection methods, *Vet.Parasitol*. 103: 183-194

The *C. elegans* Sequencing Consortium (1998), Genome sequence of the nematode *C. elegans*: a platform for investigating biology, *Science* 282: 2012-2018

Thomas JH (2006), Analysis of homologous gene clusters in *Caenorhabditis elegans* reveals striking regional cluster domains, *Genetics* 172: 127-143

Thomas JH (2007), Rapid birth-death evolution specific to xenobiotic cytochrome P450 genes in vertebrates, *PLoS.Genet*. 3: e67

Tijet N, Helvig C, Feyereisen R (2001), The cytochrome P450 gene superfamily in *Drosophila melanogaster*: annotation, intron-exon organization and phylogeny, *Gene* 262: 189-198

Troell K, Engstrom A, Morrison DA, Mattsson JG, Hoglund J (2006), Global patterns reveal strong population structure in *Haemonchus contortus*, a nematode parasite of domesticated ruminants, *Int.J.Parasitol*. 36: 1305-1316

Urquhart BL, Tirona RG, Kim RB (2007), Nuclear receptors and the regulation of drug-metabolizing enzymes and drug transporters: implications for interindividual variability in response to drugs, *J.Clin.Pharmacol*. 47: 566-578

Urquhart GM, Armour J, Duncan JL, Dunn AM, Jennings FW (1996), *Veterinary Parasitology*, Blackwell Science, Glasgow, Scotland

Vanfleteren JR, Van de Peer Y, Blaxter ML, Tweedie SA, Trotman C, Lu L, Van Hauwaert ML, Moens L (1994), Molecular genealogy of some nematode taxa as based on cytochrome c and globin amino acid sequences. *Mol Phylogenet Evol*. 3:92-101.

van Wyk JA, Malan FS (1988), Resistance of field strains of *Haemonchus contortus* to ivermectin, closantel, rafoxanide and the benzimidazoles in South Africa, *Vet.Rec*. 123: 226-228

van Wyk JA, Malan FS, Gerber HM, Alves RM (1987), Two field strains of *Haemonchus contortus* resistant to rafoxanide, *Onderstepoort J.Vet.Res*. 54: 143-146

Velík J, Szotakova B, Baliharova V, Lamka J, Savlik M, Wsol V, Snejdrova E, Skalova L (2005), Albendazole repeated administration induces cytochromes P4501A and accelerates albendazole deactivation in mouflon (*Ovis musimon*), *Res.Vet.Sci.* 78: 255-263

von Samson-Himmelstjerna G, Blackhall W (2005), Will technology provide solutions for drug resistance in veterinary helminths?, *Vet.Parasitol.* 132: 223-239

von Samson-Himmelstjerna G, Buschbaum S, Wirtherle N, Pape M, Schnieder T (2003), TaqMan minor groove binder real-time PCR analysis of beta-tubulin codon 200 polymorphism in small strongyles (Cyathostomin) indicates that the TAC allele is only moderately selected in benzimidazole-resistant populations, *Parasitology* 127: 489-496

von Samson-Himmelstjerna G, Walsh TK, Donnan AA, Carriere S, Jackson F, Skuce PJ, Rohn K, Wolstenholme AJ (2009), Molecular detection of benzimidazole resistance in *Haemonchus contortus* using real-time PCR and pyrosequencing, *Parasitology* 136: 349-358

Waller PJ (1997), Anthelmintic resistance, *Vet.Parasitol.* 72: 391-405

Waller PJ, Rydzik A, Ljungstrom BL, Tornquist M (2006), Towards the eradication of *Haemonchus contortus* from sheep flocks in Sweden, *Vet.Parasitol.* 136: 367-372

Wang L, Dankert H, Perona P, Anderson DJ (2008), A common genetic target for environmental and heritable influences on aggressiveness in *Drosophila*, *Proc.Natl.Acad.Sci.U.S.A* 105: 5657-5663

Wang Z, Gerstein M, Snyder M (2009), RNA-Seq: a revolutionary tool for transcriptomics, *Nat.Rev.Genet.* 10: 57-63

Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, Agarwala R, Ainscough R, Alexandersson M, An P, Antonarakis SE, Attwood J, Baertsch R, Bailey J, Barlow K, Beck S, Berry E, Birren B, Bloom T, Bork P, Botcherby M, Bray N, Brent MR, Brown DG, Brown SD, Bult C, Burton J, Butler J, Campbell RD, Carninci P, Cawley S, Chiaromonte F, Chinwalla AT, Church DM, Clamp M, Clee C, Collins FS, Cook LL, Copley RR, Coulson A, Couronne O, Cuff J, Curwen V, Cutts T, Daly M, David R, Davies J, Delehaunty KD, Deri J, Dermitzakis ET, Dewey C, Dickens NJ, Diekhans M, Dodge S, Dubchak I, Dunn DM, Eddy SR, Elnitski L, Emes RD, Eswara P, Eyras E, Felsenfeld A, Fewell GA, Flicek P, Foley K, Frankel WN, Fulton LA, Fulton RS, Furey TS, Gage D, Gibbs RA, Glusman G, Gnerre S, Goldman N, Goodstadt L, Grafham D, Graves TA, Green ED, Gregory S, Guigo R, Guyer M, Hardison RC, Haussler D, Hayashizaki Y, Hillier LW, Hinrichs A, Hlavina W, Holzer T, Hsu F, Hua A, Hubbard T, Hunt A, Jackson I, Jaffe DB, Johnson LS, Jones M, Jones TA, Joy A, Kamal M, Karlsson EK, Karolchik D, Kasprzyk A, Kawai J, Keibler E, Kells C, Kent WJ, Kirby A, Kolbe DL, Korf I, Kucherlapati RS, Kulbokas EJ, Kulp D, Landers T, Leger JP, Leonard S, Letunic I, Levine R, Li J, Li M, Lloyd C, Lucas S, Ma B, Maglott DR, Mardis ER, Matthews L, Mauceli E, Mayer JH, McCarthy M, McCombie WR, McLaren S, McLay K, McPherson JD, Meldrim J, Meredith B, Mesirov JP, Miller W, Miner TL, Mongin E, Montgomery KT, Morgan M, Mott R, Mullikin JC, Muzny DM, Nash WE, Nelson JO, Nhan MN, Nicol R, Ning Z, Nusbaum C, O'Connor MJ, Okazaki Y, Oliver K, Overton-Larty E, Pachter L, Parra

G, Pepin KH, Peterson J, Pevzner P, Plumb R, Pohl CS, Poliakov A, Ponce TC, Ponting CP, Potter S, Quail M, Reymond A, Roe BA, Roskin KM, Rubin EM, Rust AG, Santos R, Sapojnikov V, Schultz B, Schultz J, Schwartz MS, Schwartz S, Scott C, Seaman S, Searle S, Sharpe T, Sheridan A, Shownkeen R, Sims S, Singer JB, Slater G, Smit A, Smith DR, 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 JP, Von Niederhausern AC, Wade CM, Wall M, Weber RJ, Weiss RB, Wendl MC, West AP, Wetterstrand K, Wheeler R, Whelan S, Wierzbowski J, Willey D, Williams S, Wilson RK, Winter E, Worley KC, Wyman D, Yang S, Yang SP, Zdobnov EM, Zody MC, Lander ES (2002), Initial sequencing and comparative analysis of the mouse genome, *Nature* 420: 520-562

Waxman DJ (1999), P450 gene induction by structurally diverse xenochemicals: central role of nuclear receptors CAR, PXR, and PPAR, *Arch.Biochem.Biophys.* 369: 11-23

Waxman DJ, Azaroff L (1992), Phenobarbital induction of cytochrome P-450 gene expression, *Biochem.J.* 281 (Pt 3): 577-592

Whitton C, Daub J, Quail M, Hall N, Foster J, Ware J, Ganatra M, Slatko B, Barrell B, Blaxter M (2004), A genome sequence survey of the filarial nematode *Brugia malayi:* repeats, gene discovery, and comparative genomics, *Mol.Biochem.Parasitol.* 137: 215-227

Wienkers LC, Heath TG (2005a), Predicting in vivo drug interactions from in vitro drug discovery data, *Nat.Rev.Drug Discov.* 4: 825-833

Williams JA, Hyland R, Jones BC, Smith DA, Hurst S, Goosen TC, Peterkin V, Koup JR, Ball SE (2004), Drug-drug interactions for UDP-glucuronosyltransferase substrates: a pharmacokinetic explanation for typically observed low exposure (AUCi/AUC) ratios, *Drug Metab. Dispos.* 32: 1201-1208

Xu C, Li CY, Kong AN (2005), Induction of phase I, II and III drug metabolism/transport by xenobiotics, *Arch.Pharm.Res.* 28: 249-268

Xu M, Molento M, Blackhall W, Ribeiro P, Beech R, Prichard R (1998), Ivermectin resistance in nematodes may be caused by alteration of P-glycoprotein homolog, *Mol.Biochem.Parasitol.* 91: 327-335

Yadav M, Singh A, Rathaur S, Liebau E (2010), Structural modeling and simulation studies of Brugia malayi glutathione-S-transferase with compounds exhibiting antifilarial activity: implications in drug targeting and designing, *J.Mol.Graph.Model.* 28: 435-445

Yates DM, Portillo V, Wolstenholme AJ (2003), The avermectin receptors of *Haemonchus contortus* and *Caenorhabditis elegans*, *Int.J.Parasitol.* 33: 1183-1193

Zeng Z, Andrew NW, Arison BH, Luffer-Atlas D, Wang RW (1998), Identification of cytochrome P4503A4 as the major enzyme responsible for the metabolism of ivermectin by human liver microsomes, *Xenobiotica* 28: 313-321

Zinser EW, Wolf ML, Alexander-Bowman SJ, Thomas EM, Davis JP, Groppi VE, Lee BH, Thompson DP, Geary TG (2002), Anthelmintic paraherquamides are

cholinergic antagonists in gastrointestinal nematodes and mammals, *J.Vet.Pharmacol.Ther*. 25: 241-250

Zuckerkandl E, Pauling L (1965), Evolutionary divergence and convergence in proteins, Evolving Genes and Proteins. Academic Press, New York.