finalreport

Project code: P.PSH.0207

Submitted by: Frank Firgaira
Genetic Technologies Corporation

Date published: July 2009

ISBN: ABN 39 081 678 364

PUBLISHED BY
Meat & Livestock Australia Limited
Locked Bag 991
NORTH SYDNEY NSW 2059

Novel approaches for the control of parasitic nematodes in sheep

Meat & Livestock Australia acknowledges the matching funds provided by the Australian Government to support the research and development detailed in this publication.

Abstract

The goal of the project was to conduct fundamental research with a focus on finding new solutions to internal parasitic diseases of sheep using genomic, bioinformatic and chemical approaches. The characteristics of genes associated with reproduction and/or development were defined. The achievements provide a foundation for the development of new compounds with activity against selected nematodes. The predominant focus was on the barber's pole worm, *Haemonchus contortus*.

The key achievements of the project were:

- Construction of molecular archives for *Haemonchus contortus*;
- Microarray expression profiling of this nematode;
- Bioinformatic analysis of ESTs and prediction of functions based on comparison of molecules in current databases;
- Identification and characterization of a subset of seven molecules involved in key biological pathways and considered (based on phylogenetic analysis) to be 'specific' to nematodes;
- Isolation of full-length cDNAs and/or genes and cloned for subsequent expression
- Consolidation of a technology pipeline for drug target and drug discovery;
- Testing of a range of chemical compounds and identification of a subset with nematocidal effects.

Using the integrated technology platform developed, a multi-phase, genomic-bioinformatic pipeline was employed to identify candidate drug targets, taking advantage of nematode and mammalian sequence information available in current gene databases, and incorporating gene function data from the free-living nematode *C. elegans* to deduce a set of 'nematode-specific' molecules, whose interruption or disruption is known to adversely affect nematode development, reproduction and/or survival. The focus was on the trichostrongylid nematode *H. contortus*, which is of economic importance in Australia and worldwide. Any outcomes for this nematode also have major implications for other related nematodes of livestock.

Executive Summary

Parasites of livestock cause diseases of major economic importance worldwide. The current cost of several key parasites to agriculture represents a substantial burden on farm profitability and the livestock industry as a whole. Parasitic diseases of livestock are usually controlled using chemotherapeutic agents (anthelmintics). The excessive use of these agents has resulted in serious problems with resistance against them. Consequently, there is a major interest in the development of new classes of nematocidal compounds for controlling parasites, which has major economic, biotechnological and commercial implications.

The main objectives of the present project were:

1. To identify and characterise genes associated with reproduction and development in *Haemonchus contortus* (as a model for gastrointestinal nematodes of livestock).
2. To evaluate and validate products of key genes as targets for novel anti-parasite compounds and/or vaccines for use in sheep.

The aims were:

- To logically extend previous work and utilize the skills and infrastructure in the application of 'selective' expressed sequence tag (EST) gene libraries and genomic analyses in the *H. contortus* system; and
- To provide a foundation for developing sustainable approaches for controlling parasites of sheep and other livestock presently not controlled effectively using broad-spectrum antiparasitic drugs.

All objectives and aims were achieved. These included:

- Construction of molecular archives for *H. contortus* and complete characterisation and validation of orthologues identified previously in project AHW.022;

- Sequencing of a subset of ESTs and microarray analyses for *H. contortus*;

- Bioinformatic analysis of *H. contortus* ESTs and prediction of functions based on comparison of molecules in current databases;

- Initial characterisation of 5 pre-validated molecules for *H. contortus*;

- High throughput screen – one candidate (industry partner);

- Enhancement of the technology pipeline for drug target and drug discovery.

Using an integrated technology platform, a multi-phase, genomic pipeline was used to identify candidate drug targets and drugs, taking advantage of nematode and mammalian sequence information available in current gene databases and incorporating

gene function and genetic interaction data for *Caenorhabditis elegans* to produce a set of 'nematode-specific' genes, whose interruption or disruption adversely affects nematode development, reproduction and/or survival.

In accordance with "*Backing Australia's Ability*", the benefits flowing from the present project have been: (i) Major focus on animal health biotechnology through the development of new anti-parasite compounds and on the sustainable control of important parasites of sheep and other livestock; (ii) Increased profitability of agricultural animal production and livestock industries as well as improved welfare of agricultural animals; (iv) Consolidation of a technology platform for 'next generation' applications in the genomics of pathogens of global importance; (v) Capturing of the benefits and outcomes of fundamental research and strengthen the links between fundamental and applied research; (vi) Commitment to enhancing the quality and quantity of scientifically skilled people in biology and biotechnology.

The current project has provided the prospect of developing niche markets in the supply of key veterinary biologicals to an international market. The projected animal health market in Australia is estimated at ~ 1 billion dollars. The international market for animal health products is estimated at $10-20 billion per year. Of this, the market for antiparasitic drugs for animals is ~$5-10 billion per year. In this project, genomic, bioinformatic and chemoinformatic technologies were established for the exploration of key biochemical pathways which are specific to parasitic roundworms (Strongylida) and for the discovery of new drug targets in socio-economically important nematodes of livestock. The outcomes provide a foundation for finding new intervention strategies, focused on benefits to agricultural producers nationally and internationally. This project was a partnership among MLA, Genetic Technologies Limited (GTG) and The University of Melbourne. Within the framework of PSHIP.207, potential drug targets were defined and characterised, and three compounds with nematocidal activity were discovered and protected through patents. This outcome has significant commercial and biotechnological implications and demonstrates the effectiveness of the pipeline developed.

1. Background

Parasites of livestock have major economic impact worldwide due to the diseases they cause. The current cost of key parasites to agriculture represents a substantial burden on farm profitability. For example, the annual cost associated with parasitism in livestock in Australia has been estimated at ~1 billion dollars (McLeod, 1995). Thus, there are major economic gains to be made in Australian agriculture by enhancing the control of important parasitic diseases of livestock.

Parasitic diseases of animals are mainly controlled through the reliance on chemotherapeutic agents (anthelmintics). Even with strategic treatments, this type of control is only partially effective and expensive. The excessive use of such agents has resulted in serious problems with anthelmintic resistance. Therefore, the ongoing development of novel and improved control strategies is crucial to the sustainability of the livestock (and associated) industries (Geary et al., 2004). There is major global interest in the development of alternative means of controlling parasitic organisms, which has major economic, biotechnological and commercial implications.

The possibilities include the rational development of antiparasitic compounds and/or vaccines to specifically interrupt parasite growth, development and/or reproduction, built on a detailed understanding of disease, host-parasite relationships and the molecular biology of the parasites themselves. There is a need for more effective and sustainable solutions to major parasitic disease problems. The current research project provides a foundation for the development of new anthelmintics for international application. The projected animal health market in Australia has been estimated at $1-2 billion (based on estimates provided by the pharmaceutical companies Bayer and Pfizer). The current international market for animal health products is ~$10-20 billion per year. Of this, the market for antiparasitic drugs for animals is ~$5-10 billion per year. The economic benefits to farming and animal health should have both national and international impact.

2. Project Objectives and Milestones

Given the emerging problems with controlling parasites of livestock and the unique opportunities and prospects which genomic methods offer, the primary objectives of the present project were to:

1. Identify and characterise genes which are associated with development and survival in *Haemonchus contortus* (order Strongylida).
2. Evaluate and validate products of key genes as targets for novel anti-parasite compounds.

In achieving these primary objectives, the present project aimed to:

- To logically extend previous work and utilize the skills and infrastructure in the application of 'selective' expressed sequence tag (EST) gene libraries and genomic analyses in the *H. contortus* system; and
- To provide a foundation for developing sustainable approaches for controlling parasites of sheep and other livestock presently not controlled effectively using broad spectrum antiparasitic drugs.

The project proceeded according to the following milestones over a three year period:

Milestones and achievement criteria	
1	Complete meetings with senior personnel within at least 2 global pharmaceutical companies and complete critical assessment of project to ensure it is designed to be potentially attractive for ultimate transfer to a pharmaceutical company. Modify project plan as necessary.
2	Construction of molecular archives for *H. contortus* and complete characterisation and validation of selected *T. vitrinus* genes and gene products identified in project AHW.022.
3	Sequencing of a subset of ESTs and microarray analyses for *H. contortus*, and development of a high throughput screen for at least one gene product.
4	Bioinformatic analyses for *H. contortus*
5	Initial characterisation of five pre-validated molecules
6	Full characterisation of these molecules (5)
7	First high throughput screen – one candidate (industry partner)

3. Methodologies

In the present project, an integrated genomic-bioinformatic-chemoinformatic approach was used for the discovery of drug targets and drugs, focused on screening for genes associated with reproduction and/or development in trichostrongylid nematodes. Such genes and their products are of major interest as they represent optimal targets for specifically interfering with a parasite's survival. The approach taken was (1) to define subsets of key molecules for *Haemonchus contortus* (order Strongylida), (2) to isolate and characterise expressed sequence tags (ESTs), (3) to predict their function based on bioinformatic and genomic analyses in the model organism *C. elegans*, and then (4) to select candidate molecular targets for the discovery of novel compounds.

H. contortus was selected because of its major economic importance worldwide. This parasite provided a unique and powerful system for studying fundamental biological processes. It has a short, direct life-cycle (21 days), produces large numbers of progeny and can be readily maintained as a laboratory line. The identification of target molecules in *H. contortus* that are conserved in function in nematodes is likely to have application to other parasitic nematodes (within the order Strongylida) of livestock.

Following the selection and initial characterisation of ESTs, they were categorised using a bioinformatics approach. Partial sequences were compared with those in the GenBank™ non-redundant database, using the Basic Local Alignment Search Tool (BLAST 2.0) program from the National Center for Biotechnology Information (http://www.ncbi.nim.nih.gov/BLAST/) and BCM Search Launcher from the Human Genome Center, Baylor College of Medicine (http://dot.imgen.bcm.tmc.edu), and putative functions and roles predicted based on P-scores.

Following the selection and initial characterisation of all of the ESTs, full-length genes (for those of most biotechnological interest) were obtained and sequenced. Transcription profiles were analysed (using appropriate control genes) by microarray analysis and/or quantitative PCR. Particular attention was always paid to homologues/orthologues of the free-living nematode *C. elegans* for which extensive functional data are available (www.wormbase.org).

C. elegans represents one of the best characterised multicellular organisms. Its complete genome sequence has been determined (The *C. elegans* Sequencing Consortium, 1998). Of the ~20000 genes, ~ 60% of genes appear to be common with parasitic nematodes (Blaxter, 1998). Given that a number of key biological processes appear to be conserved among a wide range of nematodes, *C. elegans* has been considered a powerful model for parasitic nematodes (Geary et al., 2004; Gasser and Newton, 2000). Other features, including the nematode's short life cycle (three days at 25 °C), ease of propagation of well-defined lines using a simple bacterial food source (*Escherichia coli*), relatively small genome size, ability to produce clonal progeny from hermaphrodites and to cross hermaphrodites with males, make *C. elegans* a useful comparative system. As strongylid nematodes (the subject of this project) are considered to be closely related to *C. elegans* (Blaxter, 1998), the chance that a cloned gene from *H. contortus* has an orthologue in *C. elegans* is relatively high, with the exception of genes involved in host-parasite interactions. As there are no effective culture systems available for the propagation and maintenance of the entire life cycle of strongylid nematodes *in vitro*, *C. elegans* was used as the reference organism.

Production of parasites

Merino lambs (males; 8-12 weeks of age), maintained under helminth-free conditions, were infected intraruminally with 8,000-10,000 infective third-stage larvae (L3) of *H. contortus* (McMaster strain) or other trichostrongylid nematodes. The patency of the infection (24-30 days) was established by the detection of trichostrongylid eggs in the faeces using the McMaster flotation method (MAFF, 1977). For the collection of worms, infected lambs were euthanased with an overdose of pentobarbitone sodium (Lethobarb, Virbac Pty. Ltd.), administered intravenously 30 days post infection. Adult worms were collected from the abomasum or small intestine, washed extensively in (37 °C) phosphate-buffered saline (PBS) snap-frozen in liquid nitrogen for subsequent storage at -70 °C.

Isolation, purification, treatment and storage of nucleic acids

Total genomic DNA was extracted from ~0.5 g of worm material using a small-scale sodium-dodecyl-sulphate (SDS)/proteinase K extraction procedure, followed by mini-column (Wizard® Clean-Up, Promega) purification. The specific identity of the parasite material was confirmed unequivocally by PCR-amplification of the second internal transcribed spacer (ITS-2) of nuclear ribosomal DNA from genomic DNA and automated sequencing. The sequences determined were required to be identical to the ITS-2 sequence with GenBank accession no. X78803. Total RNA was extracted separately from adult female and male *H. contortus* (homogenized under liquid nitrogen using a mortar and pestle) employing the TriPure isolation reagent® (Roche Molecular Biochemicals). RNA yields were estimated spectrophotometrically, and the integrity of RNA was confirmed via the detection of discrete 18S and 28S ribosomal RNA bands on ethidium bromide-stained gels. Each RNA sample (~10 μg) was treated with 2 U of *DNase* I (Promega) and incubated at 37°C for 30 min prior to heat denaturation of the enzyme (75°C for 5 min). Both DNA and RNA samples were stored at –70 °C.

Isolation and sequencing of full-length cDNAs

The following criteria were used for the selection of ESTs for which full-length genes were isolated and characterized: [1] High microarray signal and high quality EST sequence; [2] EST (at the amino acid level) has a high identity/similarity to *C. elegans* homologues (OR other nematodes > other invertebrates > mammals); [3] Highest homology to an inferred product which plays a key biological role in *C. elegans* and is linked directly or indirectly (preferably *via* an enzyme) to a particular biochemical, physiological or molecular pathway in this nematode. Conservation suggests its presence in other parasitic nematodes; [4] Gene-silencing shown in *C. elegans*. Functionality in a free-living nematode suggests functionality in the parasitic nematode. All information regarding *C. elegans* is available in WormBase (http://www.wormbase.org/).

Full-length genes were isolated from cDNA libraries or using 5’ and 3’-rapid amplification of cDNA ends (RACE). “Full-length” cDNA libraries were constructed from adult *H. contortus* total RNA using the SMART™ PCR cDNA library construction kit

(Clontech), as described in the manufacturer's instructions with the following modifications: (i) 19 cycles were determined to be optimal for the amplification of cDNA by long distance PCR, following first strand synthesis; (ii) column chromatography of the PCR products prior to digestion with the endonuclease *Sf I* was replaced by purification using the Wizard® PCR Preps DNA purification system (Promega) to minimise cDNA loss. For library screening, 25,000 plaque forming units (pfu) were plated and lifted on to positively-charged nylon membranes (Roche Molecular Biochemicals). The membranes were pre-hybridised at 55 °C for 1 h in pre-hybridised buffer [0.25 M $NaHPO_4$, pH 6.5, 1 mM EDTA, 7 % (w/v) sodium dodecyl-sulphate (SDS) and 100 μg/ml sonicated herring sperm DNA (Roche Molecular Biochemicals)]. The denatured probe, labelled by random priming with [α-^{32}P]dCTP (GeneWorks), was added to the pre-hybridisation solution and incubated overnight at 55 °C. The filters were washed 3 times in 2x SSC (0.3 M NaCl, 30 mM Na citrate, pH 7.0), 0.1 % SDS at 60 °C for 20 min and then subjected to autoradiography for 16-24 h. Positively-hybridizing plaques were picked, the phage was eluted into 500 µl of SM buffer and re-screened by hybridization to ensure that they were clonal.

Usually, overlapping cDNA fragments were generated from total RNA from adult worms using 5' and 3'-rapid amplification of cDNA ends (RACE) using gene-specific primers, employing the SMART™ RACE cDNA Amplification Kit (BD Biosciences), according to the manufacturer's protocol. The cDNAs were ligated into the pGEM-T Easy vector (Promega), and recombinant plasmids were transformed to competent *Escherichia coli* (strain JM109) (10^8 colony forming units/μg) *via* heat shock and grown overnight at 37 °C on Luria Bertani (LB) plates containing 10 mg/ml ampicillin, 0.5 mM isopropyl-β-D-thiogalactopyranoside (IPTG) and 80 µg/ml X-gal (5-bromo-4chloro-3-indolyl-β–galactosidase. Plasmid DNA was isolated from recombinant clones and column-purified (Qiagen) from overnight cultures and sequenced in both directions using vector primers (M13 and SP6), employing Big Dye Terminator v.3.1 chemistry in an automated ABI-PRISM sequencer (Applied Biosystems). Based on the resultant sequences, oligonucleotide primers were designed to amplify the full-length gene from *H. contortus* adult cDNA, which was subsequently cloned and sequenced as described previously.

Transcriptional profiling by microarray analysis

A total of ~22,000 *H. contortus* ESTs, representing 12 libraries of various lifestages and tissues, were publicly available at www.sanger.ac.uk (The Wellcome Trust Sanger Institute) and www.nematodes.org (NemBase, Washington University). The ESTs were clustered using the analysis pipeline ESTExplorer (http://estexplorer.biolinfo.org) yielding 3743 representative sequences that were used in the computational design of oligonucleotide probes for the microarray using the web-based program eArray (http:// opengenomics.com/dnamicroarray_earray.aspx; Agilent Technologies). This selection yielded 5655 probes representing 1885 target sequences. Two probe sets were produced from the available high quality probes, one containing probes representing 52 preselected control genes which were replicated 15 times within the array, and another containing all probes meeting the selection criteria, to fill the microarray. Total RNA was extracted separately from each gender of *H. contortus* (under liquid nitrogen, employing a sterile mortar and pestle) using the TriPure isolation reagent® (Roche Molecular Biochemicals) according to the manufacturer's instructions.

Worms were obtained from independent infections of three different sheep. The integrity of the RNA was verified using a Bioanalyzer 2100 (Agilent Technologies), and the yield determined using the NanoDrop ND-1000 UV-VIS spectrophotometer (NanoDrop Technologies). Each RNA sample (~10 μg) was incubated at 37 °C for 30 min with 2 U of DNase I (Promega), prior to heat denaturation of the enzyme (75 °C for 5 min) and then frozen at -70 °C. Complementary RNA (cRNA) probes were produced from 500 ng of total RNA from female or male *H. contortus* (representing each batch produced). Each RNA was reverse transcribed using the T7 primer and MMLV reverse transcriptase. Using first strand cDNA, cRNA was then amplified and labelled simultaneously with either cyanine-3 (Cy3) or cyanine-5 (Cy5) dCTP, such that both cRNA from both genders was labeled with both dyes, using the T7 promoter primer and T7 RNA polymerase (Low RNA Input Linear Amplification Kit PLUS, Two-Color kits, Agilent). Each cRNA reaction was then purified over a RNeasy mini spin column (Qiagen), and the integrity verified using the Bioanalyzer 2100 (Agilent). ''Spike-in'' control samples were produced using the RNA Spike-in kit (Agilent). For each of three biological replicates, 500 ng each of Cy3-labelled female cRNA and Cy5-labelled male cRNA, and vice versa, were mixed and hybridized (in duplicate) to the array slides. Microarray slides were hybridized (in quadruplicate, using 250 μl of hybridization mix) at 65 °C for 17 h. Slides were then disassembled into Gene Expression (GE) wash buffer 1, washed again for 1 min with the same buffer, 1 min in prewarmed (37 °C) GE wash buffer 2 and then for 1 min in stabilisation and drying solution (Agilent). A ''dye flip'' was carried out to control for any bias in hybridisation signal between the Cy-labelled cRNA probes (produced for two distinct mRNA populations). The slides were scanned using a microarray scanner B (Agilent). Standardized data for statistical analyses were produced by loading the resultant tag image format files (tiff) into the image analysis program Feature Extraction 9.1 (Agilent Technologies, USA). The red- and green-processed signals from these data were log2-transformed and then analysed statistically using the Student's t-test in a spreadsheet (Excel). The microarray data were analysed for differential hybridization (P2.0-fold) between male and female *H. contortus* cRNAs, using a P-value of ≤0.001.

Transcriptional profiling by quantitative PCR

Double-stranded cDNA was synthesized from total RNA separately from each stage and sex of *H. contortus* using reverse transcriptase (Superscript III, Invitrogen). Briefly, 5 μg of total RNA were added to 14 μl of H_2O and 1 μl of oligo d(T)n = 12-18 primer (0.5 μg/μl), heated to 70 °C for 10 min and chilled on ice. First and second strand cDNAs were synthesized via the addition of 4 μl of first strand cDNA buffer (250 mM Tris-HCl, pH 8.3, 375 mM KCl and 15 mM $MgCl_2$), 2 μl of 0.1 M dithiothreitol, and 1 μl of 10 mM of each dNTP, followed by an incubation at 25 °C for 10 min, 42 °C for 50 min and 70 °C for 15 min. The transcripts were amplified from individual cDNAs by quantitative (q)PCR. The PCR amplification of a portion (187 bp) of the large subunit (28S) ribosomal RNA using primers 28S1/F (5'-GCATTAGCTCTCGCGTTACC-3') and 28S3/R (5'-GAGAGGGACAGCAGGTTCAC-3'), previously determined to be present equally in each developmental stage and sex in a related parasitic nematode, *Oesophagostomum dentatum*, was used as a positive control. Samples without template (no-DNA controls) were included in each qPCR run. For each sample, 0.5 μg of cDNA

was subjected to qPCR (20 µl) using the SYBR® GreenER™ qPCR SuperMix Universal (Cat. no. 11762-100, Invitrogen) in a Rotor-Gene™ 3000 thermal cycler (Corbett Research) under the following conditions: one cycle of 50 °C for 5 min and of 95 °C for 10 min, followed by 40 cycles of 95 °C for 15 s, 60 °C for 30 s and 72 °C for 30 s. Each sample was tested in triplicate, using a calibrator (28S) as well as positive and no-template controls. The specificity and identity of individual amplicons were verified by melting curve analysis and subsequent direct, automated sequencing using the same primers (individually) as employed for the PCR. Relative transcriptional differences were calculated from normalised values. The specificity and identity of individual amplicons were confirmed by direct sequencing using the same primers (separately) as employed for their amplification.

Green fluorescent protein (GFP)-based localization of transcription in C. elegans

Constructs were made using the GFP 'promoter-less' vector pPD95.75 (originally from the laboratory of Andrew Fire) of *C. elegans*. A standard protocol was used to transform *C. elegans* with each of the two constructs by microinjection. The N2 strain was co-injected into the gonad with the construct (50 ng/µl) and the plasmid pRF-4 containing the *rol-6* gene (50 ng/µl) and 100 ng/µl of plasmid DNA (pBluescript, Stratagene). For each construct, 20 young adult hermaphrodites were injected and then maintained at room temperature (22 °C) for 3 days to produce progeny. F1s with rolling behaviour ('rollers') were selected and then examined for GFP fluorescence. Individual fluorescent F1 rollers were transferred to a nematode growth medium (NGM) plate seeded with *E. coli* (strain OP50) to produce F2s. Progeny that transmitted transgenes to and beyond the F2 generation were maintained as independent transgenic lines.

RNA interference in C. elegans

RNA interference (RNAi) was performed in *C. elegans* using the feeding method. In brief, the full-length cDNAs from *C. elegans* and from *H. contortus* were each amplified by PCR from first-strand cDNA. The amplicons were purified and cloned into the RNAi plasmid vector L4440, and *E. coli* strain HT115 transformed. Overnight cultures were grown in LB containing 50 µg/ml ampicillin and seeded on to NGM agar with 1 mM IPTG and 50 µg/ml of ampicillin. Seeded plates were incubated at room temperature overnight prior to use. Five fourth-stage larvae (L4s) of the *C. elegans* wild-type (N2) strain were each placed on a seeded plate and incubated at 25 °C overnight. They were then transferred to a fresh plate, seeded with HT115. After allowing the worms to lay eggs for 5 h, they were removed from the plate. The plates were incubated at 25 °C overnight. The hatching of the eggs, development of the larvae, and visible phenotypes of the adults were subsequently examined. Each RNAi treatment was performed in triplicate. The phenotypes were recorded using a differential interference contrast (DIC) microscope.

Bioinformatic analyses

Nucleotide sequences were assembled using the program CAP3 (http://bio.ifom-firc.it/ASSEMBLY/assemble.html) and compared with those in non-redundant databases using the BLAST 2.0 suite of programs from the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/BLAST), the Sanger Centre (www.sanger.ac.uk/Projects/Celegans/) and the Parasite Genome database (www.ebi.ac.uk/parasites/parasite_blast_server.html) to confirm the identity of the genes isolated. The conceptual translation of cDNA into amino acid sequences was performed using the selection "translate", available at http://bioinformatics.org/sms/. Protein motifs were identified by scanning the databases Pfam (www.sanger.ac.uk/Software/Pfam/) and PROSITE (www.expasy.ch/tools/scnpsit1.html). Signal sequences were predicted using SignalP v2.0, available at the Center for Biological Sequence Analysis (www.cbs.dtu.dk/services/SignalP/). Amino acid sequence alignments were carried out using Clustal W and adjusted by manual inspection. The transcription profiles for *C. elegans* were obtained from existing microarray data available via WormBase (www.wormbase.org). Promoter elements in the 5'-UTR were predicted using the transcription element search system (TESS; available at the website http://www.cbil.upenn.edu/cgi-bin/tess/tess).

Amino acid sequences were subjected to phylogenetic analyses. The amino acid sequences were aligned using Clustal W, and phylogenetic analyses conducted using a maximum parsimony (MP) and neighbor-joining (NJ) methods employing PAUP* v4.0b10. Characters were weighted equally and treated as unordered. A heuristic search with tree bissection-reconnection (TBR) branch swapping was used to infer the shortest trees. The length, consistency index (C.I.), excluding uninformative characters, and retention index (R.I.) of the most parsimonious trees were recorded. A bootstrap analysis (using 1000 replicates) was conducted using heuristic searches and TBR-branch swapping with the MulTrees option to determine the relative support for clades in the consensus tree.

Genetic interactions were predicted using the method available at http://geneorienteer.org/. In brief, genomic data (regarding interactions, phenotypic, expression and gene ontology; cf. http://www.geneontology.org) from *C. elegans* gene homologues/orthologues, also incorporating data from the vinegar fly, yeast, mouse and human, were integrated using a naïve Bayesian model to predict genetic interactions among *C. elegans* genes. The predicted networks resulting from the analyses were saved in a graphic display file (gdf) format and were examined using the graph exploration system available at http://graphexploration.cond.org/. Images were labeled and saved in the joint photographic experts group (jpeg) format.

Subcloning of molecules

Full-length cDNAs encoding six genes were cloned into the glutathione S-transferase (GST) fusion protein expression vector pGEX-5X-1 or pGEX-3TEX (Amersham Biosciences) and transformed into *E. coli* BL21 (Promega). Positive clones were identified by two methods, sequencing and pilot expression experiments. Sequencing used (separately) the 5' vector primer and a 3' sequence-specific primer for each gene. This ensured that the entire gene was cloned into the plasmid and that the reading frame for protein expression was correct. Pilot expression experiments were

performed by inoculating single colonies into 3 ml of Luria-Bertani (LB) medium containing 100 μg/ml ampicillin and incubation at 37°C with vigorous shaking for 3-4 h. Prior to induction with isopropyl-β-D-thiogalactopyranoside (IPTG), a 1 ml sample was taken for use as un-induced controls and for making a glycerol stock of each clone. Cultures were induced with 0.1–0.4 mM IPTG and grown for a further 2 h at 37°C. Expression of some proteins required a second induction, using half of the initial concentration of IPTG, for a further 2 h. Cells were then collected by centrifugation at 4200 *xg* for 5 min at 4°C, washed in 2 ml of ice-cold phosphate buffered saline (PBS) and then resuspended in 1 ml of PBS. Sonication of the cells was performed for 2 x 5 sec intervals on ice, followed by centrifugation to remove cell debris. Fusion protein was isolated from the supernatant by incubation with 75 μl of 50% glutathione cross-linked to agarose. The resin was collected by centrifugation and washed five times with ice-cold PBS to remove unwanted proteins. For analysis, the fusion protein was removed from the resin by boiling for 5 min in sodium dodecyl-sulphate (SDS) polyacrylamide gel electrophoresis (PAGE) sample buffer (62.5 mM Tris, pH 6.8; 10% glycerol; 0.0125% bromophenol blue; 4%SDS), containing 5% β-mercaptoethanol. After centrifugation at 10000 *xg*, samples of 5–10 μl were analysed by SDS-PAGE and also transferred to PVDF membrane for analysis by Western blot. Membranes were blocked and incubated with a 1:500 dilution of polyclonal rabbit anti-GST antibody (Zymed Laboratories Inc. or Sigma-Aldrich Pty Ltd) and a 1:5,000 dilution of goat anti-rabbit horseradish peroxidase (HRP) conjugate (Sigma Aldrich Pty Ltd). The HRP conjugate was detected using the Pierce SuperSignal West Pico Chemiluminescent Substrate. Membranes were exposed to X-ray film for 10 s to 5 min intervals depending on the strength of the signals. Many pilot experiments were performed for each protein to define expression conditions.

Screening of chemical compounds in the H. contortus *larval development assay* (LDA)

The strain used for these assay was Haecon 5, eggs of which were provided by the Department of Primary Industries, Attwood, Victoria. This strain is known to be 66% resistant to benzimidazoles, but is susceptible to macrocyclic lactones (P. Presidente, personal communication); therefore, the susceptibility of this strain to macrocyclic lactones justifies the use of moxidectin as a positive control compound in the assay.

Ten grams of faeces were crushed and suspended in ~100 ml of sucrose solution (specific gravity 1.2), sieved (mesh size: ~ 1 mm) and transferred to a petri-dish. Strips of commercially available overhead transparency (Kodak) were placed on the surface of the suspension (to allow the eggs to stick) and removed after 45 min. The eggs were collected by washing them from the transparency strips with water into a 50 ml centrifuge tube. The tube was filled to 50 ml with water and the eggs collected by centrifugation at 1,000 x g for 10 min, the supernatant removed and the eggs collected in ~ 500 ml of water. Eggs were enumerated and the suspension adjusted with water to 200 eggs/20 ml.

Compounds were tested at five different concentrations: 6.25, 12.5, 25, 50 and 100 μM. Dilutions of each compound (10 μl in total) were performed in 1.5 ml microcentrifuge tubes, 1 ml of molten 2% agar added, the tube vortexed and the agar aliquoted (150 μl) into the wells of a 96-well microtitre plate. DMSO (1% v/v) was used in a number of wells as solvent-only controls (negative controls) whilst cydectin was used as a positive control. Concentrations of moxidectin used were: 12.5, 25, 50 and 100 μM.

Two to three replicates were performed for each concentration of each compound. Approximately 200 eggs (20 μl) were then added to each well. The plates were then sealed with 'breathable' adhesive films and then incubated overnight at 27 ^{o}C. Plates were checked the following morning and afternoon to ensure that the majority of eggs had hatched. Eggs were examined for any indication of an ovicidal effect (evidenced by a significant number of unhatched eggs compared with the control wells). Following the hatching of most eggs, 15 μl of nutritive medium was added to feed the larvae. Nutritive medium was prepared as follows: 3 ml of 10 x Earle's balanced salt solution (EBSS) [potassium chloride (KCl) 53 mg/l, sodium bicarbonate ($NaHCO_3$) 261.9 mg/l, sodium chloride (NaCl) 1172.4 mg/l, sodium phosphate monobasic (NaH_2PO_4-H_2O) 10.1 mg/ml] was added to 27 ml of yeast extract solution and the pH adjusted to 5.4-5.6 by adding bicarbonate. The yeast extract solution was prepared as follows: 1 g of yeast extract was added to 90 ml of 0.85% physiological saline and autoclaved for 20 min at 121^{o} C. Following further 6 days of incubation, the number of L3s that had developed in each well was determined. The reproducibility of "kill" was assessed on different days.

4. Results

4.1 Meetings of GTG with senior personnel of at least 2 global pharmaceutical companies to ensure that the project is designed to be potentially attractive for ultimate transfer to a pharmaceutical company (milestone 1)

Under the terms of the agreement, the first intellectual property was not expected to flow from the program until 2010. Professor Gasser's group made exceptional progress. Two provisional patent applications were lodged by GTG in November 2006, covering the identification of a novel target enzyme that, when inhibited, kills parasitic nematodes which infect sheep. Professor Gasser developed a collaboration with A/Professor Adam McCluskey of the University of Newcastle, who is an expert in studying the human equivalent of phosphatase inhibitors. He contributed to initial collaboration through chemicals consisting of two distinct chemical structures which have been shown to be safe to human cells in tissue cultures. Preliminary work showed that some chemicals were nematocidal.

Four pharmaceutical companies were approached by GTG, with the aim of assessing the potential of these products. Two companies expressed considerable interest in exploring these chemicals in more detail. Under the terms of a Non-Disclosure Agreement, they were forwarded pdf's of the two provisional patent applications which they reviewed within the following 3 weeks. They expressed an interest in testing the chemicals.

4.2 Construction of EST libraries for *H. contortus* and complete characterisation and validation of selected *T. vitrinus* genes and gene products identified in project AHW.022 (milestone 2)

Under the terms of the agreement, the milestones were to finalize the characterization of selected *Trichostrongylus vitrinus* molecules and to analyse the EST data sets for *Haemonchus contortus* (the blood-feeding barber's pole worm of sheep). Both of these milestones were achieved.

The emphasis was on logically extending previous work from project AHW.022 to *H. contortus*. The bioinformatic and genomics investigations of *H. contortus* have commenced and proceeding well. To conduct the bioinformatics component, a comprehensive workflow system for EST data management and analysis, ESTExplorer, was constructed and critically evaluated. The pipeline uses a 'distributed control approach' in which the most appropriate bioinformatics tools are implemented over different dedicated processors. Specific repeat masking and conceptual translation are in-built. ESTExplorer accepts a set of raw ESTs which can be analysed using programs selected by the user. After pre-processing and assembly, the dataset is annotated at the nucleotide and protein levels, following conceptual translation. Users may optionally provide ESTExplorer with assembled contigs for annotation purposes. Functionally annotated contigs/ESTs can be analysed individually. The overall outputs are gene ontologies, protein functional identifications in terms of mapping to protein domains and metabolic pathways.

ESTExplorer was applied successfully to annotate large EST datasets from parasitic nematodes (and other organisms) and to identify novel genes as potential targets for parasite intervention. ESTExplorer runs on a Linux cluster and is available to the scientific community. ESTExplorer has been used for the detailed analysis of all of the EST data (~22,000 ESTs representing 12 different cDNA libraries) available for *Haemonchus contortus*. Also, a detailed analysis of differential transcription between adult male and female *H. contortus* using an oligonucleotide microarray platform is being undertaken. A bioinformatic-microarray approach was employed for the analysis of representative ESTs from *H. contortus*. Following cluster analysis, ~1900 representative ESTs (rESTs) were selected, to which oligonucleotides (three per EST) were designed and spotted onto a microarray. Using a range of key controls and replicates, this microarray was hybridized with cyanine-dye labelled cRNA probes synthesized from RNA from female or male adults of *H. contortus*. Differential hybridization was displayed for ~300 of the rESTs, and a detailed analysis of the data undertaken. In spite of significant problems and delays with one of the service providers (microarray analysis), outstanding progress was made by the investigators and 'within budget'.

4.3 Sequencing of a subset of ESTs and microarray analyses for *H. contortus* and development of high throughput screen for at least one *T. vitrinus* gene product (milestone 3)

Analysis of a subset of ESTs and microarray analyses for H. contortus

The cluster analysis of the 21,967 publicly available ESTs from *H. contortus* yielded 1885 (consensus and singleton) sequences to which 5657 specific oligonucleotides (i.e., three per EST) could be designed for the array (using stringent criteria). The ESTs were classified into different categories based on bioinformatic analysis against a wide range of public databases (full list available from the authors). Of the 1885 ESTs, 60% had similarity to homologues of *C. elegans*, 10% to those of other organisms, including some parasitic nematodes, and 30% had no homologues/orthologues in any other organism.

A critical evaluation of intra-array (among three oligonucleotides) and inter-array variability (between replicates) was performed by calculating the covariance among the three oligonucleotides on the same slide and among slides. Consistent hybridization among all three probes was achieved for 1775 of the sequences (91%). Intra- and inter-array variability values were low (<10%). The Feature Extraction 9.1 (Agilent Technologies) software was used, precisely according the recommendations of the supplier, to extract and normalize the data from the microarray analysis.

Microarray analysis of the 1885 representative ESTs (rESTs) using triplicate oligonucleotides per rEST revealed differential hybridization for 301 of them (16%). Of these, 165 (55%) had a significantly greater signal intensity for female cRNA and 136 (45%) for male cRNA. The other 1584 ESTs (84%) did not have a differential hybridization signal. Of the 301 ESTs with a statistically significant higher signal intensity (2.0 to 8.0-fold) upon hybridisation with female cRNA, 135 (45%) were predicted to have *C. elegans* homologues. For 74 (~55%) of these homologues, 'non-wildtype' RNAi phenotypes, including Adl, Age, Bmd, Clr, Cyk, Dpy, Egl, 33 types of Emb, Gro, Let, Lva, Mlt, Mul, Muv, Nmo, Pch, Prl, Pvl, Ric, Rol, Rup, Sck, Sle, Spn, Ste, Stp and/or Unc (see WormBase at www.wormbase.org) have been recorded in *C. elegans*. The phenotypes Cyk, various Emb types (n = 24), Mul, Nmo, Rol and Spn were unique to the female and not represented in the male data set. These 74 homologues included molecules involved in nucleic acid synthesis and function (23%), reproduction, gametogenesis and embryonic/larval development (50%), vitellogenins (5%), protein folding (3%), mitochondrial associated proteins (4%), carbohydrate biosynthesis and metabolism (4%), locomotion and locomotory behaviour (3%), cytokinesis (3%) and a range of molecules associated with metabolism (5%). Of the 136 ESTs with statistically higher signal intensities (2.0 to 7.8-fold) for male *H. contortus*, 108 (79.4%) had *C. elegans* homologues. For 39 of these homologues, 'non-wildtype' RNAi phenotypes, including Adl, Age, Bli, Bmd, Ced, Clr, Dpy, 9 different types of Emb, Gro, Let, Lon, Lva, Mlt, Muv, Pch, Prl, Pvl, Ric, Rup, Sck, Sle, Slu, Ste, Stp and/or Unc (see WormBase: www.wormbase.org) have been recorded in *C. elegans*. The phenotypes Bli, Ced, Lon and Slu were unique to the male and not represented in the female data set. These 39 homologues included molecules involved in reproduction, gametogenesis and embryonic/larval development (40%), cell death and ageing (15%), protein production and proteolysis (10%), locomotion and locomotory behaviour (10%), nucleic acid synthesis and function (5%), sensory perception (5%), glycolysis (2.5%), potassium ion

transport (5%) and other biochemical processes (7.5%). Of all 58 gender-enriched ESTs with no known homologue in *C. elegans*, 24 ESTs (41%) had homologues in other nematodes, 4 (7%) ESTs had homologues in various other organisms, and 30 ESTs (52%) had no significant homology to any sequence presently deposited in the databases interrogated.

In order to independently verify the hybridization results from the microarray, PCR and sequencing analyses of a selected subset of ESTs (11.5%) using specific primer sets were conducted. Having verified the specificity and identity of individual amplicons, real-time PCR results were reproducible (based on multiple runs on different days) and ~92% concordant with the microarray analysis.

Development of a high throughput screen for at least one Trichostrongylus vitrinus *gene product*

Fourteen compounds synthesized in the laboratory of our collaborator A/Professor McCluskey were shown to have some activity in killing *H. contortus* and *T. vitrinus* in preliminary *in vitro* trials. Fort Dodge had expressed an interest in testing these chemicals under a Materials Transfer Agreement. Accordingly, samples of the chemicals were forwarded to Fort Dodge for testing. In the first instance, the effects of the chemicals against *C. elegans* were tested. Fort Dodge defines an active compound as one that kills 100% of the worms within 4 or 24 h at 100 ppm. Two compounds were active under this definition. These two proceeded to an assay against *T. colubriformis* in gerbils (it was not stated whether the species was *Gerbillus gerbillus* or *Meriones unguiculatus*). An active compound is defined in this test as one that removes 60% of the worms relative to untreated controls. One of the compounds removed 53% in the first experiment, but was less active in the second trial. Since it is possible that this particular compound may specifically target male worms (based on molecular evidence for *Tv-stp-1*), it is possible that all male worms (i.e. ~50% of the infecting population of worms) were removed from the infected gerbils. However, no attempt was made to establish the sex/es of the worms in the gerbils at necropsy, making it impossible to critically appraise and interprete the findings. The 14 samples were also tested in the blowfly larvae assay at 1 ppm, 10 ppm, 50 ppm, and 100 ppm. Samples that kill the larvae at 50 ppm are typically advanced to the secondary screens. Fort Dodge interpreted and concluded that none of the chemicals was active according to their definitions, but did offer to retest the samples (given the inability to interpret some of the results) or test other active compounds should we produce any.

The results of the trials enabled some prediction to be made as to which chemical structures were active and which were least/not active. This enabled A/Professor McCluskey to design a new series of chemicals for further testing. While there was initial enthusiasm by the pharmaceutical companies, unsatisfactory communication and insufficient feedback on the testing did not enable scientific conclusions to be drawn from the testing by the companies (Fort Dodge and Pfizer) and did not result in effective cooperation.

4.4 Bioinformatic analyses for *H. contortus* (milestone 4)

All transcripts showing significant differential hybridization were identified and subjected to BLASTx (NCBI: www.ncbi.nlm.nih.gov) and BLASTn (EMBL-EBI Parasite Genome Blast Server: www.ebi.ac.uk) analyses to identify putative homologues in *C. elegans,* other nematodes and other organisms (E-value of ≤ 1e-05). WormBase (www.wormbase.org) was interrogated extensively for relevant information about *C. elegans* homologues/orthologues, including RNAi phenotypes, transcriptomic, proteomic and interactome data. ESTs with homologues in *C. elegans* and other nematodes were also analysed using the KEGG Orthology-Based Annotation System (KOBAS) (kobas.cbi.pku.edu.cn), which predicted the biochemical pathways in which molecules are involved. The open reading frames (ORFs) inferred from selected ESTs with orthologues in *C. elegans* were also subjected to "secretome analysis" using the program SignalP v.2.0 www.cbs.dtu.dk/services/SignalP/), employing both the neural network and hidden Markov models to predict signal peptides and/or anchors. Also, transmembrane domains were predicted using the program TMHMM (www.cbs.dtu.dk/services/TMHMM/), and subcellular localization employing the program WoLF PSORT (http://wolfpsort.org/).

Genetic interactions were predicted based on information available for *C. elegans* orthologues of gender-enriched genes from *H. contortus*. Genomic data (regarding interactions, phenotypic, expression and gene ontology) from *C. elegans* gene homologues/orthologues, also incorporating data from the vinegar fly, yeast, mouse and human, were integrated using a naïve Bayesian model to predict genetic interactions among *C. elegans* genes. The predicted networks resulting from the analyses were saved in a graphic display file (gdf) format and were examined using the graph exploration system available at http://graphexploration.cond.org/. Images were labelled and saved in the joint photographic experts group (jpeg) format.

A complex interaction network of the *C. elegans* genes (n = 101) represented by ESTs significantly differentially transcribed between female and male *H. contortus* was predicted. A focused analysis of a subset of data was performed. Specifically, genes relating to ESTs with the highest level of transcription (> 3.0-fold) in female (n = 12) or male (n = 15) *H. contortus* were selected for the analysis. Statistically highly significant interactions were predicted for 15 of the *C. elegans* orthologous genes, of which six were represented in female and nine in male *H. contortus*. Representing the female, *C. elegans* genes C36A4.9 and C04F6.1 (*vit-5*) were both predicted to be linked to lipid transport and metabolism; C31C9.2 to amino acid transport and metabolism; F21D5.1 to carbohydrate transport and metabolism; T27E9.1 (*tag-61*) to energy production and conversion; and W07E6.2 probably associated with ribosome biogenesis and cell division in the germline. These six genes were predicted to interact directly with 83 other genes associated with embryonic and larval development (14.5%), information storage and processing (29%), cellular processes and signalling (26%) and metabolism (14.5%); the precise function of some interactors (16%) is presently unknown. Some interactors linked female with male genes or *vice versa*. Representing the male, genes T18H9.2 (*asp-2*) and H22K11.1 (*asp-3*) were both predicted to be associated with post translational modification, protein turnover and chaperones, whereas gene F14B4.2 was linked to carbohydrate transport and metabolism, F42A8.2 (*tag-55*) to energy production and conversion, C47D12.6 (*trs-1*) to translation, ribosomal structure and biogenesis,

F58G1.3 to signal transduction, C36H8.1 to a major sperm protein, T05G5.3 (*cdk-1*) to cell-cycle progression in both meiosis and mitosis, and F57F5.1 represented a cysteine protease. These nine genes were predicted to interact directly with 210 other genes associated with embryonic and larval development (6%), information storage and processing (31%), cellular processes and signaling (42%) and metabolism (11%); the precise function of some interactors (10%) is presently unknown. Some interactors linked female with male genes or *vice versa*. Eleven of the 15 genes were predicted to be linked directly or indirectly to *cdk-1* (meiosis and mitosis), predicted to be a pivotal point associated with 286 genes linked to amino acid transport and metabolism (2%), carbohydrate transport and metabolism (2%), cell cycle control, cell division and chromosome partitioning (19%), chaperones (1%), chromatin structure and dynamics (12%), coenzyme transport and metabolism (1.4%), cytoskeleton (5%), embryonic and/or larval development (6%), energy production and conversion (4%), intracellular trafficking, secretion and vesicular transport (2.4%), lipid transport and metabolism (0.5%), mitochondrial proteins (0.5%), nucleotide transport and metabolism (1%), replication, recombination and repair (6%), RNA processing and modification (2%), signal transduction (8%), transcription (7%), translation, ribosomal structure and biogenesis (4%), and unknown function (10%).

An appraisal of the present genetic interaction data of representative genes revealed that there is a close and logical interaction of key genes linked to amino acid, carbohydrate or lipid transport and metabolism; energy production and conversion; translation, ribosomal structure/biogenesis; and those associated with meiosis and/or mitosis in the germline during oogenesis or spermatogenesis. While there is presently no proof that the gene interaction network inferred from *C. elegans* data is the same as in the parasitic nematode *H. contortus*, the network does provide a useful predictive tool. The present data set showed that 60% of the enriched ESTs from *H. contortus* have a homologue in *C. elegans*, suggesting that some pathways utilized during nematode development and/or reproduction are relatively conserved.

There is major scope in exploring relatively conserved nematode-specific molecules using a combined bioinformatic-genomic approach by comparison with all available data for *C. elegans*, particularly that genome sequence data for a range of strongylid nematodes, including *H. contortus,* which will become available in the future (see http://www.genome.gov/11007951 and http://www.sanger.ac.uk/Projects/Helminths/). Nonetheless, caution was needed in the interpretation of data sets, given the biological differences between *C. elegans* and *H. contortus*. The present transcriptomic and predictive bioinformatic approach should serve as a powerful tool to assist in fundamental investigations of molecular biological pathways for 50-60% of genes from strongylid nematodes, which will support the pre-validation of molecules as possible targets for drugs against socioeconomically important parasitic nematodes.

4.5 Initial characterisation of 5 pre-validated molecules (milestone 5)

A genomic-bioinformatic approach had been applied previously to the selection of promising candidate drug targets from the ovine nematode, *Trichostrongylus vitrinus*, *via* detailed comparative analysis with *C. elegan*s (project AHW.022). This set of ESTs included groups of potential targets, including a kinase, phosphatase and translocase, thereby demonstrating the utility of the present approach. According to the agreement, we built on this work to isolate and characterize a selected subset of *H. contortus* genes and predicted gene products. Toward this milestone, five full-length cDNAs were isolated, sequenced and characterized:

Gene	cDNA (bp)	Predicted function based on comparison with *Caenorhabditis elegans*
Hc-ant-1	**894**	**Transport of ADP/ATP into and out of cells**
Hc-pbs-5	**855**	**Proteasome system**
Hc-rio-1	**1467**	**Phosphorylation of serine/threonine residues. Involved in many cellular processes**
Hc-vha-1	**2601**	**Function yet unclear**
Hc-stp-1	**951**	**Dephosphorylation of serine/threonine residues. Involved in many cellular processes**

The sequences of these cDNAs, their inferred peptide sequences and their closest matches based on database comparisons are listed in the following:

ADP/ATP translocase (*Hc-ant-1*)

ATGCCGGAGAAGTTTGATACGAAGAAGTTTTTTATCGACTTGGCCTCAGGAGGTACCGCTGCTGCCATCTCCAAGACCGC
TGTCGCCCCTATTGAGCGAGTAAAGCTTCTTCTTCAGGTACAAGATGCATCAAGCACTATTGCTGTCGATAAGCGATACA
AGGGAATTATCGATGTATTGGTCCGAGTTCCTAAAGAGCAGGGTTTCGCCGCTTTGTGGCGAGGAAACCTCGCTAACGTT
ATCCGATATTTCCCCACGCAAGCTTTGAACTTCGCCTTCAAGGATTCCTACAAGAAGATTTTCTTGGAGGGTCTTGACAA
GAAGAAGGACTTCTGGAAGTTCTTTGCCGGAAATTTGGCTTCCGGAGGTGCTGCTGGAGCTACCTCTTTGTGCTTTGTAT
ATCCCCTTGATTTTGCTCGAACCCGTCTTGCCGCTGACGTTGGCAAGGGAGCAGGTCGTGAATTCAAAGGATTACTTGAC
TGTTTGATCAAGGTTACAAAATCCGATGGTCCTATCGGTTTGTACCGTGGCTTCTTCGTCTCCGTACAAGGTATCATCAT
CTACCGTGCCGCATACTTCGGATTGTTCGATACTGCCAAGATGTTGTTGGCATCCGAAGGCAAACTCAACTTCTTCGTTG
CATGGGCCATCGCTCAGGTTGTCACGGTCAGATCAGGTATCTTGTCTTATCCTTGGGATACTGTTCGTCGTCGTATGATG
ATGCAGTCTGGAAGAAAAGATATTTTGTACAAGAACACACTTGACTGTGCCAAGAAGATCATCAAGAACGAAGGAATTGG
AGCCATGTTCAAGGGAGCATTGTCCAATGTTTTCCGTGGAACAGGAGGAGCTTTGGTTTTGGCCATCTATGACGAGATTC
AAAAATTCATTTGA

```
   1 ATGCCGGAGAAGTTTGATACGAAGAAGTTTTTTATCGACTTGGCC
     M  P  E  K  F  D  T  K  K  F  F  I  D  L  A
  46 TCAGGAGGTACCGCTGCTGCCATCTCCAAGACCGCTGTCGCCCCT
     S  G  G  T  A  A  A  I  S  K  T  A  V  A  P
  91 ATTGAGCGAGTAAAGCTTCTTCTTCAGGTACAAGATGCATCAAGC
     I  E  R  V  K  L  L  L  Q  V  Q  D  A  S  S
 136 ACTATTGCTGTCGATAAGCGATACAAGGGAATTATCGATGTATTG
     T  I  A  V  D  K  R  Y  K  G  I  I  D  V  L
 181 GTCCGAGTTCCTAAAGAGCAGGGTTTCGCCGCTTTGTGGCGAGGA
     V  R  V  P  K  E  Q  G  F  A  A  L  W  R  G
 226 AACCTCGCTAACGTTATCCGATATTTCCCCACGCAAGCTTTGAAC
     N  L  A  N  V  I  R  Y  F  P  T  Q  A  L  N
 271 TTCGCCTTCAAGGATTCCTACAAGAAGATTTTCTTGGAGGGTCTT
     F  A  F  K  D  S  Y  K  K  I  F  L  E  G  L
 316 GACAAGAAGAAGGACTTCTGGAAGTTCTTTGCCGGAAATTTGGCT
     D  K  K  K  D  F  W  K  F  F  A  G  N  L  A
 361 TCCGGAGGTGCTGCTGGAGCTACCTCTTTGTGCTTTGTATATCCC
     S  G  G  A  A  G  A  T  S  L  C  F  V  Y  P
 406 CTTGATTTTGCTCGAACCCGTCTTGCCGCTGACGTTGGCAAGGGA
     L  D  F  A  R  T  R  L  A  A  D  V  G  K  G
 451 GCAGGTCGTGAATTCAAAGGATTACTTGACTGTTTGATCAAGGTT
     A  G  R  E  F  K  G  L  L  D  C  L  I  K  V
 496 ACAAAATCCGATGGTCCTATCGGTTTGTACCGTGGCTTCTTCGTC
     T  K  S  D  G  P  I  G  L  Y  R  G  F  F  V
 541 TCCGTACAAGGTATCATCATCTACCGTGCCGCATACTTCGGATTG
     S  V  Q  G  I  I  I  Y  R  A  A  Y  F  G  L
 586 TTCGATACTGCCAAGATGTTGTTGGCATCCGAAGGCAAACTCAAC
     F  D  T  A  K  M  L  L  A  S  E  G  K  L  N
 631 TTCTTCGTTGCATGGGCCATCGCTCAGGTTGTCACGGTCAGATCA
     F  F  V  A  W  A  I  A  Q  V  V  T  V  R  S
 676 GGTATCTTGTCTTATCCTTGGGATACTGTTCGTCGTCGTATGATG
     G  I  L  S  Y  P  W  D  T  V  R  R  R  M  M
 721 ATGCAGTCTGGAAGAAAAGATATTTTGTACAAGAACACACTTGAC
     M  Q  S  G  R  K  D  I  L  Y  K  N  T  L  D
 766 TGTGCCAAGAAGATCATCAAGAACGAAGGAATTGGAGCCATGTTC
     C  A  K  K  I  I  K  N  E  G  I  G  A  M  F
 811 AAGGGAGCATTGTCCAATGTTTTCCGTGGAACAGGAGGAGCTTTG
     K  G  A  L  S  N  V  F  R  G  T  G  G  A  L
 856 GTTTTGGCCATCTATGACGAGATTCAAAAATTCATTTGA 894
     V  L  A  I  Y  D  E  I  Q  K  F  I  *
```

Blast results

ref|NP_001022799.1|	Temporarily Assigned Gene name family mem...	504	2e-141
emb|CAE73690.1|	Hypothetical protein CBG21201 [Caenorhabditis br	504	3e-141
emb|CAE60169.1|	Hypothetical protein CBG03723 [Caenorhabditis br	502	1e-140
ref|NP_501440.1|	Temporarily Assigned Gene name family member...	500	4e-140
ref|NP_501727.1|	K01H12.2 [Caenorhabditis elegans] >emb|CAA92...	499	6e-140
emb|CAA53718.1|	ADP/ATP translocase [Caenorhabditis elegans]	480	4e-134
ref|NP_491927.1|	Temporarily Assigned Gene name family member...	464	2e-129
gb|AAD30505.1|AF130365_1	ADP/ATP translocase [Ascaris suum]	447	5e-124
ref|NP_511110.1|	Adenine nucleotide translocase 2 CG1683-PA, ...	405	2e-111
sp|O27238|ADT1_ANOGA	ADP,ATP carrier protein 1 (ADP/ATP trans...	401	2e-110

Proteasome beta subunit 5 *(Hc-pbs-5)*

ATGTGGGGTTGTGGTTTTGATGAACGCGATGAAGAAAGTACGGATCTAGCAGTCATCAAGCAACAATTTATGCACGAGCCTATAATAAG
CTCCTTCGCCTTTCCTCAACCTCCTTTGGGTATTCAAGCCAAGGATTTTGTCGCAACCCATTTTGGTAAAGATGCAAAAAATATGCAGT
TTCGGAAGGGTACAACTACACTCGCTTTCATCTATGAACCGGCTACCGCTAACGACAAGGGTGGTATTGTCGTCGCGGTCGACTCGAGG
GCTTCTTCAGGAGAATACATTTCTTCGAAATCTGTTCTGAAAATCTTGGATATTGGTGATCGGATGGTTGCTACCATGGCTGGTGGCGC
TGCAGACTGTCAGTTTTGGACGAGGACAGTCGCTAAATATTGCAATTTGTTTGAACTGCGGGAAAAAACTCAAATAACTGTGAGCGCCG
CCAGCAAATACTTTGCCAACGTTCTCTACGGATACCGTGGCATGGGCCTCTCTGTGGGATCTATGATCGCTGGCTATGACAAGAGAGGG
CCACAGATTTTTAAAGTGGATAGTGATGGGGATAGATGTCAATTGCAAGTGTGCAGCGTTGGTTCAGGTTCTTTGAATGCCTACGGAAT
TCTTGATACGCATTATAAAAGAAAAATGACCGATGAAGAAGCGCTGAAGTTAGGCCGCCGGGCGATTATGCATGCCACCTACAGGGATT
CTGGATCTGGCGGTGTTTGTAATATGGTGCACATCACTCCGACAGAGAAACGTCGTCTACCACCGATTGATGTAAGCAAGTTGTGGTAT
GAGTTTTCTGCCGAACTTGGTCGTGATATTGCGTACGAACCTCGCGATGATTGA

```
      1 ATGTGGGGTTGTGGTTTTGATGAACGCGATGAAGAAAGTACGGAT
        M  W  G  C  G  F  D  E  R  D  E  E  S  T  D
     46 CTAGCAGTCATCAAGCAACAATTTATGCACGAGCCTATAATAAGC
        L  A  V  I  K  Q  Q  F  M  H  E  P  I  I  S
     91 TCCTTCGCCTTTCCTCAACCTCCTTTGGGTATTCAAGCCAAGGAT
        S  F  A  F  P  Q  P  P  L  G  I  Q  A  K  D
    136 TTTGTCGCAACCCATTTTGGTAAAGATGCAAAAAATATGCAGTTT
        F  V  A  T  H  F  G  K  D  A  K  N  M  Q  F
    181 CGGAAGGGTACAACTACACTCGCTTTCATCTATGAACCGGCTACC
        R  K  G  T  T  T  L  A  F  I  Y  E  P  A  T
    226 GCTAACGACAAGGGTGGTATTGTCGTCGCGGTCGACTCGAGGGCT
        A  N  D  K  G  G  I  V  V  A  V  D  S  R  A
    271 TCTTCAGGAGAATACATTTCTTCGAAATCTGTTCTGAAAATCTTG
        S  S  G  E  Y  I  S  S  K  S  V  L  K  I  L
    316 GATATTGGTGATCGGATGGTTGCTACCATGGCTGGTGGCGCTGCA
        D  I  G  D  R  M  V  A  T  M  A  G  G  A  A
    361 GACTGTCAGTTTTGGACGAGGACAGTCGCTAAATATTGCAATTTG
        D  C  Q  F  W  T  R  T  V  A  K  Y  C  N  L
    406 TTTGAACTGCGGGAAAAAACTCAAATAACTGTGAGCGCCGCCAGC
        F  E  L  R  E  K  T  Q  I  T  V  S  A  A  S
    451 AAATACTTTGCCAACGTTCTCTACGGATACCGTGGCATGGGCCTC
        K  Y  F  A  N  V  L  Y  G  Y  R  G  M  G  L
    496 TCTGTGGGATCTATGATCGCTGGCTATGACAAGAGAGGGCCACAG
        S  V  G  S  M  I  A  G  Y  D  K  R  G  P  Q
    541 ATTTTTAAAGTGGATAGTGATGGGGATAGATGTCAATTGCAAGTG
        I  F  K  V  D  S  D  G  D  R  C  Q  L  Q  V
    586 TGCAGCGTTGGTTCAGGTTCTTTGAATGCCTACGGAATTCTTGAT
        C  S  V  G  S  G  S  L  N  A  Y  G  I  L  D
    631 ACGCATTATAAAAGAAAAATGACCGATGAAGAAGCGCTGAAGTTA
        T  H  Y  K  R  K  M  T  D  E  E  A  L  K  L
    676 GGCCGCCGGGCGATTATGCATGCCACCTACAGGGATTCTGGATCT
        G  R  R  A  I  M  H  A  T  Y  R  D  S  G  S
    721 GGCGGTGTTTGTAATATGGTGCACATCACTCCGACAGAGAAACGT
        G  G  V  C  N  M  V  H  I  T  P  T  E  K  R
    766 CGTCTACCACCGATTGATGTAAGCAAGTTGTGGTATGAGTTTTCT
        R  L  P  P  I  D  V  S  K  L  W  Y  E  F  S
    811 GCCGAACTTGGTCGTGATATTGCGTACGAACCTCGCGATGATTGA 855
        A  E  L  G  R  D  I  A  Y  E  P  R  D  D  *
```

emb|CAE72557.1|	Hypothetical protein CBG19741 [Caenorhabditis br	458	2e-127
ref|NP_493558.1|	Proteasome Beta Subunit family member (pbs-5...	450	5e-125
ref|XP_970194.1|	PREDICTED: similar to CG12323-PA, isoform A ...	231	5e-59
gb|EAT34534.1|	proteasome subunit beta type 5,8 [Aedes aegypti]	228	3e-58
gb|AAY98499.1|	LMP X [Swiftia exserta]	225	2e-57
ref|XP_559226.2|	ENSANGP00000027573 [Anopheles gambiae str. P...	222	2e-56
ref|XP_001361733.1|	GA11556-PA [Drosophila pseudoobscura] >gb...	222	2e-56
gb|AAK92808.1|	putative proteasome epsilon chain precursor [A...	221	3e-56
ref|NP_172765.1|	PBE1 (20S proteasome beta subunit E1); pepti...	221	3e-56
dbj|BAA10931.1|	LMPX of hagfish [Myxine glutinosa]	221	3e-56

RIO protein kinase (*Hc-rio1*)

ATGGAGAATATCCTCGATATGAACGTTAACGTCATACCTGGTGGCTTAGATTCGGATTGTTATAGTGATGTCGACTCTCA
GGAAGATATTCTAGGGAATGATTATCCAAGTATAAGCCATGATAGCGACGAGGCTGATGACGACATAGCCGACTTCACTG

```
ATGAGATTGGAGATTTCACAAAGAAGTACAATGCCTTTCGCATGCACATTTCTGGACCAAACAGCCAGCACAATGCAGGC
CCAAATTTCGCAGAAACAACAGCTGATTCCACCGTTGAAAAAAAGAGGAAACGCGTCAAGGACAGAGCAGATCGTGCAAC
AGTTGAACAGGTACTAGATCCACGCACAAGACTCGTTTTGTTTCGACTTCTCCAGCGTGGAACCTTGACCAATATTCACG
GTTGCATTTCAACAGGCAAAGAAGCCAATGTTTATCACACCACTGATGAAAAAGGTTCCTTAGCAGTGAAGATATACAAA
ACGAGCATATTGACATTTAAGGATCGCGAGCGCTATGTGGCCGGTGAGTATAGATATCGAACCGGCTACTGCAAACACAA
TCCTCGAAAAATGGTAGCAGTTTGGGCGGAAAAGGAGATGCGAAATTTGTTACGAATGCATCAAGCTGGTCTCCCTGTCC
CAAAACCTCTTTTGTTAAAGGGACATGTACTGGTCACGGAATTCATTGGTCGTGATGGATGGGGTGCTCCACTTTTGAAG
AATGCCACCTTGTCATTAGAGGCTGCCGAGAAGTTATATCTTCAACTCGTACGCGACATGCGTACATTATATCGTACATG
TAAATTAGTTCATGCCGATCTCTCTGAATATAACACTCTCGTCCTTGATGATCGCTTGTTCATAATCGATGTCTCGCAGT
CTGTTGAACACGACCATCCTCACGCCTTAGATTTTCTGAGATCCGACTGTAACAACGTTTGCAAGTTCTTCAGAGGCTTG
GGAGTACCTGTGTTACCGGTATCAAAGTTATTTGAGCTCATAGTTGATCCCCTTATCAAGGATAGCGATGTCACCAATTG
GTTAGAAAAGCGAACACTTGATCCATCAGACGATGCCCTCTTTATGAATGCGTTCATCCCTCATAAACTCGATCACGTCC
TACATTTTGAGAGAGACAGTAAATTACTGAAGGCTGGTGAAGAGGCGAACAATCCGTTCCAGAATATCATTTCCAAGGTC
GATGTGTTGGGTCAAGGTTTTGAGGAGCGCGTAGTTTCTTCATCTGATGACGAATCTATTTCGCAAGGCCATTCTGACGG
CGATGGGAAGATCGTAGTGGAGACGAAAGGGAGGGCCACCAGTGGCAAACACTTCAGGGACAAAGACGAGACGCCCGAAC
AACGGAAAGTAAGAAAACAACTGATAAAGGAAGAGAAACGGGAGGCTCGGAAAACCAAAATTCCTAAGCATGTGAAGAAA
CGCGCTCATCGGCAGCACATGAAATAG
```

```
   1 ATGGAGAATATCCTCGATATGAACGTTAACGTCATACCTGGTGGC
     M  E  N  I  L  D  M  N  V  N  V  I  P  G  G
  46 TTAGATTCGGATTGTTATAGTGATGTCGACTCTCAGGAAGATATT
     L  D  S  D  C  Y  S  D  V  D  S  Q  E  D  I
  91 CTAGGGAATGATTATCCAAGTATAAGCCATGATAGCGACGAGGCT
     L  G  N  D  Y  P  S  I  S  H  D  S  D  E  A
 136 GATGACGACATAGCCGACTTCACTGATGAGATTGGAGATTTCACA
     D  D  D  I  A  D  F  T  D  E  I  G  D  F  T
 181 AAGAAGTACAATGCCTTTCGCATGCACATTTCTGGACCAAACAGC
     K  K  Y  N  A  F  R  M  H  I  S  G  P  N  S
 226 CAGCACAATGCAGGCCCAAATTTCGCAGAAACAACAGCTGATTCC
     Q  H  N  A  G  P  N  F  A  E  T  T  A  D  S
 271 ACCGTTGAAAAAAAGAGGAAACGCGTCAAGGACAGAGCAGATCGT
     T  V  E  K  K  R  K  R  V  K  D  R  A  D  R
 316 GCAACAGTTGAACAGGTACTAGATCCACGCACAAGACTCGTTTTG
     A  T  V  E  Q  V  L  D  P  R  T  R  L  V  L
 361 TTTCGACTTCTCCAGCGTGGAACCTTGACCAATATTCACGGTTGC
     F  R  L  L  Q  R  G  T  L  T  N  I  H  G  C
 406 ATTTCAACAGGCAAAGAAGCCAATGTTTATCACACCACTGATGAA
     I  S  T  G  K  E  A  N  V  Y  H  T  T  D  E
 451 AAAGGTTCCTTAGCAGTGAAGATATACAAAACGAGCATATTGACA
     K  G  S  L  A  V  K  I  Y  K  T  S  I  L  T
 496 TTTAAGGATCGCGAGCGCTATGTGGCCGGTGAGTATAGATATCGA
     F  K  D  R  E  R  Y  V  A  G  E  Y  R  Y  R
 541 ACCGGCTACTGCAAACACAATCCTCGAAAAATGGTAGCAGTTTGG
     T  G  Y  C  K  H  N  P  R  K  M  V  A  V  W
 586 GCGGAAAAGGAGATGCGAAATTTGTTACGAATGCATCAAGCTGGT
     A  E  K  E  M  R  N  L  L  R  M  H  Q  A  G
 631 CTCCCTGTCCCAAAACCTCTTTTGTTAAAGGGACATGTACTGGTC
     L  P  V  P  K  P  L  L  L  K  G  H  V  L  V
 676 ACGGAATTCATTGGTCGTGATGGATGGGGTGCTCCACTTTTGAAG
     T  E  F  I  G  R  D  G  W  G  A  P  L  L  K
 721 AATGCCACCTTGTCATTAGAGGCTGCCGAGAAGTTATATCTTCAA
     N  A  T  L  S  L  E  A  A  E  K  L  Y  L  Q
 766 CTCGTACGCGACATGCGTACATTATATCGTACATGTAAATTAGTT
     L  V  R  D  M  R  T  L  Y  R  T  C  K  L  V
 811 CATGCCGATCTCTCTGAATATAACACTCTCGTCCTTGATGATCGC
     H  A  D  L  S  E  Y  N  T  L  V  L  D  D  R
 856 TTGTTCATAATCGATGTCTCGCAGTCTGTTGAACACGACCATCCT
     L  F  I  I  D  V  S  Q  S  V  E  H  D  H  P
 901 CACGCCTTAGATTTTCTGAGATCCGACTGTAACAACGTTTGCAAG
     H  A  L  D  F  L  R  S  D  C  N  N  V  C  K
 946 TTCTTCAGAGGCTTGGGAGTACCTGTGTTACCGGTATCAAAGTTA
     F  F  R  G  L  G  V  P  V  L  P  V  S  K  L
 991 TTTGAGCTCATAGTTGATCCCCTTATCAAGGATAGCGATGTCACC
     F  E  L  I  V  D  P  L  I  K  D  S  D  V  T
1036 AATTGGTTAGAAAAGCGAACACTTGATCCATCAGACGATGCCCTC
     N  W  L  E  K  R  T  L  D  P  S  D  D  A  L
1081 TTTATGAATGCGTTCATCCCTCATAAACTCGATCACGTCCTACAT
     F  M  N  A  F  I  P  H  K  L  D  H  V  L  H
```

```
1126 TTTGAGAGAGACAGTAAATTACTGAAGGCTGGTGAAGAGGCGAAC
     F  E  R  D  S  K  L  L  K  A  G  E  E  A  N
1171 AATCCGTTCCAGAATATCATTTCCAAGGTCGATGTGTTGGGTCAA
     N  P  F  Q  N  I  I  S  K  V  D  V  L  G  Q
1216 GGTTTTGAGGAGCGCGTAGTTTCTTCATCTGATGACGAATCTATT
     G  F  E  E  R  V  V  S  S  S  D  D  E  S  I
1261 TCGCAAGGCCATTCTGACGGCGATGGGAAGATCGTAGTGGAGACG
     S  Q  G  H  S  D  G  D  G  K  I  V  V  E  T
1306 AAAGGGAGGGCCACCAGTGGCAAACACTTCAGGGACAAAGACGAG
     K  G  R  A  T  S  G  K  H  F  R  D  K  D  E
1351 ACGCCCGAACAACGGAAAGTAAGAAAACAACTGATAAAGGAAGAG
     T  P  E  Q  R  K  V  R  K  Q  L  I  K  E  E
1396 AAACGGGAGGCTCGGAAAACCAAAATTCCTAAGCATGTGAAGAAA
     K  R  E  A  R  K  T  K  I  P  K  H  V  K  K
1441 CGCGCTCATCGGCAGCACATGAAATAG 1467
     R  A  H  R  Q  H  M  K  *
```

emb|CAE60574.1|	Hypothetical protein CBG04203 [Caenorhabditis br	500	1e-139
ref|NP_001021570.2|	M01B12.5a [Caenorhabditis elegans] >gb|AA...	496	1e-138
ref|XP_793660.2|	PREDICTED: similar to RIO kinase 1 (yeast) [...	343	1e-92
ref|NP_851100.1|	RIO1 family protein [Arabidopsis thaliana] >...	337	1e-90
ref|NP_180071.1|	RIO1 family protein [Arabidopsis thaliana] >...	337	1e-90
dbj|BAD44114.1|	unknown protein [Arabidopsis thaliana]	337	1e-90
gb|AAM65700.1|	similar to extragenic suppressor of bimD6 muta...	335	3e-90
dbj|BAD12556.1|	RIO kinase [Nicotiana tabacum]	335	4e-90
ref|NP_998160.1|	RIO kinase 1 [Danio rerio] >gb|AAH45984.1| R...	330	1e-88
ref|XP_971515.1|	PREDICTED: similar to CG11660-PA, isoform A ...	326	2e-87

Vacuolar ATPase *(Hc-vha-1)*

ATGGGCTCAATCTATCGGTCGGAGGTGATGAGCCTCTGCCAAATTTTCCTTCAAACGGACTCTGCCTACCAGTGCGTGGC
AGAATTGGGAGAACTTGGCCTGGCACAGTTTCTGGATTTGAACGAGGAGCAGAATGCTTATCAAAAGAAATTTGTCAACG
AAGTACGTCGTTGTGAAGAAATGGAAAGGAAATTGCGTTTCATAGAGGACGAGGTGCAGAAGGATGACGTGGAAATAGTG
GATCACGATGAGCACATTCCAGCCCCTCAGCCAAAGAACATGGTAGAGCTCGAGGCCAATTTCGAGAAACTGGAAGAGGA
ACTGATTTCGATCAATAAGAGCACGAAGCAGTTGAAGAAGAATCATGTTCAACTCCTGGAAATGAAAGCTGTGCTTGAGA
AAGTTCAAAGCTTGTTGGACGAGTCAAAGAGAGATGCTGCTATGTCGATTAGTGAAGCATCACGAGGTGAAGCTGGTCCA
TTTACTGTTGGCATCAAGATTGATTACGAGAAGGAGCGCCGAGAAGAAACCGAGCTCAGGTTTATTACCGGAGTGATCAG
TCGCAGTAAAGTTATTTCATTTGAACGCTTCATTTGGCGTTTCTGTCGTGGTAAAGTCTTTGTGCGAACTGTCGATATCA
CTGAACAAACCGAACTTTTCGATCATGACAAATCTGATGACAAGGCAGTGTTTATACTGTTTTTCTCTGGAGACCAACTT
CGCACAAGGGTACAGAAAATCTGTGCTGGATTCCACGCCGTCATCTATAACTGTCCTGAAAATCGTGTCGAACGAGCGCA
TCTTCTTGCTCAGATTAATGGACAAGTTGGCGATATGCAAAGTGTTATCAGCAAGACGTTAGAGTATCGTCATAAAATCA
TCTTCGCAGCAGCATTGAGTGTGAAAAAGTGGTCAATTATGTTGCTTAAATTGAAATCCATTTTCCACACTCTCAATATG
TTTGCTGTCGACGTCACCTACAAATGTCTGATTGCCGAGTGTTGGATACCTACTGTAGACCTCCCACTCGTGAAAGCAGC
ACTTCGCAAAGGAACAGAACAAGCTGGATCTACCGTACACGCTGTGCTAAACGAGATGGAAACACATAAAGAGCCGCCAA
CACATTTCAAATTGAACAAATTCACTCAAGGATTCCAAAATATCGTTGACGCCTATGGTATCGCCAACTACCGAGAGGTT
AATCCTGCTCCTTGGTCGATCATCTCCTTCCCTTTCTTGTTTGCTGTTATGTTTGGTGACTCTGGACATGGAATCATCAT
GCTACTGGCAGCTCTGGCCTTCGTTATTTTTGAAAAGAAGCTGATTGCAATGAAAATCAAAGATGAGATCTTCAATACCT
TCTTCGGAGGCCGATATGTTATTCTTCTCATGGGAATATTTTCCGTCTACACTGGGTTGCTCTACAACGACATCTATTCG
AAATCTATTAACATATTCAGTTCATCATGGAAGAATCCGTATCCGCAATCACTGTTAGCGCATATGGAAGAGGAAGGCCA
CAACAACAGTCAAACATTAGATCTAACATTTCCACCAGAGTACGCTTTCGATTCAAACTTGGGGCCGTATCCGTTCGGTG
TGGATCCGGTTTGGAATATTGCCAAGAACAAGCTGAATTTCTTGAATCCAATGAAAATGAAGACCTCTATTATCGTCGGA
ATATCTCAGATGGCTTTTGGACTACTGCTTTCGCTTTGCAATCACATTCACAATCGCTCGGTGGTTGACGTATTGTTCGT
TTTCATACCACAAGTATTCTTCCTCTTCTGCATCTTCGTGTATTTATGTGTAATGGTGGTCATGAAATGGATATTTTTCT
ATGTGAAGCCGGCCTTTATATTCGGTCGTCTATATCCTGGATCATACTGTGCACCATCATTACTTATAGGGCTGATCAAT
ATGTTCATGCTTAAAGCTAGAGATCCAGGTTTTGTTCAACACATAGGTAGCGCCAACGCTACTGATAAAGTGACCATCGA
TGGAAAGAATTATACGTACGACATGTACGATCAGTGTTATCTGCAACAGTGGTATCCCAATCAGGTACTCGTCGAGGAGA
TTCTACCCCTTCTCGCTGTAGTTTCCATACCAGTTATGCTGCTCGTCAAGCCGTTCTACGTCAGGTCCTTGGCCAAGCGT
GGTCTTCCCATACCTGGCGGTCATGGTCATGGTGGTGATGAAAGTGAAGAGTTCAGTTTCGGTGATGTCATGGTCTACCA
GGCAATTCACACGATTGAATTCGCGCTTGGGTGCATCTCCCATACAGCATCCTACCTTCGACTTTGGGCTCTTTCACTGG
CTCATGCACAGCTCTCAGAAGTACTTTGGGACATGTTGCTGGCTATTGGCCTTGACATGGGTGGATGGGCTGGATCGGCT
GCAATTTTCATCCTCTACTTCTTCTTCGGAGTGTTGTCCATCTCCATTCTGATTCTGATGGAAGGGCTTTCCGCATTCCT
TCATGCACTTCGTCTACATTGGGTCGAGTTCAATTCAAAGTTCTACGGTGGTACCGGACACGCATTCGAACCGTTCCACT
TCTTGCGACTCATTCGTGTTGCTGAAGGTTTGGAGCAATAG

```
      1 ATGGGCTCAATCTATCGGTCGGAGGTGATGAGCCTCTGCCAAATT
        M  G  S  I  Y  R  S  E  V  M  S  L  C  Q  I
     46 TTCCTTCAAACGGACTCTGCCTACCAGTGCGTGGCAGAATTGGGA
        F  L  Q  T  D  S  A  Y  Q  C  V  A  E  L  G
     91 GAACTTGGCCTGGCACAGTTTCTGGATTTGAACGAGGAGCAGAAT
        E  L  G  L  A  Q  F  L  D  L  N  E  E  Q  N
    136 GCTTATCAAAAGAAATTTGTCAACGAAGTACGTCGTTGTGAAGAA
        A  Y  Q  K  K  F  V  N  E  V  R  R  C  E  E
    181 ATGGAAAGGAAATTGCGTTTCATAGAGGACGAGGTGCAGAAGGAT
        M  E  R  K  L  R  F  I  E  D  E  V  Q  K  D
    226 GACGTGGAAATAGTGGATCACGATGAGCACATTCCAGCCCCTCAG
        D  V  E  I  V  D  H  D  E  H  I  P  A  P  Q
    271 CCAAAGAACATGGTAGAGCTCGAGGCCAATTTCGAGAAACTGGAA
        P  K  N  M  V  E  L  E  A  N  F  E  K  L  E
    316 GAGGAACTGATTTCGATCAATAAGAGCACGAAGCAGTTGAAGAAG
        E  E  L  I  S  I  N  K  S  T  K  Q  L  K  K
    361 AATCATGTTCAACTCCTGGAAATGAAAGCTGTGCTTGAGAAAGTT
        N  H  V  Q  L  L  E  M  K  A  V  L  E  K  V
    406 CAAAGCTTGTTGGACGAGTCAAAGAGAGATGCTGCTATGTCGATT
        Q  S  L  L  D  E  S  K  R  D  A  A  M  S  I
    451 AGTGAAGCATCACGAGGTGAAGCTGGTCCATTTACTGTTGGCATC
        S  E  A  S  R  G  E  A  G  P  F  T  V  G  I
    496 AAGATTGATTACGAGAAGGAGCGCCGAGAAGAAACCGAGCTCAGG
        K  I  D  Y  E  K  E  R  R  E  E  T  E  L  R
    541 TTTATTACCGGAGTGATCAGTCGCAGTAAAGTTATTTCATTTGAA
        F  I  T  G  V  I  S  R  S  K  V  I  S  F  E
    586 CGCTTCATTTGGCGTTTCTGTCGTGGTAAAGTCTTTGTGCGAACT
        R  F  I  W  R  F  C  R  G  K  V  F  V  R  T
    631 GTCGATATCACTGAACAAACCGAACTTTTCGATCATGACAAATCT
        V  D  I  T  E  Q  T  E  L  F  D  H  D  K  S
    676 GATGACAAGGCAGTGTTTATACTGTTTTTCTCTGGAGACCAACTT
        D  D  K  A  V  F  I  L  F  F  S  G  D  Q  L
```

```
 721 CGCACAAGGGTACAGAAAATCTGTGCTGGATTCCACGCCGTCATC
     R  T  R  V  Q  K  I  C  A  G  F  H  A  V  I
 766 TATAACTGTCCTGAAAATCGTGTCGAACGAGCGCATCTTCTTGCT
     Y  N  C  P  E  N  R  V  E  R  A  H  L  L  A
 811 CAGATTAATGGACAAGTTGGCGATATGCAAAGTGTTATCAGCAAG
     Q  I  N  G  Q  V  G  D  M  Q  S  V  I  S  K
 856 ACGTTAGAGTATCGTCATAAAATCATCTTCGCAGCAGCATTGAGT
     T  L  E  Y  R  H  K  I  I  F  A  A  A  L  S
 901 GTGAAAAAGTGGTCAATTATGTTGCTTAAATTGAAATCCATTTTC
     V  K  K  W  S  I  M  L  L  K  L  K  S  I  F
 946 CACACTCTCAATATGTTTGCTGTCGACGTCACCTACAAATGTCTG
     H  T  L  N  M  F  A  V  D  V  T  Y  K  C  L
 991 ATTGCCGAGTGTTGGATACCTACTGTAGACCTCCCACTCGTGAAA
     I  A  E  C  W  I  P  T  V  D  L  P  L  V  K
1036 GCAGCACTTCGCAAAGGAACAGAACAAGCTGGATCTACCGTACAC
     A  A  L  R  K  G  T  E  Q  A  G  S  T  V  H
1081 GCTGTGCTAAACGAGATGGAAACACATAAAGAGCCGCCAACACAT
     A  V  L  N  E  M  E  T  H  K  E  P  P  T  H
1126 TTCAAATTGAACAAATTCACTCAAGGATTCCAAAATATCGTTGAC
     F  K  L  N  K  F  T  Q  G  F  Q  N  I  V  D
1171 GCCTATGGTATCGCCAACTACCGAGAGGTTAATCCTGCTCCTTGG
     A  Y  G  I  A  N  Y  R  E  V  N  P  A  P  W
1216 TCGATCATCTCCTTCCCTTTCTTGTTTGCTGTTATGTTTGGTGAC
     S  I  I  S  F  P  F  L  F  A  V  M  F  G  D
1261 TCTGGACATGGAATCATCATGCTACTGGCAGCTCTGGCCTTCGTT
     S  G  H  G  I  I  M  L  L  A  A  L  A  F  V
1306 ATTTTTGAAAAGAAGCTGATTGCAATGAAAATCAAAGATGAGATC
     I  F  E  K  K  L  I  A  M  K  I  K  D  E  I
1351 TTCAATACCTTCTTCGGAGGCCGATATGTTATTCTTCTCATGGGA
     F  N  T  F  F  G  G  R  Y  V  I  L  L  M  G
1396 ATATTTTCCGTCTACACTGGGTTGCTCTACAACGACATCTATTCG
     I  F  S  V  Y  T  G  L  L  Y  N  D  I  Y  S
1441 AAATCTATTAACATATTCAGTTCATCATGGAAGAATCCGTATCCG
     K  S  I  N  I  F  S  S  S  W  K  N  P  Y  P
1486 CAATCACTGTTAGCGCATATGGAAGAGGAAGGCCACAACAACAGT
     Q  S  L  L  A  H  M  E  E  E  G  H  N  N  S
1531 CAAACATTAGATCTAACATTTCCACCAGAGTACGCTTTCGATTCA
     Q  T  L  D  L  T  F  P  P  E  Y  A  F  D  S
1576 AACTTGGGGCCGTATCCGTTCGGTGTGGATCCGGTTTGGAATATT
     N  L  G  P  Y  P  F  G  V  D  P  V  W  N  I
1621 GCCAAGAACAAGCTGAATTTCTTGAATCCAATGAAAATGAAGACC
     A  K  N  K  L  N  F  L  N  P  M  K  M  K  T
1666 TCTATTATCGTCGGAATATCTCAGATGGCTTTTGGACTACTGCTT
     S  I  I  V  G  I  S  Q  M  A  F  G  L  L  L
1711 TCGCTTTGCAATCACATTCACAATCGCTCGGTGGTTGACGTATTG
     S  L  C  N  H  I  H  N  R  S  V  V  D  V  L
1756 TTCGTTTTCATACCACAAGTATTCTTCCTCTTCTGCATCTTCGTG
     F  V  F  I  P  Q  V  F  F  L  F  C  I  F  V
1801 TATTTATGTGTAATGGTGGTCATGAAATGGATATTTTTCTATGTG
     Y  L  C  V  M  V  V  M  K  W  I  F  F  Y  V
1846 AAGCCGGCCTTTATATTCGGTCGTCTATATCCTGGATCATACTGT
     K  P  A  F  I  F  G  R  L  Y  P  G  S  Y  C
1891 GCACCATCATTACTTATAGGGCTGATCAATATGTTCATGCTTAAA
     A  P  S  L  L  I  G  L  I  N  M  F  M  L  K
1936 GCTAGAGATCCAGGTTTTGTTCAACACATAGGTAGCGCCAACGCT
     A  R  D  P  G  F  V  Q  H  I  G  S  A  N  A
1981 ACTGATAAAGTGACCATCGATGGAAAGAATTATACGTACGACATG
     T  D  K  V  T  I  D  G  K  N  Y  T  Y  D  M
2026 TACGATCAGTGTTATCTGCAACAGTGGTATCCCAATCAGGTACTC
     Y  D  Q  C  Y  L  Q  Q  W  Y  P  N  Q  V  L
2071 GTCGAGGAGATTCTACCCCTTCTCGCTGTAGTTTCCATACCAGTT
     V  E  E  I  L  P  L  L  A  V  V  S  I  P  V
2116 ATGCTGCTCGTCAAGCCGTTCTACGTCAGGTCCTTGGCCAAGCGT
     M  L  L  V  K  P  F  Y  V  R  S  L  A  K  R
2161 GGTCTTCCCATACCTGGCGGTCATGGTCATGGTGGTGATGAAAGT
     G  L  P  I  P  G  G  H  G  H  G  G  D  E  S
2206 GAAGAGTTCAGTTTCGGTGATGTCATGGTCTACCAGGCAATTCAC
     E  E  F  S  F  G  D  V  M  V  Y  Q  A  I  H
```

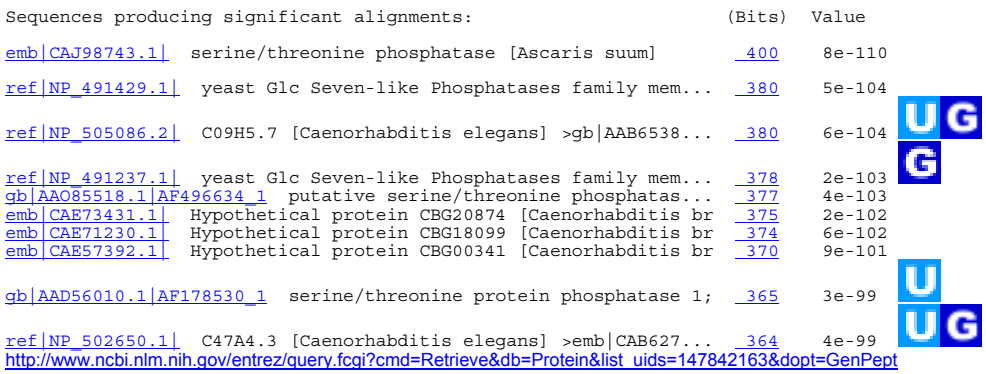
```
2251 ACGATTGAATTCGCGCTTGGGTGCATCTCCCATACAGCATCCTAC
     T  I  E  F  A  L  G  C  I  S  H  T  A  S  Y
2296 CTTCGACTTTGGGCTCTTTCACTGGCTCATGCACAGCTCTCAGAA
     L  R  L  W  A  L  S  L  A  H  A  Q  L  S  E
2341 GTACTTTGGGACATGTTGCTGGCTATTGGCCTTGACATGGGTGGA
     V  L  W  D  M  L  L  A  I  G  L  D  M  G  G
2386 TGGGCTGGATCGGCTGCAATTTTCATCCTCTACTTCTTCTTCGGA
     W  A  G  S  A  A  I  F  I  L  Y  F  F  F  G
2431 GTGTTGTCCATCTCCATTCTGATTCTGATGGAAGGGCTTTCCGCA
     V  L  S  I  S  I  L  I  L  M  E  G  L  S  A
2476 TTCCTTCATGCACTTCGTCTACATTGGGTCGAGTTCAATTCAAAG
     F  L  H  A  L  R  L  H  W  V  E  F  N  S  K
2521 TTCTACGGTGGTACCGGACACGCATTCGAACCGTTCCACTTCTTG
     F  Y  G  G  T  G  H  A  F  E  P  F  H  F  L
2566 CGACTCATTCGTGTTGCTGAAGGTTTGGAGCAATAG 2601
     R  L  I  R  V  A  E  G  L  E  Q  *
```

emb|CAE59507.1|	Hypothetical protein CBG02894 [Caenorhabditis br	1211	0.0
ref|NP_496436.1|	Vacuolar H ATPase family member (vha-6) [Cae...	1211	0.0
ref|NP_001023021.1|	UNCoordinated family member (unc-32) [Cae...	860	0.0
ref|NP_001023019.1|	UNCoordinated family member (unc-32) [Cae...	856	0.0
ref|NP_001023020.1|	UNCoordinated family member (unc-32) [Cae...	856	0.0
ref|NP_001023018.1|	UNCoordinated family member (unc-32) [Cae...	852	0.0
ref|NP_001023017.1|	UNCoordinated family member (unc-32) [Cae...	852	0.0
ref|NP_001023022.1|	UNCoordinated family member (unc-32) [Cae...	848	0.0
emb|CAE58454.1|	Hypothetical protein CBG01592 [Caenorhabditis br	767	0.0
ref|NP_501399.1|	Vacuolar H ATPase family member (vha-5) [Cae...	761	0.0

Serine/threonine phosphatase *(Hc-stp-1)*

ATGGACCCTACTCAATTGATTACTAACCTACTTAATGTGGGTCTCCCAGATAAGGGACTCACGAAAACGGTTTCTGAAAA
CGATATAATGGAAGTACTAGGTAAAGCTCGTGAGATGTTTCTGTCTCAACCTCCGATGGTAGAACTTGATTCACCTGTTA
AAATATGTGGCGATACTCATGGGCAGTATATTGACCTTTTAAGGCTTTTTAACAAGGGTGGATTTCCACCACTATCGAAT
TACTTGTTCTTGGGTGATTATGTGGATAGAGGGAAGCAAAATCTTGAAGTAATACTCCTAATGATTGCATATAAGCTGAG
GTTTCCGAAGAATTTCTTCCTGCTACGAGGAAACCACGAATGTGCTAACGTCAACCGTGCCTACGGATTCTACGAAGAAT
GCAATAGGCGCTATCAAAGCCAACGAATGTGGCAGGCATTTCAGGACGTTCTCTGTGTAATGCCCCTGACAGCCCTTGTT
AGTGATAAAATTCTCTGTATGCATGGTGGACTGTCACCTCACTTGCAGTCATTGGATCAACTCAGAAATATAACCAGACC
TACTGACGCACTGGGAGCAACTCTGGAAATGGACCTACTTTGGGCCGATCCAGTGATTGGGCTGAACGGCTTCCAGGCAA
ACATTCGTGGTGCTTCATATGGATTTGGGCCCGATATTCTGGCTAAGTACTGTCAACTCCTCAACATCGACTTAGTTGCT
CGAGCCCACCAAGTTGTCCAAGATGGCTACGAGTTCTTTGGAGGGAGAAAGCTTGTGACTATTTTTTCGGCTCCTCATTA
TTGTGGCCAATTCGATAATGCCGCTGCCATGATGACAGTTGATGAAAACTTGCAGTGTTCGTTCGATGCTTTTCGCCCTT
CCTGTGCAAAACCTCAGCCAAAGATTGTGGCCACTTCCATGGGATCTCCTGGTGCTCCACCTTGTCAATAA

```
     1   M  D  P  T  Q  L  I  T  N  L  L  N  V  G  L  P  D  K  G  L
     1  ATGGACCCTACTCAATTGATTACTAACCTACTTAATGTGGGTCTCCCAGATAAGGGACTC

    21   T  K  T  V  S  E  N  D  I  M  E  V  L  G  K  A  R  E  M  F
    61  ACGAAAACGGTTTCTGAAAACGATATAATGGAAGTACTAGGTAAAGCTCGTGAGATGTTT

    41   L  S  Q  P  P  M  V  E  L  D  S  P  V  K  I  C  G  D  T  H
   121  CTGTCTCAACCTCCGATGGTAGAACTTGATTCACCTGTTAAAATATGTGGCGATACTCAT

    61   G  Q  Y  I  D  L  L  R  L  F  N  K  G  G  F  P  P  L  S  N
   181  GGGCAGTATATTGACCTTTTAAGGCTTTTTAACAAGGGTGGATTTCCACCACTATCGAAT

    81   Y  L  F  L  G  D  Y  V  D  R  G  K  Q  N  L  E  V  I  L  L
   241  TACTTGTTCTTGGGTGATTATGTGGATAGAGGGAAGCAAAATCTTGAAGTAATACTCCTA

   101   M  I  A  Y  K  L  R  F  P  K  N  F  F  L  L  R  G  N  H  E
   301  ATGATTGCATATAAGCTGAGGTTTCCGAAGAATTTCTTCCTGCTACGAGGAAACCACGAA

   121   C  A  N  V  N  R  A  Y  G  F  Y  E  E  C  N  R  R  Y  Q  S
   361  TGTGCTAACGTCAACCGTGCCTACGGATTCTACGAAGAATGCAATAGGCGCTATCAAAGC

   141   Q  R  M  W  Q  A  F  Q  D  V  L  C  V  M  P  L  T  A  L  V
   421  CAACGAATGTGGCAGGCATTTCAGGACGTTCTCTGTGTAATGCCCCTGACAGCCCTTGTT

   161   S  D  K  I  L  C  M  H  G  G  L  S  P  H  L  Q  S  L  D  Q
   481  AGTGATAAAATTCTCTGTATGCATGGTGGACTGTCACCTCACTTGCAGTCATTGGATCAA

   181   L  R  N  I  T  R  P  T  D  A  L  G  A  T  L  E  M  D  L  L
   541  CTCAGAAATATAACCAGACCTACTGACGCACTGGGAGCAACTCTGGAAATGGACCTACTT

   201   W  A  D  P  V  I  G  L  N  G  F  Q  A  N  I  R  G  A  S  Y
   601  TGGGCCGATCCAGTGATTGGGCTGAACGGCTTCCAGGCAAACATTCGTGGTGCTTCATAT

   221   G  F  G  P  D  I  L  A  K  Y  C  Q  L  L  N  I  D  L  V  A
   661  GGATTTGGGCCCGATATTCTGGCTAAGTACTGTCAACTCCTCAACATCGACTTAGTTGCT

   241   R  A  H  Q  V  V  Q  D  G  Y  E  F  F  G  G  R  K  L  V  T
   721  CGAGCCCACCAAGTTGTCCAAGATGGCTACGAGTTCTTTGGAGGGAGAAAGCTTGTGACT

   261   I  F  S  A  P  H  Y  C  G  Q  F  D  N  A  A  A  M  M  T  V
   781  ATTTTTTCGGCTCCTCATTATTGTGGCCAATTCGATAATGCCGCTGCCATGATGACAGTT

   281   D  E  N  L  Q  C  S  F  D  A  F  R  P  S  C  A  K  P  Q  P
   841  GATGAAAACTTGCAGTGTTCGTTCGATGCTTTTCGCCCTTCCTGTGCAAAACCTCAGCCA

   301   K  I  V  A  T  S  M  G  S  P  G  A  P  P  C  Q  -
   901 AAGATTGTGGCCACTTCCATGGGATCTCCTGGTGCTCCACCTTGTCAATAA
```



```
Sequences producing significant alignments:                          (Bits)  Value

emb|CAJ98743.1|  serine/threonine phosphatase [Ascaris suum]          400    8e-110

ref|NP_491429.1|  yeast Glc Seven-like Phosphatases family mem...     380    5e-104

ref|NP_505086.2|  C09H5.7 [Caenorhabditis elegans] >gb|AAB6538...     380    6e-104

ref|NP_491237.1|  yeast Glc Seven-like Phosphatases family mem...     378    2e-103
gb|AAO85518.1|AF496634_1  putative serine/threonine phosphatas...     377    4e-103
emb|CAE73431.1|  Hypothetical protein CBG20874 [Caenorhabditis br     375    2e-102
emb|CAE71230.1|  Hypothetical protein CBG18099 [Caenorhabditis br     374    6e-102
emb|CAE57392.1|  Hypothetical protein CBG00341 [Caenorhabditis br     370    9e-101

gb|AAD56010.1|AF178530_1  serine/threonine protein phosphatase 1;     365    3e-99

ref|NP_502650.1|  C47A4.3 [Caenorhabditis elegans] >emb|CAB627...     364    4e-99
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=Protein&list_uids=147842163&dopt=GenPept
```

As recommended in the independent assessment of the previous project (AHW.022), we extended previous work on *Trichostrongylus*, to isolate and partially characterize 5 pre-validated molecules (i.e., *Hc-ant-1, Hc-pbs-5, Hc-rio1, Hc-vha-1, Hc-stp-1*). Given the intellectual property position, clearly, an emphasis was placed on exploring *Hc-stp-1* further. In addition to these outcomes, we substantially strengthened the capacity of our team through the appointment of a bioinformatician (Ross Hall BSc, Dip Comp Sci), who worked full-time on improving our *in silico* resources and bioinformatic capacity. The appointment of the principal investigator (Gasser) to the Board of WormBase (www.wormbase.org) and the development of collaborations with other international groups further enhanced our team's capacity to effectively mine EST data and identify nematode-specific drug targets. Moreover, extensive progress has been in the scientific training of postdoctoral staff and graduate students, who published numerous articles and presented at conferences. The expertise in genomics, bioinformatics and molecular biology was substantially enhanced during this phase of the project.

4.6. Full characterisation of five pre-validated molecules (milestone 6)

A genomic-bioinformatic approach had been applied previously to the selection of promising candidate drug targets from the ovine nematode, *Trichostrongylus vitrinus*, *via* detailed comparative analysis with *Caenorhabditis elegans* (project AHW.022). This set of ESTs included groups of potential targets, including a kinase, phosphatase and translocase, thereby demonstrating the utility of the present approach. According to the agreement, we built on this work to isolate and characterize a selected subset of *H. contortus* genes and predicted gene products. The approach used was essentially the same as used to achieve milestone 5. Extending previous work, three additional genes (**marked in blue text**) were isolated and characterized. *H. contortus* genes isolated and cloned into the plasmid vectors pGEM-T-Easy and/or pGEX. The full list of genes is listed in the following:

Gene	cDNA (bp)	Estimated size (GST fusion) (kDa)	Predicted function based on comparison with *Caenorhabditis elegans*
Hc-ant-1	894	100 kDa	Transport of ADP/ATP into and out of cells
Hc-pbs-5	855	96 kDa	Proteasome system
Hc-rio-1	1467	146 kDa	Phosphorylation of serine/threonine residues. Involved in many cellular processes
Hc-vha-1	2601	243 kDa	Function yet unclear
Hc-stp-1	951	61 kDa	Dephosphorylation of serine/threonine residues. Involved in many cellular processes
Hc-cc-1	**339**	**53 kDa**	**Electron transport**
Hc-als-1	**2862**	**265 kDa**	**Binding and stabilization of insulin-like growth factor.**
Hc-rpn-3	**1527**	**84 kDa**	**Regulatory subunit of 26 proteasome. Involved in degradation of ubiquitinated proteins**

The sequences of the three new cDNAs, their inferred peptide sequences and their closest "matches", based on database comparisons, are listed in the following:

Cytochrome *c* (*Hc-cc-1*)

ATGGGCGATATTCCCGAGGGAGACTACGAGAAGGGCAAAAAGGTCTTCAAGCAAAGATGTCTTCAATGTCACGTTGTCGACTCGAAAGCT
ACCAAGACAGGCCCAACCCTTCACGGAATCATTGGACGTAAATCTGGAACAGTTGAAGGGTTCGACTACTCCGCCGCCAACAAAAACAAG
GGAGTAATATGGTCACGTGAAACATTGTTCGAGTACCTCCTGAATCCTAAGAAGTACATCCCTGGAACGAAGATGGTGTTCGCTGGATTG
AAGAAAGCCGACGAGCGAGCGGATCTCATTAAATACATTGAAGTAGAATCAGCAAAGCCTGCCAGCTAG

```
1   M  G  D  I  P  E  G  D  Y  E  K  G  K  K  V  F  K  Q  R  C
1  ATGGGCGATATTCCCGAGGGAGACTACGAGAAGGGCAAAAAGGTCTTCAAGCAAAGATGT

21   L  Q  C  H  V  V  D  S  K  A  T  K  T  G  P  T  L  H  G  I
61  CTTCAATGTCACGTTGTCGACTCGAAAGCTACCAAGACAGGCCCAACCCTTCACGGAATC

41   I  G  R  K  S  G  T  V  E  G  F  D  Y  S  A  A  N  K  N  K
121 ATTGGACGTAAATCTGGAACAGTTGAAGGGTTCGACTACTCCGCCGCCAACAAAAACAAG

61   G  V  I  W  S  R  E  T  L  F  E  Y  L  L  N  P  K  K  Y  I
181 GGAGTAATATGGTCACGTGAAACATTGTTCGAGTACCTCCTGAATCCTAAGAAGTACATC

81   P  G  T  K  M  V  F  A  G  L  K  K  A  D  E  R  A  D  L  I
241 CCTGGAACGAAGATGGTGTTCGCTGGATTGAAGAAAGCCGACGAGCGAGCGGATCTCATT

101  K  Y  I  E  V  E  S  A  K  P  A  S  -
301 AAATACATTGAAGTAGAATCAGCAAAGCCTGCCAGCTAG
```

```
ref|NP_500629.1|  E04A4.7 [Caenorhabditis elegans] >gb|AAB9203...   211    1e-53
sp|P19974|CYC_CAEEL  Cytochrome c                                    210    2e-53
dbj|BAA11131.1|  type-1 cytochrome c [Ascaris suum]                  209    4e-53
sp|P92504|CYC1_ASCSU  Cytochrome c type-1                            209    4e-53
emb|CAE58578.1|  Hypothetical protein CBG01744 [Caenorhabditis br    206    3e-52
emb|CAE63947.1|  Hypothetical protein CBG08529 [Caenorhabditis br    191    9e-48

ref|NP_506156.1|  ZC116.2 [Caenorhabditis elegans] >sp|Q23240|...    188    1e-46
sp|P92505|CYC2_ASCSU  Cytochrome c type-2 >dbj|BAA11132.1| typ...    167    2e-40
gb|AAY86487.1|  cytochrome C [Dermacentor variabilis] >gb|AAY8...    150    2e-35

ref|XP_967227.1|  PREDICTED: similar to CG17903-PA isoform 1 [...    148    1e-34
```

http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=Protein&list_uids=67083909&dopt=GenPept

Although cytochrome *c* genes (*cyt c*) and proteins (CYT C) have been relatively well studied in mammals, almost nothing is known about them in parasitic helminths. These molecules were investigated in detail in both *Haemonchus contortus* (barber's pole worm) and *Trichostrongylus vitrinus* (black scour worm). The *cyt c* gene (512 bp) of *H. contortus* had one intron and encoded a transcript of 345 nucleotides, whilst that of *T. vitrinus* (792 bp) had two introns and encoded a transcript of 360 nucleotides. The inferred proteins (designated *Hc*-CYT C and *Tv*-CYT C, respectively) shared nucleotide and amino acid identities of 78% and 85%, respectively. The alignment of these and other CYT C sequences from nematodes, flatworms, insects and mammals identified conserved motifs associated with CYT C oxidase- and reductase- as well as haem-binding. One residue (histidine-26) was conserved for mammals, whilst this residue was absent from all nematodes; the functional significance of this mutation is not yet known. Both phylogenetic analysis and protein modelling revealed that CYT C proteins of nematodes are structurally distinct from those of mammals and other organisms, indicating their potential as targets for parasite intervention.

Acid labile subunit *(Hc-als-1)*

ATGTTCCGGCTCATTCACCTGCACTTCAACAAATTCCTGCTCCTCCTTGTCATCTGGTCCAGCTCTGTCGCTGCCGATGACTGTCCAATT
CTTCCGTCGACTTGTGTCTGTAACAAAGAGGCCACAGGTCGGATCTCGATTACTTGTGATGGCGCTCATCTCCCCAGTCTCTTTAAGGAT
ATCGGTGGGGCAAGTATCGACCGATTGCGGATCTCGAAATGTGCGCAGCCAGTAGTGGATGTATTGCCCACTGGTCCAATTCGATCATTG
GCTTTAACGAACTGCCAGATTACCACCATTAACCCAGAGATCTTTCGAAGTGTGGAATCAACGTTGGAGGAGTTGAATTTGTCGAATAAT
AGCCTGAAAAATGTCCCTCTTCTTGGAAACATGTCCCAACTGCGATCGTTGAATTTGTTTGGAAACAAGGTGAAAGAGGTTCCTGAAGGT
GTATTCTCTGGTCTATCGCAACTTCGAATGCTTCGATTGGAGAACAATGAGATCTGTTCGCTGTCGCCAACTTCGTTATCGGAGGTCAAA
CAGAAGATTGAGCTACTTGATCTTAGTGGGAACTGTTTTGCAAGCGTACCAGCCCAAGCTCTCAGAAATAGCATCAACCTGATGTATCTC
GATTTGTCCGACAACAAGATCTCTGAGATCAACAACTTCGAGCTGATGAACTTGCCGCAGTTGAAAGAGCTTCGATTACAAAACAACCAA
CTCCCCGAGATTCACCCGATGGCGTTCATGAATGTTCCTCAGCTACAGTATCTCTACTTGAAGGACAATTTGATCGGAAGTCTGGACGGA
AATCGTCTTCAGGCTTTTCATCGTCTCGAAATACTCGATGTCACCAATAATCTCTTACTAAAATTACCGGAGTTGAAAGATCTAACCAGT
CTCAAACAAGTTCGCTTGGACGGGAATTTGATTGAAAAAATTGACACACTTGCCTTCTCTAACAACCCGAACCTTCAACTAATAAGTATA
CAGGATAACAATATCGCTCAAATAGCGAAGAATAGTTTCGATTCACTGGATCAGCTCATGGTTTTGCTCTTGTCGAACAACTCAATTCAG
AAGCTAGAGAGAGGAATGCTTGATGGGATGCGAAATCTGCAACAACTGAATCTTCGCAATAACTCATTGACGGCAGTCGATGAATATAGT
TTTGCTTCATTACGATTTCTAACGACTTTGGATTTAGCACATAATGAACTAGTGAAAATCGGCAAGAGAACCTTTTCTCATCAGTCGAAG
CTGTTCTGGTTGGATTTATCGAACAACAAGCTGACTGGTTTTGAGGAAGGAACATTTGACACAAAAATCGCTAACCTTCTACTTGATGGG
AATCACTTAATATGTGACGATGATTTCGATTGGTTTGTCCGATATTTGGTTACAAATCGAGTACGCACATTCTTGCCGTTCCAAACTGAG
ATCACCTGTGCAGGTCCTGAAAAATACGCTGGTGTCAGGCTCAAAGATCTGATGATCAAAAAAGCCAATGACACTTTGACGGAGGGCATG
AGGTCTCTCGGATTCAATAGCGATCAAAGTCAGAACTCACTTCTGTCGGCTCTATTGCCAGGATTTCGGCCCCCTCCGGCGGCTTCTGCT
AATGGTGGTGGCGCTCTGGGACAAGCCGCTGCTGGTTTGCCAATTTTAAGCACTCTTTCACAGGCGATCCCATCAATGCGAAATATGCCT
GGTCTCGATATGACGGCTGCCGCTGCTGCTGCTGGTGGTGCTCAGCCAGCTAATCCGAACCTGAATACGGCTATTGAACAAAAGTTCACC
GCTCCTCTCGTCCGTTTCGCCACTGGCGGCCAGCCAGTTCAAAGTGACATCGAGCAGCTGATTCAGTCGATACCAAACTTTGTCGTCAAT
GTCCCCGGAATGGGCAATGTCGATATCAGCAAGTTGGATCCGAATCTCATTGCTCACGTACTTCGTGGCGGTCAGATTCCTGGTATACCA
AAAGAAACTCTGGATGGTATCGTTCAACAATACATGTTGAAAATGCATGAAGCGGCTGCAGCGGCTCAAAATGGTGCTCCACTGAAGAAT
GGTTCGAAGTTTTTGCCACCAATCGAGAAATTCCCTCCAGATATTGTCGCCTCCGTTTTTCAAGGTAAAAATCTGCCCGGATTGAATGAA
GAGCAAACAAAAGTGATCAAGGACTACTACACGGCACGCCTTCCTGTCCCACCTGCAGCCGTCCCGGAGAGAATTCCAGCCATGGCCAAT
ACTCCACCCGTACAATACACCACAAGAAACACTACTGCGCCGCCTTTCGTATTCTCGAAGGACCTGCTCGACATGCTGAAGATACTACGC
GTTGGTTACAATTTGAGCAAGATTCCAGCCGAGATAATCACCGCGTTTTCTCGTGGCGAAGTTCCTGATCTGCGTAAGTTGCCGACCGAT
CTTATCGAGCATTTCAAGTCGAACAGCGAAAAACTTGGCTTGATATTCAATCGGGTGGCTAAGGCTAATGCGTCTCTAGAAGAAATGCTA
GCGAAGATGCCTAAGTTCGAAAAGCCAGTGCTTTCCACATTCTCACCGTATGACATCAATCACCTCTCAAACGACATGATTCACGAGGAG
GAAGAAGCTGCACGAGCAGCAAGGATGAGGGTATACACGGCAATCGCCTTGGGACTTGTTGGTGCGGCAACCGTGGTAATTCTCGGCTTT
TTTGCTGTCTACATTAAGAAGCAAAGAGAAGCGAAAGAAAGGAGAGGGCCGATGACAACACCGCGCTGTGCATTTCCGGGTGCACCATTA
GCGAACTCTACCATGCGCAACCCACCACGTTCGCATTCGTCCAGTCATTTGCACTTGCGGACTAATACGTAG

```
     1 ATGTTCCGGCTCATTCACCTGCACTTCAACAAATTCCTGCTCCTC
       M  F  R  L  I  H  L  H  F  N  K  F  L  L  L
    46 CTTGTCATCTGGTCCAGCTCTGTCGCTGCCGATGACTGTCCAATT
       L  V  I  W  S  S  S  V  A  A  D  D  C  P  I
    91 CTTCCGTCGACTTGTGTCTGTAACAAAGAGGCCACAGGTCGGATC
       L  P  S  T  C  V  C  N  K  E  A  T  G  R  I
   136 TCGATTACTTGTGATGGCGCTCATCTCCCCAGTCTCTTTAAGGAT
       S  I  T  C  D  G  A  H  L  P  S  L  F  K  D
   181 ATCGGTGGGGCAAGTATCGACCGATTGCGGATCTCGAAATGTGCG
       I  G  G  A  S  I  D  R  L  R  I  S  K  C  A
   226 CAGCCAGTAGTGGATGTATTGCCCACTGGTCCAATTCGATCATTG
       Q  P  V  V  D  V  L  P  T  G  P  I  R  S  L
   271 GCTTTAACGAACTGCCAGATTACCACCATTAACCCAGAGATCTTT
       A  L  T  N  C  Q  I  T  T  I  N  P  E  I  F
   316 CGAAGTGTGGAATCAACGTTGGAGGAGTTGAATTTGTCGAATAAT
       R  S  V  E  S  T  L  E  E  L  N  L  S  N  N
   361 AGCCTGAAAAATGTCCCTCTTCTTGGAAACATGTCCCAACTGCGA
       S  L  K  N  V  P  L  L  G  N  M  S  Q  L  R
   406 TCGTTGAATTTGTTTGGAAACAAGGTGAAAGAGGTTCCTGAAGGT
       S  L  N  L  F  G  N  K  V  K  E  V  P  E  G
   451 GTATTCTCTGGTCTATCGCAACTTCGAATGCTTCGATTGGAGAAC
       V  F  S  G  L  S  Q  L  R  M  L  R  L  E  N
   496 AATGAGATCTGTTCGCTGTCGCCAACTTCGTTATCGGAGGTCAAA
       N  E  I  C  S  L  S  P  T  S  L  S  E  V  K
   541 CAGAAGATTGAGCTACTTGATCTTAGTGGGAACTGTTTTGCAAGC
       Q  K  I  E  L  L  D  L  S  G  N  C  F  A  S
   586 GTACCAGCCCAAGCTCTCAGAAATAGCATCAACCTGATGTATCTC
       V  P  A  Q  A  L  R  N  S  I  N  L  M  Y  L
   631 GATTTGTCCGACAACAAGATCTCTGAGATCAACAACTTCGAGCTG
       D  L  S  D  N  K  I  S  E  I  N  N  F  E  L
   676 ATGAACTTGCCGCAGTTGAAAGAGCTTCGATTACAAAACAACCAA
       M  N  L  P  Q  L  K  E  L  R  L  Q  N  N  Q
   721 CTCCCCGAGATTCACCCGATGGCGTTCATGAATGTTCCTCAGCTA
```

```
     L  P  E  I  H  P  M  A  F  M  N  V  P  Q  L
 766 CAGTATCTCTACTTGAAGGACAATTTGATCGGAAGTCTGGACGGA
     Q  Y  L  Y  L  K  D  N  L  I  G  S  L  D  G
 811 AATCGTCTTCAGGCTTTTCATCGTCTCGAAATACTCGATGTCACC
     N  R  L  Q  A  F  H  R  L  E  I  L  D  V  T
 856 AATAATCTCTTACTAAAATTACCGGAGTTGAAAGATCTAACCAGT
     N  N  L  L  L  K  L  P  E  L  K  D  L  T  S
 901 CTCAAACAAGTTCGCTTGGACGGGAATTTGATTGAAAAAATTGAC
     L  K  Q  V  R  L  D  G  N  L  I  E  K  I  D
 946 ACACTTGCCTTCTCTAACAACCCGAACCTTCAACTAATAAGTATA
     T  L  A  F  S  N  N  P  N  L  Q  L  I  S  I
 991 CAGGATAACAATATCGCTCAAATAGCGAAGAATAGTTTCGATTCA
     Q  D  N  N  I  A  Q  I  A  K  N  S  F  D  S
1036 CTGGATCAGCTCATGGTTTTGCTCTTGTCGAACAACTCAATTCAG
     L  D  Q  L  M  V  L  L  L  S  N  N  S  I  Q
1081 AAGCTAGAGAGAGGAATGCTTGATGGGATGCGAAATCTGCAACAA
     K  L  E  R  G  M  L  D  G  M  R  N  L  Q  Q
1126 CTGAATCTTCGCAATAACTCATTGACGGCAGTCGATGAATATAGT
     L  N  L  R  N  N  S  L  T  A  V  D  E  Y  S
1171 TTTGCTTCATTACGATTTCTAACGACTTTGGATTTAGCACATAAT
     F  A  S  L  R  F  L  T  T  L  D  L  A  H  N
1216 GAACTAGTGAAAATCGGCAAGAGAACCTTTTCTCATCAGTCGAAG
     E  L  V  K  I  G  K  R  T  F  S  H  Q  S  K
1261 CTGTTCTGGTTGGATTTATCGAACAACAAGCTGACTGGTTTTGAG
     L  F  W  L  D  L  S  N  N  K  L  T  G  F  E
1306 GAAGGAACATTTGACACAAAAATCGCTAACCTTCTACTTGATGGG
     E  G  T  F  D  T  K  I  A  N  L  L  L  D  G
1351 AATCACTTAATATGTGACGATGATTTCGATTGGTTTGTCCGATAT
     N  H  L  I  C  D  D  D  F  D  W  F  V  R  Y
1396 TTGGTTACAAATCGAGTACGCACATTCTTGCCGTTCCAAACTGAG
     L  V  T  N  R  V  R  T  F  L  P  F  Q  T  E
1441 ATCACCTGTGCAGGTCCTGAAAAATACGCTGGTGTCAGGCTCAAA
     I  T  C  A  G  P  E  K  Y  A  G  V  R  L  K
1486 GATCTGATGATCAAAAAAGCCAATGACACTTTGACGGAGGGCATG
     D  L  M  I  K  K  A  N  D  T  L  T  E  G  M
1531 AGGTCTCTCGGATTCAATAGCGATCAAAGTCAGAACTCACTTCTG
     R  S  L  G  F  N  S  D  Q  S  Q  N  S  L  L
1576 TCGGCTCTATTGCCAGGATTTCGGCCCCCTCCGGCGGCTTCTGCT
     S  A  L  L  P  G  F  R  P  P  P  A  A  S  A
1621 AATGGTGGTGGCGCTCTGGGACAAGCCGCTGCTGGTTTGCCAATT
     N  G  G  G  A  L  G  Q  A  A  A  G  L  P  I
1666 TTAAGCACTCTTTCACAGGCGATCCCATCAATGCGAAATATGCCT
     L  S  T  L  S  Q  A  I  P  S  M  R  N  M  P
1711 GGTCTCGATATGACGGCTGCCGCTGCTGCTGCTGGTGGTGCTCAG
     G  L  D  M  T  A  A  A  A  A  A  G  G  A  Q
1756 CCAGCTAATCCGAACCTGAATACGGCTATTGAACAAAAGTTCACC
     P  A  N  P  N  L  N  T  A  I  E  Q  K  F  T
1801 GCTCCTCTCGTCCGTTTCGCCACTGGCGGCCAGCCAGTTCAAAGT
     A  P  L  V  R  F  A  T  G  G  Q  P  V  Q  S
1846 GACATCGAGCAGCTGATTCAGTCGATACCAAACTTTGTCGTCAAT
     D  I  E  Q  L  I  Q  S  I  P  N  F  V  V  N
1891 GTCCCCGGAATGGGCAATGTCGATATCAGCAAGTTGGATCCGAAT
     V  P  G  M  G  N  V  D  I  S  K  L  D  P  N
1936 CTCATTGCTCACGTACTTCGTGGCGGTCAGATTCCTGGTATACCA
     L  I  A  H  V  L  R  G  G  Q  I  P  G  I  P
1981 AAAGAAACTCTGGATGGTATCGTTCAACAATACATGTTGAAAATG
     K  E  T  L  D  G  I  V  Q  Q  Y  M  L  K  M
2026 CATGAAGCGGCTGCAGCGGCTCAAAATGGTGCTCCACTGAAGAAT
     H  E  A  A  A  A  A  Q  N  G  A  P  L  K  N
2071 GGTTCGAAGTTTTTGCCACCAATCGAGAAATTCCCTCCAGATATT
     G  S  K  F  L  P  P  I  E  K  F  P  P  D  I
2116 GTCGCCTCCGTTTTTCAAGGTAAAAATCTGCCCGGATTGAATGAA
     V  A  S  V  F  Q  G  K  N  L  P  G  L  N  E
2161 GAGCAAACAAAAGTGATCAAGGACTACTACACGGCACGCCTTCCT
     E  Q  T  K  V  I  K  D  Y  Y  T  A  R  L  P
2206 GTCCCACCTGCAGCCGTCCCGGAGAGAATTCCAGCCATGGCCAAT
     V  P  P  A  A  V  P  E  R  I  P  A  M  A  N
2251 ACTCCACCCGTACAATACACCACAAGAAACACTACTGCGCCGCCT
     T  P  P  V  Q  Y  T  T  R  N  T  T  A  P  P
```

```
2296 TTCGTATTCTCGAAGGACCTGCTCGACATGCTGAAGATACTACGC
     F  V  F  S  K  D  L  L  D  M  L  K  I  L  R
2341 GTTGGTTACAATTTGAGCAAGATTCCAGCCGAGATAATCACCGCG
     V  G  Y  N  L  S  K  I  P  A  E  I  I  T  A
2386 TTTTCTCGTGGCGAAGTTCCTGATCTGCGTAAGTTGCCGACCGAT
     F  S  R  G  E  V  P  D  L  R  K  L  P  T  D
2431 CTTATCGAGCATTTCAAGTCGAACAGCGAAAAACTTGGCTTGATA
     L  I  E  H  F  K  S  N  S  E  K  L  G  L  I
2476 TTCAATCGGGTGGCTAAGGCTAATGCGTCTCTAGAAGAAATGCTA
     F  N  R  V  A  K  A  N  A  S  L  E  E  M  L
2521 GCGAAGATGCCTAAGTTCGAAAAGCCAGTGCTTTCCACATTCTCA
     A  K  M  P  K  F  E  K  P  V  L  S  T  F  S
2566 CCGTATGACATCAATCACCTCTCAAACGACATGATTCACGAGGAG
     P  Y  D  I  N  H  L  S  N  D  M  I  H  E  E
2611 GAAGAAGCTGCACGAGCAGCAAGGATGAGGGTATACACGGCAATC
     E  E  A  A  R  A  A  R  M  R  V  Y  T  A  I
2656 GCCTTGGGACTTGTTGGTGCGGCAACCGTGGTAATTCTCGGCTTT
     A  L  G  L  V  G  A  A  T  V  V  I  L  G  F
2701 TTTGCTGTCTACATTAAGAAGCAAAGAGAAGCGAAAGAAAGGAGA
     F  A  V  Y  I  K  K  Q  R  E  A  K  E  R  R
2746 GGGCCGATGACAACACCGCGCTGTGCATTTCCGGGTGCACCATTA
     G  P  M  T  T  P  R  C  A  F  P  G  A  P  L
2791 GCGAACTCTACCATGCGCAACCCACCACGTTCGCATTCGTCCAGT
     A  N  S  T  M  R  N  P  P  R  S  H  S  S  S
2836 CATTTGCACTTGCGGACTAATACGTAG 2862
     H  L  H  L  R  T  N  T  *
```

ref|NP_492348.1|	K07A12.2 [Caenorhabditis elegans] >emb|CAB03...	840	0.0
emb|CAE68896.1|	Hypothetical protein CBG14867 [Caenorhabditis br	833	0.0
emb|CAE63200.1|	Hypothetical protein CBG07555 [Caenorhabditis br	162	9e-38
ref|NP_510425.1|	SYnthetic lethal with Mec family member (sym...	158	2e-36
gb|AAQ23582.1|	RE27764p [Drosophila melanogaster]	132	1e-28
ref|NP_649228.1|	CG5195-PA [Drosophila melanogaster] >gb|AAF5...	132	1e-28
ref|XP_001352983.1|	GA18727-PA [Drosophila pseudoobscura] >gb...	129	1e-27
ref|NP_510426.1|	SYnthetic lethal with Mec family member (sym...	128	2e-27
ref|XP_001147414.1|	PREDICTED: leucine rich repeat containing 15	127	5e-27
gb|AAI01065.1|	Leucine rich repeat containing 15 [Homo sapien...	125	1e-26

26S proteasome regulatory subunit *(Hc-rpn-3)*

ATGGCCCCGATGGCAAATCCTGCTCCGCCTGTGGAGAAAATGGAAGTTGATGAACCGGCAAAAGCGGAGGAGAAGGAAACACCCAAAGAC
CTGAACGCAATCGCCGTAGAAAACATCAAAGAACATTGTGCTCTGCTCGACAAAGGGGATGTTCATTTTGTGGGGCGCGTATTACAAGTC
GTGCATAAAACCAGAAAGCAGTGTAATGCTGAGGTTTTGCACCGCTTGCTGTCAAGCCACTTGACTTCACCATCGTTGCACAAGGATGCC
CTCCTAGCATGGGTGCCGTCCACCACTCCCGTGGCCATGGAAGTGGACACAAAACTTACTCCATCTTCAACACAGCTATCTTCCTCTACA
AAATCTCCTCGACCAACTCGCAAAACAGCAACTCCCTCCGCCGAATCAGAACTTTACATTCACCTTCTTGTTCTTCTTTACCTTGTGGAT
GGGAATAGGTTAGACGACGCAAAGAAATGTGCAGAGATGTTAATTGCTCGTGCTGACGCACTTGATAAACGATCTCTTGATCCTATGGTG
GCGAAGGCTATTTTCTATCTTGCTCTAATCTACGAGCGACAGCACAAACTGCACGAACTTGGCTGCTTCCTCAACTCTCGGTTGCGCACA
GCAACCCTTCGTCATCAACGAGAATCACAGGCAGTGCTCATATACACTCTTCTTCGTTGTTATCTTATCAATAAGCAGTTCCAATCTGCA
GCGAAGCTTGTTTCAAAGGTTACGTTCCCGGGAAGGAACCTAGCTACTATGATCTCGTCGGTTTTTGTACTACAGCGCATAAAAGCTCGG
CAGCTTGATACACGCTGCAGCAGGGTACTTTTGCAAGCAATGCGCAAGGCACCACAAGAGCCGGCTATTGGTTTTAAGCCAGAATGTGCA
GAAATGGGTTATTGTTATTGGTCTGCTTCAGGGAGAGATACCAGAAAAGGTCGAATCTTCCGTCAACCAATCTATAGAAAATGTCTGCAT
CCGTACTTGGAGTTGACTCAACCCGTTCGAACTGGGGATCTAGCGCAATTCAACAACTTGGTGAAACGTCATGGTCCTATCTTCGAAAAA
GACGAAACGCTTACATTGATCGTTCGTCTTCGGCAAAATGTCATCAAAACTGCAGTCAAACAAATATCGCTGGCTTATTCTCGAATCACG
ATCAAGGACATCGCAAAAAAGTTGTGCATGAGCAATGAGGTTGAGACAGAGTATATGGTAGCGAAAGCTATAGCGGATGGTGCAATTGAT
GCAGTGATTACGTGTGATACAAAAGATGGTGCTAGGTTCATGAGGAGCTCCGAAACAGTCAATGTTTACACTACGACTGAGCCACAATTG
CATTTTGACAGTAGGATTAGGTATTGCTTAGAGCTGCATAATCAAGCCGTTAAGGCGCTCCGATTTCCACCGAAAAACAAGTGTGATGTG
GAGAGCATAGAAGCTCAACGGGAAAGAGAGCAACAGGAGTTGGAGTTTGCGAAGGAGATGGCTGATGAGGATGACGAAGATTTCTGA

```
     1   M  A  P  M  A  N  P  A  P  P  V  E  K  M  E  V  D  E  P  A
     1  ATGGCCCCGATGGCAAATCCTGCTCCGCCTGTGGAGAAAATGGAAGTTGATGAACCGGCA

    21   K  A  E  E  K  E  T  P  K  D  L  N  A  I  A  V  E  N  I  K
    61  AAAGCGGAGGAGAAGGAAACACCCAAAGACCTGAACGCAATCGCCGTAGAAAACATCAAA

    41   E  H  C  A  L  L  D  K  G  D  V  H  F  V  G  R  V  L  Q  V
   121  GAACATTGTGCTCTGCTCGACAAAGGGGATGTTCATTTTGTGGGGCGCGTATTACAAGTC

    61   V  H  K  T  R  K  Q  C  N  A  E  V  L  H  R  L  L  S  S  H
   181  GTGCATAAAACCAGAAAGCAGTGTAATGCTGAGGTTTTGCACCGCTTGCTGTCAAGCCAC

    81   L  T  S  P  S  L  H  K  D  A  L  L  A  W  V  P  S  T  T  P
   241  TTGACTTCACCATCGTTGCACAAGGATGCCCTCCTAGCATGGGTGCCGTCCACCACTCCC

   101   V  A  M  E  V  D  T  K  L  T  P  S  S  T  Q  L  S  S  S  T
   301  GTGGCCATGGAAGTGGACACAAAACTTACTCCATCTTCAACACAGCTATCTTCCTCTACA

   121   K  S  P  R  P  T  R  K  T  A  T  P  S  A  E  S  E  L  Y  I
   361  AAATCTCCTCGACCAACTCGCAAAACAGCAACTCCCTCCGCCGAATCAGAACTTTACATT

   141   H  L  L  V  L  L  Y  L  V  D  G  N  R  L  D  D  A  K  K  C
   421  CACCTTCTTGTTCTTCTTTACCTTGTGGATGGGAATAGGTTAGACGACGCAAAGAAATGT

   161   A  E  M  L  I  A  R  A  D  A  L  D  K  R  S  L  D  P  M  V
   481  GCAGAGATGTTAATTGCTCGTGCTGACGCACTTGATAAACGATCTCTTGATCCTATGGTG

   181   A  K  A  I  F  Y  L  A  L  I  Y  E  R  Q  H  K  L  H  E  L
   541  GCGAAGGCTATTTTCTATCTTGCTCTAATCTACGAGCGACAGCACAAACTGCACGAACTT

   201   G  C  F  L  N  S  R  L  R  T  A  T  L  R  H  Q  R  E  S  Q
   601  GGCTGCTTCCTCAACTCTCGGTTGCGCACAGCAACCCTTCGTCATCAACGAGAATCACAG

   221   A  V  L  I  Y  T  L  L  R  C  Y  L  I  N  K  Q  F  Q  S  A
   661  GCAGTGCTCATATACACTCTTCTTCGTTGTTATCTTATCAATAAGCAGTTCCAATCTGCA

   241   A  K  L  V  S  K  V  T  F  P  G  R  N  L  A  T  M  I  S  S
   721  GCGAAGCTTGTTTCAAAGGTTACGTTCCCGGGAAGGAACCTAGCTACTATGATCTCGTCG

   261   V  F  V  L  Q  R  I  K  A  R  Q  L  D  T  R  C  S  R  V  L
   781  GTTTTTGTACTACAGCGCATAAAAGCTCGGCAGCTTGATACACGCTGCAGCAGGGTACTT

   281   L  Q  A  M  R  K  A  P  Q  E  P  A  I  G  F  K  P  E  C  A
   841  TTGCAAGCAATGCGCAAGGCACCACAAGAGCCGGCTATTGGTTTTAAGCCAGAATGTGCA

   301   E  M  G  Y  C  Y  W  S  A  S  G  R  D  T  R  K  G  R  I  F
   901  GAAATGGGTTATTGTTATTGGTCTGCTTCAGGGAGAGATACCAGAAAAGGTCGAATCTTC
```

```
 321   R  Q  P  I  Y  R  K  C  L  H  P  Y  L  E  L  T  Q  P  V  R
 961  CGTCAACCAATCTATAGAAAATGTCTGCATCCGTACTTGGAGTTGACTCAACCCGTTCGA

 341   T  G  D  L  A  Q  F  N  N  L  V  K  R  H  G  P  I  F  E  K
1021  ACTGGGGATCTAGCGCAATTCAACAACTTGGTGAAACGTCATGGTCCTATCTTCGAAAAA

 361   D  E  T  L  T  L  I  V  R  L  R  Q  N  V  I  K  T  A  V  K
1081  GACGAAACGCTTACATTGATCGTTCGTCTTCGGCAAAATGTCATCAAAACTGCAGTCAAA

 381   Q  I  S  L  A  Y  S  R  I  T  I  K  D  I  A  K  K  L  C  M
1141  CAAATATCGCTGGCTTATTCTCGAATCACGATCAAGGACATCGCAAAAAAGTTGTGCATG

 401   S  N  E  V  E  T  E  Y  M  V  A  K  A  I  A  D  G  A  I  D
1201  AGCAATGAGGTTGAGACAGAGTATATGGTAGCGAAAGCTATAGCGGATGGTGCAATTGAT

 421   A  V  I  T  C  D  T  K  D  G  A  R  F  M  R  S  S  E  T  V
1261  GCAGTGATTACGTGTGATACAAAAGATGGTGCTAGGTTCATGAGGAGCTCCGAAACAGTC

 441   N  V  Y  T  T  T  E  P  Q  L  H  F  D  S  R  I  R  Y  C  L
1321  AATGTTTACACTACGACTGAGCCACAATTGCATTTTGACAGTAGGATTAGGTATTGCTTA

 461   E  L  H  N  Q  A  V  K  A  L  R  F  P  P  K  N  K  C  D  V
1381  GAGCTGCATAATCAAGCCGTTAAGGCGCTCCGATTTCCACCGAAAAACAAGTGTGATGTG

 481   E  S  I  E  A  Q  R  E  R  E  Q  Q  E  L  E  F  A  K  E  M
1441  GAGAGCATAGAAGCTCAACGGGAAAGAGAGCAACAGGAGTTGGAGTTTGCGAAGGAGATG

 501   A  D  E  D  D  E  D  F  -
1501  GCTGATGAGGATGACGAAGATTTCTGA
```

Sequences producing significant alignments:	(Bits)	Value
emb|CAE71252.1| Hypothetical protein CBG18132 [Caenorhabditis br	427	1e-117
ref|NP_498869.1| proteasome Regulatory Particle, Non-ATPase-l...	421	6e-116
ref|XP_319831.2| ENSANGP00000016528 [Anopheles gambiae str. P...	288	8e-76
gb|AAC18058.1| diphenol oxidase-A2 [Anopheles gambiae]	286	2e-75
ref|XP_623947.1| PREDICTED: similar to Probable 26S proteasom...	284	1e-74
gb|EAT35739.1| 26S proteasome regulatory subunit S3 [Aedes aegyp	283	2e-74
dbj|BAB71019.1| unnamed protein product [Homo sapiens]	282	5e-74
ref|NP_001026533.1| proteasome 26S non-ATPase subunit 3 [Gall...	281	6e-74
ref|XP_784007.1| PREDICTED: similar to PSMD3 [Strongylocentro...	279	4e-73
emb|CAG33053.1| PSMD3 [Homo sapiens]	278	5e-73

http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=Protein&list_uids=126308136&dopt=GenPept

4.7 Screening of chemical compounds (milestone 7)

**** Preliminary screening of compounds on third-stage larvae (L3s) of four parasitic nematodes***

The aim here was to undertake a preliminary evaluation of a possible nematocidal effect of any of 131 chemical compounds using L3s of four different species of trichostrongylid nematodes in a "liquid" assay.

Parasites and compounds: Lambs (8-12 weeks of age), maintained under helminth-free conditions, were infected intraruminally with infective third-stage larvae (L3) of *Trichostrongylus vitrinus* or *Haemonchus contortus*. The patency of the infection (~24 days after inoculation) was established *via* the detection of strongylid eggs in the faeces using a standard McMaster flotation method. The L3 stage was collected from the faeces of experimentally infected sheep and stored in water at 4-9 ^{o}C, depending on the species, until needed. The L3 stages of *Teladorsagia circumcincta* and *Trichostrongylus colubriformis* were provided by Dr Gareth Hutchinson (Elizabeth Macarthur Agricultural Institute, Department of Primary Industries, Woodbridge Road, Menangle, NSW 2568). Stock solutions of each compound were prepared by suspending the compound in 100% dimethyl sulphoxide (DMSO) to a final concentration of 100 mM.

Microplate assay: Approximately 50 L3s of each species were added (in 95 µl of water) to individual wells of a 96-well microtitre plate. All compounds were tested initially at a concentration of 5 mM, by adding 5 µl of compound (dissolved in 100% DMSO). The plates were incubated overnight at 22-24 ^{o}C (room temperature). Plates were examined under a binocular microscope, and the compounds showing "kill" of L3 were determined. Concentrations of > 5 mM were not tested, as DMSO alone kills larvae when the proportion exceeds 5% of the total volume. Many of the compounds were predicted to have limited solubility in aqueous solution, thus requiring the presence of organic solvent. Additional controls included wells with and without various concentrations of DMSO alone. For each compound, the % "kill" was estimated by subtracting the number of dead L3s in a well from the number of dead L3s in the DMSO (negative)-control well. Table 4.7.1 provides a summary of the compounds which, in this preliminary assessment, had some effect on one or more of the said parasitic nematodes of small ruminants.

Table 4.7.1: Preliminary results from the screening of compounds against four species parasitic nematodes of small ruminants. Total numbers of compounds tested against each species are given in parentheses. Some effect on larvae (mortality of ≥ 50%) was recorded by comparison to a "negative control" (DMSO without compound) for the following compounds:

Compounds*			
***T. vitrinus* (81)**	***H. contortus* (121)**	***T. colubriformis* (130)**	***T. circumcincta* (126)**
BS1066	SS-TH1-48	BS2015	BS 2010
BS1089	THTP19-1	BS1064	BS 1068
SS-TH1-12	THTP-31-3	CgNorC d14	BS 1065
SS-TH1-13	THTP-33	SS-TH1-18	CgNorC d18
SS-TH1-14	THTP-39	SS-TH1-69	SS-TH1-68
SS-TH1-15	THTP-63	SS-TH1-79	SS-TH1-79
SS-TH1-21	THTPA6-2	THTP-3	SS-TH1-73
SS-TH1-70	CgNorC d14	THTP-23-2	SS-TH1-56
SS-TH1-71		THTP-50	SS-TH1-10
SS-TH1-79		THTP-63	SS-TH1-22/28
		THTP-50	SS-TH1-08
		THTP-76	THTP-83
		THTP-83	THTP-3
		THTP-87	THTP-80
		THTPA-4-2	THTP-30-2
			THTP-27

* Some compounds formed heavy precipitates or crystals.

**** Extended, preliminary screening of compounds on third-stage larvae (L3s) of* Trichostrongylus colubriformis**

The aim was to undertake a preliminary evaluation of a possible nematocidal effect of any of a set of 131 compounds using L3s of *T. colubriformis* using 5 mM or a ten-fold serial titration of 5 mM to 50 nM of each compound.

Larvae: The L3 stage of *T. colubriformis* (*via* Elizabeth Macarthur Agricultural Institute, Department of Primary Industries, Woodbridge Road, Menangle, NSW 2568). The larvae were stored at 4^{o} C until required.

Chemicals: One hundred and thirty one chemicals were tested (Table 4.7.2). Each chemical was provided, with details on molecular weight and the mass of chemical in each tube. Using this information, each chemical was resuspended in 100% dimethyl sulphoxide (DMSO) to a final concentration of 100 mM.

Table 4.7.2: Codes of the 131 compounds tested using *T. colubriformis* L3s

BS1050	BS2004	CgNorC d4	SS-TH1-45	SS-TH1-85	THTP-30-1	THTP-81
BS1064	BS2005	CgNorC d5	SS-TH1-46	SS-TH1-87	THTP-30-2	THTP-82
BS1065	BS2006	CgNorC d6	SS-TH1-48	SS-TH1-88	THTP-31-1	THTP-83
BS1066	BS2007	CgNorC d7	SS-TH1-50	SS-TH1-91	THTP-31-2	THTP-84
BS1067	BS2008	CgNorC d8	SS-TH1-56	SS-TH1-92	THTP-31-3	THTP-86
BS1068	BS2009	CgNorC d9	SS-TH1-58	SS-TH1-93	THTP-33	THTP-87
BS1069	BS2010	SS-TH1-05	SS-TH1-60	SS-TH1-97	THTP-39	THTP-89
BS1071	BS2011	SS-TH1-08	SS-TH1-65	THTP-13-2	THTP-50	THTP-91
BS1074	BS2012	SS-TH1-09	SS-TH1-66	THTP-16-2	THTP-51	THTPA-4-2
BS1078	BS2013	SS-TH1-10	SS-TH1-67	THTP-18-2	THTP-52	THTPA-5-2
BS1080	BS2015	SS-TH1-12	SS-TH1-68	THTP-18-3	THTP-53	THTPA-6-2
BS1083	CgNorC d1	SS-TH1-13	SS-TH1-69	THTP-19-1	THTP-55	
BS1085	CgNorC d10	SS-TH1-14	SS-TH1-70	THTP-19-2	THTP-56	
BS1086	CgNorC d11	SS-TH1-15	SS-TH1-71	THTP-20-1	THTP-57	
BS1088	CgNorC d13	SS-TH1-18	SS-TH1-76	THTP-23-1	THTP-63	
BS1089	CgNorC d14	SS-TH1-20	SS-TH1-77	THTP-23-2	THTP-64	
BS1091	CgNorC d15	SS-TH1-21	SS-TH1-79	THTP-26	THTP-73	
BS2001	CgNorC d16	SS-TH1-22/28	SS-TH1-81	THTP-27	THTP-74	
BS2002	CgNorC d17	SS-TH1-35	SS-TH1-82	THTP-3	THTP-76	
BS2003	CgNorC d18	SS-TH1-43	SS-TH1-83	THTP-3	THTP-80	

Microplate assay: Approximately 50 L3s of each species were added (in 95 µl of water) to the wells of a 96-well microtitre plate. All compounds were tested initially at concentrations of 5 mM and then, in a subsequent round of testing, in ten-fold dilutions to 50 nM. The plates were incubated overnight at room temperature. Plates were examined under a binocular microscope, and the compounds showing “kill” of L3s were determined. Concentrations of >5 mM were not tested, as DMSO alone killed larvae when the proportion exceeded 5% of the total volume. Many of the compounds had limited solubility in aqueous solution, thus requiring the presence of the organic solvent (i.e. DMSO). Additional controls included wells with or without various concentrations of DMSO alone. In Table 4.7.3 is a summary of the findings.

Table 4.7.3: Compounds shown to kill (≥ 25%) *T. colubriformis* L3s, following serial dilution

Compound code
CgNorC d11
CgNorC d8
SS-TH1-97
THTP-23-2
THTP-33
THTP-50
THTP-55
THTP-56
THTP-64
THTP-76
THTP-82
THTP-86
THTP-91

**** Preliminary screening of compounds on* Trichostrongylus colubriformis *third-stage larvae (L3s) and* Caenorhabditis elegans *adults on agar plates***

Aim: To test compounds in a solid agar assay L3s of *T. colubriformis* or the free-living nematode *Caenorhabditis elegans.*

Parasite: The L3 stage of *T. colubriformis* was provided by Gareth Hutchinson (Elizabeth Macarthur Agricultural Institute, Department of Primary Industries, Woodbridge Road, Menangle, NSW 2568). Larvae were kept at 4^{o} C until required. Initial stocks of *C. elegans* were provided by Dr Warwick Grant (Department of Genetics, LaTrobe University). The strain was cultured on MYOB plates on a lawn of *Escherichia coli* (OP50 strain) at 23^{o} C until the adult stage was reached.

Chemicals: Eighty-one compounds were tested (Table 4.7.4). Individual compounds were prepared by adding 2.5 mg into a 1.5 ml microcentrifuge tube and adding DMSO to a final concentration of 100 mM. An aliquot of this stock solution was then diluted (1/10) in DMSO to be used as a "working solution" of 10 mM. Cydectin^{TM} (a commercial form of moxidectin, Fort Dodge) was used as a positive control.

Table 4.7.4: Codes of the compounds tested on *T. colubriformis* L3s and *C. elegans* adults in agar plate assays

5,6 NorC	*CgLD49*	CgNorC d17	SS-TH1-14
BS1066	*CgLD5*	CgNorC d18	SS-TH1-14
Cg Skwd-5,6-NorCd1	*CgLD50*	CgNorC d2	SS-TH1-18
Cg Skwd-5,6-NorCd2	*CgLD52*	CgNorC d4	SS-TH1-22/28
Cg Skwd-NorCd1	*CgLD53*	CgNorC d5	SS-TH1-56
Cg UDA d1	*CgLD54*	CgNorC d6	SS-TH1-68
Cg UDA d2	*CgLD56*	CgNorC d7	SS-TH1-71
Cg UDA d3	*CgLD57*	CgNorC d8	THTP-19-1
Cg UDA d4	*CgLD58*	CgNorC d9	THTP-1A
Cg UDA d5	*CgLD60*	MT-NOVO-02	THTP-23-2
CgLD1	*CgLD63*	MT-NOVO-03	THTP-30-2
CgLD12	*CgLD64*	MT-NOVO-04	THTP-39
CgLD13	*CgLD8*	MT-RO-01	THTP-63
CgLD19	CgNorC d1	MT-RO-02	THTP-63
CgLD2	CgNorC d10	MT-RO-03	THTP-83
CgLD26	CgNorC d11	MT-RO-04	THTP-87
CgLD38	CgNorC d12	NorC	THTPA-4-2
CgLD39	CgNorC d13	Novo-6	THTPA-5-2
CgLD44	CgNorC d14	SS-TH1-10	
CgLD46	CgNorC d15	SS-TH1-12	
CgLD48	CgNorC d16	SS-TH1-13	

Methods: The approach is based on methods described in previous studies (Burns et al., 2006 and Kwok et al., 2006). In brief, 80 µl of the solvent or the appropriate dilution of anthelmintic were added to 8 ml of molten MYOB (~ 45 ^{o}C), vortexed and poured into two 60 mm Petri-dishes. This solution was allowed to set and dry for around 15 min. Approximately 50 *T. colubriformis* L3s or *C. elegans* adults were then added (in a maximum volume of 10 µl water) to each 60 mm Petri-dish. Two replicates were performed for each concentration of compound. The negative control plates contained 0.5% DMSO, whilst the positive control plates contained moxidectin (final concentration of 1.8 or 3.5 mM). Each plate was then incubated at 22 ^{o}C overnight and scored for % mortality.

Results: Using this approach, none of the compounds showed toxicity to *T. colubriformis.* One compound, MT-NOVO-02, showed ≥ 50% toxicity in *C. elegans* on two different days.

**** Screening of chemical compounds using an optimised LDA for* H. contortus (see methods)**

A total of 120 compounds were tested or retested multiple times using an industry standard larval development assay (LDA) based on that of Gill et al. (1995) (see Table 4.7.5). Individual compounds were prepared by dissolving 2.5 mg in DMSO (in a 1.5 ml microcentrifuge tube) to a final concentration of 100 mM. An aliquot of this stock solution was then diluted to 10 mM. Cydectin (a commercial form of moxidectin manufactured by Fort Dodge) was used as a positive control compound at 12.5, 25, 50 and 100 μM concentrations.

Table 4.7.5: Codes of the compounds tested or retested against *H. contortus* in a larval development assay (LDA).

5,6 NorC	*CgLD44*	CgNorC d13	Cg UDA d5	MT-CN-02	SS-TH1-10
BS1066	*CgLD46*	CgNorC d14	CgUDA d7a	MT-CN-03	SS-TH1-12
Cg Skwd-5,6-NorCd1	*CgLD48*	CgNorC d15	CgUDA d7b	MT-CN-05	SS-TH1-13
Cg Skwd-5,6-NorCd2	*CgLD49*	CgNorC d16	CgUDA d9a	MT-CN-06	SS-TH1-14
Cg Skwd-NorCd1	*CgLD5*	CgNorC d17	CgUDA d9b	MT-CN-07	SS-TH1-18
CgAMcd2	*CgLD50*	CgNorC d18	CgUDA d10a	MT-NOVO-00	SS-TH1-22/28
CgAMcd25	*CgLD52*	CgNorC d2	CgUDA d10b	MT-NOVO-02	SS-TH1-56
CgAMcd33	*CgLD53*	CgNorC d4	Cg UDA d11a	MT-NOVO-05 (racemic)	SS-TH1-68
CgAMcd34	*CgLD54*	CgNorC d5	Cg UDA d11b	MT-NOVO-05a (pure)	SS-TH1-71
CgAMcd4	*CgLD56*	CgNorC d6	Cg UDA d12a	MT-NOVO-06	THTP-1A
CgAMcd47	*CgLD57*	CgNorC d7	Cg UDA d12b	MT-NOVO-07	THTP-23-2
CgAMcd7	*CgLD58*	CgNorC d8	Cg UDA d13b	MT-NOVO-07a	THTP-30-2
CgIminC	*CgLD60*	CgNorC d9	Cg UDA d13b	MT-NOVO-08	THTP-39
CgLD1	*CgLD63*	CgTNorC p32	Cg UDA d14a	MT-NOVO-09 (racemic)	THTP-55
CgLD12	*CgLD64*	Cg UDA d1	Cg UDA d14b	MT-NOVO-09a (pure)	THTP-63
CgLD13	*CgLD8*	Cg UDA d1a	Cg UDA d17a	MT-NOVO--11a	THTP-82
CgLD19	CgNorC d1	Cg UDA d1b	Cg UDA d17b	MT-NOVO--11b	THTP-83
CgLD26	CgNorC d10	Cg UDA d2	Cg UDA T1	MT-RO-03	THTP-87
CgLD38	CgNorC d11	Cg UDA d3	Cg UDA T5	NorC	THTPA-4-2
CgLD39	CgNorC d12	Cg UDA d4	MT-CN-01	Novo-6	THTPA-5-2

Parasite

The strain used for these assay was Haecon 5, eggs of which were provided by the Department of Primary Industries, Attwood, Victoria. As stated previously, this strain is known to be 66% resistant to benzimidazoles, but is susceptible to macrocyclic lactones (P. Presidente, personal communication, October 2008); therefore, the susceptibility of this strain to macrocyclic lactones, justified the use of moxidectin as a positive control compound in the assay.

Isolation of H. contortus eggs from sheep faeces

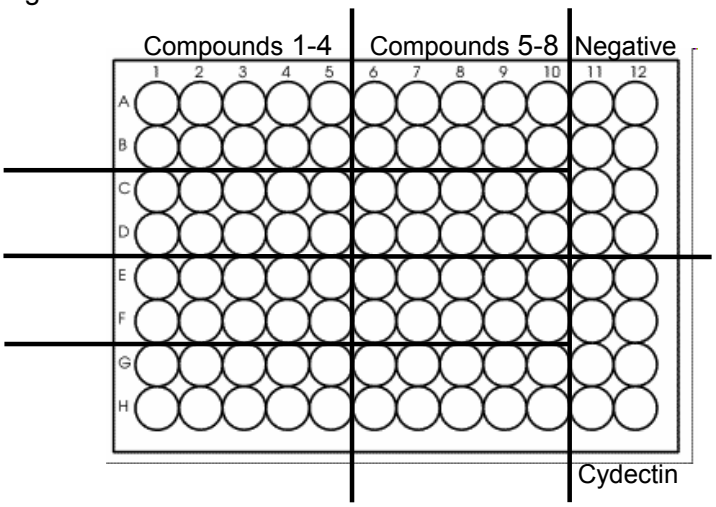
Ten grams of faeces were crushed and suspended in ~100 ml of sucrose solution (specific gravity 1.2), sieved (mesh size: ~ 1 mm) and transferred to a petri-dish. Strips of commercially available overhead transparency (Kodak) were placed on the surface of the suspension (to allow the eggs to stick) and removed after 45 min. The eggs were collected by washing them from the transparency strips with water into a 50 ml centrifuge tube. The eggs were then collected by

centrifugation at 1,000 x g for 5 min, the supernatant removed and the eggs collected in ~ 500 μl of water. Eggs were enumerated and the suspension adjusted with water to 200 eggs/20 μl.

Dilution and preparation of compounds in solid agar

Inhibitors were initially tested at five different concentrations: 6.25, 12.5, 25, 50 and 100 μM. Compounds found to be toxic were further tested at 2.5, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90 and 100 μM. Dilutions of each compound (10 μl in total) were performed in 1.5 ml microcentrifuge tubes, 1 ml of molten 2% agar added, the tube vortexed and the agar aliquoted (150 μl) into the wells of a 96-well microtitre plate. DMSO (1% v/v) was used in a number of wells as solvent-only controls (negative controls) whilst cydectin was used as a positive control. Concentrations of moxidectin used were: 12.5, 25, 50 and 100 μM. Two replicates were performed for each concentration of each compound (Figure 4.7.1 shows the plate set-up). Approximately 200 eggs (20 μl) were then added to each well. The plates were then sealed with adhesive, breathable lids and incubated overnight at 27 ^{o}C. Plates were checked the following morning and afternoon to ensure that the majority of eggs had hatched. Eggs were examined for any indication of an ovicidal effect (evidenced by a significant number of unhatched eggs compared with the control wells). Following the hatching of most eggs, 15 μl of nutritive medium was added to feed the larvae. Nutritive medium was prepared as follows: 3 ml of 10 x Earle's balanced salt solution (EBSS) [potassium chloride (KCl) 53 mg/l, sodium bicarbonate ($NaHCO_3$) 261.9 mg/l, sodium chloride (NaCl) 1172.4 mg/l, sodium phosphate monobasic (NaH_2PO_4-H_2O) 10.1 mg/ml) was added to 27 ml of yeast extract solution and the pH adjusted to 5.4-5.6 by adding bicarbonate. The yeast extract solution was prepared as follows: 1 g of yeast extract was added to 90 ml of 0.85% physiological saline and autoclaved for 20 min at 121^{o} C. Following further 6 days of incubation, the number of L3s that had developed in each well was determined. The reproducibility of "kill" was assessed on different days.

Figure 4.7.1.



Results

One hundred and twenty compounds (Table 4.7.5) were tested against *H. contortus.* Seven compounds, namely CgNorC_d12, MT-CN-01, MT-CN-03, MT-CN-05, MT-CN-07, SS-TH1-12 and SS-TH1-14, killed ~100% of *H. contortus* larvae in the LDA. Only compounds covered by the patents (CgNorC d12, SS-TH1-12 and SS-TH1-14) were pursued further. All compounds were screened at least three times in preliminary rounds of testing, and some were tested up to 5 times. Compounds CgNorC_d12 and SS-TH1-12 showed ~100% kill at levels as low as 70 μM (upon separate serial titrations at 10 μM increments). The LD_{50} values estimated for these compounds were: 20-30 μM for CgNorC d12, 30-40 μM for SS-TH1-12 and SS-TH1-14, and 10 μM for the MT-CN compounds (MT-CN-01, MT-CN-03, MT-CN-05 and MT-CN-07). There was no evidence of an ovicidal effect for any of the compounds. The MT-CN compounds were used as positive controls in the assay, as these compounds had been tested by previously Ali et al. (2007) and found to be toxic to *H. contortus.* Testing of compounds against *H. contortus* continues within the framework of ARC LP0882285.

5. Success in Achieving Objectives and Milestones

All of the objectives and milestones of the present project were met.

The project focused on working toward new solutions to parasitic diseases problems of animals, leading to improved livestock production and decreased losses. Although there is relatively little information on parasite genomes, substantial information is available on the free-living nematode, *C. elegans*. Given that the complete genome sequence of *C. elegans* has been determined, a well-defined model system is in place to test the function of orthologous genes from parasitic nematodes. This is of paramount importance because parasitic nematodes cannot be propagated and maintained effectively in *in vitro* culture systems. Therefore, we selectively isolated genes associated with reproduction and development, and made connections between the parasite genes (its genome) and their function in *C. elegans*. This molecular genetics approach coupled with the use of modern technologies allow investigations into the complex parasite system and the factors which determine the phenotype of the 'treated' worms to be determined.

The present project has enabled all participants to establish unique infrastructure, skills and expertise in genomics, bioinformatics and drug target discovery. The key goal of this project was the identification of key molecules associated predominantly with development and/or reproduction in economically important parasites. The project was extended to genomic and post-genomic work, with a view toward developing biotechnological products, and discovered three synthetic chemical compounds with nematocidal activity in *H. contortus* in a larval development assay (LDA). This outcome was protected by patents and is of major commercial importance to the industry partners. The project also continually maintained an emphasis on research and the development of the scientific skills base required for the 'next generation' biotechnology industry. The use of this integrated genomic-bioinformatic platform extended logically to the barber's pole worm, *Haemonchus contortus*.

The socio-economic benefits flowing from the project are: (i) Enhanced focus on animal health biotechnology through the development of safe anti-parasite compounds; (ii) Improved and sustainable control of important parasites of sheep and other livestock; (iii) Increased profitability of agricultural animal production; (iv) Improved welfare of agricultural animals; (v) Improved international marketability and consumer acceptance, due to reduced use of conventional anti-parasitic drugs during production; (vi) Reduced risk of food animals contributing to the pool of drug resistance, and thus improved sustainability of agriculture; (vii) Development of a technology platform for further applications in genomics and post-genomics of pathogens of global significance; (viii) Capturing of the benefits and outcomes of fundamental research and strengthen the links between fundamental and applied research; (ix) Enhanced quality and quantity of scientifically skilled people in biology and biotechnology.

It was a major priority of the project to conduct high quality science, to internationalise the research, in order to increase global visibility of MLA, the participants and Australian science, and, importantly, to enhance scientific training, in order to improve opportunities for Australian researchers. An important outcome was the establishment of a skills base for the long-term training of scientists, thus providing them with career opportunities and prospects in the biotechnology industry in Australia. Many committed scientists, including Drs Bronwyn Campbell, Min Hu, Shoba Ranganathan, Paul Sternberg, Weiwei Zhong, Ian Beveridge, Stuart Ralph, Maria Doyle, Elida Rabelo, James Lok, Sia Nikolaou, Shivashankar Nagaraj and Mark Pellegrino, worked on this project.

6. Impact on Meat and Livestock Industry – Current and Future

The present project led to conceptual advances in the knowledge and understanding of molecular biological processes in parasites (order Strongylida) and key parasitic diseases of livestock through the application of targeted genomic and bioinformatic technologies. These improved insights into molecular aspects of reproduction and development will lead to novel and innovative methods of parasite control, with significant biotechnological and commercial outcomes in the longer term. Importantly, the present project led to a consolidated pipeline for the selection and pre-validation of candidate drug targets in parasitic nematodes, particularly *Haemonchus contortus*, the 'barber's pole worm of sheep.

In the very near future, we plan to boost agricultural nematology through an intense effort to sequence the genomes and transcriptomes of a range of parasitic nematodes of paramount socioeconomic and agricultural importance. The impact of these parasites is major in livestock due to clinical and subclinical infections and production losses associated with disease and deaths. Most parasites are relatively host-specific, but some cross host species boundaries. The control of these nematodes, particularly strongylids, has become extremely difficult worldwide due to widespread drug resistance to all major classes of anthelmintics. There is a significant, active research community working on strongylid and ascaridoid nematodes (clades V and III, respectively) but work on enoplids (I and II) is limited. These nematode groups have been the subject of substantial research efforts on epidemiology, control strategies, drug efficacy trials and drug resistance as well as vaccine and diagnostic test development. Some are key models of parasitism as well as major pathogens in their own right. Well-defined strains of these parasites are available in a number of key laboratories. Extensive expertise is available via a range of laboratories worldwide, with which we have links, including California Institute of Technology (Caltech), the Veterinary Schools in Hannover (Prof. Schnieder), Vienna (Prof. Joachim) and Copenhagen (Prof. Thamsborg) as well as the Ghent University (Dr. Peter Geldhof), Moredun Research Institute (Dr. Alasdair Nisbet), and the Institut National de Recherche Agronomique (INRA; Dr. Herve Hoste).

The genome information generated will have immediate utility, aiding in the identification of novel target molecules for the control of these parasites by vaccination or anthelmintics. In addition, these datasets will be invaluable for characterizing developmental processes and the mechanisms of drug resistance. Sequence information will prove an invaluable resource to the global scientific community for comparative genomics to seek common genes linked to this resistance problem. Having genome sequences available also opens the door to comprehensive and meaningful investigations of host-parasite relationships, evolutionary relationships, gene expression and applied areas, such as drug development and diagnostic tests, and biotechnological outcomes of socioeconomic value. The benefits would be to accelerate the development of new and innovative methods for the control of these parasites in livestock, defining the genetic mechanisms underlying drug resistance and susceptibility, with the prospect of extending the useful life of existing drugs as well as providing the means for meaningful whole animal studies. Importantly, detailed insights into the molecular developmental processes could lead to specific intervention strategies, which would have a significant impact on herd or flock health through the accumulated reduction in pasture contamination and hence the ingestion of infective parasite stages. Thus, detailed knowledge of the genome will also benefit studies focused on novel methods of parasite control through new compounds or vaccines, with major commercial implications.

7. Conclusions and Recommendations

Sequencing projects for parasites have focused predominantly on the use of a genome-wide EST approach. Besides “house-keeping” and structural genes isolated employing this approach, some genes relating to potential drug targets have been identified in EST data sets. These include genes encoding antioxidant and de-toxifying enzymes, proteinases, proteinase inhibitors, cyclophilins, neurotransmitter receptors, transporters and nuclear hormone receptors. However, such an approach can have limitations in that many of the sequences obtained are abundantly represented, thus generating redundant sequence information of limited scientific value. This limitation can be overcome by employing the targeted approach employed in the present project. This integrated genomic-bioinformatic approach has provided first insights into molecular developmental and reproductive processes in selected parasitic nematodes and provides a platform for exploring the functional roles of key molecules as well as the discovery of new anthelmintic targets.

The technological advances made in this project provide exciting prospects for investigating the molecular aspects of development, survival and reproduction in a range of parasitic nematodes as well as for developing new methods of parasite control. Particularly when whole genome data sets become available for parasitic nematodes (particularly *H. contortus*). Since relatively few developmentally-regulated molecules have been characterized for parasitic nematodes of animals, EST data sets provide a basis for transcriptional analysis by microarray, whole genome sequencing and/or assessing gene function. However, considerable improvements are still needed in the processing and bioinformatic analyses of large data sets. The availability of *C. elegans* and its complete genome sequence, and information on gene function in this nematode, provide a platform for comparative analyses of homologues/orthologues from parasitic nematodes, because there seems to be a relatively close evolutionary relationship between *C. elegans* and strongylid nematodes and because most parasites cannot be maintained effectively and/or propagated *in vitro*. Recent investigations also indicate that functional genomic approaches, such as the RNAi, will be useful for exploring developmental processes in parasites, although, clearly, improved approaches are required for the *in vitro* cultivation of parasitic nematodes, particularly *H. contortus*, and for the recording/scoring of RNAi phenotypes. Also the application of proteomic technology will allow the analysis of differentially expressed and developmentally regulated proteins from small amounts of parasite material, which will enable an important link between the regulation of transcription and translation. Hence, the combined use of improved genomic, bioinformatic and proteomic approaches to focus on development in parasitic nematodes should improve our understanding of the molecular biology of key processes, including moulting, invasion and establishment in the host, hypobiosis, as well as tissue-specific gene expression, sexual differentiation, maturation and behaviour. Such an integrated, ‘systems biology’ approach will facilitate the identification of new anti-parasitic drug targets linked to key biological or biochemical pathways. The recommendation is to substantially enhance research on the genomes of parasitic nematodes in order to discovery these key pathways.

Nematodes are major parasites of livestock and crops, and a few are used for the biological control of insect pests. The sheer number of species, the relatively small research communities working on them, and the historical emphasis of large genome centres focused primarily on organisms related to human health and disease have slowed the molecular research of the Nematoda. The plan of Professor Gasser and his team is to boost agricultural nematology through an intense effort to sequence the genomes and transcriptomes of a range of parasitic nematodes of paramount socioeconomic and agricultural importance. The effort is planned to provide high quality genomic sequences, automated and experimental annotation of the genomic data, as well as *in silico* systems for the display of genomic information in browsers, such as WormBase’s GBrowse and the UCSC Genome Browser. In addition, it will provide detailed transcriptomic data for

approximately two-thirds of the genes in each species of nematode studied. The transcriptomic data will be used to assist in the assembly of genomes and the determination of gene structures. The genome annotations will provide, for the first time, an extremely solid foundation for the large-scale prediction of new drug targets for a wide range of pathogens, based on an improved understanding of global biochemical pathways and molecular interactions. The recent success in comparative analysis of microbial genomes to infer metabolism is the intellectual precedent for the present project. Large-scale genomic sequence data will make a wide range of functional genomic experiments possible, such as RNAi, DNA-mediated transformation, microarray and gene expression level measurements, proteomics, and single nucleotide polymorphism (SNP) detection, which have not been feasible until now. The extensive data from this future project will be analysed using WormBase, a stably funded public database of nematode genomics and biology. Importantly, the datasets for individual nematodes will stimulate research in both fundamental nematology (e.g., detailed molecular biological, biochemical, phylogenetic, and physiological research) and applied nematology (e.g., development of drugs, vaccines, other new intervention strategies as well as diagnostic tools).

We have the molecular biology expertise, sequencing capabilities, genome assembly, RNA annotation, computational annotation, database and genome browser display capabilities in place at California Institute of Technology, the main collaborating institution. We have extensive expertise in experimental biology as well as animal systems to fully exploit the outcomes from the genome sequencing effort. We will be analyzing and utilizing these genomes, interpreting the data in the context of the biology of the parasites, epidemiology, ecology, evolution, host-parasite relationships and disease. We will select a range of species of nematodes that parasitize livestock. Our target organisms will be amongst the most economically significant known, accounting for huge livestock and production losses due to the diseases they cause. We have balanced the desire for a complete survey, with the power that comes from particular, close comparisons. Thus, will select 2-4 members of various clades to allow detailed comparative genomic-bioinformatic analyses. Although we will be generating draft rather than finished genomes, they should be of lasting value: it took almost 10 years for *C. elegans* to go from an excellent, publishable draft genome to a complete sequence; and, even now, the *C. elegans* genome sequence still gets updated as errors are identified through intense analyses of individual genes. We will prioritize the genomes to be sequenced by inherent interest, availability and suitability of material, as well as comparative genomic and phylogenetic value, the availability of model systems for fundamental and applied studies, as well as major potential for applied and biotechnological outcomes of socioeconomic value. MLA is invited to be part of this next, exciting programme (RBG).

8. References

Ali A (2007) Bioorg Med Chem Lett 17, 993-997.
Blaxter ML (1998) Science 282, 2041-2046.
Blaxter ML et al. (1998) Nature 392, 71-75.
Burns AR et al. (2006) Nature Protoc 1, 1906-1914.
Gasser RB, Newton SE (2000) Int J Parasitol 30, 509-534.
Geary TG et al. (2004) Trends Parasitol 20, 449-455.
Gill J et al. (1995) Int J Parasitol 25, 463-470.
Kwok TCY et al. (2006) Nature 441, 91-95.
McLeod RS (1995) Int J Parasitol 25, 1363-1367.

9. Acknowledgements

We are very grateful to the Meat and Livestock Australia (MLA) for support. Collaborators who supported the programme (mentioned in the text) are also gratefully acknowledged.