



Expressed sequence tags from life cycle stages of *Trichinella spiralis*: Application to biology and parasite control

Makedonka Mitreva^{a,*}, Judith Appleton^b,
James P. McCarter^{a,c}, Douglas P. Jasmer^d

^a Genome Sequencing Center, Department of Genetics, Washington University School of Medicine,
4444 Forest Park Boulevard, St. Louis, MO 63108, USA

^b James A. Baker Institute for Animal Health, College of Veterinary Medicine, Cornell University,
Hungerford Hill Road, Ithaca, NY 14853, USA

^c Divergence Inc., 893 North Warson Road, St. Louis, MO 63141, USA

^d Department of Veterinary Microbiology and Pathology, Washington State University,
Pullman, WA 99164-7040, USA

Abstract

While the approach taken to date to study *Trichinella* spp., involves mainly characterization of individual genes of interest, we initiated a genomics approach as an antecedent to more complete genome sequencing. Our approach involves use of expressed sequence tags (ESTs) obtained from three life cycle stages of *Trichinella spiralis*; adult worms (AD), mature muscle larvae (ML) and immature L1 larvae (immL1, also known as newborn larvae) (Mitreva et al., 2004a) to improve the technical capacity for research on *Trichinella* spp. and to generate information that will aid prospective development of relevant hypotheses. In this review, we will summarize findings of our EST analysis and discuss how they relate to topics mentioned above. The foundation laid by this data will also contribute toward development of a more substantial genomic database and technical capacity to dissect molecular interactions between vertebrate hosts and *Trichinella* spp.

© 2005 Elsevier B.V. All rights reserved.

Keywords: *Trichinella spiralis*; Expressed sequence tags; Genomics

1. Background

Trichinella spp. are intracellular parasites of vertebrates with an entire life cycle confined to the host. These parasites alternate between enteric and

muscle phases of infection and several species are known to cause mortality and/or morbidity in humans. The parasites establish remarkable intracellular interactions with different host cell types when adapting to these diverse habitats (intestinal mucosal epithelial cells, striated skeletal muscle cells). While a wide range of changes are induced in host cells by infections with *Trichinella* spp., these are poorly characterized with regard to mechanisms involved and

* Corresponding author. Tel.: +1 314 2861118;
fax: +1 314 2861810.

E-mail address: mmitreva@watson.wustl.edu (M. Mitreva).

significance to parasite survival. Furthermore, the role of the parasite in regulating these host cell changes remains largely unresolved. While *Trichinella spiralis* is the topic of this paper, other *Trichinella* spp. exhibit biological differences that encompass host specificity, phenotypic characteristics induced in infected host cells and interactions with the host immune system, to name a few. Additionally, each of these parasites must migrate extracellularly through host tissues in order to infect the various host cells they inhabit. Mechanisms involved in this migration are unresolved. The topics mentioned above are only a few of the mysteries about *Trichinella* spp. that need to be solved. A number of technical obstacles exist that impede progress on many of these issues, including a paucity of information on the genes that comprise *Trichinella* spp.

Moving beyond *Trichinella* spp., the vast majority of molecular information available for nematodes comes from secernentean clades III–V (Blaxter et al., 1998; Dorris et al., 1999). As recently as August 2000, the only adenophorean nematode (clades I and II) sequences deposited in dbEST were 149 ESTs from *Trichuris muris*. As of January 2004 that number has risen to 16,232, the majority from Washington University and collaborators (12,950 ESTs, www.nematode.net, Wylie et al., 2003) and Edinburgh University and Sanger Centre (2714 ESTs, www.nematodes.org, Parkinson et al., 2004). While adenophorean orders such as the Trichocephalida (*Trichinella*, *Trichuris*, *Capillaria*), Mermithida, Dorylaimida, and Mononchida remain largely unexplored territory for genomic studies, the new ESTs from *T. spiralis* (Mitreva et al., 2004a) provided an entry point toward a deep understanding of the phylum nematoda.

2. Estimation of the number of *T. spiralis* genes represented by the dataset

ESTs from the investigated stages were sorted in order to estimate the total number of genes represented (Fig. 1). NemaGene clustering was used (McCarter et al., 2003) to organize into contigs all ESTs representing nearly identical transcripts. Contigs were then grouped into non-redundant clusters, which were expected to include alleles or splice isoforms of a gene. Rarely, clusters may also include highly related gene family members. Finally, the predicted level of

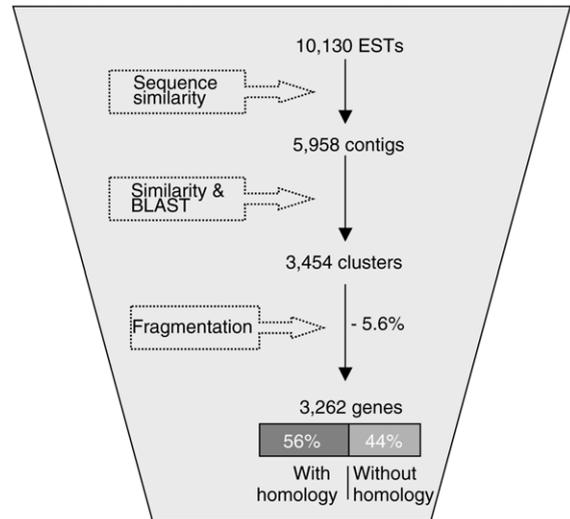


Fig. 1. Organization applied to the *T. spiralis* ESTs to improve data quality and to estimate the total number of represented genes.

fragmentation (non-overlapping clusters representing the same gene; Mitreva et al., 2004b) was calculated and subtracted from the total gene estimate. Hence, the 10,130 ESTs obtained from immL1, ML and AD, were organized into 3454 clusters (Fig. 1). Discounting for ‘fragmentation’ the estimated gene number for this EST set is 3262 (Mitreva et al., 2004a). Based on known genomic information from *Caenorhabditis elegans* (The *C. elegans* Sequencing Consortium, 1998), this number could represent 17% of all the *T. spiralis* genes.

This set of clusters and corresponding predicted proteins supported a variety of analyses with relevance to several areas of *T. spiralis* research, including phylogenetic relationships within the phylum nematoda, as described below.

3. Functional considerations

Using the non-redundant catalog of *T. spiralis* genes, BLAST homology searches indicated that 56% had homology to proteins from other species (Fig. 1), while 44% lacked any BLAST homology and were placed in the category of ‘novel’ proteins. The number of predicted novel proteins may be an overestimate, but the data lead us to expect that a large volume of new information will result from a more complete analysis of the *T. spiralis* genome.

In addition, potential functional information on translated cluster sequences was obtained by several other analyses, alone or in combination. Thirty-four percent (1172) of translated clusters had homology to InterPro protein domains (Apweiler et al., 2001) and 18% mapped to Gene Ontology pathways (The Gene Ontology Consortium, 2000). Additional information came from enzyme commission (E.C.) numbers coupled with mapping to metabolic pathways (Kyoto Encyclopedia of Genes and Genomes; Kanehisa and Goto, 2000). Collectively, predicted proteins provide information for a wide-spectrum of putative functions, which can be used for development of hypotheses relating to cellular and biochemical characteristics of these parasites.

Because the different investigated stages of *T. spiralis* perform distinct roles in the life cycle, proteins that distinguish these stages may convey important functional information. Stage-specific information was integrated with other functional analyses, including methods to identify prospective secreted proteins. This was done because proteins secreted from the parasite have potential for mediating interactions with host cells or products, including immune interactions. We found that a high percentage of clusters (87%) were represented by only one of the three stages (Fig. 2). A limited EST sample size might explain this result for some clusters. In the case of clusters with high representation by ESTs, 17 of the 25

most abundant clusters were exclusive to one or two of the three stages investigated. The high representation increases the probability for identification of bona fide developmentally regulated genes. In the case of the TspE1 protein antigen (Arasu et al., 1994), our data agreed with previous observations on expression by ML and AD, but not immL1. In the case of the p43 DNase II-like protein (MacLea et al., 2003), expression of that gene appears to be restricted to ML (Vassilatis et al., 1992). However, expression of a diverse family of p43 homologs, all distinct from p43, was detected exclusively in immL1. Because TspE1 and p43 are secreted, they are suspected of contributing to the host–parasite interaction, either in muscle cells or during the enteric phase of infection. In this context, the newly recognized p43 homologs expressed by immL1 raise numerous questions about possible functions of these proteins in the life cycle. While other developmental differences were suggested (Mitreva et al., 2004a), this brief account indicates the value of the general approach in elucidating genes that are developmentally regulated during the *T. spiralis* life cycle.

A wide range of other predicted proteins were identified with potential significance to host–parasite interactions, developmental regulation, metabolism, etc. Some of the issues that these findings touched on are summarized below. Homologs were identified for enzymes involved in glycan synthesis. These enzymes are of high interest, given the unusual properties of *T. spiralis* glycans (Appleton and Romaris, 2001) and their relationship to host immune responses (Denkers et al., 1990). About 50 predicted kinases and phosphatases were identified, which besides potential significance for intracellular signaling, might also be relevant to the unique extracellular nucleotide metabolism discovered in ML (Smith et al., 2000). Parasite enzymes that mediate or modulate proteolytic functions warrant interest because of their potential in facilitating tissue migration, other interactions with host cells and protection against digestion by the host (inhibitors). A spectrum of predicted proteinases (serine, cysteine, metallo and aspartic) and proteinase inhibitors (serpins and cystatins) were identified. Some of these are potentially stage-specific, and some related proteolytic activities have been detected in excretory and secretory products of this parasite. A range of predicted antioxidant enzymes were identi-

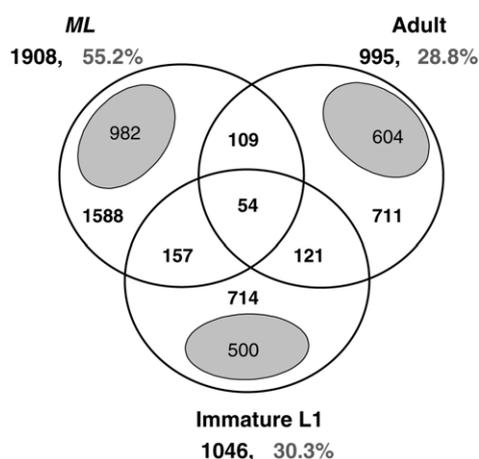


Fig. 2. Distribution of clusters based on life stage of origin: (○) number of singletons, it is expected that further sampling of the libraries will increase representation by multiple stages.

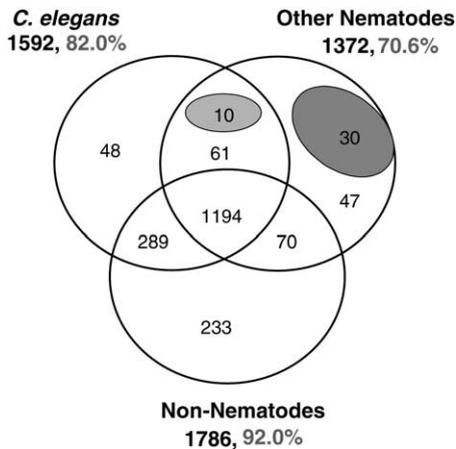


Fig. 3. Distribution of BLAST matches facilitating prioritizing targets for anthelmintic and nematocide development: (●) Clade I-specific and (○) nematode-specific with *C. elegans* homolog having RNAi phenotype.

fied, which could be important in protecting the parasite from reactive oxygen or nitrogen species of host or parasite origin.

The data mentioned above provide rich source of information that is relevant to current and future research on *T. spiralis*.

4. Phylogenetic relationships

BLAST analysis identified several classes of predicted protein sequences with potential phylogenetic significance (Fig. 3), which in some cases may relate to parasite control strategies. One class involved proteins with homology across diverse clades of nematodes (61 clusters), e.g. inclusive of clades I and V, but nevertheless were restricted to the phylum nematoda (nematode-specific). This class is likely to include proteins that represent ancestral features of nematodes, and hence determinants in the evolution of this specific animal group. Therefore, proteins in this class may be important for investigations on broad-spectrum control strategies.

A second class was proteins that are broadly conserved across diverse nematode species, whether restricted to this phylum or not. This class also represents potential targets that are attractive for parasite control strategies. In this context and the

preceding examples, RNA interference investigations in *C. elegans* have identified numerous genes that are required for survival or normal function of this nematode (Kamath et al., 2003). Of those, we found 463 *C. elegans* genes that encode protein homologs of predicted *T. spiralis* proteins. Since *T. spiralis* homologs in this set may have similar significance for function or survival, this example identifies one approach that could accelerate discovery of new nematode targets for parasite control.

A third class involved proteins (30 clusters) with homology restricted to clade I nematodes, and specifically *Trichinella* spp. and *Trichuris* spp. Proteins in this class are of interest because they may include molecular determinants with specific importance in the evolution of clade I nematodes, or subgroups of those nematodes. In relation to this hierarchical evaluation of proteins along nematode phylogenetic lines, the *T. spiralis* proteins described above as ‘novel’ could include specific adaptations confined to this parasite, or other species in the genus. Hence, these interspecies comparisons may provide a molecular outline for nematode evolution at several levels.

Clade I nematodes are thought to have diverged very early from many other animal and plant parasitic nematodes (Blaxter et al., 1998). Hence, gene and protein data available from *T. spiralis* may aid predictions on molecular conservation and diversity across the phylum. Our results support that view and provide expectations that additional genomic sequencing of clade I nematodes will significantly clarify genomic content and organization among species that comprise this phylum.

5. Future directions

The genomic approach involving generation and analysis of *T. spiralis* ESTs produced many benefits on which we can build. While clearly only a beginning, these advances have stimulated high expectations that sequencing the genome of *T. spiralis* and related *Trichinella* spp. will be extremely useful. In this context, *T. spiralis* was selected for genomic sequencing (<http://www.genome.gov/12511858>), which will be accomplished through the Genome Sequencing Center at Washington University in St. Louis (<http://>

www.genome.gov/page.cfm?pageID=10002154).

Selection of *T. spiralis* was prompted by our EST analysis and the expected value of clade I genomic information to clarify evolutionary limits of diversity within the phylum nematoda. This genomic information will have important implications regarding development of nematicides against plant and animal pathogens. The data will also make available specific parasite proteins to directly test those implications. Concepts on molecular evolution of nematodes and parasites are expected to progress significantly by integrating comparative genomics and biological observations among species from diverse nematode clades. For *T. spiralis* specifically, more complete genomic data will greatly enhance proteomics approaches to investigate parasite proteins that localize to relevant host compartments, e.g. host muscle cell nuclei, intestinal lumen and associated cells, or tissues through which the parasite migrates. Genomic data offer a means to develop microarray techniques that will efficiently assess expression of genes comprising the genome during the entire development of the parasite. Existing data from investigations involving ESTs and single parasite genes heightens anticipation of this potential. Finally, a genomic database for *T. spiralis* represents a first step toward comparative genomics among species of this genus. This prospect is particularly appealing given the differences in host–parasite interactions, species specificity and epidemiology demonstrated by these pathogens.

References

- Appleton, J.A., Romaris, F., 2001. A pivotal role for glycans at the interface between *Trichinella spiralis* and its host. *Vet. Parasitol.* 101, 249–260.
- Apweiler, R., Attwood, T.K., Bairoch, A., Bateman, A., Birney, E., Biswas, M., Bucher, P., Cerutti, L., Corpet, F., Croning, M.D., Durbin, R., Falquet, L., Fleischmann, W., Gouzy, J., Hermjakob, H., Hulo, N., et al., 2001. The InterPro database, an integrated documentation resource for protein families, domains and functional sites. *Nucleic Acids Res.* 29, 37–40.
- Arasu, P., Ellis, L.A., Iglesias, R., Ubeira, F.M., Appleton, J.A., 1994. Molecular analysis of antigens targeted by protective antibodies in rapid expulsion of *Trichinella spiralis*. *Mol. Biochem. Parasitol.* 65, 201–211.
- Blaxter, M.L., De Ley, P., Garey, J.R., Liu, L.X., Scheldeman, P., Vierstraete, A., Vanfleteren, J.R., Mackey, L.Y., Dorris, M., Frisse, L.M., Vida, J.T., Thomas, W.K., 1998. A molecular evolutionary framework for the phylum nematoda. *Nature* 392, 71–75.
- Denkers, E.Y., Wassom, D.L., Hayes, C.E., 1990. Characterization of *Trichinella spiralis* antigens sharing an immunodominant, carbohydrate-associated determinant distinct from phosphorylcholine. *Mol. Biochem. Parasitol.* 41, 241–249.
- Dorris, M., De Ley, P., Blaxter, M.L., 1999. Molecular analysis of nematode diversity and the evolution of parasitism. *Parasitol. Today* 15, 188–193.
- Kamath, R.S., Fraser, A.G., Dong, Y., Poulin, G., Durbin, R., Gotta, M., Kanapin, A., Le Bot, N., Moreno, S., Sohrmann, M., Welchman, D.P., Zipperlen, P., Ahringer, J., 2003. Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. *Nature* 421, 231–237.
- Kanehisa, M., Goto, S., 2000. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28, 27–30.
- MacLea, K.S., Krieser, R.J., Eastman, A., 2003. A family history of deoxyribonuclease II: surprises from *Trichinella spiralis* and *Burkholderia pseudomallei*. *Gene* 305, 1–12.
- McCarter, J., Dautova Mitreva, M., Martin, J., Dante, M., Wylie, T., Rao, U., Pape, D., Bowers, Y., Theising, B., Murphy, C.V., Kloek, A.P., Chiapelli, B.J., Clifton, S.W., Bird, M.D., Waterston, R., 2003. Analysis and functional classification of transcripts from the nematode *Meloidogyne incognita*. *Genome Biol.* 4 (R26), 1–19.
- Mitreva, M., Jasmer, D.P., Appleton, J., Martin, J., Dante, M., Wylie, T., Clifton, S.W., Waterston, R.H., McCarter, J.P., 2004a. Gene discovery in the adenophorean nematode *Trichinella spiralis*: an analysis of transcription from three life cycle stages. *Mol. Biochem. Parasitol.* 137, 277–291.
- Mitreva, M., McCarter, J.P., Martin, J., Dante, M., Wylie, T., Chiapelli, B., Pape, D., Clifton, S.W., Nutman, T.B., Waterston, R.H., 2004b. Comparative Genomics of gene expression in the parasitic and free-living nematodes *Strongyloides stercoralis* and *Caenorhabditis elegans*. *Genome Res.* 14, 209–220.
- Parkinson, J., Whitton, C., Schmid, R., Thomson, M., Blaxter, M., 2004. NEMBASE: a resource for parasitic nematode ESTs. *Nucleic Acids Res.* 32, 427–430.
- Smith, V.P., Selkirk, M.E., Gounaris, K., 2000. A reversible protein phosphorylation system is present at the surface of infective larvae of the parasitic nematode *Trichinella spiralis*. *FEBS Lett.* 483, 104–108.
- The *C. elegans* Sequencing Consortium, 1998. Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* 282, 2012–2018.
- The Gene Ontology Consortium, 2000. Gene ontology: tool for the unification of biology. *Nat. Genet.* 25, 25–29.
- Vassiliatis, D.M., Despommier, D., Misek, D., Polvere, R., Gold, A.M., Van der Ploeg, L.H.T., 1992. Analysis of a 43-kDa glycoprotein from the intracellular parasitic nematode *Trichinella spiralis*. *J. Biol. Chem.* 267, 18459–18465.
- Wylie, T., Martin, J., Dante, M., Mitreva, M., Clifton, S.W., Chinwalla, A., Waterston, R.H., Wilson, R.K., McCarter, J.P., 2003. Nematode.net: a tool for navigating sequences from parasitic and free-living nematodes. *Nucleic Acids Res.* 32, 423–426.