JOURNAL OF NEMATOLOGY

VOLUME 32 DECEMBER 2000 NUMBER 4

Journal of Nematology 32(4):331–333. 2000.

Nematode Gene Sequences, December 2000 Update

James P. McCarter,[1] David McK. Bird,[2] Sandra W. Clifton,[1] and Robert H. Waterston[1]

High-throughput sequencing is revolutionizing molecular nematology by providing the sequences of thousands of genes never before characterized. The most rapid and cost-effective route to gene discovery for nematode genomes is the generation of expressed sequence tags (ESTs), single-pass reads from random cDNA library clones that provide 300–600 nucleotides of sequence from a gene. Projects are currently under way at Washington University's Genome Sequencing Center that will generate 225,000 ESTs from 14 nematode species by 2003. Additionally, the Sanger Centre and Edinburgh University will produce 80,000 ESTs from seven species. New sequences are immediately submitted to the dbEST (database of expressed sequence tags) division of GenBank. By the completion of these efforts, we anticipate the identification of more than 80,000 new nematode genes.

Here we present a brief progress report on publicly available ESTs from nematodes. *Caenorhabditis elegans* has long been a focus of sequencing efforts, and 111,639 ESTs are available from *Caenorhabditis* species. Currently, 72,764 ESTs are available from 19 nematode species beyond *Caenorhabditis,* including seven human parasites, five animal parasites, five plant parasites, and two free-living bacteriovores (Table 1). The majority of these ESTs were generated in 1999–2000. EST sequences are publicly available from dbEST at GenBank and a number of parasite-specialized Web sites (Table 2).

A major factor contributing to the popularity of ESTs is the versatility of the data. ESTs can be used for new gene discovery, gene discovery by similarity searching, genomic physical mapping, determination of gene coding regions and splice isoforms, determination of a species' codon bias, design of gene expression biochips and microarrays, determination of evolutionary relationships between species, identification of protein and domain families, and identification of single nucleotide polymorphisms (SNPs). Strategies for using ESTs, as well as discussions of the strengths and weaknesses of EST data, are available from reviews (Marra et al., 1998; Blaxter et al., 1999; McCarter et al., 2000a).

Nematologists also benefit greatly from the availability of the complete genome sequence of *C. elegans* and the annotation of its genes (The *C. elegans* Sequencing Consortium, 1998). The essentially complete sequence of *C. elegans* published in 1998 was composed of 97 megabases with 19,099 predicted protein encoding genes. Gap filling to date has brought the total genome to just over 100 finished megabases (Genome Sequencing Center, unpubl. data) with 19,705 predicted proteins including 405 splice variants (Sanger Centre Wormpep Release 33, 2000). A number of small gaps remain. Ad-

Received for publication .

Parasitic nematode EST sequencing at Washington University is supported by NIH-NIAID research grant AI 46593 to Robert Waterston and NSF Plant Genome award 0077503 to David Bird (PI) and co-PIs Sandra Clifton, Joseph Kieber, Charles Opperman, and Jeffrey Thorne. James McCarter is a Merck Postdoctoral Fellow of the Helen Hay Whitney Foundation.

[1] Genome Sequencing Center, Department of Genetics, Box 8501, Washington University School of Medicine, St. Louis, MO 63108, USA.

[2] Plant Nematode Genetics Group, Department of Plant Pathology, North Carolina State University, Raleigh, NC 27695, USA.

The authors thank members of the Genome Sequencing Center EST lab, especially Deana Pape, John Martin, Todd Wylie, and Brandi Chiapelli.

E-mail: mccarter@genetics.wustl.edu

This paper was edited by B. C. Hyman.

TABLE 1. 21 Nematode species have more than 50 ESTs registered in the GenBank dbEST database, December 2000.

Nematode species	Number of ESTs	Major source	Publications (ESTs analyzed)
Caenorhabditis elegans	109,215	NIG, Japan[1]	Waterston et al., 1992 (1,517) McCombie et al., 1992 (720)
Brugia malayi	22,392	WHO FilGenNet[2,3,6]	Blaxter et al., 1996 (596)
Onchocerca volvulus	13,802	Smith College, MA[3]	Unnasch and Williams, 2000
Strongyloides stercoralis	10,922	GSC, St. Louis, MO[4]	Moore et al., 1996 (55)
Meloidogyne incognita	6,626	GSC, St. Louis, MO[4]	McCarter et al., 2000b (5,713)
Ancylostoma caninum	5,546	GSC, St. Louis, MO[4]	
Pristionchus pacificus	4,989	GSC, St. Louis, MO[4]	
Caenorhabditis briggsae	2,424	GSC, St. Louis, MO[5]	
Haemonchus contortus	2,399	Univ. Edinburgh[6] & Sanger Centre, UK,[7] Wash. State, WA[8]	
Heterodera glycines	1,506	GSC, St. Louis, MO[4]	
Meloidogyne javanica	1,208	GSC, St. Louis, MO[4]	
Globodera rostochiensis	894	Scottish Crop & Wageningen[9]	Popeijus et al., 2000 (894)
Ascaris suum	588	Univ. Edinburgh, UK[6]	
Toxocara canis	519	Univ. Edinburgh, UK[10]	Tetteh et al., 1999 (519)
Zeldia punctata	378	GSC, St. Louis, MO[4]	
Trichuris muris	301	Univ. Edinburgh, UK[6]	
Necator americanus	211	Univ. Edinburgh, UK[6]	Daub et al., 2000 (211)
Litomosoides sigmodontis	198	Univ. Edinburgh, UK[6]	
Wuchereria bancrofti	131	Smith College, MA[3]	
Globodera pallida	94	Scottish Crop & Wageningen[9]	Popeijus et al., 2000 (94)
Onchocerca ochengi	60	Smith College, MA[3]	
Total ESTs	184,403		
Total Non-*Caenorhabditis*	72,764		

[1] Korhara et al., National Institute of Genetics, Mishima, Japan.
[2] World Health Organization Filarial Genome Network.
[3] Williams et al., Dept. of Biology, Smith College, Northampton, MA USA.
[4] McCarter et al., Genome Sequencing Center, Washington Univ. School of Medicine, St. Louis, MO USA.
[5] Marra et al., Genome Sequencing Center, Washington Univ. School of Medicine, St. Louis, MO USA.
[6] Blaxter et al., Institute of Cell, Animal, and Population Biology, Univ. of Edinburgh, Edinburgh, UK.
[7] Sanger Centre, Hinxton, UK.
[8] Jasmer et al., Dept. of Veterinary Microbiology and Pathology, Washington State Univ., Pullman, WA.
[9] Jones et al., Nematology Dept., Scottish Crop Research Inst., Dundee, UK, and Laboratory of Nematology, Wageningen Univ., Wageningen, The Netherlands.
[10] Maizels et al., Institute of Cell, Animal, and Population Biology, Univ. of Edinburgh, Edinburgh, UK.

TABLE 2. Selected Web resources for nematode EST access.

EST resources	URLs
GenBank dbEST	www.ncbi.nlm.nih.gov/dbEST
Blaxter Lab ESTs & Clusters	www.nematodes.org
Genome Sequencing Center ESTs & NemaGene	www.nematode.net
EMBL Parasite Genome Server	www.ebi.ac.uk/parasites/parasite-genome.html
The Filarial Genome Network	nema.cap.ed.ac.uk/fgn/ests.html or circuit.neb.com/fgn/ests.html
More Extensive Links	elegans.swmed.edu/Nematodes/

The Blaxter Lab EST page provides access to EST data sets from *Haemonchus contortus, Ascaris suum, Necator americanus, Litomosoides sigmodontis,* and *Trichuris muris* as well as data from *Trichinella spiralis* and *Loa loa* not yet available in dbEST. A BLAST server with data from 18 species is also available. Nematode.net, a collaboration between the Washington University Genome Sequencing Center and North Carolina State University is a project to provide comprehensive access to nematode ESTs, including search capacity, ftp downloads, and sequence trace files. Nematode.net will also provide access to the NemaGene Index of nematode genes built by clustering overlapping ESTs. Currently available information includes ESTs and NemaGene version 1.0 clusters from *Meloidogyne incognita.* The European Molecular Biology Lab (EMBL) Parasite Genome Server provides BLAST search access to the *Brugia malayi* ESTs as well as several protozoan parasites. *Brugia malayi* data, as well as information about the *B. malayi* EST project, is also provided by the Filarial Genome Network page.

ditionally, 10.1 megabases of the *C. briggsae* genome is in finished form (Genome Sequencing Center, unpubl. data). The majority of genes identified to date in parasitic

TABLE 3. Selected Web resources for *C. elegans* genome access.

EST resources	URLs
Wormbase	www.wormbase.org
Wormpep, Sanger Centre	www.sanger.ac.uk/Projects/C_elegans/wormpep
WormPD	www.proteome.com/databases/index.html
C. elegans Project & BLAST Server at Sanger Centre	www.sanger.ac.uk/Projects/C_elegans
or Genome Sequencing Center	or genome.wustl.edu/gsc/C_elegans/elegans.shtml
C. elegans WWW Server	elegans.swmed.edu

Wormbase, an outgrowth of Acedb, is a database of all genetic and genomic information for *C. elegans,* including genetic and physical maps. Wormpep is an up-to-date list of all predicted proteins from *C. elegans* maintained by the Sanger Centre. WormPD from Proteome Inc. is a *C. elegans* protein database with extensive literature annotation. Both the Sanger Centre and the Washington University Genome Sequencing Center (GSC) maintain information on the *C. elegans* genome project including a *C. elegans*-specific BLAST server and access to ftp downloads. The GSC maintains similar information for *C. briggsae.* The *C. elegans* WWW Server provides comprehensive links for *C. elegans,* including a literature search engine.

nematodes have homologues in *C. elegans.* For example, BLAST analysis revealed that 66% of *Meloidogyne incognita* genes have a *C. elegans* homologue ($E < 10^{-5}$) (McCarter et al., 2000a). Key *C. elegans* genome resources are shown in Table 3.

An awareness of available nematode sequence data can save time and effort in the lab as well as greatly affect plans for experimental design. Periodic updates on the status of nematode gene sequence will be provided over the next several years as nematode EST data sets rapidly expand.

LITERATURE CITED

Blaxter, M., M. Aslett, D. Guiliano, J. Daub, and the Filarial Genome Project. 1999. Parasitic helminth genomics. Parasitology 118:S39–S51.

Blaxter, M. L., N. Raghavan, I. Ghosh, D. Guiliano, W. Lu, S. A. Williams, B. Slatko, and A. L. Scott. 1996. Genes expressed in *Brugia malayi* infective third-stage larvae. Molecular and Biochemical Parasitology 77:77–93.

Daub, J., A. Loukas, D. I. Pritchard, M. Blaxter. 2000. A survey of genes expressed in adults of the human hookworm, *Necator americanus.* Parasitology 120:171–184.

Marra, M. A., L. Hillier, and R. H. Waterston. 1998. Expressed sequence tags—ESTablishing bridges between genomes. Trends in Genetics 14:4–7.

McCarter, J., P. Abad, J. T. Jones, and D. Bird. 2000a. Rapid gene discovery in plant-parasitic nematodes via expressed sequence tags. Nematology, in press.

McCarter, J., D. McK. Bird, U. Rao, A. Kloek, S. Clifton, D. Pape, J. Martin, T. Wylie, S. Eddy, R. Waterston, and the Washington University GSC EST Team. 2000b. Progress toward high throughput gene discovery in parasitic nematodes: Initial findings from *Meloidogyne incognita* EST sequencing. Journal of Nematology, in press.

McCombie, W. R., M. D. Adams, J. M. Kelley, M. G. Fitzgerald, T. R. Utterback, M. Khan, M. Dubnick, A. R. Kerlavage, J. C. Venter, and C. Fields. 1992. *Caenorhabditis elegans* expressed sequence tags identify gene families and potential disease gene homologues. Nature Genetics 1:124–131.

Moore, T. A., S. Ramachandran, A. A. Gam, F. A. Neva, W. Lu, L. Saunders, S. A. Williams, and T. B. Nutman. 1996. Identification of novel sequences and codon usage in *Strongyloides stercoralis.* Molecular and Biochemical Parasitology 79:243–248.

Popeijus, H., V. C. Blok, L. Cardle, J. Bakker, M. S. Phillips, J. Helder, G. Smant, J. T. Jones. 2000. Analysis of genes expressed in second-stage juveniles of the potato cyst nematodes *Globodera rostochiensis* and *G. pallida* using the expressed sequence tag approach. Nematology, in press.

Tetteh, K. K., Loukas, C. Tripp, and R. M. Maizels. 1999. Identification of abundantly expressed novel and conserved genes from the infective larval stage of *Toxocara canis* by an expressed sequence tag strategy. Infection and Immunity 67:4771–4779.

The *C. elegans* Sequencing Consortium. 1998. Genome sequence of the nematode *C. elegans:* A platform for investigating biology. Science 282:2012–2018.

Unnasch, T. R., and S. A. Williams. 2000. The genomes of *Onchocerca volvulus.* International Journal of Parasitology 30:543–452.

Waterston, R., C. Martin, M. Craxton, C. Huynh, A. Coulson, L. Hillier, R. Durbin, P. Green, R. Shownkeen, N. Metzstein, T. Hawkins, R. Wilson, M. Berks, Z. Du, K. Thomas, J. Thierry-Mieg, and J. Sulston. 1992. A survey of expressed genes in *Caenorhabditis elegans.* Nature Genetics 1:114–123.