JOURNAL OF NEMATOLOGY

VOLUME 34 JUNE 2002 NUMBER 2

Journal of Nematology 34(2):71–74. 2002.

Nematode Gene Sequences, Update for June 2002

James P. McCarter,[1,3] Sandra W. Clifton,[1] David McK. Bird,[2] 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 to 600 nucleotides of sequence from a gene. Projects are currently under way at Washington University's Genome Sequencing Center that will generate 235,000 5′ ESTs from approximately 25 nematode species by 2003 (119,448 to date). Additionally, the Sanger Institute and Edinburgh University are producing 80,000 ESTs from seven species (10,772 to date). New sequences are immediately submitted to the dbEST (database of expressed sequence tags) division of GenBank and are also available from a number of parasite-specialized Web sites (Table 1). Strategies for using ESTs, as well as discussions of the strengths and weaknesses of EST data, are available from reviews (Blaxter et al., 1999; Marra et al., 1998; McCarter et al., 2000a; Parkinson et al., 2001).

Here we present a brief progress report on publicly available ESTs from nematodes. Since our last update in December 2000 (McCarter et al., 2000b), 179,968 new nematode-derived ESTs have been submitted to dbEST including 94,073 from parasites. *Caenorhabditis elegans* has long been a focus of sequencing efforts (The *C. elegans* Sequencing Consortium, 1998), and 193,692 ESTs are available from *Caenorhabditis* species (Kohara, 1996; McCombie et al., 1992; Waterston et al., 1992). Currently, 170,679 ESTs are available from 28 nematode species beyond *Caenorhabditis,* including 7 human parasites, 12 animal parasites, 7 plant parasites, and 2 free-living bacteriovores (Table 2). The majority of these ESTs were generated in 1999–2002. With the exception of *Caenorhabditis* species and *Brugia malayi*, ESTs dominate the available sequence data for nematodes with 31-fold the number of conventionally submitted nucleotide and protein sequences in GenBank. Because ESTs are redundant with common mRNAs highly represented, the 170,679 ESTs from nematodes beyond *Caenorhabditis* likely represent 50,000–70,000 genes. For example, 12,269 ESTs from *Onchocerca volvulus* have been clustered to form 4,208 groups (Williams et al., 2002). We have clustered 3,979 ESTs from *Trichinella spiralis* to form 1,880 groups. The *Trichinella* clusters along with those from five other species are searchable at www.nematode.net/Nemagene.

Available EST data from plant-parasitic nematodes derive from root-knot nematodes (four species, 25,900 ESTs) (Dautova et al., 2001) and cyst nematodes (three species, 12,093 ESTs) (Popeijus et al., 2000). To date, stage representation is limited to cDNA libraries made from eggs and second-stage juveniles. A goal for future EST generation from plant parasites is to increase representation from other life-cycle stages including adult males and dissected juvenile and adult females. There are no publicly funded EST projects focused on migratory endoparasitic or ectoparasitic nematodes, nor is there yet a funded project aimed at obtaining the complete genome sequence of a plant-parasitic nematode.

The 123,387 ESTs from human and animal parasitic nematodes provide generally better stage coverage than is available from plant parasites, and more analyses of these sequences have been completed (Blaxter et al., 1996; Blaxter, 2000; Daub et al., 2000; Hoekstra et al., 2000; Lizotte-Waniewski et al., 2000; Maizels et al., 2000; Moore et al., 1996; Tetteh et al., 1999; Unnasch and Williams, 2000). For both *Brugia malayi* and *Onchocerca volvulus,* ESTs have been generated from six stage-specific libraries. Many species have representation from two or more stages. A unique resource has been generated from *Ascaris suum,* where the adult parasite's large size allows the dissection of individual organs—a procedure that is difficult for most nematodes. Tissue-specific cDNA libraries have been constructed and ESTs sequenced from muscle and nerve cord (684 ESTs); female head (2,572), male head (2,388); female intestine (3,028); male intestine (2,415); female ovary - germinal zone (2,250), differentiation zone (500), and maturation zone (4,160); and male testis-germinal zone (1,608). Moving beyond ESTs, the continuing drop in sequencing cost is now making full genome sequencing from parasitic nematodes plausible, at least for draft quality sequence. Recently, the National Institutes of

Received for publication 18 February 2002.
[1] Genome Sequencing Center, Department of Genetics, Box 8501, Washington University School of Medicine, St. Louis, MO 63108.
[2] Plant Nematode Genetics Group, Department of Plant Pathology, North Carolina State University, Raleigh, NC 27695.
[3] Divergence Inc., 892 North Warson Road, St. Louis, MO 63141.
E-mail: mccarter@genetics.wustl.edu
This paper was edited by B. C. Hyman.

TABLE 1. Selected Web resources for nematode EST access.

EST resources	URLs
GenBank dbEST	www.ncbi.nlm.nih.gov/dbEST
Genome Sequencing Center ESTs & Clusters	www.nematode.net
Blaxter Lab ESTs & Clusters	http://nema.cap.ed.ac.uk/index.html
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	www.nematode.net/Links

Health-National Institute of Allergy and Infectious Diseases (NIH-NIAID) has approved funding for The Institute for Genomic Research and collaborators to generate 5× coverage of the *Brugia malayi* genome by sequencing of paired-end reads and BAC ends (www.tigr.org/tdb/e2k1/bma1/).

Nematologists 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; Fraser et al., 2000; Jones et al., 2000; Kim et al., 2001; Stein et al., 2001). 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 101 finished megabases (Genome Sequencing Center, unpubl. data), with 20,448 predicted proteins including 823 splice variants (Wellcome Trust Sanger Institute Wormpep Release 78, April 26, 2002). A number of small gaps remain. The majority of genes identified to date in parasitic nematodes have homologues in *C. elegans*. For example, BLASTX analy-

TABLE 2. 30 Nematode species have more than 50 ESTs registered in the GenBank dbEST database, June 2002.

Nematode species	ESTs 3/97	ESTs 12/00	ESTs 6/02	Other GenBank entries 6/02	Major EST sources
Caenorhabditis elegans	30,196	109,215	191,268	87,591	1, 2, 11
Ascaris suum	0	588	24,492	348	2, 3, 6
Brugia malayi	7,496	22,392	22,439	18,337	3, 4, 5, 2
Onchocerca volvulus	310	13,802	14,922	777	5, 2
Strongyloides stercoralis	57	10,922	11,392	54	2
Meloidogyne incognita	0	6,626	10,899	148	2, 7
Pristionchus pacificus	703	4,989	8,818	15	2
Strongyloides ratti	0	0	8,645	23	2
Parastrongyloides trichosuri	0	0	7,963	3	2
Ancylostoma caninum	0	5,546	7,656	93	2
Meloidogyne hapla	0	0	6,157	18	2
Globodera rostochiensis	0	894	5,934	75	2, 7, 8
Meloidogyne javanica	22	1,208	5,600	41	2
Ostertagia ostertagi	0	0	5,591	184	2, 3, 6
Haemonchus contortus	0	2,399	4,906	497	3, 6, 9, 10
Heterodera glycines	0	1,506	4,327	183	2
Trichinella spiralis	0	0	4,247	141	2
Toxocara canis	8	519	3,920	106	2, 3
Meloidogyne arenaria	0	0	3,334	37	2
Ancylostoma ceylanicum	0	0	2,690	58	2
Caenorhabditis briggsae	2,424	2,424	2,424	519	2
Trichuris muris	0	301	2,125	3	3, 6
Globodera pallida	0	94	1,832	121	7, 8
Necator americanus	0	211	961	125	3, 6
Nippostrongylus brasiliensis	0	0	734	32	3
Zeldia punctata	0	378	391	5	2
Teladorsagia circumcincta	0	0	315	119	3, 6
Litomosoides sigmodontis	0	198	198	33	3
Wuchereria bancrofti	119	131	131	71	5
Onchocerca ochengi	0	60	60	13	5
Dirofilaria immitis	0	0	Pending	161	2
Pratylenchus penetrans	0	0	Pending	19	2
Total Sequences	41,335	184,403	364,371	109,950	
Total Non-*Caenorhabditis*	11,139	72,764	170,679	21,840	

1. National Institute of Genetics, Mishima, Japan.
2. Genome Sequencing Center, Washington University School of Medicine, St. Louis, MO USA.
3. Institute of Cell, Animal, and Population Biology, University of Edinburgh, Edinburgh, UK.
4. World Health Organization Filarial Genome Network.
5. Department of Biology, Smith College, Northampton, MA USA.
6. The Wellcome Trust Sanger Institute, Hinxton, UK.
7. Laboratory of Nematology, Wageningen University, Wageningen, The Netherlands.
8. Nematology Department, Scottish Crop Research Institute, Dundee, UK.
9. Institute for Animal Science and Health, Lelystad, The Netherlands.
10. Department of Veterinary Microbiology and Pathology, Washington State University, Pullman, WA USA.
11. The Institute for Genomic Research, Rockville, MD USA.

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

EST resource	URL
Wormbase	www.wormbase.org
Wormpep, Sanger Centre	www.sanger.ac.uk/Projects/C_elegans/wormpep
C. elegans WWW Server	elegans.swmed.edu
C. elegans Project & BLAST Server at Sanger Institute or Genome Sequencing Center	www.sanger.ac.uk/Projects/C_elegans or http://genome.wustl.edu/projects/celegans/

sis reveals that 66% of *Meloidogyne incognita* EST clusters have a *C. elegans* homologue ($E < 10^{-5}$). Key *C. elegans* genome resources are shown in Table 3. Additionally, in 2001 the Genome Sequencing Center at Washington University and the Wellcome Trust Sanger Institute each sequenced approximately 1 million whole genome shotgun reads from *C. briggsae* providing >10× coverage of this ~100-Mb genome (13 Mb had already been finished). A draft assembly of the whole *C. briggsae* genome is available for blast searching at http://genome.wustl.edu/projects/cbriggsae, and comparisons with syntenic stretches of the *C. elegans* genome have begun (Kent and Zahler, 2000; Sanger Institute and the Washington University Genome Sequencing Center, in preparation).

Using available nematode sequence data can save time and effort in the laboratory as well as greatly affect plans for experimental design. We will continue to provide periodic updates on the status of nematode gene sequencing over the next several years as the EST and whole genome data sets continue their rapid expansion.

ACKNOWLEDGMENTS

Nematode EST sequencing at Washington University is supported by NIH-NIAID research grant AI 46593 to Robert Waterston; NSF Plant Genome award 0077503 to David Bird (PI) and co-PIs Sandra Clifton, Joseph Kieber, Charles Opperman, and Jeffrey Thorne; an MRC research grant to Mark Viney; and a Max Planck Institute grant to Ralf Sommer. James McCarter was supported by a Merck Postdoctoral Fellowship from the Helen Hay Whitney Foundation. We would like to thank members of the Genome Sequencing Center EST lab, especially Deana Pape, John Martin, Todd Wylie, Brandi Chiapelli, and Claire Murphy, and the many collaborators who have generously provided nematode materials for cDNA library production (www.nematode.net/Collaborators/), especially Al Scott for supplying dissected *Ascaris* tissues.

LITERATURE CITED

Blaxter, M. 2000. Genes and genomes of *Necator americanus* and related hookworms. International Journal of Parasitology 30:347–355.

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.

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

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

Dautova, M., M. N. Rosso, P. Abad, F. J. Gommers, J. Bakker, and G. Smant. 2001. Single pass cDNA sequencing—a powerful tool to analyze gene expression in preparasitic juveniles of the southern root-knot nematode *Meloidogyne incognita.* Nematology 3:129–139.

Fraser, A. G., R. S. Kamath, P. Zipperlen, M. Martinez-Campos, M. Sohrmann, and J. Ahringer. 2000. Functional genomic analysis of *C. elegans* chromosome I by systematic RNA interference. Nature 408: 325–330.

Hoekstra, R., A. Visser, M. Otsen, J. Tibben, J. A. Lenstra, and M. H. Roos. 2000. EST sequencing of the parasitic nematode *Haemonchus contortus* suggests a shift in gene expression during transition to the parasitic stages. Molecular and Biochemical Parasitology 110: 53–68.

Jones, S. J. M., D. L. Riddle, A. T. Pouzyrev, V. E. Velculescu, L. Hillier, S. R. Eddy, S. L. Stricklin, D. L. Baillie, R. Waterston, and M. A. Marra. 2001. Changes in gene expression associated with developmental arrest and longevity in *Caenorhabditis elegans.* Genome Research 11:1346–1352.

Kent, W. J., and A. M. Zahler. 2000. Conservation, regulation, synteny, and introns in a large-scale *C. briggsae–C. elegans* genomic alignment. Genome Research 10:1115–1125.

Kim, S. K., J. Lund, M. Kiraly, K. Duke, M. Jiang, J. M. Stuart, A. Eizinger, B. N. Wylie, and G. S. Davidson. 2001. A gene expression map for *Caenorhabditis elegans.* Science 293:2087–2092.

Kohara, Y. 1996. Large-scale analysis of *C. elegans* cDNA. Tanpakushitsu Kakusan Koso 41:715–720.

Lizotte-Waniewski, M., W. Tawe, D. B. Guiliano, W. Lu, J. Liu, S. A. Williams, and S. Lustigman. 2000. Identification of potential vaccine and drug target candidates by expressed sequence tag analysis and immunoscreening of *Onchocerca volvulus* cDNA libraries. Infection and Immunity 68:3491–3501.

Maizels, R. M., K. K. A. Tetteh, and A. Loukas. 2000. *Toxocara canis:* Genes expressed by the arrested infective larval stage of a parasitic nematode. International Journal of Parasitology 30:495–508.

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

McCarter, J. Abad, J. T. Jones, and D. Bird. 2000a. Rapid gene discovery in plant-parasitic nematodes via expressed sequence tags. Nematology 2:719–731.

McCarter, J. P., D. McK. Bird, S. W. Clifton, and R. H. Waterston. 2000b. Nematode gene sequences, December 2000 update. Journal of Nematology 32:331–333.

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.

Parkinson, J., C. Whitton, D. Guiliano, J. Daub, and M. Blaxter. 2001. 200,000 nematode expressed sequence tags on the Net. Trends in Parasitology 17:394–396.

Popeijus, H., V. C. Blok, L. Cardle, J. Bakker, M. S. Phillips, J. Helder, G. Smant, and 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 2:567–574.

Stein, L., P. Sternberg, R. Durbin, J. Thierry-Mieg, and J. Spieth. 2001. Wormbase: Network access to the genome and biology of *Caenorhabditis elegans*. Nucleic Acids Research 29:82–86.

Tetteh, K. K. A., A. 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.

Unnasch, T. R., and S. A. Williams. 2000. The genomes of *Onchocerca volvulus*. International Journal of Parasitology 30:543–552.

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.

Williams, S. A., S. J. Laney, M. Lizotte-Waniewski, and L. A. Bierwert. 2002. The Riverblindness Genome Project. Trends in Parasitology 18:86–90.