Research

Signal sequence analysis of expressed sequence tags from the nematode *Nippostrongylus brasiliensis* and the evolution of secreted proteins in parasites

Yvonne M Harcus*, John Parkinson*‡, Cecilia Fernández*§, Jennifer Daub*, Murray E Selkirk†, Mark L Blaxter* and Rick M Maizels*

Addresses: *Institute of Cell, Animal and Population Biology, University of Edinburgh, Edinburgh, EH9 3JT, UK. †Department of Biological Sciences, Imperial College London, London SW7 2AZ, UK. ‡Current address: Program in Genetics and Genomic Biology, Hospital for Sick Children, University Avenue, Toronto, Ontario M5G 1X8, Canada. §Current address: Facultad de Química, Cátedra de Inmunología, Universita de la Republica, Montevideo 11300, Uruguay.

Correspondence: Rick M Maizels. E-mail: rick.maizels@ed.ac.uk

Abstract

Background: Parasitism is a highly successful mode of life and one that requires suites of gene adaptations to permit survival within a potentially hostile host. Among such adaptations is the secretion of proteins capable of modifying or manipulating the host environment. *Nippostrongylus brasiliensis* is a well-studied model nematode parasite of rodents, which secretes products known to modulate host immunity.

Results: Taking a genomic approach to characterize potential secreted products, we analyzed expressed sequence tag (EST) sequences for putative amino-terminal secretory signals. We sequenced ESTs from a cDNA library constructed by oligo-capping to select full-length cDNAs, as well as from conventional cDNA libraries. SignalP analysis was applied to predicted open reading frames, to identify potential signal peptides and anchors. Among 1,234 ESTs, 197 (~16%) contain predicted 5' signal sequences, with 176 classified as conventional signal peptides and 21 as signal anchors. ESTs cluster into 742 distinct genes, of which 135 (18%) bear predicted signal-sequence coding regions. Comparisons of clusters with homologs from *Caenorhabditis elegans* and more distantly related organisms reveal that the majority (65% at $P < e^{-10}$) of signal peptide-bearing sequences from *N. brasiliensis* show no similarity to previously reported genes, and less than 10% align to conserved genes recorded outside the phylum Nematoda. Of all novel sequences identified, 32% contained predicted signal peptides, whereas this was the case for only 3.4% of conserved genes with sequence homologies beyond the Nematoda.

Conclusions: These results indicate that secreted proteins may be undergoing accelerated evolution, either because of relaxed functional constraints, or in response to stronger selective pressure from host immunity.

Background

A central tenet of parasitology is that parasites must secrete biologically active mediators that modify or customize their niche within the host in order to survive immune attack. Such secretions have long been the focus of biochemical and immunological analyses [1-4]. With larger-scale genomic approaches now possible, a screen can be designed in which the characteristic signal sequences, necessary for proteins to exit the eukaryotic cell via the secretory pathway, can be identified by bioinformatic methods [5-9]. We describe here an analysis of this nature, applied to a widely used model system, *Nippostrongylus brasiliensis*, the gastrointestinal nematode of rats [10-12].

N. brasiliensis biology encapsulates many key aspects of parasite infection and immunology. It is a multicellular metazoan belonging to the phylum Nematoda, which together with the platyhelminth groups (Cestoda and Trematoda) are collectively known as helminths. Helminth infections are typically accompanied by a polarized type-2 (Th2) immune response, characterized by IgE antibody production, eosinophilia and mastocytosis [13-15]. *N. brasilensis* drives extremely strong Th2 responses [16], and this bias can be reproduced with secreted proteins collected from parasites *in vitro* [17]. More than 100 secreted proteins have been found by two-dimensional SDS-PAGE analysis (Y.H. and R.M.M., unpublished work), and among those experimentally verified are acetylcholinesterases [18-20], cysteine proteases [21,22], and a hydrolase that degrades an important host inflammatory mediator, platelet activating factor [23,24].

The molecular biological analysis of *N. brasiliensis* genes and gene products is at a very early stage. Secreted and intracellular globins have been characterized [25], and genes for both secretory [26,27] and neuronal [28] acetylcholinesterases cloned. A recombinant cystatin (cysteine protease inhibitor) has been shown functionally to inhibit host antigen-processing pathways [29]. Structural genes for both tubulin [30] and a keratin-like protein [31] have been described, and an α-crystallin-like small heat-shock protein (Hsp20) has been reported [32]. However, these studies on individual genes have yet to be complemented by higher-throughput molecular analyses. The potential of *N. brasiliensis* as an experimental system for functional genomics has been greatly enhanced by the demonstration of successful RNAi knockdown in this species [33].

The genomes of parasitic nematode species are between 60 and 250 megabases (Mb) in size [34], and there are more than 20 species of medical, veterinary and scientific importance [35]. Over the past decade, the most tractable way of applying genomics to this group of organisms has been by expressed sequence tag (EST) projects [36]. Large-scale EST sequencing of the human filarial parasite *Brugia malayi* [37,38] has been followed by similar studies in the sheep intestinal worm *Haemonchus contortus* [39], human hookworms [40], the river-blindness parasite *Onchocerca volvulus* [41], and important plant-parasitic species such as *Meloidogyne incognita* [42]. Smaller projects have added *Litomosoides sigmodontis* [43], *Toxocara canis* [44] and many other related species to the available database of parasitic nematode sequences [36]. In designing a study on *N. brasiliensis*, we wished to focus on the potential for secreted proteins that may interact with the host immune system. We therefore conducted an EST project that included a cDNA library specifically enriched for full-length inserts [45], allowing analysis of amino-terminal signal peptides to be carried out.

The evolutionary history of secreted immunomodulators is likely to be that of recent adaptation from ancestral genes which fulfilled other functions in free-living ancestors. Comparative studies on nematodes can take advantage of full-genome information available for the free-living species *Caenorhabditis elegans* [46] and *C. briggsae* [47], which are quite closely related to *N. brasiliensis* [48]. If rapid evolution of secreted gene products was required for efficient parasitism, this may be evident in greater diversity among signal peptide-bearing sequences than among genes coding for non-secreted proteins. We report here our results that support this hypothesis.

Results and discussion

A high proportion of *N. brasiliensis* ESTs encode proteins with predicted signal sequences

A total of 1,234 ESTs were collected from adult *N. brasiliensis* cDNA libraries constructed either by conventional means or by an oligo-capping method to select full-length cDNAs [45]. A full analysis of these has been posted on our website [49]. ESTs were then analyzed by SignalP, which predicted that 16.0% of total ESTs (197/1,234) contained either 5' signal peptide sequences (176/1,234) or signal anchors (21/1,234, Table 1). The oligo-capped cDNA library yielded a notably higher proportion of sequences with predicted signal peptides (20.4%) than did conventional cDNA libraries (10.1%).

The dataset was then clustered to account for multiple ESTs from highly expressed genes, and ESTs were assigned to 742 clusters, including 567 singletons. The proportion of clusters bearing potential signal sequences remained high (135/742; 18.2%), confirming that the dataset is not skewed by over-representation of a few abundant transcripts. The overall proportion of cDNAs encoding predicted signal peptides is within the 15-25% range estimated by analyis of whole-genome sequence data [50]. Of all predicted signal-sequence-bearing clones or clusters from *N. brasiliensis*, around 90% were classified as conventional signal peptides associated with export and secretion into the extracellular environment. The remaining approximately 10% were identified as potential signal anchors, in which the hydrophobic amino-terminal segment is retained, without cleavage, as a transmembrane domain for type II plasma membrane proteins [7].

Table 1

Analysis of transcripts represented in conventional and oligo-capped cDNA libraries

	Conventional cDNA libraries	Oligo-capped cDNA library
Total sequences providing peptide predictions	734	500
In-frame ATG followed by ≥ 99-nucleotide open reading frame (ORF)	567 (77.2%)	430 (86.0%)
Predicted ORF length (average)	114.6	101.5
% Signal peptide or signal anchor	SP: 74 (10.1%)	SP: 102 (20.4%)
	SA: 16 (2.2%)	SA: 5 (1.0%)
% Spliced leader	0	37 (7.4%)

Presence of *trans*-spliced leaders in *N. brasiliensis*

All nematodes undergo *trans*-splicing at the 5' end of a proportion of their mRNA transcripts; a short leader sequence is added upstream of the initiation codon. The leader is normally a 22-nucleotide sequence termed SL1 [51]. The precise SL1 sequence is highly conserved throughout the phylum, although the degree to which transcripts are *trans*-spliced varies between different nematode species [52]. To evaluate the prominence of SL1-*trans*-splicing in *N. brasiliensis*, we searched the 1,234 ESTs with the 3' 14 nucleotides of SL1, to allow for any minor truncation of cDNAs. Only 37 matches were found, all from the oligo-capped cDNA library (from 500 ESTs, giving a frequency of 7.4%); a few clones from the conventional libraries had 10 or fewer nucleotides identical to the SL1 sequence at their 5' termini. Although the overall frequency of *trans*-splicing in *N.brasiliensis* is not yet known, this level is well below those of other species, such as *C. elegans*. Moreover, transcripts bearing the spliced leader (and its unique tri-methylguanosine cap) are, in certain species, under-represented by the method we used to selectively amplify full-length mRNAs [45]. Hence the true extent of *trans*-splicing may be higher than the proportion evident in the current dataset.

N. brasiliensis sequences show closest similarity to those of other trichostrongyles

N. brasiliensis is a stronglylid nematode, closely related to veterinary parasites such as *Haemonchus contortus* and *Teladorsagia* (previously *Ostertagia*) *circumcincta* in the Superfamily Trichostrongyloidea, and within the Order Strongylida which includes human hookworm pathogens *Ancylostoma duodenale* and *Necator americanus* [53]. The closest free-living taxa to the Strongylida are members of the Rhabditina, including *C. elegans*, and both are grouped in Clade V of the Nematoda, on the basis of small subunit rRNA sequence analysis [48].

A more objective technique for visualizing the evolutionary relationships between species for which large datasets are available is to use SimiTri, which plots in two-dimensional space the relative similarities of gene sequences between one species (*N. brasiliensis*) and three comparators [54]. As shown in Figure 1a, *N. brasiliensis* sequences group slightly closer to *Haemonchus* than to *Ancylostoma*, consistent with the relationship described above. Likewise, in Figure 1b, *N. brasiliensis* sequences group more towards *Teladorsagia* than *Necator*.

A compilation of the *N. brasiliensis* clusters, for which assigned homologs exist in protein databases, is presented in Table 2. Many sequences with high similarities to biosynthetic, structural, signaling and regulatory pathway proteins can readily be identified, corresponding to predicted nuclear or cytoplasmic proteins. Interestingly, multiple clusters encode categories of genes which are prominent in other nematode parasites, such as the five clusters encoding homologs of *Ancylostoma* secreted protein [2], five clusters of C-type and S-type lectins [55] and seven clusters for cysteine proteinases [56].

Proteins bearing signal sequences are less evolutionarily conserved

The set of 742 clusters was then divided into three categories according to their similarity to existing database sequences. 'Conserved' genes were defined as those with similarities to any non-nematode database entry above a given cutoff score; 'nematode-specific' genes were similar only to sequences from *C. elegans* or other nematode species, and 'novel' showed no similarity to any existing entry. BLASTX cutoff scores of 50 (P < e^{-6}) and 80 (P < e^{-10}) were both used to define these categories at different levels. Using the more stringent criterion, roughly one third (27-37%) of clusters fell into each category (Figure 2a), while the lower cutoff resulted in approximately half (48%) being classified as conserved, with the remainder evenly divided between nematode-specific (25%) and novel (27%).

The distribution of clusters containing signal sequences was, however, remarkably skewed towards the novel category. Because the primary classification of 92 novel genes was
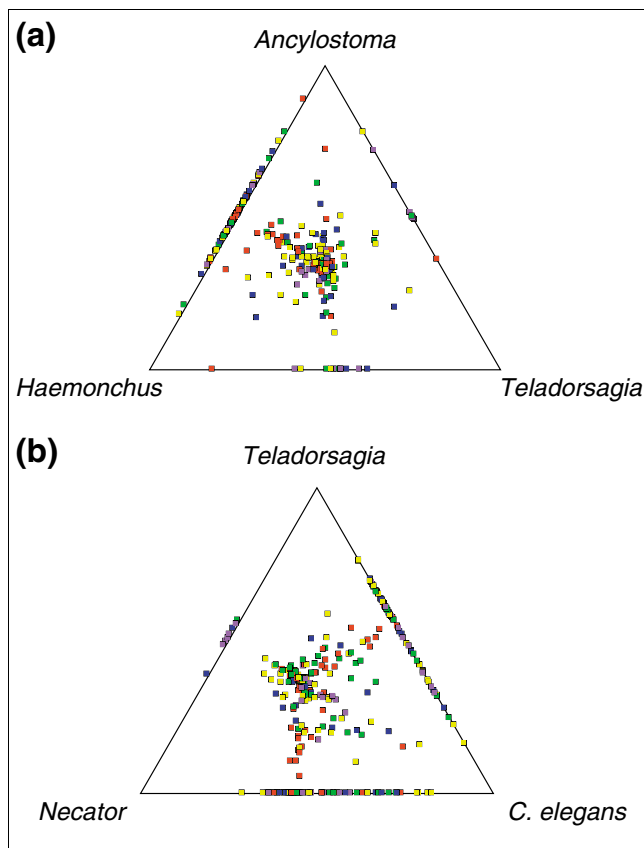
(a)

Ancylostoma

Haemonchus   *Teladorsagia*

(b)

Teladorsagia

Necator   *C. elegans*

Figure 1
Similarity of *N. brasiliensis* ESTs to sequences from other nematodes.
SimiTri [54] was used to plot 736 *N. brasiliensis* EST clusters against related
species database entries. For each consensus sequence associated with the
736 Nippo clusters, a BLAST was performed against a series of different
databases. Each tile in the graphic represents a unique consensus sequence
and its relative position is computed from the raw BLAST scores derived
above (with a cutoff of ≥ 50). Hence each tile's position shows its degree
of sequence similarity to each of the three selected databases. Sequences
showing similarity to only one database are not shown. Sequences showing
sequence similarity to only two databases appear on the lines joining the
two databases. Tiles are colored by their highest TBLASTX score to each
of the databases: red ≥ 300; yellow ≥ 200; green ≥ 150, blue ≥ 100 and
purple < 100. **(a)** SimiTri plot showing sequence similarity relationships
between *N. brasiliensis* consensus sequences and database entries of
Ancylostoma caninum/duodenale ESTs (20,177 entries, 386 hits),
Haemonchus contortus ESTs (22,337 entries, 384 hits) and *Teladorsagia
circumcincta* ESTs (5,300 entries, 264 hits). Database comparisons were
performed using TBLASTX. **(b)** SimiTri plot showing sequence similarity
relationships between *N. brasiliensis* consensus sequences and database
entries of *Necator americanus* ESTs (4,821 entries, 244 hits), *Teladorsagia
circumcincta* ESTs (5,300 entries, 264 hits), and *C. elegans* wormpep (21,600
entries, 466 hits). Database comparisons were performed using TBLASTX
for *N. americanus* and *T. circumcincta*, while *C. elegans* wormpep
comparions used BLASTX.

based on 5' EST sequences, all clusters initially designated as
novel signal-sequence positive were further scrutinized. In 72
cases, clusters read through to a 3' poly(A) tail (either single
reads from clones of 700 or fewer nucleotides or overlapping
ESTs with at least one poly(A) tail present); in 20 cases, where
no poly(A) tail was observed, 3' sequencing was carried out.

Of these, three showed database homologies from 3' sequence
and were reclassified as conserved, and two showed no
poly(A) tail and were excluded from further analysis as pre-
sumed internal fragments. The remaining 15 clusters showed
overlap between 3' and 5' cluster reads, without revealing any
additional similarities. Thus, a total of 87 clusters were veri-
fied as novel signal-sequence positive.

Taking this more rigorously defined subset, some 65% (87/
133) of sequences are predicted to encode either signal pep-
tides or signal anchors when classified as novel at the higher
cutoff (49% at the lower level), and only 4% were found in the
conserved category (7% at the lower cutoff). Moreover, 32% of
all novel sequences contained a signal peptide or anchor,
compared to 18% of nematode-specific and only 3.4% of
conserved.

Although the latter category will include many structural and
housekeeping proteins for which secretion is unlikely to con-
fer a selective advantage, the data suggest that nematode
secreted proteins have diversified more rapidly than those
that do not enter the secretory pathway.

This association between signal peptides and novel proteins
may be falsely amplified where, for example, conserved
domains are sufficiently distant from the amino terminus to
have been omitted from EST sequences. Equally, some clones
will have been sequenced from truncated transcripts, and a
proportion of those erroneously classified as encoding non-
signal sequence bearing proteins. However, neither of these
considerations seems likely to account for the very large dis-
parity in signal sequence frequency between the three catego-
ries we describe. A more general caveat with these analyses is
that SignalP is a fallible prediction tool, with an accuracy of
70% or less when applied to non-mammalian sequences [6].
There is no reason, however, to expect that false-positive
assignations would occur disproportionately in the novel
group rather than the conserved, and the conclusion drawn
here would remain valid over a wide range of prediction
accuracies.

**Has there been evolutionary acquisition of signal
peptides?**
The subset of signal-peptide-encoding *N. brasiliensis* clusters
with similarity to predicted genes from *C. elegans* with either
assigned function or of no known function was then identi-
fied. Examples of each category are given in Table 3. Some
nine clusters were identified as bearing signal-peptide
sequences, where in each case the *C. elegans* homologs
appear not to possess a signal-pepide motif. Five of these
clusters represent globins, which have previously been noted
to possess signal peptides in *N. brasiliensis* even though the
C. elegans paralogs do not [25,57]. One cluster (NBC00028)
is almost identical to the recorded cuticular isoform precursor
(P51536), but four additional clusters represent new mem-
bers of this family in *N. brasiliensis* bearing signal peptides.

Table 2

ESTs from adult cDNAs with known homologs, classified by function

Cluster number	Conventional cDNAs	Oligo-capped cDNAs	*P*	Accession	Description
Proteases/proteosome/ubiquitin					
NBC00018	2	0	1e-33	S66528	26S proteinase regulatory complex, non-ATPase chain (*Drosophila melanogaster*)
NBC00030	2	0	8e-56	U41556	Cysteine protease CPR-6 (*Caenorhabditis elegans*)
NBC00086	1	0	3e-29	A48454	Cathepsin B-like cysteine proteinase (*Ostertagia ostertagi*)
			5e-28	D48435	Cysteine proteinase AC-3 (*Haemonchus contortus*)
NBC00168	1	0	2e-42	NM_065563	Calpain thiol protease (*Caenorhabditis elegans*)
NBC00198	1	0	7e-60	NM_073736	Cysteine protease (*legumain, asparaginyl endopeptidase*) (*Caenorhabditis elegans*)
NBC00204	3	0	2e-32	NM_072733	Protease (*aspartic*) (*Caenorhabditis elegans*)
NBC00231	2	0	5e-90	NM_064106	Serine carboxypeptidase (*Caenorhabditis elegans*)
NBC00307	1	0	2e-32	NM_015277	Ubiquitin-protein ligase NEDD4-like; neural precursor (*Homo sapiens*)
NBC00311	1	0	5e-31	NM_073736	Cysteine protease (*legumain, asparaginyl endopeptidase*) (*Caenorhabditis elegans*)
NBC00352	2	0	6e-31	NM_065253	Ubiquitin (*Caenorhabditis elegans*)
NBC00348	1	0	2e-83	A48145	Ubiquitin-conjugating enzyme, UBC-2 (*Caenorhabditis elegans*)
NBC00362	1	0	1e-76	S17521	Multicatalytic endopeptidase complex (*proteasome*) zeta chain (*Caenorhabditis elegans*)
NBC00368	1	0	9e-13	LCE_ORYLA	Low choriolytic enzyme precursor (*zinc metalloprotease*) (*Oryzias latipes*)
NBC00377	1	0	3e-75	PSA4_CAEEL	Proteasome subunit, alpha type 4, PAS-3 (*Caenorhabditis elegans*)
NBC00459	2	1	2e-26	NM_072733	Protease (*aspartic*) (*Caenorhabditis elegans*)
NBC00469	1	0	7e-17	NM_060215	Zinc metalloprotease (*Caenorhabditis elegans*)
NBC00509	1	1	4e-71	AL161503	Polyubiquitin, UBQ10 (*Arabidopsis thaliana*)
NBC00664	0	1	5e-09	NM_074798	Cathepsin-like (*cysteine*) protease (*Caenorhabditis elegans*)
NBC00670	0	1	3e-18	S17435	Polyubiquitin 6 (*Helianthus annuus*)
NBC00772	0	1	4e-24	NM_003352	Sentrin, ubiquitin-like small protein (*Gallus gallus*)
NBC00783	0	1	2e-89	U41556	Cysteine protease CPR-6 (*Caenorhabditis elegans*)
NBC00828	0	1	9e-63	NC_003424	Pad1 protein; 26S proteasome subunit (*Schizosaccharomyces pombe*)
Enzymes (other than proteases)					
NBC00045	2	0	2e-92	NM_065870	Fructose-biphosphate aldolase (*Caenorhabditis elegans*)
NBC00049	1	0	9e-50	NM_070783	Lipase (*Caenorhabditis elegans*)
NBC00066	2	1	7e-76	NM_074348	Peptidyl-prolyl *cis-trans* isomerase (*Caenorhabditis elegans*)
NBC00079	1	0	2e-35	NM_058712	Helicase (*Caenorhabditis elegans*)
NBC00102	1	0	7e-37	NM_074031	Peroxidase-like (*Caenorhabditis elegans*)
NBC00139	1	0	8e-29	NM_060074	Hexokinase (*Caenorhabditis elegans*)
NBC00143	1	0	4e-66	ADHX_MYXGL	Alcohol dehydrogenase class III (*Caenorhabditis elegans*)
NBC00147	1	0	6e-19	XM_087230	Similar to Uridine phosphorylase (UDRPase) (*Homo sapiens*)
NBC00157	1	0	3e-13	XM_058660	Similar to Protein tyrosine phosphatase 1E (*Homo sapiens*)
NBC00173	1	0	5e-72	AJ440747	Protein disulphide isomerase 1 (*Ostertagia ostertagi*)
NBC00183	1	0	3e-56	T46280	Isocitrate dehydrogenase, NADP+, cytosolic (*Homo sapiens*)
NBC00189	1	0	1e-21	XM_129069	Similar to Acetyltransferase (GNAT) family (*Mus musculus*)
NBC00212	1	0	6e-57	NM_016100	N-terminal acetyltransferase complex ard1 subunit (*Homo sapiens*)
NBC00283	1	0	4e-27	NM_012088	6-phosphogluconolactonase (*Homo sapiens*)
NBC00285	1	0	2e-47	LDHA_ANGRO	L-lactate dehydrogenase A chain (*Anguilla rostrata*)
NBC00290	1	0	3e-17	I55976	Dihydrolipoamide S-acetyltransferase (*Rattus norvegicus*)

Table 2 *(Continued)*

ESTs from adult cDNAs with known homologs, classified by function

NBC00292	1	0	1e-40	NM_006223	Peptidyl-prolyl *cis/trans* isomerase (*Homo sapiens*)
NBC00304	1	0	4e-12	NM_073341	Glucose-1-dehydrogenase (*Caenorhabditis elegans*)
NBC00309	1	0	1e-18	NM_066225	Hydroxymethylglutaryl-coA reductase (*Caenorhabditis elegans*)
NBC00326	1	0	1e-65	NM_065761	Protein phosphatase 2A (*Caenorhabditis elegans*)
NBC00337	1	0	2e-60	GMD1_CAEEL	Probable GDP-mannose 4,6 dehydratase 1 (*Caenorhabditis elegans*)
NBC00353	1	0	2e-56	NM_065537	ATP synthase B chain (*Caenorhabditis elegans*)
NBC00378	1	0	2e-43	NM_073253	Acetyltransferase (GNAT) family (*Caenorhabditis elegans*)
NBC00382	1	0	4e-49	NM_063827	Phospholipase A2 (*Caenorhabditis elegans*)
NBC00389	2	0	1e-48	NM_058626	Phosphotransferase (*Caenorhabditis elegans*)
NBC00404	1	0	2e-76	NM_064078	Glucosamine-fructose-6-phosphate aminotransferase (*Caenorhabditis elegans*)
NBC00413	1	0	6e-22	NM_078324	AMP-activated protein kinase (*Caenorhabditis elegans*)
NBC00427	1	0	2e-20	NC_003423	3-oxoacyl-(acyl-carrier-protein)-synthase (*Schizosaccharomyces pombe*)
NBC00475	1	0	3e-42	NM_065313	Serine/threonine protein phosphatase (*Caenorhabditis elegans*)
NBC00483	1	0	4e-25	NM_059984	Phospholipase, similar to ADRAB-b (*Caenorhabditis elegans*)
NBC00504	1	0	7e-65	AF292096	Protein kinase AIRK2 (*Xenopus laevis*)
NBC00508	1	2	5e-64	PPCK_HAECO	Phosphoenolpyruvate carboxykinase (*Haemonchus contortus*)
NBC00528	1	0	5e-66	PPCK_HAECO	Phosphoenolpyruvate carboxykinase (*Haemonchus contortus*)
NBC00561	0	7	1e-54	NDKB_RAT	Nucleoside diphosphate kinase B (*Rattus norvegicus*)
NBC00713	0	1	1e-08	XM_140038	Similar to tau-tubulin kinase (*Mus musculus*)
NBC00729	0	2	4e-21	NM_079041	Flap endonuclease 1 (*Drosophila melanogaster*)
NBC00743	0	1	3e-64	G3P_BRUMA	Glyceraldehyde 3-phosphate dehydrogenase (*Brugia malayi*)
NBC00745	0	1	4e-13	NM_068436	Casein kinase (*Caenorhabditis elegans*)
NBC00689	0	3	2e-17	CLYC_CAEEL	Serine hydroxymethyltransferase MEL-32 (*Caenorhabditis elegans*)
NBC00696	0	2	2e-15	NM_000414	Hydroxysteroid (17-beta) dehydrogenase 4 (*Homo sapiens*)
NBC00770	0	1	3e-45	NM_066907	Serine/threonine kinase, casein kinase-like (*Caenorhabditis elegans*)
NBC00777	0	1	8e-21	OAZ_PRIPA	Ornithine decarboxylase antizyme (*Pristionchus pacificus*)
NBC00796	0	1	8e-52	XM_125017)	Putative lysophosphatidic acid acyltransferase (*Mus musculus*)
NBC00802	0	1	4e-49	NM_078623	Enoyl Coenzyme A hydratase, short chain 1 (*Rattus norvegicus*)

Structural

NBC00056	1	0	4e-58	NM_071024	Actin depolymerizing factor (*Caenorhabditis elegans*)
NBC00062	1	0	1e-11	NM_006400	Dynactin 2; dynactin complex 50 kD subunit; dynamitin (*Homo sapiens*)
NBC00078	2	0	0	NM_059538	Calponin (*Caenorhabditis elegans*)
NBC00097	1	0	1e-42	MLR1_CAEEL	Myosin regulatory light chain 1 (*Caenorhabditis elegans*)
NBC00142	1	0	2e-76	S53776	Beta-tubulin isotype 1 (*Haemonchus contortus*)
NBC00172	2	0	0	NM_073416	Actin (*Caenorhabditis elegans*)
NBC00224	1	0	2e-40	NM_063850	Troponin C (*Caenorhabditis elegans*)
NBC00239	4	1	2e-39	NM_077559	Collagen (*Caenorhabditis elegans*)
NBC00241	2	0	2e-47	NM_069715	Collagen (*Caenorhabditis elegans*)
			6e-47	NM_077291	Cuticular collagen (*Caenorhabditis elegans*)
NBC00246	1	1	3e-19	NM_077087	Troponin I (*Caenorhabditis elegans*)
NBC00287	2	0	2e-61	MLR1_CAEEL	Myosin regulatory light chain 1 (*Caenorhabditis elegans*)
NBC00360	1	1	3e-30	NM_145671	Actinfilin (*Rattus norvegicus*)
NBC00396	1	0	2e-67	MYSP_CAEEL	Paramyosin (*Caenorhabditis elegans*)
NBC00403	1	0	3e-32	NM_077291	Cuticular collagen (*Caenorhabditis elegans*)
NBC00418	1	0	6e-27	NM058881	Calponin (*Caenorhabditis elegans*)
NBC00430	1	0	3e-11	NM_011722	Dynactin 6; p27 dynactin subunit (*Mus musculus*)

Table 2 *(Continued)*

ESTs from adult cDNAs with known homologs, classified by function						
NBC00526	1	0	2e-44	NM_060857	Profilin (*Caenorhabditis elegans*)	
NBC00552	0	1	9e-47	MYSP_CAEEL	Paramyosin (*Caenorhabditis elegans*)	
NBC00569	0	1	1e-23	NM_060369	Alpha crystallin B chain (*Caenorhabditis elegans*)	
NBC00749	0	1	3e-43	NM_060857	Profilin (*Caenorhabditis elegans*)	
Embryo/egg/mating etc						
NBC00068	3	0	1e-25	VIT5_CAEEL	Vitellogenin 5 precursor (*Caenorhabditis elegans*)	
NBC00161	1	0	2e-15	VIT5_CAEEL	Vitellogenin 5 precursor (*Caenorhabditis elegans*)	
NBC00397	1	9	7e-61	MS10_CAEEL	Major Sperm Protein 10 (*Caenorhabditis elegans*)	
NBC00523	1	0	4e-69	XM_038960	Similar to preimplantation protein 3 (*Homo sapiens*)	
NBC00585	0	5	2e-30	NM_076467	Vitellogenin (*Caenorhabditis elegans*)	
NBC00611	0	1	1e-25	NM_060189	Placental protein 11 (*Caenorhabditis elegans*)	
Transporters/receptors/lectins and other binding proteins						
NBC00027	2	0	9e-17	NM_062882	Lectin, *C-type* (*Caenorhabditis elegans*)	
			5e-15	NM_076712	Asialoglycoprotein receptor (*C-type lectin*) (*Caenorhabditis elegans*)	
NBC00110	1	0	4e-17	NC_001263	Acyl-CoA-binding protein (*Deinococcus radiodurans*)	
NBC00118	1	0	4e-41	T31073	Multidrug resistance P-glycoprotein (*Haemonchus contortus*)	
NBC00128	3	0	1e-92	NM_067381	ADP/ATP carrier protein/translocase (*Caenorhabditis elegans*)	
NBC00167	1	0	2e-12	NM_130415	Lysosomal amino acid transporter 1 (*Rattus norvegicus*)	
NBC00175	1	0	7e-15	A48925	Mannose receptor (*C-type lectin*), macrophage (*Mus musculus*)	
NBC00319	1	0	8e-15	NXT2_HUMAN	NTF2-related export protein 2 (p15-2 protein) (*Homo sapiens*)	
NBC00324	2	0	7e-15	AJ243873	Galectin (*S-type lectin*) (*Haemonchus contortus*)	
NBC00340	1	0	2e-61	NM_077246	Galectin (*S-type lectin*) LEC-10 (*Caenorhabditis elegans*)	
NBC00355	1	0	8e-21	NM_059527	Fatty acid-binding protein LBP-6 (*Caenorhabditis elegans*)	
NBC00363	1	0	6e-48	NM_016208	Vacuolar protein sorting 28 homolog (*Homo sapiens*)	
NBC00583	0	5	4e-35	NM_065836	Low density lipoprotein receptor (*Caenorhabditis elegans*)	
NBC00593	0	2	2e-26	NM_059525	Fatty acid-binding protein LBP-6 (*Caenorhabditis elegans*)	
NBC00752	0	1	3e-08	NM_059071	Acetylcholine receptor UNV-38 (*Caenorhabditis elegans*)	
NBC00766	0	1	7e-44	POR2_MELGA	Voltage-dependent anion-selective channel protein 2 (VDAC-2) (*Meleagris gallopavo*)	
NBC00808	0	1	6e-53	NM_072174	Calreticulin precursor (*Caenorhabditis elegans*)	
NBC00838	0	1	1e-78	NM_063349	T-complex protein, delta subunit (cytosolic chaperonin CCT-4) (*Caenorhabditis elegans*)	
Signaling						
NBC00207	1	0	0	RAB2_LYMST	RAS-Related protein RAB-2 (*Lymnea stagnalis*)	
NBC00252	1	0	8e-97	NM_070558	RAS-like GTP-binding protein RhoA (*Caenorhabditis elegans*)	
NBC00297	1	0	4e-17	NM_009106	Rhotekin (*Mus musculus*)	
NBC00312	1	0	4e-46	A35350	Protein kinase C inhibitor (*Bos bovis*)	
NBC00269	1	0	1e-43	NM_058274	RAS-related protein RAB-11 (*Caenorhabditis elegans*)	
NBC00282	1	0	9e-25	NP_741191	A kinase anchor protein 1 (*Caenorhabditis elegans*)	
NBC00395	1	0	2e-29	NM_07328	RAS-like GTP-binding protein (*cdc42-like*) (*Caenorhabditis elegans*)	
NBC00436	1	0	2e-44	NM_070985	Calmodulin (*Caenorhabditis elegans*)	
NBC00462	1	0	2e-13	SSRP_DROME	Single-strand recognition protein (SSRP) (Chorion-factor 5) (*Drosophila melanogaster*)	
NBC00409	1	0	1e-16	NM_019746	Programmed cell death 5/TFAR19 protein (*Mus musculus*)	
NBC00440	1	0	3e-72	S43599	SNF5 homolog R07E5.3 (*Caenorhabditis elegans*)	
NBC00510	1	0	2e-28	XM_129572	Calcyclin (*S100 family*) binding protein (*Mus musculus*)	
NBC00629	0	1	1e-20	NM_026297	RAB (RAS oncogene family-like 3) (*Mus musculus*)	

Table 2 *(Continued)*

ESTs from adult cDNAs with known homologs, classified by function

NBC00648	0	1	3e-20	NM_002624	Prefoldin 5 isoform alpha; myc modulator-1; c-myc binding protein (*Homo sapiens*)
NBC00727	0	1	3e-17	AB091687	TGF-beta induced apotosis protein 3 (*Mus musculus*)
NBC00768	0	1	3e-18	NM_078471	TGF-beta-1 induced anti-apoptotic factor 1 isoform 1 (*Homo sapiens*)
NBC00829	0	1	1e-42	A49146	Developmental regulator WNT-4 (*Xenopus laevis*)
NBC00841	0	1	1e-31	NM_012453	Transducin (beta)-like 2, isoform 1 (*Homo sapiens*)

DNA-related/transcription/DNA binding/regulation

NBC00024	1	0	1e-37	NM_003752	Eukaryotic translation initiation factor 3, subunit 8 (*Homo sapiens*)
NBC00048	1	0	1e-28	NM_069150	Glycine-rich RNA-binding protein (*Caenorhabditis elegans*)
			5e-21	NM_007007	Cleavage and polyadenylation specific factor 6 (*Homo sapiens*)
NBC00050	1	0	2e-12	HEXP_LEIMA	DNA-binding protein HEXBP (Hexamer-binding protein) (*Leishmania major*)
NBC00055	1	1	2e-24	NM_060622	RNA recognition motif (RRM, RBD, or RNP domain) (*Caenorhabditis elegans*)
NBC00090	2	1	0	NM_066119	Elongation factor 1-alpha (*Caenorhabditis elegans*)
NBC00099	1	0	2e-30	NM_067248	Splicing factor (*Caenorhabditis elegans*)
NBC00170	1	0	2e-56	NM_011304	RuvB DNA helicase -like protein 2 (*Mus musculus*)
NBC00181	1	0	4e-13	NM_001698	AU RNA-binding protein/enoyl-Coenzyme A hydratase (*Homo sapiens*)
NBC00192	1	0	2e-26	NM_060622	RNA recognition motif (RRM, RBD, or RNP domain) (*Caenorhabditis elegans*)
NBC00210	1	0	3e-15	NM_018403	Transcription factor (SMIF gene) (*Homo sapiens*)
NBC00267	1	0	4e-20	T2EB_XENLA	Transcription initiation factor IIE, beta subunit (*Xenopus laevis*)
NBC00321	1	0	1e-16	NM_033224	Purine-rich element binding protein B (*Homo sapiens*)
NBC00280	1	0	3e-58	NM_006578	Guanine nucleotide-binding protein, beta-5 subunit (*Homo sapiens*)
NBC00350	1	0	6e-40	DPOD_DROME	DNA polymerase delta catalytic subunit (*Drosophila melanogaster*)
NBC00366	2	0	6e-79	NM_066119	Elongation factor 1-alpha (*Caenorhabditis elegans*)
NBC00370	1	0	1e-17	NM_031992	Eukaryotic translation initiation factor 4H, isoform 2 (*Homo sapiens*)
NBC00374	1	2	2e-53	NM_070415	Elongation factor 1-beta/delta chain (*Caenorhabditis elegans*)
NBC00480	1	0	3e-21	NM_061014	Regulator of chromosome condensation, RCC1 (*Caenorhabditis elegans*)
NBC00543	0	2	5e-23	NM_065536	Zinc finger, C3HC4 type (RING finger) (*Caenorhabditis elegans*)
NBC00577	0	7	2e-31	NP_872244	Translation elongation factor EFT-4 (*Caenorhabditis elegans*)
NBC00600	0	1	3e-74	NM_063406	Initiation factor 5A (*Caenorhabditis elegans*)
NBC00630	0	1	9e-39	SFR4_MOUSE	Splicing factor, arginine/serine-rich 4 (*Mus musculus*)
NBC00764	0	1	4e-16	XM_132357	Similar to Translation Initiation factor EIF-2B alpha (*Mus musculus*)
NBC00776	0	1	6e-27	SN2L_CAEEL	Potential global transcription activator SNF2L (*Caenorhabditis elegans*)
NBC00791	0	1	5e-38	NM_001207	Basic transcription factor 3 (*Homo sapiens*)
NBC00816	0	1	2e-24	S3B2_HUMAN	Splicing factor 3B subunit 2 (Spliceosome associated protein 145) (*Homo sapiens*)

Other homologs of interest

NBC00025	1	0	3e-16	AF352714	HC40 putative secretory protein precursor (ASP homolog) (*Haemonchus contortus*)
NBC00065	1	0	6e-20	AA063577	Secreted protein 5 precursor (ASP homolog) (*Ancylostoma caninum*)
NBC00095	1	0	8e-59	GLB2_NIPBR	Myoglobin (body wall isoform globin) (*Nippostrongylus brasiliensis*)
NBC00103	1	0	9e-12	DIM1_CAEEL	Protein dim-1 (2D-page protein spot 8) (*Caenorhabditis elegans*)

Table 2 (*Continued*)

ESTs from adult cDNAs with known homologs, classified by function

NBC00029	1	0	5e-17	NM_001545	Immature colon carcinoma transcript 1 (*Homo sapiens*)
NBC00141	1	0	2e-35	NM_018984	Slingshot 1 (*Homo sapiens*)
NBC00160	1	0	5e-12	NM_053810	Synaptosomal-associated protein, 29kD (*Rattus norvegicus*)
NBC00199	1	0	9e-39	AF278538	Nucleosome assembly protein 1 (*Xenopus laevis*)
NBC00256	2	0	2e-09	NM_075227	Transthyretin-like family (*Caenorhabditis elegans*)
NBC00293	1	0	7e-08	NC_003424	F-box protein (*Schizosaccharomyces pombe*)
NBC00399	1	0	2e-22	NM_076443	Calumenin, calcium-binding protein (*Caenorhabditis elegans*)
NBC00429	1	0	4e-14	XM_122362	Chromobox homolog 2 (*Drosophila* Pc class) (*Mus musculus*)
NBC00491	1	0	3e-21	NM_076885	Thrombospondin (*Caenorhabditis elegans*)
NBC00518	1	0	3e-73	T37461	Mago nashi-like protein (*Caenorhabditis elegans*)
NBC00544	0	1	2e-45	NM_061213	Alpha-2-macroglobulin family (*Caenorhabditis elegans*)
NBC00560	0	1	1e-35	NM_021305	SEC61, alpha subunit 2 (*Saccharomyces cerevisiae*)
NBC00705	0	1	3e-31	DVA1_DICVI	DVA-1 nematode polyprotein allergen precursor (NPA) (*Dictyocaulus viviparus*)
			2e-12	ABA1_ASCSU	ABA-1 nematode polyprotein allergen precursor (Body fluid allergen-1) (*Ascaris suum*)
NBC00753	0	1	4e-10	AF089728	Ancylostoma-secreted protein 2 precursor, ASP-2 (*Ancylostoma caninum*)
NBC00755	0	1	2e-40	TCPB_CAEEL	T-complex protein 1, beta subunit (CCT-beta) (*Caenorhabditis elegans*)
NBC00757	0	1	2e-68	1432_SCHMA	14-3-3 Protein homolog 2 (14-3-3-2) (*Schistosoma mansoni*)
NBC00803	0	1	3e-09	ASP_ANCCA	Ancylostoma secreted protein (*ASP-1*) precursor (*Ancylostoma caninum*)
			3e-09	AF079521	Ancylostoma-secreted protein 1 precursor (*ASP-1 homolog*) (*Necator americanus*)
NBC00827	0	1	3e-14	NM_070108	Testis-specific protein TPX-1 like (*ASP homolog*) (*Caenorhabditis elegans*)

The table gives, for each numbered cluster, the highest homolog with a functional description where available; in a number of cases a *C. elegans* homolog exists with a higher similarity, but has no description. Similarities to entries described as 'hypothetical proteins' are excluded, as are heat-shock proteins, cytochromes, mitochondrial and ribosomal products. Where *C. elegans* protein description is ambiguous (for example, protease, lectin), further descriptors added manually are italicized. Different clusters may derive from a single gene if sequences are non-overlapping; for example, NBC00198 and NBC00311 align to different segments of the *C. elegans* protease gene NM_073736. This table does not include *N. brasiliensis* gene products discovered previously and/or reported by other laboratories. All entries for this species are aggregated on the NEMBASE website.

In contrast, a distinct globin (NBC00095) closely related to the known body-wall isoform (P51535) lacks a predicted signal peptide. Hence, gene duplication may have predated the development in some globin forms, of a secretory function.

In these cases, and in the four additional examples given in Table 3, it is possible that pre-existing genes have been adapted for secretion or membrane expression in order to promote parasitism. Acquisition of secretory signals may not, in evolutionary terms, be demanding, in view of the report that approximately 20% of protein-coding fragments from *Saccharomyces cerevisiae* can function as a signal peptide [58]. In the case of the globins, conversion to the secretory pathway (as well as gene multiplication) may be interpreted as a physiological adaptation to the environment within the

mammalian gastrointestinal tract [57]. Whether any of the four remaining genes in this category might have undergone a similar evolutionary process to counter immune attack is unknown at this stage.

Similar findings have previously been reported in individual genes from other nematode parasites. In *B. malayi*, the microfilarial secreted serpin gene (*Bm-spn-2*) is homologous to eight *C. elegans* genes, none of which encodes a signal peptide [59]. Likewise, the extracellular glutathione-*S*-transferase gene, *Ov-gst-1*, of *Onchocerca volvulus* has acquired a signal-peptide sequence [60], as has a gene for keratin-like protein (KLP) in *N. brasiliensis* itself [31]. Hence, conversion of key gene products to secretory function may be a common adaptive strategy for parasitic organisms.
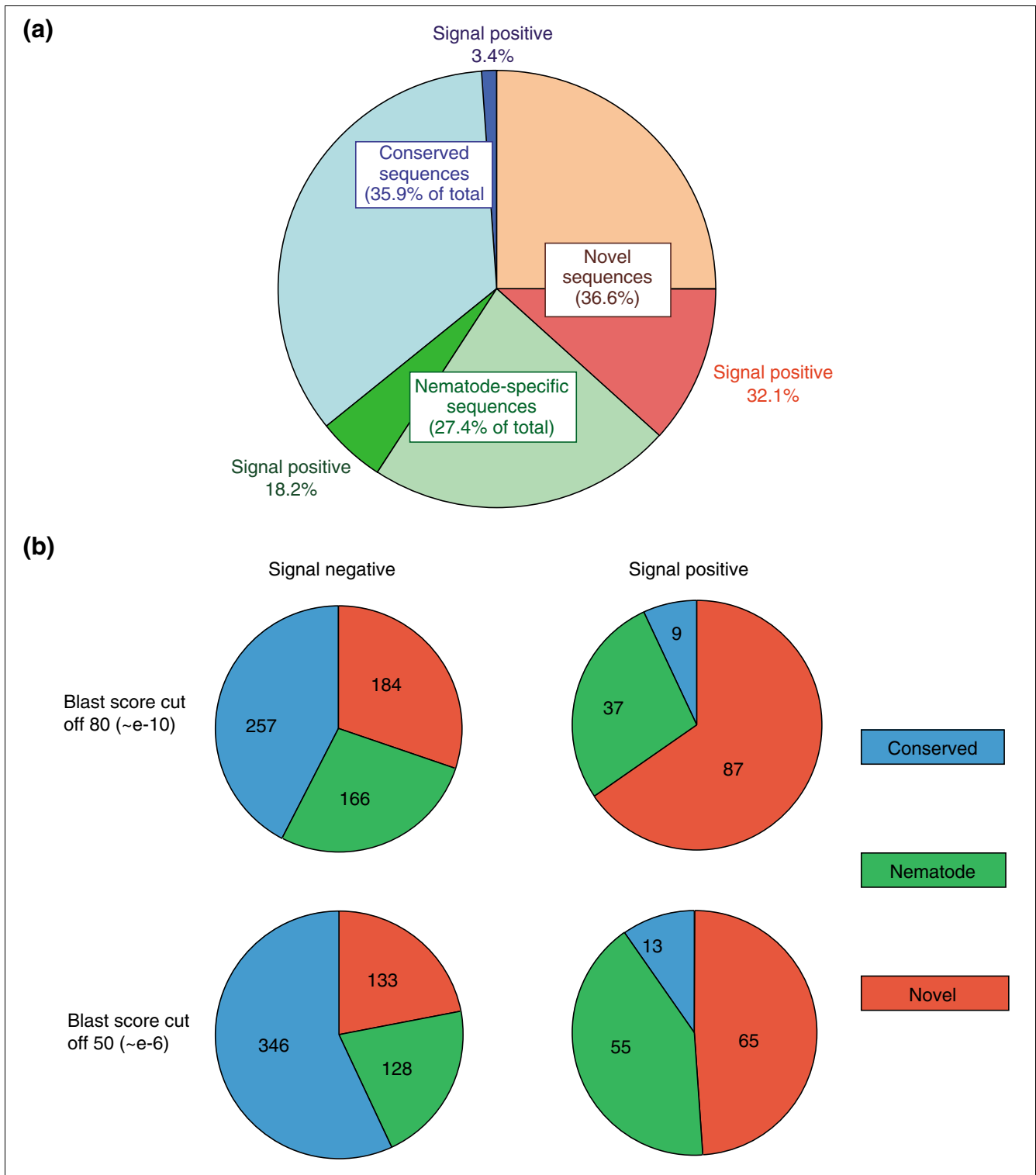
Figure 2
Proportion of ESTs predicted to encode signal sequences. **(a)** EST sequences were classified as conserved (similarities to non-nematode database entries), nematode-specific (similarities only to *C. elegans* or other nematode sequences), or novel (no similarities to existing entries), using a cutoff score of 80 in BLASTX ($P < e^{-10}$). The number of ESTs bearing potential signal sequences was then calculated and the results are shown here. **(b)** Effects of relaxing cutoff scores on distribution of signal peptide-containing predicted gene products among conserved, nematode-specific and novel categories. Numbers of clusters in each category are given for cutoffs of 80 ($P <e^{-10}$), as used in (a), and 50 ($P <e^{-6}$).

Table 3

ESTs from adult cDNAs with predicted amino-terminal signal peptides and with homologs in *C. elegans*

Cluster	Score	*P*	Con- ventional cDNAs	Oligo- capped cDNAs	Worm- pep ID	SignalP criteria	SignalP scores				Signal in *C. elegans*?	Description of *C. elegans* gene
							C-p	Amino acids	SP-p	SP?		
(a) Signal peptides predicted in both *N. brasiliensis* and *C. elegans*												
NBC00012	86	6e-18	4	0	CE20223	YYYYS	0.533	16	1.000	Y	Y	Unknown (similar to NBC00237)
NBC00031	80	3e-16	2	2	CE17924	YYYYS	0.932	18	0.999	Y	Y	Unknown
NBC00237	84	5e-17	1	2	CE20223	YYYYS	0.671	19	1.000	Y	Y	Unknown (similar to NBC00012)
NBC00258	145	1e-35	1	0	CE00133	YYYYS	0.524	19	0.999	Y	Y	FAR-1 fatty acid/ retinol-binding protein
NBC00266	129	6e-31	1	0	CE19630	YYYYS	0.662	20	1.000	Y	Y	Unknown
NBC00314	147	3e-36	1	1	CE03639	YYYYS	0.708	19	0.987	Y	Y	Transthyretin-like family
NBC00327	94	2e-20	1	0	CE00906	YYYYS	0.542	25	0.998	Y	Y	Unknown
NBC00336	138	2e-33	1	0	CE23545	YYYYS	0.903	17	1.000	Y	Y	Unknown
NBC00354	91	4e-21	4	0	CE16530	YYYYS	0.511	17	0.943	Y	Y	Unknown
NBC00472	215	8e-57	1	0	CE04886	YYYYS	0.319	15	0.999	Y	Y	Signal sequence receptor
NBC00487	55	7e-09	1	0	CE05972	YYYYS	0.979	21	0.988	Y	Y	Unknown
NBC00495	51	3e-07	1	1	CE13171	YYYYS	0.566	19	0.999	Y	Y	Transthyretin-like family
NBC00502	176	3e-45	1	0	CE32298	YYYYS	0.634	20	1.000	Y	Y	Ectonucleotide pyrophosphatase/ phosphodiesterase
NBC00592	80	1e-15	0	3	CE17924	YYYYS	0.920	16	1.000	Y	Y	Unknown
NBC00606	81	4e-16	0	2	CE02454	YYYYS	0.399	20	1.000	Y	Y	Similar to *O. volvulus* hypodermal antigen Ov-17
NBC00615	207	3e-54	0	1	CE04533	YYYYS	0.995	18	1.000	Y	Y	LBP-1 fatty acid- binding protein
NBC00616	61	3e-10	0	1	CE20257	YYYYS	0.754	19	0.993	Y	Y	Unknown
NBC00633	153	4e-38	0	1	CE03639	YYYYS	0.450	17	1.000	Y	Y	Transthyretin-like family
NBC00641	145	1e -35	0	1	CE33289	YYYYS	0.219	19	0.930	Y	Y	Unknown
NBC00643	102	2e-22	0	2	CE27850	YYYYS	0.961	17	0.999	Y	Y	Unknown
NBC00706	50	9e-07	0	1	CE06014	YYYYS	0.466	20	1.000	Y	Y	Unknown
NBC00720	12	3e-30	0	1	CE16958	YYYYS	0.967	19	0.998	Y	Y	NLP-13 neuropeptide
NBC00742	60	3e-10	0	1	CE16731	YYYYS	0.880	21	0.993	Y	Y	Unknown
NBC00748	50	4e-07	0	1	CE02932	YYYYS	0.804	17	0.998	Y	Y	Transthyretin-like family
NBC00767	79	7e-16	0	1	CE31662	YYYYS	0.559	17	1.000	Y	Y	Unknown

Table 3 *(Continued)*

ESTs from adult cDNAs with predicted amino-terminal signal peptides and with homologs in *C. elegans*

(b) Signal peptides predicted in *N. brasiliensis* but not *C. elegans*

NBC00028	104	1e-23	1	1	CE00431	YYYYS	0.731	18	0.999	Y	N	Globin
NBC00124	128	8e-31	1	1	CE00431	YYYYS	0.731	18	0.999	Y	N	Globin
NBC00144	195	7e-51	1	0	CE29663	YYNYS	0.866	19	0.963	Y	N	Transport-secretion protein
NBC00197	143	8e-35	3	6	CE00431	YYYYS	0.557	16	1.000	Y	N	Globin
NBC00272	144	2e-35	1	0	CE32475	YYNYS	0.262	22	0.513	Y	N	Unknown
NBC00328	147	4e-36	3	4	CE00431	YYYYS	0.523	17	0.999	Y	N	Globin
NBC00581	122	7e-29	0	1	CE00431	YYYYS	0.404	21	0.998	Y	N	Globin
NBC00601	93	5e-20	0	1	CE30218	YYYYS	0.535	34	0.944	Y	N	Unknown
NBC00607	159	4e-40	0	1	CE29597	YYNYS	0.529	18	0.786	Y	N	Unknown

Entries in table do not match numbers in Figure 2, which includes predicted signal anchors. SignalP criteria are C-score (raw cleavage site score); S-score (signal peptide score); Y-score (combined cleavage site score); mean S score; and assignation as signal peptide (S as in all entries above); otherwise A for signal anchor or N for neither). SignalP scores are as follows: C-p: probability of predicted cleavage site being correct; amino acids: length of predicted signal peptide in amino acids; SP-p: probability of existence of signal peptide; SP?: overall prediction for signal peptide. Note that NBC00028 is almost identical to the cuticular globin of *N. brasiliensis* (P51536), and NBC00197 and NBC00328 are closely related, whereas NBC0124 and NBC00581 are more similar to, but not identical to, the body-wall form of globin (P51535).

Conclusions

Our study raises both methodological and evolutionary questions. First, it remains to be determined how valid is the assumption that signal sequences reflect secretion into the parasite environment. Clearly, this notion must be qualified in a metazoan parasite, because many such proteins will remain on the cell surface or be sorted to extracellular and extracytosolic compartments within the worm. However, the extent to which signal-peptide-bearing proteins are truly exported by these multicellular organisms will be clarified by current proteomic analyses on proteins secreted by the same adult-stage parasites as were used to construct the cDNA libraries. The same studies will answer a further methodological caveat: proteins can be secreted by non-signal-sequence-dependent pathways, and we have no information on the extent to which parasites may avail themselves of this possibility. One example already exists, of the macrophage migration inhibitory factor homolog of *B. malayi* which is exported despite lacking a signal peptide [61,62].

On a broader platform, we have addressed the question of whether secreted proteins of parasitic nematodes show accelerated evolution, and our results indicate that this is the case. The predominance of predicted secreted proteins in the novel class prevents us, at this stage, from discerning whether rapid evolution was consequent upon acquiring secretory status, or if the more divergent gene products were those most advantageous to co-opt into secretion. Parallel studies on other parasitic nematodes would now clarify these and additional issues. Have genes for parasite secreted proteins indeed acquired signal peptides, or have free-living lineages lost these motifs in the genes in question? Is more rapid diversification of secreted proteins a specific feature of parasiticatodes, or can a similar phenomenon be observed in comparisons between divergent free-living organisms (such as *C. elegans* and *C. briggsae*)? These questions are now under study.

Materials and methods
Parasite material

N. brasiliensis was maintained in Sprague-Dawley rats as previously described [10,63]. For cDNA synthesis, adult worms were recovered from gastrointestinal contents 5 or 6 days following subcutaneous injection of 3,000 infective L3 larvae. Adults were recovered by Baermannization in saline at 37°C, washed 6 × in saline and 6 × in RPMI1640 containing 100 μg/ml penicillin and 100 U/ml streptomycin. Worms were incubated with 10% gentamicin for 20 min and then washed a further 6 × in RPMI1640 with antibiotics before immersion in Trizol for mRNA preparation.

cDNA libraries

Conventional libraries were constructed in Uni-Zap (Stratagene) and propagated in pBluescript SK+ from mixed adult worm mRNA as previously described [27]. To construct an oligo-capped cDNA library, the technique of Fernández [45] was followed. mRNA was isolated from 1 ml of packed adult *N. brasiliensis* (approximately 10,000 worms) homogenized in 10 ml Trizol (Gibco Life Technologies). The homogenate was centrifuged (12,000*g*, 10 min), and the supernatant extracted with chloroform before isopropanol precipitation of RNA from the aqueous phase. mRNA was then purified with PolyA Purist oligo-dT cellulose (Ambion). Following dephosphorylation with calf intestinal phosphatase, mRNA was treated with tobacco acid pyrophosphatase to remove the 7-methylguanosine terminal cap on full-length mRNAs, leaving these with a reactive phosphate group. These were then

adducted with the GeneRacer oligonucleotide (Invitrogen). Reverse transcription of mRNA was primed with a tagged oligo-dT (*Not*I primer-adapter). In this way, full-length transcripts contained specific extension sequences (5' Gene Racer and 3' oligo-dT tag) amenable to PCR amplification. Following PCR, products were ligated at both ends to *Sal*I adapters, so that subsequent digestion with *Not*I provided inserts with cohesive ends to be directionally cloned into *Not*I/*Sal*I-digested pSPORT1 vector.

EST sequencing
The library was used to transform DH10B *Escherichia coli* by electroporation, plated on ampicillin agar petri dishes, and colonies picked for sequencing. All colonies picked were grown overnight in 96-well plates, which were used to provide template samples for PCR before being directly archived. PCR reactions used M13 forward and reverse primers, and following shrimp alkaline phosphatase/exonuclease I treatment, products were directly sequenced with T7 primer on ABI automated sequencers. Archived clones are available on request from R.M.M. Where 3' sequencing was required, T3 primer was used.

Bioinformatics
Raw sequence trace data were processed to screen out vector and linking sequence, to remove low-quality sequence, and to trim poly(dA) tails using an in-house software solution. The resulting sequences were annotated with similarity information and library details and submitted to dbEST. To identify the nonredundant set of putative gene objects, sequences were clustered on the basis of sequence similarity using the CLOBB program [64]. Consensus sequences representing the putative gene objects were then generated from clusters containing more than one sequence using the assembly program phrap (Phil Green, University of Washington; available from [65]). Clusters containing only a single sequence ('singletons') and the consensuses generated from clusters containing more than one sequence ('clusters') were then subjected to the following BLAST analyses: BLASTN against a nonredundant DNA database (GenBank); BLASTX against a nonredundant protein database (SwissProt-trEMBL) and BLASTN against dbEST. Results from these analyses are available from our online database - NEMBASE [49]. Peptide predictions were performed on individual sequences using the program DEcoder [66]. Where DEcoder was unable to predict a peptide, ESTscan [67] was used. SignalP V2.0 [6] was used to predict the presence of secretory signal peptides and signal anchors for each of the predicted proteins. Peptides were defined as bearing a signal peptide if both the hidden Markov model (HMM) predicted the presence of a secretory leader and three of the four parameters defined by the neural network model (C-score, Y-score, S-score and S-mean, as described in legend to Table 3) were fulfilled. Signal anchors were predicted if both the HMM predicted a signal anchor and two of the four criteria specified by the neural network model were fulfilled. Selected clones were subject to compar-

ative analysis with database entries from *C. elegans* and other species. Alignments were made using Clustal X within MacVector 7.0 (Oxford Molecular) and the SignalP V2.0 web server [68] was used to chart hydrophobicity and potential cleavage sites in predicted protein sequences.

Cross-taxon similarity analysis
The relative similarity between *N. brasiliensis* EST sequences and those from the related parasitic nematodes *Ancylostoma caninum/duodenale*, *Haemonchus contortus* and *Teladorsagia circumcincta* were plotted with the SimiTri program [54], downloadable from [69].

References
1. Lightowlers MW, Rickard MD: **Excretory-secretory products of helminth parasites: effects on host immune responses.** *Parasitology* 1988, **96:**S123-S166.
2. Hawdon JM, Jones BF, Hoffman DR, Hotez PJ: **Cloning and characterization of *Ancylostoma*-secreted protein. A novel protein associated with the transition to parasitism by infective hookworm larvae.** *J Biol Chem* 1996, **271:**6672-6678.
3. Maizels RM, Gomez-Escobar N, Gregory WF, Murray J, Zang X: **Immune evasion genes from filarial nematodes.** *Int J Parasitol* 2001, **31:**889-898.
4. Yatsuda AP, Krijgsveld J, Cornelissen AWCA, Heck AJ, De Vries E: **Comprehensive analysis of the secreted proteins of the parasite *Haemonchus contortus* reveals extensive sequence variation and differential immune recognition.** *J Biol Chem* 2003, **278:**16941-16951.
5. von Heijne G: **A new method for predicting signal sequence cleavage sites.** *Nucleic Acids Res* 1986, **14:**4683-4690.
6. Nielsen H, Engelbrecht J, Brunak S, von Heijne G: **Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites.** *Protein Eng* 1997, **10:**1-6.
7. Nielsen H, Brunak S, von Heijne G: **Machine learning approaches for the prediction of signal peptides and other protein sorting signals.** *Protein Eng* 1999, **12:**3-9.
8. Menne KM, Hermjakob H, Apweiler R: **A comparison of signal sequence prediction methods using a test set of signal peptides.** *Bioinformatics* 2000, **16:**741-742.
9. Chou KC: **Prediction of protein signal sequences.** *Curr Protein Pept Sci* 2002, **3:**615-622.
10. Maizels RM, Meghji M, Ogilvie BM: **Restricted sets of parasite antigens from the surface of different stages and sexes of the nematode *Nippostrongylus brasiliensis*.** *Immunology* 1983, **48:**107-121.
11. Finkelman FD, Shea-Donohue T, Goldhill J, Sullivan CA, Morris SC, Madden KB, Gause WC, Urban JF Jr: **Cytokine regulation of host defense against parasitic gastrointestinal nematodes: lessons from studies with rodent models.** *Annu Rev Immunol* 1997, **15:**505-533.
12. Maizels RM, Holland MJ: **Parasite immunity: pathways for expelling intestinal parasites.** *Curr Biol* 1998, **8:**R711-R714.
13. Maizels RM, Bundy DAP, Selkirk ME, Smith DF, Anderson RM: **Immunological modulation and evasion by helminth parasites in human populations.** *Nature* 1993, **365:**797-805.
14. MacDonald AS, Araujo MI, Pearce EJ: **Immunology of parasitic helminth infections.** *Infect Immun* 2002, **70:**427-433.
15. Maizels RM, Yazdanbakhsh M: **Regulation of the immune response by helminth parasites: cellular and molecular mechanisms.** *Nat Rev Immunol* 2003, **3:**733-743.
16. Urban JF Jr, Madden KB, Svetic A, Cheever A, Trotta PP, Gause WC, Katona IM, Finkelman FD: **The importance of Th2 cytokines in

protective immunity to nematodes. *Immunol Rev* 1992, **127:**205-220.

17. Holland MJ, Harcus YM, Riches PL, Maizels RM: **Proteins secreted by the parasitic nematode *Nippostrongylus brasiliensis* act as adjuvants for Th2 responses.** *Eur J Immunol* 2000, **30:**1977-1987.

18. Ogilvie BM, Rothwell TLW, Bremner KC, Schnitzerling HJ, Nolan J, Keith RK: **Acetylcholinesterase secretion by parasitic nematodes. I. Evidence for secretion by a number of species.** *Int J Parasitol* 1973, **3:**589-597.

19. Blackburn CC, Selkirk ME: **Characterisation of the secretory acetylcholinesterases from adult *Nippostrongylus brasiliensis*.** *Mol Biochem Parasitol* 1992, **53:**79-88.

20. Grigg ME, Tang L, Hussein AS, Selkirk ME: **Purification and properties of monomeric (G$_1$) forms of acetylcholinesterase secreted by *Nippostrongylus brasiliensis*.** *Mol Biochem Parasitol* 1997, **90:**513-524.

21. Healer J, Ashall F, Maizels RM: **Characterization of proteolytic enzymes from larval and adult *Nippostrongylus brasiliensis*.** *Parasitology* 1991, **103:**305-314.

22. Kamata I, Yamada M, Uchikawa R, Matsuda S, Arizono N: **Cysteine protease of the nematode *Nippostrongylus brasiliensis* preferentially evokes an IgE/IgG1 antibody response in rats.** *Clin Exp Immunol* 1995, **102:**71-77.

23. Blackburn CC, Selkirk ME: **Inactivation of platelet activating factor by a putative acetylhydrolase from the gastrointestinal nematode parasite *Nippostrongylus brasiliensis*.** *Immunology* 1992, **75:**41-46.

24. Grigg ME, Gounaris K, Selkirk ME: **Characterization of a platelet-activating factor acetylhydrolase secreted by the nematode parasite *Nippostrongylus brasiliensis*.** *Biochem J* 1996, **317:**541-547.

25. Blaxter ML, Ingram L, Tweedie S: **Sequence, expression and evolution of the globins of the parasitic nematode *Nippostrongylus brasiliensis*.** *Mol Biochem Parasitol* 1994, **68:**1-14.

26. Hussein A, Harel M, Selkirk M: **A distinct family of acetylcholinesterases is secreted by *Nippostrongylus brasiliensis*.** *Mol Biochem Parasitol* 2002, **123:**125-134.

27. Hussein AS, Chacón MR, Smith AM, Tosado-Acevedo R, Selkirk ME: **Cloning, expression, and properties of a nonneuronal secreted acetylcholinesterase from the parasitic nematode *Nippostrongylus brasiliensis*.** *J Biol Chem* 1999, **274:**9312-9319.

28. Hussein AS, Grigg ME, Selkirk ME: ***Nippostrongylus brasiliensis*: characterisation of a somatic amphiphilic acetylcholinesterase with properties distinct from the secreted enzymes.** *Exp Parasitol* 1999, **91:**144-150.

29. Dainichi T, Maekawa Y, Ishii K, Zhang T, Nashed BF, Sakai T, Takashima M, Himeno K: **Nippocystatin, a cysteine protease inhibitor from *Nippostrongylus brasiliensis*, inhibits antigen processing and modulates antigen-specific immune response.** *Infect Immun* 2001, **69:**7380-7386.

30. Tang L, Prichard RK: **Comparison of the properties of tubulin from *Nippostrongylus brasiliensis* with mammalian brain tubulin.** *Mol Biochem Parasitol* 1988, **29:**133-140.

31. Shibui A, Takamoto M, Shi Y, Komiyama A, Sugane K: **Cloning and characterization of a novel gene encoding keratin-like protein from nematode *Nippostrongylus brasiliensis*.** *Biochim Biophys Acta* 2001, **1522:**59-61.

32. Tweedie S, Grigg ME, Ingram L, Selkirk ME: **The expression of a small heat shock protein homologue is developmentally regulated in *Nippostrongylus brasiliensis*.** *Mol Biochem Parasitol* 1993, **61:**149-154.

33. Hussein AS, Kichenin K, Selkirk ME: **Suppression of secreted acetylcholinesterase expression in *Nippostrongylus brasiliensis* by RNA interference.** *Mol Biochem Parasitol* 2002, **122:**91-94.

34. Hammond MP, Bianco AE: **Genes and genomes of parasitic nematodes.** *Parasitol Today* 1992, **8:**299-305.

35. Muller R: *Worms and Human Disease* Wallingford, UK: CABI Publishing; 2002.

36. Parkinson J, Mitreva M, Hall N, Blaxter M, McCarter JP: **400,000 nematode ESTs on the Net.** *Trends Parasitol* 2003, **19:**283-286.

37. Williams SA, Lizotte-Waniewski MR, Foster J, Guiliano D, Daub J, Scott AL, Slatko B, Blaxter ML: **The filarial genome project: analysis of the nuclear, mitochondrial and endosymbiont genomes of *Brugia malayi*.** *Int J Parasitol* 2000, **30:**411-419.

38. Blaxter M, Daub J, Guiliano D, Parkinson J, Whitton C, Filarial Genome Project: **The *Brugia malayi* genome project: expressed sequence tags and gene discovery.** *Trans R Soc Trop Med Hyg* 2002, **96:**7-17.

39. Hoekstra R, Visser A, Otsen M, Tibben J, Lenstra JA, Roos MH: **EST sequencing of the parasitic nematode *Haemonchus contortus* suggests a shift in gene expression during transition to the parasitic stages.** *Mol Biochem Parasitol* 2000, **110:**53-68.

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

41. Lizotte-Waniewski M, Tawe W, Guiliano DB, Lu W, Liu J, Williams SA, Lustigman S: **Identification of potential vaccine and drug target candidates by expressed sequence tag analysis and immunoscreening of *Onchocerca volvulus* larval cDNA libraries.** *Infect Immun* 2000, **68:**3491-3501.

42. McCarter JP, Mitreva MD, Martin J, Dante M, Wylie T, Rao U, Pape D, Bowers Y, Theising B, Murphy CV, *et al.*: **Analysis and functional classification of transcripts from the nematode *Meloidogyne incognita*.** *Genome Biol* 2003, **4:**R26.

43. Allen JE, Daub J, Guilliano D, McDonnell A, Lizotte-Waniewski M, Taylor D, Blaxter M: **Analysis of genes expressed at the infective larval stage validate the utility of *Litomosoides sigmodontis* as a murine model for filarial vaccine development.** *Infect Immun* 2000, **68:**5454-5458.

44. Tetteh KKA, Loukas A, Tripp C, Maizels RM: **Identification of abundantly-expressed novel and conserved genes from infective stage larvae of *Toxocara canis* by an expressed sequence tag strategy.** *Infect Immun* 1999, **67:**4771-4779.

45. Fernández C, Gregory WF, Loke P, Maizels RM: **Full-length-enriched cDNA libraries from *Echinococcus granulosus* contain separate populations of oligo-capped and *trans*-spliced transcripts and a high level of predicted signal peptide sequences.** *Mol Biochem Parasitol* 2002, **122:**171-180.

46. The *C. elegans* Genome Consortium: **Genome sequence of *Caenorhabditis elegans*: a platform for investigating biology.** *Science* 1998, **282:**2012-2018.

47. Stein LD, Bao Z, Blasiar D, Blumenthal T, Brent MR, Chen N, Chinwalla A, Clarke L, Clee C, Coghlan A, *et al.*: **The genome sequence of *Caenorhabditis briggsae*: a platform for comparative genomics.** *PLoS Biol* 2003, **1:**E45.

48. Blaxter ML, De Ley P, Garey JR, Liu LX, Scheldeman P, Vierstraete A, Vanfleteren JR, Mackey LY, Dorris M, Frisse LM, *et al.*: **A molecular evolutionary framework for the phylum Nematoda.** *Nature* 1998, **392:**71-75.

49. **Blaxter lab nematode genomics** [http://www.nematodes.org]

50. Liu J, Rost B: **Comparing function and structure between entire proteomes.** *Protein Sci* 2001, **10:**1970-1979.

51. Nilsen TW: **Trans-splicing of nematode premessenger RNA.** *Annu Rev Microbiol* 1993, **47:**413-440.

52. Blaxter M, Liu L: **Nematode spliced leaders - ubiquity, evolution and utility.** *Int J Parasitol* 1996, **26:**1025-1033.

53. Anderson RC: *Nematode Parasites of Vertebrates: Their Development and Transmission* Wallingford, UK: CAB International; 1992.

54. Parkinson J, Blaxter M: **SimiTri - visualizing similarity relationships for groups of sequences.** *Bioinformatics* 2003, **19:**390-395.

55. Loukas A, Maizels RM: **Helminth C-type lectins and host-parasite interactions.** *Parasitol Today* 2000, **16:**333-339.

56. Tort J, Brindley PJ, Knox D, Wolfe KH, Dalton JP: **Proteinases and associated genes of parasitic helminths.** *Adv Parasitol* 1999, **43:**161-266.

57. Blaxter ML: **Nemoglobins: divergent nematode globins.** *Parasitol Today* 1993, **9:**353-360.

58. Kaiser CA, Preuss D, Grisafi P, Botstein D: **Many random sequences functionally replace the secretion signal sequence of yeast invertase.** *Science* 1987, **235:**312-317.

59. Zang X, Maizels RM: **Serine proteinase inhibitors from nematodes and the arms race between host and pathogen.** *Trends Biochem Sci* 2001, **26:**191-197.

60. Sommer A, Nimtz M, Conradt HS, Brattig N, Boettcher K, Fischer P, Walter R, Liebau E: **Structural analysis and antibody response to the extracellular glutathione S-transferases from *Onchocerca volvulus*.** *Infect Immun* 2001, **69:**7718-7728.

61. Pastrana DV, Raghavan N, FitzGerald P, Eisinger SW, Metz C, Bucala R, Schleimer RP, Bickel C, Scott AL: **Filarial nematode parasites secrete a homologue of the human cytokine macrophage migration inhibitory factor.** *Infect Immun* 1998, **66:**5955-5963.

62. Zang XX, Taylor P, Meyer D, Wang JM, Scott AL, Walkinshaw MD, Maizels RM: **Homologues of human macrophage migration inhibitory factor from a parasitic nematode: gene cloning, protein activity and crystal structure.** *J Biol Chem* 2002, **277:**44261-44267.

63.   Camberis M, Le Gros G, Urban J Jr: **Animal model of *Nippos-trongylus brasiliensis* and *Heligmosomoides polygyrus*.** In *Current Protocols in Immunology* Edited by: Coico R. New York: John Wiley and Sons; 2003:19.12.11-19.12.27.

64.   Parkinson J, Guiliano DB, Blaxter M: **Making sense of EST sequences by CLOBBing them.** *BMC Bioinformatics* 2002, **3:**31.

65.   **The Phred/Phrap/Consed system home page**   [http://www.phrap.org]

66.   Fukunishi Y, Hayashizaki Y: **Amino acid translation program for full-length cDNA sequences with frameshift errors.** *Physiol Genomics* 2001, **5:**81-87.

67.   Iseli C, Jongeneel CV, Bucher P: **ESTScan: a program for detecting, evaluating, and reconstructing potential coding regions in EST sequences.** *Proc Int Conf Intell Syst Mol Biol* 1999:138-148.

68.   **SignalP server**   [http://www.cbs.dtu.dk/services/SignalP-2.0]

69.   **Index of /SimiTri**   [http://www.nematodes.org/SimiTri]

comment

reviews

reports

deposited research

refereed research

interactions

information