Available online at www.sciencedirect.com
SCIENCE DIRECT
ELSEVIER
International Journal for Parasitology 36 (2006) 829–839
www.elsevier.com/locate/ijpara

# Identification and analysis of genes expressed in the adult filarial parasitic nematode *Dirofilaria immitis*☆

Yong Yin [a,*], John Martin [a], James P. McCarter [a,b], Sandra W. Clifton [a], Richard K. Wilson [a], Makedonka Mitreva [a]

[a] *Department of Genetics, Genome Sequencing Center, Washington University School of Medicine, Box 8501, St Louis, MO 63108, USA*
[b] *Divergence Inc., 893 North Warson Road, St Louis, MO 63141, USA*

Received 23 November 2005; received in revised form 28 February 2006; accepted 2 March 2006

**Abstract**

The heartworm *Dirofilaria immitis* is a filarial parasitic nematode infecting dogs and other mammals worldwide causing fatal complications. Here, we present the first large-scale survey of the adult heartworm transcriptome by generation and analysis of 4005 expressed sequence tags, identifying about 1800 genes and expanding the available sequence information for the parasite significantly. *Brugia malayi* genomic data offered the most valuable information to interpret heartworm genes, with about 70% of *D. immitis* genes showing significant similarities to the assembly. Comparative genomic analyses revealed both genes common to metazoans or nematodes and genes specific to filarial parasites that may relate to parasitism. Characterization of abundant transcripts suggested important roles for genes involved in energy generation and antioxidant defense in adults. In particular, we proposed that adult heartworm likely adopted an anaerobic electron transfer-based energy generation system distinct from the aerobic pathway utilized by its mammalian host, making it a promising target in developing next generation macrofilaricides and other treatments. Our survey provided novel insights into the *D. immitis* transcriptome and laid a foundation for further comparative studies on biology, parasitism and evolution within the phylum Nematoda.


*Keywords:* Dirofilaria immitis; Heartworm; Parasite; Expressed sequence tags (EST's); Comparative genomics; Filarial nematode; *Caenorhabditis elegans*; Energy generation

## 1. Introduction

The dog heartworm *Dirofilaria immitis* belongs to the clade III filarial parasitic nematodes (Blaxter et al., 1998), slender roundworms that are parasitic in blood and tissue of vertebrates and have life cycle stages within intermediate insect hosts. Heartworm has a worldwide distribution in temperate and tropical climates, infecting dogs, its definitive host, as well as cats, ferrets and other mammals including humans (Knight, 1977, 1987). L1 larvae (microfilariae, ~300 μm long), which are able to survive 1–3 years circulating in the host bloodstream, develop into the infective L3 stage (L3i) after entering a mosquito during its feeding. After entering another host when the mosquito feeds again, the L3i larvae molt and develop further into adults (Knight, 1977, 1987). Adult females (up to 27 cm long) and males (~17 cm) normally reside in the host pulmonary arteries and right ventricles. Their life span in dogs is at least 5–7 years, with the number of adults reaching up to 250 in a single dog, resulting in various lung and heart diseases such as severe pulmonary arterial inflammation and congestive heart failure that can be fatal (Knight, 1977, 1987). In the United States, heartworm infection has been found in dogs native to all 50 states. A 2001 survey of over 18,000 veterinary clinics reported heartworm positive tests for more than 240,000 dogs and 3000 cats (McCall, 2005), with the actual incidence likely much higher. Heartworm test kits and macrocyclic lactone anthelmintic drugs used to prevent the infection are among the best-selling products in companion animal veterinary medicine, even though only half of dogs are estimated to receive any preventive treatments. Currently, the arsenical melarsomine dihydrochloride is the only Federal

☆ Note: EST sequences are available from GenBank, EMBL, and DDJB under the accession numbers BQ454296–BQ457429, BQ481967–BQ482740, BU587123–BU587163, and CD285915–CD285970. The sequences are also available at www.nemtode.net.

* Corresponding author. Tel.: +1 314 286 1118; fax: +1 314 286 1810.
*E-mail address:* yyin@watson.wustl.edu (Y. Yin).

0020-7519/$30.00 
doi:10.1016/j.ijpara.2006.03.002

Department of Agriculture-approved drug to eliminate adult heartworms in infected dogs (McTier et al., 1994). Both new preventive strategies such as vaccines and safer curative macrofilaricides are needed.

To date, studies of the *D. immitis* transcriptome have been limited to individual genes, with only 112 *D. immitis* gene sequences available in public databases before the current study (April 2005). A comprehensive expressed sequence tag (EST) study would greatly accelerate the identification of heartworm genes and prediction of their putative functions based on primary sequence similarities, helping us to better understand its biology and eventually develop next generation drugs and vaccines. In addition, various human endemic diseases are caused by related filarial nematodes. Lymphatic filariasis, a disease carried by 120 million people worldwide and recognized as the second leading cause of permanent and long-term disability, is caused by the infection with *Wuchereria bancrofti*, *Brugia malayi* and *Brugia timori* (Molyneux et al., 2003). Another filarial nematode *Onchocerca volvulus* is the causative agent of river blindness (onchocerciasis) infecting 18 million individuals worldwide (Molyneux et al., 2003). A large-scale survey and analysis of the *D. immitis* transcriptome will complement the data being generated from the human parasites in various EST and genome sequencing projects (Blaxter et al., 1999, 2002; Williams et al., 2000, 2002; Ghedin et al., 2004), setting a stage for comparative genomic studies among veterinary and human filarial parasites.

Here we report, for the first time, a large-scale survey of the adult *D. immitis* transcribed genome, identifying about 1800 genes by sequencing and analyzing 4005 ESTs. Comparative genomic approaches were used to interpret heartworm genes using sequence data from other species, especially those generated in the *B. malayi* Genome Project (Ghedin et al., 2004). In addition, those comparisons identified putative proteins specific to nematodes or individual nematode clades. Abundant transcripts were examined into more detail, highlighting special aspects of adult heartworm physiology.

## 2. Materials and methods

### 2.1. Source material, library construction, and EST sequencing

Adult heartworms were collected at necropsy of a euthanized adult dog obtained from an animal shelter near College of Veterinary Medicine at North Carolina State University. The live worms, including both males and females, were rinsed in 1× sterile PBS, frozen in liquid $N_2$ immediately upon recovery and stored at −80 °C. Pulverization of four to five whole worms of mixed sexes was performed using an alloy steel mortar and pestle set (Fisher Scientific). The absolute and relative numbers of males and females used are not known. No special measures were taken to remove developing embryos and larvae from the sample; therefore a small number of embryonic transcripts were expected to be included in the EST sampling. Poly(A)+RNA was isolated, two cDNA libraries were constructed (using either the splice leader 1 (SL1)-based or SMART (Clontech Laboratories) cDNA library construction systems) from which 815 and 3190 ESTs were generated by 5′-end sequencing, respectively, as described previously (McCarter et al., 2003; Mitreva et al., 2004a,b).

### 2.2. Sequence analysis and functional assignment

Examination, processing, and clustering of the EST sequences were performed as described previously (McCarter et al., 2003; Mitreva et al., 2004a). A maximum likelihood-based program ESTFREQ (W. Gish, unpublished) was used to estimate the complexity of the EST libraries; TRANSLATE (S. Eddy, unpublished) was used for predicting open reading frames (ORFs) with default parameters. Databases used for sequence comparison were: *B. malayi* genome draft assembly (October 2005), WORMPEP (v140), GenBank (15 April 2005), non-redundant protein database NR (15 April 2005) and genome of *Wolbachia* strain TRS of *B. malayi* (15 April 2005). WU-BLASTX ($S=100$ $M=$PAM120 $V=0$ $W=4$ $T=17$) was used to query translated nucleotide sequences against protein databases and WU-TBLASTX ($Q=10$ $R=2$ gapw=10 filter=seg+xnu hspsepsmax=5000 gapsepsmax=5000) for querying translated nucleotide sequences versus translated nucleotide databases in all six reading frames. Expectation value (*E*-value) of $1.0\,e^{-5}$ was used as the cut-off to accept sequence similarities in both BLAST searches. Default parameters for InterProScan (Quevillon et al., 2005) were used to query against InterPro (Release 9.0) (Mulder et al., 2005) in assigning Gene Ontology terms (Release 2,00,509) (The Gene Ontology Consortium, 2000). An *E*-value cut-off of $1.0\,e^{-10}$ reported by WU-BLASTX against the GeneDB (Release 34.0) from Kyoto Encyclopedia of Genes and Genomes (KEGG) was used for metabolic pathway association, the top match and all the matches within a range of 30% of the top score, if meeting the cut-off, were accepted for KEGG association (Bono et al., 1998; Kanehisa and Goto, 2000; Kanehisa et al., 2004). TargetP (v1.1) was used to identify *D. immitis* clusters with putative signal peptides for secretion and predictions within reliability class 1 were accepted (Emanuelsson et al., 2000); such clusters were further screened to identify transmembrane (TM) regions by TMHMM (v2.0) (Krogh et al., 2001). A *D. immitis* protein was assigned as secreted if either no TM region was identified or only one TM domain was predicted and its First60 value was between 10 and 30.

## 3. Results and discussion

### 3.1. Properties and clustering of ESTs

As part of a larger effort to examine ESTs from about 30 parasitic nematodes (Mitreva et al., 2005), we constructed two *D. immitis* adult cDNA libraries using different protocols and sequenced in total 4005 ESTs from the 5′ end, of which 3999 passed our automated screen and manual inspection. They were submitted to the dbEST division of GenBank immediately upon generation in 2002 and 2003. Their average length is 405 nucleotides, totaling 1.62 million bases, which represents