Published in final edited form as:

Parasitology. 2008 July ; 135(8): 869–880. doi:10.1017/S0031182008004472.

Advances in the sequencing of the genome of the adenophorean nematode *Trichinella spiralis*

M. MITREVA^{1,*} and D. P. JASMER²

¹Genome Sequencing Center, Department of Genetics, Washington University School of Medicine, St Louis, MO 63108, USA

²Department of Veterinary Microbiology and Pathology, College of Veterinary Medicine, Washington State University, Pullman, WA 99164-7040, USA

SUMMARY

The adenophorean nematodes are evolutionarily distant from other species in the phylum Nematoda. Interspecific comparisons of predicted proteins have supported such an ancient divergence. Accordingly, *Trichinella spiralis* represents a basal nematode representative for genome sequencing focused on gaining a deeper insight into the evolutionary biology of nematodes. In addition, molecular characteristics that are conserved across the phylum could be of great value for control strategies with broad application. In this review, we describe and summarize progress that has been made on the sequencing and analysis of the *T. spiralis* genome. The genome sequence was used in preliminary analyses for the investigation of specific questions relating to the biology of *T. spiralis* and, more generally, to parasitic nematodes. For instance, we evaluated an unusually large DNase II-like protein family, predicted proteins of prospective interest in the parasite-host muscle cell interaction, anthelmintic targets and prospective intestinal genes, the encoded proteins (potentially) linked to immunological control against other nematodes. The results are discussed in relation to characteristics that are broadly conserved among evolutionary distant nematodes. The results lead to expectations that this genome sequence will contribute to advances in research on *T. spiralis* and other parasitic nematodes.

Keywords

Trichinella spiralis; expressed sequence tags; ESTs; genome; sequencing; intestine; host-parasite interactions

INTRODUCTION

In this review, we summarize the progress that has been made on the sequencing and analysis of the genome of *Trichinella spiralis*. Genome information from *T. spiralis* has immediate application to both intraspecific (*T. spiralis*) and interspecific investigations into the biology of parasitic nematodes. Major groups belonging to the phylum Nematoda have been categorized into 5 clades (I to V; Blaxter *et al.* 1998). The bulk of genomic data represents clade V-nematodes, which includes the free-living nematode *Caenorhabditis elegans*, and Clade III-nematodes in relation to filarial parasites. *T. spiralis* is a nematode of clade I, the most distant evolutionarily from clade V. Hence, the genome sequence of *T. spiralis* will allow an important survey of representatives that span the phylogenetic

extremes of the Nematoda. There is keen interest in identifying biological features that uniquely define the nematodes. Insights into unique aspects of host-nematode interactions could assist in developing control strategies with broad applicability. Knowledge of characteristics that are broadly conserved, whether unique or not to nematodes, have considerable potential. In addition, comparative genomic analyses provide useful information that should be able to explain adaptations responsible for the radiation of nematode lineages and species. Thus, studies of genomic data are expected to provide exciting insights into molecular determinants of parasitism. The preliminary analyses presented herein provide examples of such applications.

The biological characteristics of *T. spiralis* and related species within this genus are complex, diverse and challenging to study. The importance of the genome sequence to advance progress in this area has been discussed (Mitreva and Jasmer, 2006). A few examples of contemporary biological topics will be mentioned here to illustrate some expected benefits.

A part of the diversity amongst *Trichinella* spp. stems from the dichotomy of two distinct evolutionary lineages within the genus, the encapsulated lineage and the non-encapsulated lineage. These two lineages diverged an estimated 15–20 million years ago (Zarlenga *et al.* 2006). The designation of the encapsulated form relates to the morphological characteristic of the collagen capsule, which develops early during the muscle stage of the infection to surround host muscle cells infected by first-stage larvae (L1). In contrast, non-encapsulated species do not provoke the production of the capsule. While the collagen capsule appears to be host derived (Haehling *et al.* 1995; Polvere *et al.* 1997), cellular mechanisms responsible for its development and maintenance are not well understood. Nevertheless, the differential induction of this structure is associated with other dichotomies between lineages, including major changes induced in host muscle cells, relative cold tolerance of muscle larvae and disparate immunological interactions with their hosts (Stewart *et al.* 1985). It is possible that the collective differences that distinguish the two lineages represent biologically inter-related consequences of a basal adaptation that led to the initial diversification.

Consequently, comparative genomics across the two lineages is expected to provide significant evolutionary and other biological insights. The genetic characteristics of both newborn larvae (referred to here as immature L1, immL1), which migrate and invade host muscle cells, and the L1 muscle larvae (ML) that develop within muscle cells are two stages relevant to interactions with the infected muscle cells.

The enteric stage of infection is equally fascinating, in that *T. spiralis* and presumably other *Trichinella* species, invade and simultaneously occupy multiple host intestinal epithelial cells (Despommier, 1983) as appears to be the case for other nematodes in the Trichuroidea. Developmental stages of *Trichinella* that inhabit the small intestinal niche include ML digested from muscle, which develop through to the adult (Ad). The stages presently most amenable to research during the intestinal phase of infection are the ML and Ad, as there is a rapid development (days) from L1 to L5 and the previous need for *in vivo* infections to initiate that development. Advances in *in vitro* culture of intestinal stages may help alleviate this obstacle (Gagliardo *et al.* 2002). While the specific advantages conferred by this unusual mode of multi-intracellular infection are unresolved, the invasion process involves exquisite coordination in the parasite among sensory and motor control as well as mechanisms of host cell recognition and penetration. It is presently unclear whether the parasite actively manipulates the host epithelial cell for its own purposes, or simply extracts benefit (sustenance and protection) as it passes through.

An extracellular existence characterizes ML following its release from muscle tissue and prior to intestinal cell invasion. ImmL1 also have an extracellular existence prior to the

invasion of muscle cells. After their release from adult females, immL1 must translocate from the small intestine into the skeletal musculature of the host. This journey involves the entry into the circulatory system, transport and the invasion of muscle cells. In this sense, the immL1 is the only stage of infection that is extracellularly exposed to the peripheral immune system. In contrast to ML, which rapidly invade epithelial cells, immL1 are produced and released for a prolonged period during a single infection. Hence, there is potential for acquired immunity to develop against this stage, which can be induced (Marti *et al.* 1987). The prolonged extracellular existence in a peripheral tissue compartment may demand adaptations that are unique to immL1.

In addition to the basic biological interactions with host cells invaded by *Trichinella* species, there are numerous accounts of interactions with the host immune system during the enteric, muscle and migratory phases of the infection. Documented interactions relate to protective immune responses, immunopathological changes, immunosuppression, polarization of responses related to host cytokines and specific parasite antigens, modulation of responses against other pathogens, and immune responses of diagnostic value (summarized by Mitreva and Jasmer, 2006). Some advances have been made through knowledge of parasite molecules involved in some of these interactions, but we are at an early stage in identifying specific molecules and the mechanisms involved.

From the considerations presented in this article, the diversity of host niches occupied by *Trichinella* spp., known interactions with the host immune system and cells invaded by the parasite lead to an expectation of numerous molecular and cellular adaptations that facilitate success of these parasite life-cycles. Information on the genomes of *Trichinella* species should provide a useful platform for the discovery of adaptations and dissection of mechanisms that mediate these interactions. Genomic investigations will also assist studies of the biodiversity and ecology of *Trichinella* spp. Although recognized as having broad host ranges, differences in host compatibility characterize individual species representing both encapsulated and non-encapsulated forms. These differences among species have important links to their distribution and dynamics of transmission in the environment. With an improved understanding of the phylogeny within the genus (Zarlenga *et al.* 2006), comparative genomics may provide insights into the molecular determinants of these differences and hence basic information on the more global epidemiology of trichinellosis.

EXPRESSED SEQUENCE TAG (EST) DATASETS

The genomic analysis of *T. spiralis* began with the generation of expressed sequence tag (EST) databases for 3 life-cycle stages; immL1, ML and Ad. Results from this work have been published (Mitreva *et al.* 2004) and reviewed (Mitreva *et al.* 2005), such that they are not considered here, except for selected applications. Some useful statistics include: (i) 10 130 ESTs were generated and condensed into 3454 gene clusters (overlapping ESTs are condensed into contigs and contigs are grouped into clusters of related sequences); (ii) ESTs were conservatively predicted to represent 3262 genes; (iii) EST representation for stages was 3234 (immL1, 1046 clusters), 4028 (ML, 1908 clusters) and 2891 (Ad, 995 clusters); and (iv) 12.6% of clusters (which likely includes housekeeping genes) were recognized as similar among multiple developmental stages. The percentage of expressed genes that are common among these stages is expected to be much higher, and will require deeper sampling. In addition to observations summarized in published accounts, the EST data clearly supported the ancient phylogenetic divergence of *T. spiralis* (clade I) from other nematodes in clades III, IV and V of the Nematoda. Gene sequences apparently unique to *T. spiralis* by comparison with data for *C. elegans* and other nematodes and organisms, indicated possible lineage or species-specific characteristics. ESTs also identified stage-specific characteristics in an unexpected way. For instance, a family of ESTs encoding

DNase II-like protein sequences in the immL1 stage, displayed extensive diversity by comparison with related sequences expressed by a prototypical DNase-II like protein (p43), which is secreted from ML (Vassilatis *et al.* 1992; discussed below). Comparisons with other species showed that 56.2% of all clusters of translated molecules had significant similarity with proteins from other organisms, including nematodes, whereas 4.5% had similarity restricted exclusively to other nematodes. These data support the proposal that sequences from a clade I-nematode should contribute to our understanding of nematode biology at multiple phylogenetic and intraspecific levels. Therefore, based on findings derived from this EST project, the value of the genome sequence from a clade I-nematode was substantiated. The EST data, in the context of the rich biological history of *Trichinella* spp., provided a foundation for the sequencing of the genome of *T. spiralis*. The EST and clusters of sequences provide an essential starting point for the analysis of the *T. spiralis* genome sequence. In addition to their use as probes for assessing, for example, the quality of the genome sequence, the content of specific genes and gene families, EST data can also be used effectively for developing algorithms for gene models (start sites, intron/exon boundaries) to be used for large-scale gene discovery.

THE GENOME SEQUENCING PROCESS

The rapid increase in sequence data submitted to public databases in the past 10 years reflects the substantial improvements in DNA sequencing techniques. DNA sequencing procedures date back to the mid-1970's; of the two early techniques, Maxam-Gilbert chemical analysis method (Maxam and Gilbert, 1977, 1980) and Sanger's enzymatic procedure, employing chain-terminating nucleotides (Sanger *et al.* 1977), only the latter has been used routinely for high throughput sequencing. Since the introduction of these techniques, numerous technological refinements in DNA sequencing have been reported, aimed at producing more information per sequencing run, and increasing throughput. These refinements and advances have contributed to a continuous decline in the costs of sequencing and rapid data collection, making it more feasible to carry out sequencing projects on 'neglected parasites', such as *T. spiralis*.

The two most frequently used approaches for the sequencing of genomes include the 'clone-by-clone' and the 'whole genome' methods. For the clone-by-clone method, a genetic resource is needed to map genes, and the DNA sequence is derived from a known region. In whole genome shotgun (WGS) sequencing, the DNA is first broken into fragments and then sequenced at random and assembled together by looking for overlaps (Sanger *et al.* 1982). To date, this approach has dominated in the sequencing of the *T. spiralis* genome. The advantage of WGS is that it requires no prior mapping. Experience tells us that whole-genome shotgun sequencing leaves gaps. To close these gaps a primer-directed sequencing (pre-finishing) is performed (<http://genome.wustl.edu/platforms.cgi?id=5>), which reduces the number of contigs and resolves low quality regions.

Fig. 1 shows the major steps involved in the sequencing of the genome of *T. spiralis*. The strategy required a bacterial artificial chromosome (BAC) fingerprint map, BAC end-sequencing, 8-fold sequence coverage in plasmids, end sequence of a filamentous phage-plasmid (fosmid) library at 0.3-fold coverage, followed by 2 rounds of directed sequence improvement. The line represents the present status of the *T. spiralis* genome project. The project is currently focused on the final assembly and annotation.

For the annotation of the *T. spiralis* genome, the first step was to identify repeat elements (e.g. Bao and Eddy, 2002) in the final assembly, which enables masking of the genome prior to 'gene calling'. Different protein-coding prediction algorithms usually have concordance for predicting genes, but they disagree on the grouping of exons into genes (Reese *et al.*

2000). Generally, a combined approach is used to minimize the number of occasions on which different gene prediction algorithms give different results for the same region of the genome. Genes can be predicted using several evidence-based prediction programs, such as Genewise (Birney *et al.* 2004), FGENESH_PLUS (Salamov and Solovyev, 2000), Eannot (Ding *et al.* 2004), and *ab initio* predictions, such as Fgenesh (Salamov and Solovyev, 2000), GeneMark (Besemer and Borodovsky, 2005) and Snap (Korf, 2004) and the Ensembl annotation pipeline (Clamp *et al.* 2003). The available ESTs from *T. spiralis* (Mitreva *et al.* 2004) will provide valuable information when used *via* the evidence-based gene prediction algorithms to determine exon/intron and gene boundaries as well as splice-isoforms. The long-term goal of the project is to produce a fully annotated genome of *T. spiralis*, which offers the greatest utility for researchers.

STATUS OF THE *T. SPIRALIS* GENOME

In December 2007, a total of 4 168 619 traces were available in the National Center for Biotechnology Information (NCBI) trace archive from *T. spiralis* generated using different sequencing substrates : 3 940 023 are genomic sequences, 29 360 BAC end reads, and 199 236 fosmid hybrid vector end sequences.

The assembly of the *T. spiralis* genome is an on-going process and the initial assembly is expected to be improved after gap closure by primer-directed reactions (see Fig. 1). The initial assembly was generated using the PCAP package (Huang *et al.* 2006) on the paired-end sequences from plasmid, fosmid and BAC end clones. Ninety-six % (3 389 303/3 534 683) of the reads were placed in the assembly with 97% (55 472 619/56 779 425) of the sequences having output contig quality scores of Q20 (Table 1). The assembled genome size (56.8 Mb) was smaller than estimates published earlier (Searcy and MacInnis, 1970), but was similar to a recent genome size estimate of $1C=71.3\pm11.2$ Mb, based on flow-sorted nuclei, stained with propidium iodide (Spencer Johnston, unpublished). The data provided an average coverage of 35x, and the genome is represented by 6262 large regions (supercontigs) comprised of ordered and oriented fragments (contigs) interspersed with gaps. The majority of the sequence generated (78%) is in 14 supercontigs of size 1Mb (Fig. 2); 9% of the contigs are larger than 10 kb, covering 77% of the genome (Fig. 3). Primer-directed sequence improvement (automated closure) and a BAC-based physical map, intended to close gaps and improve the assembly, are underway. The automated closure will target the gaps within the supercontigs, in addition to regions of low sequence coverage. The clones which contain these regions are automatically selected, primers are generated, and these areas are to be sequenced using a 'primer walking' approach.

With the availability of an extensive amount of the genomic sequence of *T. spiralis*, a preliminary assessment was made of the quality and utility of the data. While some chromosomal and cDNA sequences already existed for a number of *T. spiralis* genes, we chose to utilize the EST database to begin this assessment, due to the large representation of genes in this database. Consensus nucleotide sequences for contigs of overlapping EST sequences were used [contigs are then grouped into clusters, as described (Mitreva *et al.* 2004)] in BLASTN analyses. Approximately 95% of all contigs had matches to the genome sequence equal to or greater than a cut-off value of 99% identity. This level of agreement indicates that a high percentage of the *T. spiralis* genome has been sequenced. Gene ontology analysis of *T. spiralis* ESTs implicated protein sequences in a broad range of cellular functions and pathways. With improved gene prediction programs, the stage is set to make useful predictions of functional aspects of molecules.

PRELIMINARY FINDINGS FROM THE GENOMIC DATA

In the following section, we address some selected questions that have evolved from this and related projects, the rationale for which is provided within each section. It is important to note that the results described (Table 2) are preliminary and based on a draft genome sequence for *T. spiralis*; therefore, findings are subject to some change over time.

DNase II-like genes

DNase II-like genes can be used to illustrate how the genome sequence of *T. spiralis* can provide extensive information on biological characteristics of this parasite. Although not initially recognized as one, the secreted ML p43 protein is a DNase II homologue (MacLea *et al.* 2003) and has been investigated in the context of host-parasite interactions (Vassiliadis *et al.* 1992). The analysis of EST data has identified related DNase II-like genes expressed by immL1 displaying unusually high diversity. The cluster TS00013.cl is composed of 29 contigs, in which the diversity of ESTs comprising contigs in this cluster approached 20% (Mitreva *et al.* 2004). To determine whether this diversity was reflected at the genomic level, the p43 protein sequence was used to screen the genome by TBLASTN (protein query *versus* translated database). The probe (DGAWIVHTVPGFKARTGYSWPAAEIAKGHLLICMTIAKTQINAI) represented the translation of an exon, as indicated by an identical match from a genome BLAST analysis using the entire p43 protein. The probe identified 40 distinct genome hits with scores of 1×10^{-10} or greater (Table 2). Other examples showed that scores were only slightly higher, but $< 1 \times 10^{-5}$, a standard cut-off for homology. Numerous matches were found to localize consecutively to multiple single-genome contigs, which suggests contiguous gene amplification. To assess the possibility of sequencing anomalies, several other probe sequences were used in similar searches. An example involved a search using a β -tubulin sequence from the trichostrongylid (clade V) nematode *Haemonchus contortus*. There was a single, high match with this sequence and several other sequences derived from the EST database (TBLASTX, translated nucleotide query *versus* translated database). The variable contig sequences that comprise TS00013.cl were then used and identified perfect or almost perfect matches in the genome. There are also 15 related cDNA sequences from immL1 that are in the public database (Liu *et al.* direct submission; <http://www.ncbi.nlm.nih.gov/entrez/viewer.fcgi?db=nucleotide&val=63095190>). Therefore, the diversity appears to be authentic and amplification of DNase II-like genes for expression by immL1 is a distinct possibility. The related p43 protein expressed by the ML stage is a secreted product, but only minor diversity of p43 ESTs is evident in this stage (Mitreva *et al.* 2004).

In relation to this diversity, it may be significant that the immL1 stage migrates extracellularly to infect muscle cells, while being released by Ad over a prolonged course of enteric infection. Hence, the potential for acquired immunity against the immL1 stage is high. Secreted DNases could be beneficial for the migration of the parasite through host tissues, and amino acid diversity could dampen host immune responses to these proteins. Functional relationships of these diverse DNase II-like molecules with tissue migration and evasion of the host immunity are intriguing in this context. Interestingly, no apparent diversity of this kind was found for related BLAST hits from genome sequences of *C. elegans*, *H. contortus* or filarial nematodes, or EST sequences from other nematode species (data not shown). This apparent, phylogenetically restricted gene amplification and diversification parallels the situation found for intestinal cathepsin B-like cysteine proteases expressed on the apical surface of *H. contortus* intestine (Jasmer *et al.* 2004). Such examples could reflect seminal adaptations that have promoted parasitism for given species or lineages. When related to other biological information (e.g., expression and tissue location), the functional importance of these observations may be better resolved.

Interactions with the host

There are relatively few candidate proteins or bio-synthetic products from *Trichinella* species that are known to mediate interactions with host cells infected by these parasites. In this context, the application of molecular approaches and genomic sequence data are likely to contribute to the discovery of relevant molecules. The application of mass spectrometry will identify parasite proteins, isolated in a number of contexts. Amino acid sequences translated from *T. spiralis* EST data have already proven valuable for identification of proteins in excretory-secretory products (ESP) from ML released from infected muscle (Robinson and Connolly, 2005). ESP from ML could mediate interactions with host cells in either the infected muscle cell or host intestinal cell compartments. The identification of nuclear antigens that localize to host muscle cell nuclei during the muscle phase of the infection might also be facilitated by mass spectrometric approaches. There are increasing numbers of proteins in ESP, for which information on function has become available, and secreted enzymes involved in nucleotide metabolism provide one interesting case (Gounaris and Selkirk, 2005). The screening of the genome for examples of these enzymes can rapidly assess the size, content and diversity of the family under investigation. Tyvelose-containing glycans engage in significant interactions with host antibody responses (Denkers *et al.* 1990; Carlisle *et al.* 1991), but it is uncertain whether glycans play more fundamental roles in the host-parasite interaction. One approach to investigate other roles is to inhibit glycan synthesis by genetic or chemical means, which may become possible *via* the identification of enzymes involved in glycan synthesis. Putative enzymes of relevance were identified by EST analysis (Mitrevva *et al.* 2004) and genomic data should enhance prospects.

The muscle stage of infection involves the loss of the differentiated state and cell-cycle repositioning in host muscle cells (Jasmer, 1993). Parasite-derived helix-loop-helix (HLH) antagonists of basic-HLH muscle differentiation factors have received attention as possible mediators of this effect (Vassilatis *et al.* 1992; Connolly *et al.* 1996; Jasmer and Kwak, 2006). In this regard, one *T. spiralis* EST was identified that had some similarity to the mammalian HLH inhibitor of differentiation (TS00545.cl; Mitrevva *et al.* 2004), whereas another displayed similarity to cyclin D (TS02454.cl), which is a G1 cyclin which can cause an inhibition of muscle differentiation and influence cell cycle position (Table 2). In each case, these ESTs identified a single locus in the genome sequence. However, neither of the translated proteins had classic signal peptides. While these examples illustrate how genomic data may aid in the identification of proteins which mediate infected cell changes, these specific proteins do not yet meet expected criteria for consideration as likely mediator candidates.

Comparative analyses with data for other nematodes

Comparative genomics offers an important application of genome data to elucidate broadly conserved nematode characteristics of potential value for the development of new methods of parasite control. The availability of a clade I-nematode genome establishes a critical anchor point, which together with existing clade III- and V-nematode genomes provides preliminary coverage of the phylum. The genome sequence of *T. spiralis* is of importance in this context. The integration of genome and EST data from multiple nematode species from the different clades will lead to a rapid expansion of knowledge and application of these resources. We present selected examples to illustrate some benefits of these data.

A useful example is β -tubulin, which appears to be a primary cellular target for benzimidazoles, a group of anthelmintics (Martin, 1997). Benzimidazoles are variably effective against *T. spiralis*, with more reliable efficacy against enteric stages and the early muscle stage of infection (Campbell and Denham, 1983). There was a single, high quality match between one sequence in the *T. spiralis* genome using a β -tubulin protein sequence

from benzimidazole sensitive *H. contortus* (KVSDTVVEPY NATLSVHQLVENTDETFCIDNEALYDICFRTLKLTNP, Accession number AAA29168). Phenylalanine (phe) at position 200 (underlined) has been associated with benzimidazole sensitivity, whereas a tyrosine substitution at this site is linked to resistance (Kwa *et al.* 1995). The apparent *T. spiralis* β -tubulin sequence has a Phe at this position, which may explain *T. spiralis* susceptibility to benzimidazole treatment under some circumstances. Since a single gene was identified, other factors may explain the inconsistent susceptibility among life-cycle stages of *T. spiralis*. This example could be extended to other applications related to known or prospective targets for anthelmintics. For instance, glutamate-gated chloride channel proteins (GluCl) are implicated in conferring nematode sensitivity to macrocyclic lactones, but ivermectin is relatively ineffective against *T. spiralis* (Campbell and Denham, 1983). Sequence and structural comparisons among these proteins from other nematode species, including *T. spiralis* orthologues, could provide valuable functional information. A GluCl α probe from *H. contortus* gave at least 9 matches to sequences in the *T. spiralis* genome (Table 2). The *T. spiralis* genome sequences have now been used in broader analyses of cys-loop gated ion channels (Williamson *et al.* 2007).

Molecules in the intestine of *T. spiralis*

The classification of proteins involved in basic functions of nematode tissues provides an important dimension to functionally analyse nematode genomes. Resultant information is expected to identify broadly conserved characteristics that could have applications to parasite intervention. Molecules in the intestine of nematodes represent one example. The luminal surface of the intestine (apical membrane) is an important interface with the host molecules. Research on *H. contortus* has emphatically demonstrated that the intestine from this parasite has application to parenteral immunization, mucosal immunization and chemotherapy (Jasmer *et al.* 2000, 2007; Knox and Smith, 2001). While the application of these observations to other parasitic nematodes is of interest, direct investigation of the intestine and other tissues is complicated either by the small size of many parasitic nematodes or constraints with the supply of parasite material. These obstacles exist for many nematodes of interest. We were interested in assessing the use of genome data to facilitate studies of particular tissues.

The examples that follow draw heavily on progress made with *H. contortus* because of the information available for this species. Nevertheless, we have initiated comparative genomics with intestinal ESTs and other data from species that are evolving as models for this research. The major focus has been on two parasites, *H. contortus*, and *A. suum*, and the free-living nematode *C. elegans*, due to the experimental advantages each has to offer. We have compared 41 106 intestinal ESTs from the two parasites and 5065 intestinal genes from *C. elegans* which grouped in 5586 protein families (unpublished data). Valuable progress has also been made with hookworms (Ranjit *et al.* 2006).

Apical intestinal membrane (AIM) proteins from *H. contortus* are expected to have fundamental roles in nutrient digestion and acquisition as well as in immune responses. Accordingly, AIM proteins provide a specific example for discussion here. The best known AIM proteins include a diverse group of proteases, which includes aspartyl (pepsin-like, PEPs), cysteine (cathepsin B-like, CBLs), and metallo (MEPs) protease classes (Table 2). In *H. contortus*, these AIM proteases appear to reside in a large complex or complexes on the apical membrane surface (Smith *et al.* 1999), perhaps functioning as a unit (a digestosome) in nutrient digestion. In addition to these proteases, the GA1 AIM proteins from *H. contortus* are of interest because they induced significant protective immunity against this parasite (Jasmer, 1993; Jasmer *et al.* 1993), although their functions are presently not known.

Pertinent questions regarding this group of molecules include (i) how conserved are the AIM proteins, and (ii) do they have application for the control of other parasitic nematodes? Some observations (Jolodar *et al.* 2004; Mitreva and Jasmer, unpublished) indicate that proteases related to AIM proteins are expressed in the intestine in at least some of the nematodes of clades I and III. To extend these observations, the *T. spiralis* genome was interrogated using protein sequence probes for each of the AIM proteins discussed (summarized in Table 2). Different search strategies were used, as dictated by the genomic characteristics of each protein or protein family. For instance, when a large number of prospective genes was identified using complete protein sequences, short probes representing exons were used to simplify the analysis and allow enumeration. In some cases, when apparent *T. spiralis* homologues were identified, the *T. spiralis* sequence identified was subsequently used to interrogate the genome.

Intestinal CBLs in *H. contortus* comprise a large family of diverse genes and proteins. An estimated 16% of *H. contortus* adult intestinal ESTs encoded CBL sequences (Jasmer *et al.* 2004). The relative abundance and diversity of these CBL ESTs are unique compared with other nematodes evaluated across the phylum. Accurate estimates are not yet available, but the number of distinct CBL loci could reach many tens of genes. The *H. contortus* genome sequence is incomplete, but a preliminary analysis supports this proposal. Two major CBL subfamilies (referred to as CBL clades) were identified based on the presence of a classic form of the occluding loop (CBL clade 1) or an altered version of this loop (CBL clade 2) (Rehman and Jasmer, 1999). In contrast, CBL ESTs were identified from *A. suum* intestinal RNA that represented 0.1% of those ESTs and encoded a classic occluding loop sequence. An estimated 12 CBL genes were found in the *C. elegans* genome, some of which are known to be expressed in the intestine (for summary, see Rehman and Jasmer, 1999) and encode a mix of classic and modified occluding loops. Intestinal CBLs are also known from other nematodes (summarized by Jasmer *et al.* 2004). Using a CBL probe for a clade 1-occluding loop, only 1 strong match was found in *T. spiralis* genomic sequences, while a CBL probe for a clade 2-occluding loop identified 6 *T. spiralis* genomic matches. Since CBLs are expressed in the intestine of diverse nematode species, some or all of the *T. spiralis* genomic sequences may represent ‘intestinal genes’. Specific reagents can now be produced to determine whether CBLs represent a conserved intestinal component for clade I-nematodes. While it seems likely that CBLs represent a broadly conserved characteristic of the intestine across the phylum Nematoda, the term ‘conserved’ is conditional. For instance, the relative representation of CBL mRNA and diversity does not seem to be conserved, different CBL lineages appear to have undergone expansion among species, and no CBLs were identified for filarial nematodes. Nevertheless, even the diverse CBLs expressed in intestinal cells of *H. contortus* are of interest as vaccine candidates (Redmond and Knox, 2006), and may have significance for other parasitic nematodes. From these considerations, CBL proteins appear to represent a highly plastic family, the composition of which has been ‘customized’ extensively during the evolution of nematodes.

Intestinal aspartyl proteases (PEPs) have been investigated for a number of nematode species. There are at least 2 related but distinct intestinal examples found in *H. contortus*, HcPEP-1 and HcPEP-2 (Longbottom *et al.* 1997). Related proteases have been described from the intestine of *C. elegans* (see Tcherepanova *et al.* 2000) and *Ancylostoma caninum* (see Williamson *et al.* 2004). Evidence for 2 aspartyl proteases exists in intestinal ESTs from *A. suum* (unpublished data), and an intestinal aspartyl protease has been reported for *Onchocerca volvulus* (see Jolodar *et al.* 2004). The HcPEP-1 sequence was used to probe *T. spiralis* EST clusters, and the translated sequence of the ‘best cluster match’ was used to probe the *T. spiralis* genome. Two examples were identified, which were similarly identified when the HcPEP-1 sequence was used to probe the *T. spiralis* genome. Hence, we identified at least 2 putative aspartyl proteases, which may be intestinal proteases. It was noted that the

protein sequences for clade I-nematodes were divergent from the clade V-nematodes. We have also noticed substantial diversity in these protein sequences among clade V-nematodes. The low number of prospective genes in the *T. spiralis* genome is consistent with observations made for the other nematodes.

MEPs comprise a family of AIM metalloproteases in *H. contortus* (see Knox *et al.* 2003). The *H. contortus* MEP-1b sequence identified a best match from the *T. spiralis* genome, which was one long continuous sequence of 708 amino acid residues. This sequence was used to probe the *T. spiralis* genome, which identified an additional series of segmental sequences with spacing, suggesting separation by introns. In addition, multiple other high similarity hits were obtained, the large number of which complicated the interpretation of the results. It is possible that numerous metalloprotease genes were detected. The best matches represent prospective intestinal MEPs, which can now be investigated.

Although the functions are unknown, GA1 proteins of *H. contortus* have been shown to induce significant immunity against challenge infection (Jasmer *et al.* 1996). The *H. contortus* GA1 polyprotein p100^{GA1} is evidently processed to produce 2 closely related AIM proteins, p46^{GA1} and p52^{GA1} (Jasmer *et al.* 1996); p52^{GA1} has an apparent C-terminal glycosylinositolphospholipid anchor sequence and its membrane attachment is phospholipase C-sensitive. In contrast, there are two p52^{GA1} homologues in *C. elegans* (see Rehman and Jasmer, 1998, and unpublished observations), which appear to have undergone a tandem duplication but lack a polyprotein organization. While not yet identified in *A. suum*, GA1 homologues have been detected in EST data for *T. spiralis*, *A. caninum* and *Trichuris muris* (unpublished observation). A single contig match was found in the *T. spiralis* genome with a combined length (over exons) which was equivalent to a single GA1 protein, and the highest similarity was to p52^{GA1}. The *T. spiralis* sequence had similarity to bacterial TolB proteins, as did the GA1 proteins. The highest TolB similarities for these proteins located to a region encoding a β -propeller domain, but have not yet clarified the putative function of the GA1 proteins. While functions remain unknown, the GA1 AIM proteins do appear to have ancient origins within the phylum Nematoda. This perspective may be important given the immunological significance of these proteins. Also, the broader phylogenetic conservation that can be attributed to GA1 proteins should stimulate interest in establishing their functional roles.

In summary, the analysis of genes linked to intestinal tissues identified a number of prospective proteins in *T. spiralis* that are predicted to function as AIM proteins. While non-intestinal homologues may have been identified, similarities with molecules from *H. contortus* support the proposal that at least some of the emphasized *T. spiralis* sequences do possess related intestinal functions. This prediction can now be investigated at both the molecular and the cellular level. This example demonstrates how rapid advancements can be made in the nematode system, such as for molecules involved in nutrient digestion and acquisition. AIM proteins and cellular processes that transport these proteins to the intestinal surface have been implicated as targets for control of clade V-nematodes. Genes identified from *T. spiralis* should support experiments to determine how general these inferences might be. A much larger database of 'intestinal genes' is under development (Mitreva and Jasmer, unpublished), intended to support much more extensive applications of this kind. With the ongoing development of relevant databases, parallel applications seem eminently achievable for other parasite tissues, stages, cells and cellular pathways.

CONCLUSIONS

The genome sequence for *T. spiralis* is nearing completion. This progress combined with a comprehensive EST database offers valuable resources for investigating the intriguing

biology of *T. spiralis* and related species. The standing of *T. spiralis* in clade I means that anchor points of genome sequences now exist that span the phylum (clades I, III and V), which facilitates comparative analysis across the phylum. Consequently, the *T. spiralis* genome sequence provides a valuable dimension to dissect the biology of parasitic nematodes and elucidate their evolution. Selected examples of molecules were used to assess anticipated applications of the genome sequence in relation to *T. spiralis* and other nematodes. The integration of EST and genome data provided insights into lineage and or species-specific characteristics, such as the *T. spiralis* genes encoding DNase II-like proteins and *H. contortus* genes encoding CBL proteins. The unique characteristics of these families of genes and gene products in each species may reflect key adaptations to parasitism. Applications of the genome sequence to discovering potential drug targets is particularly valuable in the context of similarities and differences in target sequences, genome organization and gene content. This proposal is clearly supported by data presented here and in published accounts, drawing on this and other genome sequences. A useful demonstration of pan-phylum application was the identification of a substantial set of prospective AIM protein genes for a clade I-nematode, based on similarities to a clade V-nematode. While suggesting that multiple AIM proteins are conserved in a pan-phylum context, reagents can also be designed to test this possibility at the molecular and cellular levels. Consequently, emerging tools provide unique prospects to support pan-phylum investigations of molecules, such as those involved in nutrient digestion in nematodes. The genome sequence of *T. spiralis* provides an important anchor point for coverage that now spans the phylum. Recognized benefits of this coverage will be enhanced upon the completion of the genomes for multiple additional nematode species currently being sequenced (Mitrevu *et al.* 2007).

Acknowledgments

Funding for the sequence characterization of the *Trichinella* genome is being provided by the National Human Genome Research Institute (NHGRI), National Institutes of Health (NIH).

References

- Bao Z, Eddy SR. Automated de novo identification of repeat sequence families in sequenced genomes. *Genome Research*. 2002; 12:1152–1155. [PubMed: 12176921]
- Besmer J, Borodovsky M. GeneMark: web software for gene finding in prokaryotes, eukaryotes and viruses. *Nucleic Acids Research*. 2005; 33:W451–W454.
- Birney E, Clamp M, Durbin R. GeneWise and Genomewise. *Genome Research*. 2004; 14:988–995. [PubMed: 15123596]
- Blaxter ML, De Ley P, Garey JR, Liu LX, Scheldeman P, Vierstraete A, Vanfleteren JR, Mackey LY, Dorris M, Frisse LM, Vida JT, Thomas WK. A molecular evolutionary framework for the phylum Nematoda. *Nature*, London. 1998; 392:71–75. [PubMed: 9510248]
- Campbell, WC.; Denham, DA. Chemotherapy. In: Campbell, WC., editor. *Trichinella and Trichinosis*. Plenum Press; New York: 1983. p. 335-366.
- Carlisle MS, McGregor DD, Appleton JA. The role of the antibody Fc region in rapid expulsion of *Trichinella spiralis* in suckling rats. *Immunology*. 1991; 74:552–558. [PubMed: 1769702]
- Clamp M, Andrews D, Barker D, Bevan P, Cameron G, Chen Y, Clark L, Cox T, Cuff J, Curwen V, Down T, Durbin R, Eyraas E, Gilbert J, Hammond M, Hubbard T, Kasprzyk A, Keefe D, Lehvaslaiho H, Iyer V, Melsopp C, Mongin E, Pettett R, Potter S, Rust A, Schmidt E, Searle S, Slater G, Smith J, Spooner W, Stabenau A, Stalker J, Stupka E, Ureta-Vidal A, Vastrik I, Birney E. Ensembl 2002: accommodating comparative genomics. *Nucleic Acids Research*. 2003; 31:38–42. [PubMed: 12519943]
- Connolly B, Trenholme K, Smith DF. Molecular cloning of a myoD-like gene from the parasitic nematode, *Trichinella spiralis*. *Molecular and Biochemical Parasitology*. 1996; 81:137–149. [PubMed: 8898330]

- Denkers EY, Wassom DL, Hayes CE. Characterization of *Trichinella spiralis* antigens sharing an immunodominant, carbohydrate-associated determinant distinct from phosphorylcholine. *Molecular and Biochemical Parasitology*. 1990; 41:241–249. [PubMed: 1697936]
- Despommier, DD. Biology. In: Campbell, WC., editor. *Trichinella and Trichinosis*. Plenum Press; New York: 1983. p. 75–151.
- Ding L, Sabo A, Berkowicz N, Meyer RR, Shotland Y, Johnson MR, Pepin KH, Wilson RK, Spieth J. EAnnot: a genome annotation tool using experimental evidence. *Genome Research*. 2004; 14:2503–2509. [PubMed: 15574829]
- Gagliardo LF, McVay CS, Appleton JA. Molting, ecdysis, and reproduction of *Trichinella spiralis* are supported *in vitro* by intestinal epithelial cells. *Infection Immunology*. 2002; 70:1853–1859.
- Gounaris K, Selkirk ME. Parasite nucleotide-metabolizing enzymes and host purinergic signalling. *Trends in Parasitology*. 2005; 21:17–21. [PubMed: 15639736]
- Haehling E, Niederkorn JY, Stewart GL. *Trichinella spiralis* and *Trichinella pseudospiralis* induce collagen synthesis by host fibroblasts *in vitro* and *in vivo*. *International Journal for Parasitology*. 1995; 25:1393–1400. [PubMed: 8719950]
- Huang X, Yang SP, Chinwalla AT, Hillier LW, Minx P, Mardis ER, Wilson RK. Application of a superword array in genome assembly. *Nucleic Acids Research*. 2006; 34:201–205. [PubMed: 16397298]
- Jasmer DP. *Trichinella spiralis* infected skeletal muscle cells : arrest in G2/M is associated with the loss of muscle gene expression. *Journal of Biological Chemistry*. 1993; 121:785–793.
- Jasmer DP, Dautova Mitreva M, McCarter JP. mRNA sequences for *Haemonchus contortus* intestinal cathepsin B-like cysteine proteases display an extreme in abundance and diversity compared with other adult mammalian parasitic nematodes. *Molecular and Biochemical Parasitology*. 2004; 137:297–305. [PubMed: 15383300]
- Jasmer DP, Kwak D. Differentiation and fusion of skeletal muscle cells expressing the p43 protein from *Trichinella spiralis* muscle larvae. *Experimental Parasitology*. 2006; 112:67–75. [PubMed: 16330028]
- Jasmer DP, Lahmers K, Brown WC. *Haemonchus contortus* intestine : a prominent source of mucosal antigens. *Parasite Immunology*. 2007; 29:139–151. [PubMed: 17266741]
- Jasmer DP, Perryman LE, Conder GA, Crow S, McGuire T. Protective immunity to *Haemonchus contortus* induced by immunoaffinity isolated antigens that share a phylogenetically conserved carbohydrate gut surface epitope. *Journal of Immunology*. 1993; 151:5450–5460.
- Jasmer DP, Perryman LE, McGuire TC. *Haemonchus contortus* GA1 antigens: Related, phospholipase C-sensitive, apical gut membrane proteins encoded as a polypeptide and released from the nematode during infection. *Proceedings of the National Academy of Sciences, USA*. 1996; 93:8642–8647.
- Jasmer DP, Yao C, Rehman A, Johnson S. Multiple lethal effects induced by a benzimidazole anthelmintic in the anterior intestine of the nematode *Haemonchus contortus*. *Molecular and Biochemical Parasitology*. 2000; 105:81–90. [PubMed: 10613701]
- Jolodar A, Fischer P, Buttner DW, Miller DJ, Schmetz C, Brattig NW. *Onchocerca volvulus*: expression and immunolocalization of a nematode cathepsin D-like lysosomal aspartic protease. *Experimental Parasitology*. 2004; 107:145–156. [PubMed: 15363940]
- Knox DP, Redmond DL, Newlands GF, Skuce PJ, Pettit D, Smith WD. The nature and prospects for gut membrane proteins as vaccine candidates for *Haemonchus contortus* and other ruminant trichostrongyloids. *International Journal for Parasitology*. 2003; 33:1129–1137. [PubMed: 13678629]
- Knox DP, Smith WD. Vaccination against gastrointestinal nematode parasites of ruminants using gut-expressed antigens. *Veterinary Parasitology*. 2001; 100:21–32. [PubMed: 11522403]
- Korf I. Gene finding in novel genomes. *BMC Bioinformatics*. 2004; 5:59. [PubMed: 15144565]
- Kwa MS, Veenstra JG, Van Dijk M, Roos MH. Beta-tubulin genes from the parasitic nematode *Haemonchus contortus* modulate drug resistance in *Caenorhabditis elegans*. *Journal of Molecular Biology*. 1995; 246:500–510. [PubMed: 7877171]
- Longbottom D, Redmond DL, Russell M, Liddell S, Smith WD, Knox DP. Molecular cloning and characterisation of a putative aspartate proteinase associated with a gut membrane protein complex

- from adult *Haemonchus contortus*. Molecular and Biochemical Parasitology. 1997; 88:63–72. [PubMed: 9274868]
- MacLea KS, Krieser RJ, Eastman A. A family history of deoxyribonuclease II: surprises from *Trichinella spiralis* and *Burkholderia pseudomallei*. Gene. 2003; 305:1–12. [PubMed: 12594037]
- Marti HP, Murrell KD, Gamble HR. *Trichinella spiralis*: immunization of pigs with newborn larval antigens. Experimental Parasitology. 1987; 63:68–73. [PubMed: 3803533]
- Martin RJ. Modes of action of anthelmintic drugs. Veterinary Journal. 1997; 154:11–34.
- Maxam AM, Gilbert W. A new method for sequencing DNA. Proceedings of the National Academy of Sciences, USA. 1977; 74:560–564.
- Maxam AM, Gilbert W. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods in Enzymology. 1980; 65:499–560. [PubMed: 6246368]
- Mitreva M, Appleton J, McCarter JP, Jasmer DP. Expressed sequence tags from life cycle stages of *Trichinella spiralis*: application to biology and parasite control. Veterinary Parasitology. 2005; 132:13–17. [PubMed: 15993542]
- Mitreva, M.; Jasmer, DP. Community, T. C. e. R, editor. Biology and genome of *Trichinella spiralis*; Wormbook. 2006. p. 1-21. <http://www.wormbook.org>
- Mitreva M, Jasmer DP, Appleton J, Martin J, Dante M, Wylie T, Clifton SW, Waterston RH, McCarter JP. Gene discovery in the adenophorean nematode *Trichinella spiralis*: an analysis of transcription from three life cycle stages. Molecular and Biochemical Parasitology. 2004; 137:277–291. [PubMed: 15383298]
- Mitreva M, Zarlenga D, McCarter JP, Jasmer DP. Parasitic nematodes – from genomes to control. Veterinary Parasitology. 2007; 148:31–42. [PubMed: 17560034]
- Polvere RI, Kabbash CA, Capo VA, Kadan I, Despommier DD. *Trichinella spiralis*: synthesis of type IV and type VI collagen during nurse cell formation. Experimental Parasitology. 1997; 86:191–199. [PubMed: 9225769]
- Ranjit N, Jones MK, Stenzel DJ, Gasser RB, Loukas A. A survey of the intestinal transcriptomes of the hookworms, *Necator americanus* and *Ancylostoma caninum*, using tissues isolated by laser microdissection microscopy. International Journal for Parasitology. 2006; 36:701–710. [PubMed: 16545815]
- Redmond DL, Knox DP. Further protection studies using recombinant forms of *Haemonchus contortus* cysteine proteinases. Parasite Immunology. 2006; 28:213–219. [PubMed: 16629707]
- Reese MG, Hartzell G, Harris NL, Ohler U, Abril JF, Lewis SE. Genome Annotation Assessment in *Drosophila melanogaster*. Genome Research. 2000; 10:483–501. [PubMed: 10779488]
- Rehman A, Jasmer DP. A tissue specific approach for analysis of membrane and secreted protein antigens from *Haemonchus contortus* gut and its application to diverse nematode species. Molecular and Biochemical Parasitology. 1998; 97:55–68. [PubMed: 9879887]
- Rehman A, Jasmer DP. Defined characteristics of cathepsin B-like proteins from nematodes: inferred functional diversity and phylogenetic relatedness. Molecular and Biochemical Parasitology. 1999; 102:297–310. [PubMed: 10498185]
- Robinson MW, Connolly B. Proteomic analysis of the excretory-secretory proteins of the *Trichinella spiralis* L1 larva, a nematode parasite of skeletal muscle. Proteomics. 2005; 5:4525–4532. [PubMed: 16220533]
- Salamov AA, Solovyev VV. *Ab initio* gene finding in *Drosophila* genomic DNA. Genome Research. 2000; 10:516–522. [PubMed: 10779491]
- Sanger F, Niklen S, Coulson A. DNA sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Sciences, USA. 1977; 74:5463–5467.
- Sanger F, Coulson AR, Hong GF, Hill DF, Petersen GB. Nucleotide sequence of bacteriophage lambda DNA. Journal of Molecular Biology. 1982; 162:729–773. [PubMed: 6221115]
- Searcy DG, MacInnis AJ. Measurements by DNA renaturation of the genetic basis of parasitic reduction. Evolution. 1970; 24:796–806.
- Smith SK, Pettit D, Newlands GF, Redmond DL, Skuce PJ, Knox DP, Smith WD. Further immunization and biochemical studies with a protective antigen complex from the microvillar membrane of the intestine of *Haemonchus contortus*. Parasite Immunology. 1999; 21:187–199. [PubMed: 10320616]

- Stewart GL, Wood B, Boley RB. Modulation of host response by *Trichinella pseudospiralis*. *Parasite Immunology*. 1985; 7:223–233. [PubMed: 4011298]
- Tcherepanova I, Bhattacharyya L, Rubin CS, Freedman JH. Aspartic proteases from the nematode *Caenorhabditis elegans*. Structural organization and developmental and cell-specific expression of asp-1. *Journal of Biological Chemistry*. 2000; 275:26359–26369. [PubMed: 10854422]
- Vassilatis DM, Despommier D, Misek D, Polvere R, Gold AM, Van Der Ploeg LHT. Analysis of a 43-kDa glycoprotein from the intracellular parasitic nematode *Trichinella spiralis*. *Journal of Biological Chemistry*. 1992; 267:18459–18465. [PubMed: 1382055]
- Williamson AL, Lecchi P, Turk BE, Choe Y, Hotez PJ, McKerrow JH, Cantley LC, Sajid M, Craik CS, Loukas A. A multienzyme cascade of hemoglobin proteolysis in the intestine of blood-feeding hookworms. *Journal of Biological Chemistry*. 2004; 279:35950–35957. [PubMed: 15199048]
- Williamson SM, Walsh TK, Wolstenholme AJ. The cys-loop ligand-gated ion channel gene family of *Brugia malayi* and *Trichinella spiralis*: a comparison with *Caenorhabditis elegans*. *Invertebrate Neuroscience*. 2007; 7:219–226. [PubMed: 17952476]
- Zarlenga DS, Rosenthal BM, La Rosa G, Pozio E, Hoberg EP. Post-Miocene expansion, colonization, and host switching drove speciation among extant nematodes of the archaic genus *Trichinella*. *Proceedings of the National Academy of Sciences, USA*. 2006; 103:7354–7359.

Fig. 1. Outline of a genome sequencing process. The horizontal line demarcates the current status of the *Trichinella spiralis* genome project. The dotted line indicates a clone-by-clone approach which was not undertaken when sequencing the *T. spiralis* genome.

Fig. 2.
Distribution of bp per supercontig size bin: cumulative percentage of bp in each size bin.
Example: 10% of all bp in the assembly (~5.8 Mb) are in supercontigs of size 1–10 kb.

Fig. 3.

Distribution of contigs per size, and coverage per contig size group. For example, 90% of the contigs are smaller than 10 kb, but they cover only 23% of the genome. Hence, 77% of the genome is in 816 contigs.

Table 1*Trichinella spiralis* assembly characteristics^a

Category	Contigs	Supercontigs
Total number	9282	6262
Total bases (million)	56.8	—
Total Q20 bases	55.5	—
Average length (bp)	6117	9067
Maximum length (bp)	308 109	11 900 814
Major (>1000 bp) number	5540	3117
N50 ^b length (bp)	60 168	3 383 625
N50 number	270	5

^a Assembly version v-1.0; 03/22/2006.^b The N50 length is the length such that 50% of the sequence is contained in segments of length X or greater.

Table 2

Summary of *Trichinella spiralis* genome hits obtained with various protein probes, as described in the text

Protein	Ts genome hits	Similarity scores ^a
Tsp43 (DNase II-like)	40	<10
Ts00545.cl (ID-like)	1	40
TsS02454 (cyclinD-like)	1	130
Hc-tubulin	1	19
HcGluCl	9	9 to 82
<i>AIM Proteins</i>		
HcCBLs		
clade 1	1	31
clade 2	6	5 to 24
HcPEPs	2	8,39
HcMEPs	Many	8 to 48
HcGA1	1	55

^aOnly the negative exponent for the similarity score is shown.

Hc, *Haemonchus contortus*; Ts, *Trichinella spiralis*.