

200 000 nematode expressed sequence tags on the Net

Expressed sequence tags (ESTs) are single-pass sequence reads made from randomly selected cDNA clones, which represent the expressed genes of an organism. EST analysis is an efficient and cost-effective method for sampling the genes expressed by an organism or tissue. ESTs have been a focus of eukaryotic parasite genome initiatives for several years¹ (see the Parasite Genome web server at <http://www.ebi.ac.uk/parasites/parasite-genome.html>), and parasitic organism ESTs make up a significant portion of the dbEST subsection of GenBank^{2,3}. EST data sets can be mined for useful or interesting content using standard similarity-based search tools such as BLAST. In the field of molecular parasitology, such approaches have led to the discovery of many new potential drug targets and vaccine candidates. However, EST data sets also contain important additional types of information.

The abundance of a particular EST sequence in the database for an organism will reflect its steady-state expression level. As cDNA libraries from which ESTs derive might have been constructed from isolated tissues or life-cycle stages, the EST content per library can also be informative as to tissue- or stage-specific, or tissue- or stage-regulated expression. In addition, by comparing EST profiles from related organisms, a phylogenetic profile of conservation at the sequence and expression pattern level can be built up for each gene and its homologues.

One of the problems inherent in the analysis of ESTs is the quantity of data available (e.g. there are more than 22 000 ESTs from the human filarial nematode *Brugia malayi*^{2,4}), which precludes simple browsing as a means to analysis. In addition, individual ESTs are single-pass sequences of unverified quality, and can therefore include base-calling and other errors. Hence, to make EST data sets available in a user-friendly form, it is necessary to build databases that hold pre-computed analyses that are amenable to searching with complex queries (e.g.

show me the unique genes from organism X that have an expression profile restricted to stage Y). ESTs represent the expressed portions of the genome, and thus comparison to the corresponding genome sequence can allow grouping of ESTs by which genomic segment or gene they are derived from, and the genomic sequence can be used as a coordinating key in analysis.

More than 109 000 ESTs have been submitted for the free-living nematode *Caenorhabditis elegans*⁵⁻⁷. The 100 Mb genome of *C. elegans* has been completely sequenced⁸, and these ESTs have proven utility in defining and confirming the boundaries of predicted genes, and in showing where alternative splicing of pre-mRNAs is found^{9,10}. The EST data set has been integrated into the genome database for *C. elegans* (Wormbase), and additional analyses are made public at the National Institute of Genetics in Japan (Table 1). The *C. elegans* ESTs, by comparison to the genome sequence, represent over half of the predicted 20 000 genes⁹. The EST sequences, and the clones from which they derive, are integral parts of several post-genomics analyses in *C. elegans*, including genome-wide double-stranded RNA interference (RNAi) screens for gene function¹⁰, and microarray analysis of gene expression¹¹.

Large numbers of ESTs are being generated from parasitic nematodes (Table 1)¹²⁻¹⁸. Two large projects, one at the University of Edinburgh (Blaxter Nematode Genetics Laboratory, BaNG) and the other at Washington University Genome Sequencing Center (GSC), St Louis, USA (under the direction of Jim McCarter)^{19,20}, will produce 10 000–90 000 ESTs each for 14 species of nematode parasites of humans, animals and plants. The ESTs are deposited in the public databases (GenBank dbEST) and can be accessed through the standard database query systems (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi> and <http://srs.ebi.ac.uk/>). In the absence of sequenced genomes on which to organise the EST data, both projects are developing databases [NemaGene at GSC and NEMBASE at BaNG (<http://nema.cap.ed.ac.uk/nematodeESTs/nembase.html>)] to aid users in access to the analysed data.

NEMBASE

NEMBASE is a relational database based on the popular SQL (originally derived as an acronym for structured query language) database system that permits access to ESTs from multiple nematode species. NEMBASE presents this data through a forms-based Web browser interface. For each species, the entire public DNA sequence data is downloaded (ESTs, cDNAs and genomic DNA). The sequences are compared to one another, and sequences that show significant identity are grouped together to form a cluster. The clustering process attempts to account for any of the technical or methodological errors that can arise in EST data sets, such as chimaerism and low quality. Each cluster is given a unique identifier, and is presumed to be made up of sequences that derive from a single gene. The clustered sequences are then aligned to each other and a consensus sequence predicted. The consensus sequence is then used to perform database searches using standard algorithms. The unique cluster identifier is maintained through successive incremental updates, and can therefore be used to track individual genes²¹. The database includes information on individual ESTs as well as on the clusters and associated sequence similarity information. Human curation of the cluster and autoannotation information is ongoing, and is retained in the database system.

Currently, five data sets are available on the NEMBASE server at <http://www.nematodes.org>. The *B. malayi* genome project data set of 22 441 ESTs has been clustered and analysed. Data sets including BaNG ESTs from *Haemonchus contortus* (2961 sequences), *Trichuris muris* (752), *Necator americanus* (911) and *Ascaris suum* (686) are also presented. The database can be queried using simple text searches [e.g. 'are there sequences with significant similarity to proteins with this identifier?' (which could be 'kinase', 'globin', 'growth factor' or any other text string)]. The query interface allows you to ask for only those database hits scoring better than a given cut-off. It is possible to restrict the search by library or stage, so that genes expressed in one tissue or life-cycle stage can be identified. A dedicated BLAST server also permits sequence similarity searching of the clustered data sets.

Table 1. Nematode expressed sequence tag (EST) sequencing projects (data at 26/04/2001)

Species	Description	Sequencing centre ^a	No. of ESTs Planned deposited ^b	EST total	Database URLs	Comments
<i>Caenorhabditis elegans</i>	Free living	NIG, GSC, TIGR and Sanger	109 215	^d	Wormbase http://www.wormbase.org and http://www.ddbj.nig.ac.jp/c-elegans/html/CE_INDEX.html	
<i>Caenorhabditis briggsae</i>	Free living	GSC	2424	^d	http://genome.wustl.edu/gsc/Projects/briggsae.shtml	The <i>C. briggsae</i> ESTs are also incorporated into Wormbase http://www.wormbase.org
<i>Brugia malayi</i>	Human lymphatic filarial parasite	FGP	22 441	^d	NEMBASE http://www.nematodes.org	Additional ESTs from Jasmer ^c and Roos ¹⁸ ; 20 000 ESTs will also be generated from <i>Teladorsagia circumcincta</i> by BaNG
<i>Onchocerca volvulus</i>	Human river blindness filarial parasite	FGPRC	14 608	^d		
<i>Litomosoides sigmodontis</i>	Rodent model filarial parasite	BaNG	198	^d		
<i>Haemonchus contortus</i>	Sheep gut parasite	BaNG	2749	20 000		
<i>Necator americanus</i>	Human hookworm	BaNG	900	20 000		
<i>Trichuris muris</i>	Murine model for human threadworm	BaNG	751	20 000		
<i>Ascaris suum</i>	Swine gut parasite	BaNG and GSC	588	30 000		5000 ESTs will also be generated from the human parasite <i>A. lumbricoides</i>
<i>Globodera rostochiensis</i>	Potato cyst nematode	SCRI	894	^d		Preliminary analysis at NEMBASE http://www.nematodes.org 94 ESTs generated for <i>G. pallida</i> Preliminary analysis at NEMBASE http://www.nematodes.org
<i>Toxocara canis</i>	Canine gut parasite	Maizels and GSC	519	5000		
<i>Trichinella spiralis</i>	Muscle parasite	GSC and BaNG	3766	11 500		Additional data in NEMBASE http://www.nematodes.org
<i>Ancylostoma caninum</i>	Dog hookworm	GSC	5625	11 500		11 500 ESTs will also be generated from the human hookworm <i>A. duodenale</i>
<i>Strongyloides stercoralis</i>	Human gut parasite	GSC	10 979	11 500	http://genome.wustl.edu/est/PNP_esthmpg.html and http://www.nematode.net	
<i>Strongyloides ratti</i>	Rodent model gut parasite	GSC	636	21 500		
<i>Meloidogyne incognita</i>	Root knot plant parasite	GSC	6626	91 500		There are also 1223 ESTs from <i>M. javanica</i>
<i>Heterodera glycines</i>	Soy bean cyst plant parasite	GSC	1772	11 500		
<i>Pristionchus pacificus</i>	Free living	GSC	4989	15 000		There are also 378 ESTs from the free-living cephalobe <i>Zeldia punctata</i>

^aAbbreviations: NIG, Y. Kohara, National Institute of Genetics, Mishima, Japan; GSC, Genome Sequencing Center, Washington University School of Medicine, St Louis, MO, USA; TIGR, The Institute for Genome Research, Gaithersburg, MA, USA; Sanger, The Sanger Centre, Hinxton Genome Campus, Cambridge, UK; FGP, The Filarial Genome Project (see <http://nema.cap.ed.ac.uk/fgn/filgen1.html> for participant information); FGPRC, S. Williams, The Filarial Genome Project Resource Center, Smith College, Northampton, MA, USA; BaNG, M. Blaxter, Nematode Genetics Laboratory, Institute of Cell, Animal and Population Biology, University of Edinburgh, UK; SCRI, J. Jones, Scottish Crop Research Institute, Invergowrie, Dundee, UK; Maizels, R. Maizels, Institute of Cell, Animal and Population Biology, University of Edinburgh, UK.

^bESTs deposited as of 9 March 2001.

^cUnpublished dbEST submissions from Douglas Jasmer, University of Washington, Pullman, WA, USA.

^dNo additional sequencing planned.

In addition, BaNG has set up a Nematode Blast Server that includes sequence data from many additional species of nematodes (http://nema.cap.ed.ac.uk/ncbi_blast.html), where searches can be performed on individual species' or (phylogenetically grouped) sets of species' sequences.

Within the next two years, over 250 000 additional nematode ESTs will be deposited in public databases. NEMBASE will develop and expand with these developments, particularly to include additional species. Future enhancements will include integrated phylogenetic profiles cross-comparing all the nematode EST data sets. The site will develop into a central clearing-house for the analysis of gene expression in nematodes that infect humans, animals and crops, and aid the development of new anti-nematode treatments.

Acknowledgements

Nematode genomics work in BaNG is supported by the Medical Research Council and the Wellcome Trust. The authors thank their collaborators at the Sanger Centre (Bart Barrell, Neil Hall, Mike Quail and Barbara Harris) and at GSC (Jim McCarter), and their research colleagues who have supplied nematode materials and libraries (Alan Scott, Rick Maizels, David Knox, David Pritchard, Richard Grencis, Doug Jasmer, Bernadette Connolly and Tim Geary).

John Parkinson*

Claire Whitton

David Guilliano

Jen Daub

Mark Blaxter

Institute of Cell, Animal and Population Biology, Ashworth Laboratories, Kings Buildings, University of Edinburgh, Edinburgh, UK EH9 3JT.

*e-mail: john.parkinson@ed.ac.uk

References

- Blaxter, M.L. and Aslett, M.A. (1997) Internet resources for the parasite genome projects. *Trends Genet.* 13, 40–41
- Blaxter, M.L. *et al.* (1999) Parasitic helminth genomics. *Parasitology* 118, S39–S51
- Johnston, D.A. *et al.* (1999) Genomics and the biology of parasites. *BioEssays* 21, 131–147
- Williams, S.A. *et al.* (2000) The filarial genome project: analysis of the nuclear, mitochondrial and endosymbiont genomes of *Brugia malayi*. *Int. J. Parasitol.* 30, 411–419
- McCombie, W.R. *et al.* (1992) *Caenorhabditis elegans* expressed sequence tags identify gene families and potential disease gene homologues. *Nat. Genet.* 1, 124–131
- Waterston, R. *et al.* (1992) A survey of expressed genes in *Caenorhabditis elegans*. *Nat. Genet.* 1, 114–123
- Kohara, Y. (1996) Large scale analysis of *C. elegans* cDNA. *Tanpakushitsu Kakusan Koso* 41, 715–720
- The *C. elegans* Genome Sequencing Consortium (1998) Genome sequence of *Caenorhabditis elegans*: a platform for investigating biology. *Science* 282, 2012–2018
- Reboul, J. *et al.* (2001) Open-reading-frame sequence tags (OSTs) support the existence of at least 17,300 genes in *C. elegans*. *Nat. Genet.* 27, 332–336
- Maeda, I. *et al.* (2001) Large-scale analysis of gene function in *Caenorhabditis elegans* by high-throughput RNAi. *Curr. Biol.* 11, 171–176
- Hill, A.A. *et al.* (2000) Genomic analysis of gene expression in *C. elegans*. *Science* 290, 809–812
- Daub, J. *et al.* (2000) A survey of genes expressed in adults of the human hookworm *Necator americanus*. *Parasitology* 120, 171–184
- Blaxter, M.L. *et al.* (1996) Genes expressed in *Brugia malayi* infective third stage larvae. *Mol. Biochem. Parasitol.* 77, 77–96
- Lizotte-Waniewski, M. *et al.* (2000) Identification of potential vaccine and drug target candidates by expressed sequence tag analysis and immunoscreening of *Onchocerca volvulus* larval cDNA libraries. *Infect. Immun.* 68, 3491–3501
- Allen, J.E. *et al.* (2000) Analysis of genes expressed at the infective larval stage validates utility of *Litomosoides sigmodontis* as a murine model for filarial vaccine development. *Infect. Immun.* 68, 5454–5458
- Moore, T.A. *et al.* (1996) Identification of novel sequences and codon usage in *Strongyloides stercoralis*. *Mol. Biochem. Parasitol.* 79, 243–248
- Maizels, R.M. *et al.* (2000) *Toxocara canis*: genes expressed by the arrested infective larval stage of a parasitic nematode. *Int. J. Parasitol.* 30, 495–508
- Hoekstra, R. *et al.* (2000) EST sequencing of the parasitic nematode *Haemonchus contortus* suggests a shift in gene expression during transition to the parasitic stages. *Mol. Biochem. Parasitol.* 110, 53–68
- McCarter, J.P. *et al.* Rapid gene discovery in plant parasitic nematodes via expressed sequence tags. *Nematology* (in press)
- McCarter, J.P. *et al.* Nematode gene sequences, Update for December 2000. *J. Nematol.* (in press)
- Blaxter, M.L. *et al.* (1997) A unified nomenclature for filarial genes. *Parasitol. Today* 13, 416–417

ParaSite

Nematodes and the Neem tree on the Net

Nematode parasites of flies?

John Potter (Agriculture and Agri-Food, Canada) asked the Nematode Discussion List for information about parasites of fungus gnats (*Bradysia* spp.) or of shore flies (*Scatella* spp.). Brian Clark (University of Nebraska-Lincoln, USA) suggested checking papers by D. Gouge [e.g. Gouge D.H. and Hague N.G. (1995) The susceptibility of different species of sciarid flies to entomopathogenic nematodes. *J. Helminthol.* 69, 313–318, which mentions *Steinernema feltiae*, a parasite of six species of *Sciarid* flies, and two *Heterorhabditis* spp. Apparently, adult sciarids infected by *S. feltiae* can disperse nematodes to nematode-free compost]. Suzanne Wainwright has used *S. feltiae* for many years to control fly larvae in nurseries and interscapes (*sic*) and was about to start field trials to see how *S. carpocapsae* compared with it. From replies received, Potter concluded that fungus gnats, but not shore-flies, have been well-studied. 'I guess mushrooms are more significant than rotting seaweed...', he said.

Malaria discussion group

Ideas for new research – how discussions can degenerate

In May 2000, Dave Sarshalom (a third-year medical student from Venezuela) searched the Web in vain. P. Krishnan, a 79-year-old Emeritus Consultant from Tamilnadu Hospital, Madras, India, following his own observations, came up with an age-old idea – ultimately, it is the immune system that will repay study, especially to explain how patients recover and why they relapse. This requires fieldwork. Ruth Sponsler supported him, favouring a study of immune enhancement, particularly by vitamin supplementation, perhaps combined with vector control; for example, by bednets. Tina Skinner-Adams (University of Western Australia) said that vitamin A is important for immune function and supplements might be required in populations whose serum retinol is low. Indeed, experiments in her lab suggest that vitamin A and other retinoids might inhibit parasite growth directly. 'This action of retinol is synergistic with some antimalarials, but antagonistic with others'. Mike Hollingdale (then at Leeds University, UK) referred to interesting work that suggests that clinical episodes, spleen enlargement and parasite density