

39. I. Greenwald, *Genes Dev.* **12**, 1751 (1998); D. Levitan and I. Greenwald, *Nature* **377**, 351 (1995); X. Li and I. Greenwald, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 12204 (1997).
40. M. Hengartner, in *C. elegans II*, D. L. Riddle *et al.*, Eds. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1997), pp. 383–416.
41. K. Kempthues and S. Strome, in *ibid.*, pp. 335–360; R. Schnabel and J. R. Priess, in *ibid.*, pp. 361–382; S. Guo and K. Kempthues, *Curr. Opin. Genet. Dev.* **6**, 408 (1996).
42. F. Slack and G. Ruvkun, *Annu. Rev. Genet.* **31**, 611 (1998); R. Lee *et al.*, *Cell* **75**, 843 (1993).
43. R. Johnson and C. Tabin, *Cell* **81**, 313 (1995); T. R. Burglin, *Curr. Biol.* **6**, 1047 (1996). The HOG genes identified by Burglin bear the probable intein-like autoproteolysis domain that is also present in hedgehog and may also recognize and bind to sterols, but these genes do not bear the many other features of vertebrate and invertebrate hedgehog orthologs.
44. P. Kuwabara, personal communication.
45. B. J. Meyer, in *C. elegans II*, D. L. Riddle *et al.*, Eds. (Cold Spring Harbor Laboratory, Cold Spring Harbor, 1997), pp. 209–240.
46. P. G. Okkema and J. Kimble, *EMBO J.* **10**, 171 (1991).
47. T. M. Barnes and J. Hodgkin, *ibid.* **15**, 4477 (1996); P. Aza-Blanc *et al.*, *Cell* **89**, 1043 (1997).
48. Y. T. Ip and M. Levine, *Curr. Opin. Genet. Dev.* **14**, 672 (1994).
49. X. S. Hou and N. Perrimon, *Trends Genet.* **13**, 105 (1997).
50. A. Fire *et al.*, *Nature* **391**, 806 (1998).
51. G. Jansen *et al.*, *Nature Genet.* **17**, 119 (1997).
52. We have in general referenced reviews and apologize to all those whose work we could not cite due to space constraints. We are indebted to many *C. elegans* and *Drosophila* developmental geneticists for comments on the manuscript and for communicating unpublished results.

# Caenorhabditis elegans Is a Nematode

Mark Blaxter

## REVIEW

*Caenorhabditis elegans* is a rhabditid nematode. What relevance does this have for the interpretation of the complete genome sequence, and how will it affect the exploitation of the sequence for scientific and social ends? Nematodes are only distantly related to humans and other animal groups; will this limit the universality of the *C. elegans* story? Many nematodes are parasites; can knowledge of the *C. elegans* sequence aid in the prevention and treatment of disease?

In terms of numbers of described species, the arthropods dominate the known metazoan life on Earth. Although the number of described species of nematode is only ~20,000, estimates of the actual number range from 40,000 to 10 million. The high estimates are based on repeated sampling of single marine habitats and are supported by surveys of terrestrial faunas (1). Nematodes are also numerically abundant, attaining millions of individuals per square meter (2). *Caenorhabditis elegans* is therefore a representative of a diverse and successful group of animals.

How do the molecular, physiological, and developmental mechanisms used by *C. elegans*—as revealed by the *C. elegans* genome sequence and by the equally important genetic and developmental biological work carried out in the last 30 years (3)—relate to those used by other animals? Although there are undoubtedly nematode-specific components to the *C. elegans* basic body plan, some recent studies indicate that signaling systems have been recruited wholesale to perform new functions as if they are self-contained cassettes that can be exchanged with little functional consequence (4). At a higher level, though, the patterns and processes used by *C. elegans* to build its body are a product of adaptive evolution over millions of years. Thus, the phylogenetic position of *C. elegans* with respect to other animals is of importance in deciphering the modes and tempos of evolution of these processes (5).

For example, if a gene [such as a particular nuclear hormone receptor subtype (4)] is found in both the fruit fly *Drosophila* and *C. elegans*, does this imply that it will most likely also be present in the human genome? If *C. elegans*' ancestor diverged before the vertebrate-arthropod split, the answer will be yes. If, as has been suggested, nematodes are more closely related to arthropods than to vertebrates (see below), similarities between *Drosophila* and *C. elegans* may merely reflect their common ancestry. Is *C. elegans* representative of a primitive metazoan, or is it a highly derived organism?

### *C. elegans*' Place in the Tree of Life

The application of the *C. elegans* project to the understanding of other animals, and of humans in particular, is compromised by the deep

phyletic separation of the nematodes from other groups. Current best estimates of the time of divergence range from 1200 million to 600 million years ago (6). There are about 35 animal groups whose body plans are distinct enough to warrant elevation to phylum status (7). After 130 years of phylogeny (8), the interrelationships of the animal phyla are still the subject of vigorous debate, and the position of the Nematoda within the animals is far from clear. The integration of molecular and morphological analyses is required to resolve these long-standing problems (9).

Morphological phylogenies have usually indicated that the pseudocoelomate nematodes arose early in animal evolution, as part of a radiation of "aschelminth" phyla, predating the split into protostome groups (annelids, arthropods, mollusks, and others) and deuterostome groups (chordates, brachiopods, and others) (Fig. 1A) (10, 11). This scheme suggests that nematodes are equally distant from both arthropods and vertebrates. Cladistic analyses of developmental and morphological traits have resulted in a reassessment of this unresolved phylogeny. Nielsen (7) proposed that the nematodes, along with four other pseudocoelomate phyla (nematomorphs, priapulids, kinorhynchans, and loriciferans), form a monophyletic group of animals with an introvert (extensible, spined anterior organ), no locomotory cilia, and a cuticle that is shed at periodic molts. Nematodes are recognized as protostomes, animals where the mouth is formed from the embryonic blastopore. This feature is not particularly evident in *C. elegans*, where the embryo is a dense mass of cells and the blastopore is not distinct, but is in other nematodes (12). In Nielsen's phylogeny, therefore, nematodes are slightly more closely related to arthropods than they are to vertebrates.

Molecular phylogenetic analyses of the position of the Nematoda with respect to other phyla were initially compromised by the use of *C. elegans* as a marker nematode taxon. The genes of *C. elegans* appear to have undergone accelerated molecular evolution relative to those of many other animals. This relative rate difference resulted in the (probably) artifactual placement of the origin of *C. elegans* (and with it, by association, all of the nematodes) very early in metazoan molecular phylogenies. This phenomenon has meant that the nematodes have been left out of such analyses until recently. Sequencing of small subunit ribosomal RNA genes from additional species of nematode has yielded taxa with reduced apparent rates, and these sequences can be used to place nematodes more robustly within the metazoa (13, 14). The results of these studies are surprising and challenge the view that nematodes branched off before the arthropod-vertebrate split. Two major rearrangements are proposed. The arthropods are removed from a close relationship to the annelids, and a new high-level taxon, of animals that shed a cuticle by ecdysis (the Ecdysozoa), is proposed to include arthropods, nematodes, and their allies (Fig. 1C) (14). The Ecdysozoa hypothesis is not universally accepted, as it

The author is at the Institute of Cell, Animal, and Population Biology, University of Edinburgh, Edinburgh EH9 3JT, UK.

contradicts some morphological evidence, but it is eminently testable with other genes.

Genome sequencing of model organisms has allowed larger data sets, encompassing many genes, to be used to examine nematode-animal relationships (15). The analyses are equivocal concerning arthropod-nematode-vertebrate relationships, but again suffer from relative rate effects due to accelerated evolution in both arthropod and nematode branches. The slowest-evolving genes tend to support an arthropod-nematode association. As sequence accumulates from other species [and particularly other species of nematode (16)], these hypotheses will be tested more rigorously.

### C. elegans and Other Nematode Species

*Caenorhabditis elegans* is not the most important nematode on our planet. From the human perspective, that prize probably goes to *Ascaris lumbricoides*, the large gut roundworm that infects more than 1 billion people worldwide, causing malnutrition and obstructive bowel disease (16). Close behind are the human hookworms (*Ancylostoma duodenale* and *Necator americanus*), blood-sucking strongylid parasites that infect more than 600 million today and were once the scourge of the southern United States. These parasites are transmitted by water contamination; others are spread by biting arthropod vectors (for example, the causative agents of human lymphatic filariasis, *Wuchereria bancrofti* and *Brugia malayi*) or by eating contaminated food (for example, the pork trichina worm *Trichinella spiralis*). The plant-parasitic root-knot nematodes (*Meloidogyne* spp.) cause hundreds of billions of dollars of crop production loss worldwide, and thus contribute significantly to malnutrition and disease. Other plant parasitic nematodes (*Xiphinema* and *Trichodorus* species) are ectoparasites that transmit devastating plant viruses. Hence, it is important that the *C. elegans* genome project yields an improved understanding of other nematodes, so as to enable the development of control strategies to alleviate their effects on human populations (17).

Application of molecular phylogenetic methods (18) has led to a reappraisal of the interrelationships of the accepted nematode orders and revealed a surprising depth and diversity in many

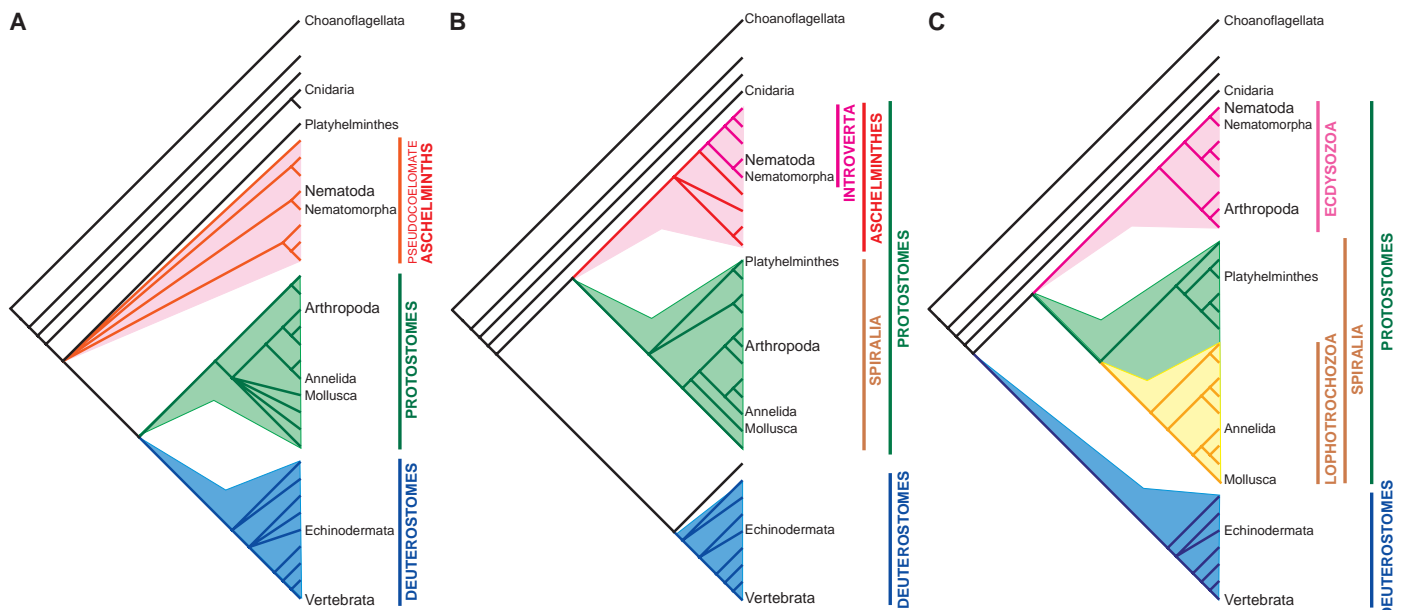
groups. [Our new analysis is summarized and explained in Fig. 2 (16).] The new analyses fit well with many morphological (12) and developmental (19) characters, but debate on their validity is still vigorous. The molecular phylogeny can be used to direct research programs by defining stepping stones across the phylum to get from a target of interest in *C. elegans* to a parasite with major economic effects. For example, the animal parasitic Strongylida (including the human hookworms *Ancylostoma* and *Necator*) are robustly placed within the Rhabditida, and *C. elegans* is likely to be an excellent model for these important pathogens. Genetic resistance to current anti-nematode drugs is on the rise, and the development of novel control strategies, perhaps involving nematode-specific neurotropic agents (20) or disrupting sex determination or embryonic pathways, is a priority (21).

Genome-wide analysis of parasitic nematodes is still in its infancy but is already yielding dividends (22). One of the frustrations of working with parasitic organisms, particularly those of humans, is that they are hard to grow. Genetic and transgenic analysis is much more difficult. Thus, the opportunity afforded by *C. elegans* as a tractable testbed for gene function is attractive. A gene of interest can be identified, its *C. elegans* homolog found, the function of the homolog investigated exhaustively, and the results then transferred to the parasite.

### Nematode-Specific Genes

The *C. elegans* genome sequence predicts 18,600 genes (23). Comparison of the whole of the coding potential of the *C. elegans* genome with that of other (non-nematode) organisms reveals that ~58% of the genes appear to be nematode-specific. A proportion of these nematode-specific genes have been functionally identified by genetic analyses, and many (34% of the total) form families with other nematode genes. What are these nematode-specific elaborations and inventions doing? Even within the 42% of genes with homologs in other phyla, there are still specific (perhaps nematode-specific) variations, such as novel juxtapositions of protein modules, or wholesale amplification of particular gene families (4, 24, 25).

The genes that have no clear homologs will derive from four



**Fig. 1.** The relationships of the animal phyla. Three hypotheses of these relationships are represented (10); each has different implications for the expected similarity of the *C. elegans* genome to other species of medical or research importance. (A) A phylogeny based on traditional morphological criteria (10). Nematodes are part of a basal radiation of pseudocoelomic

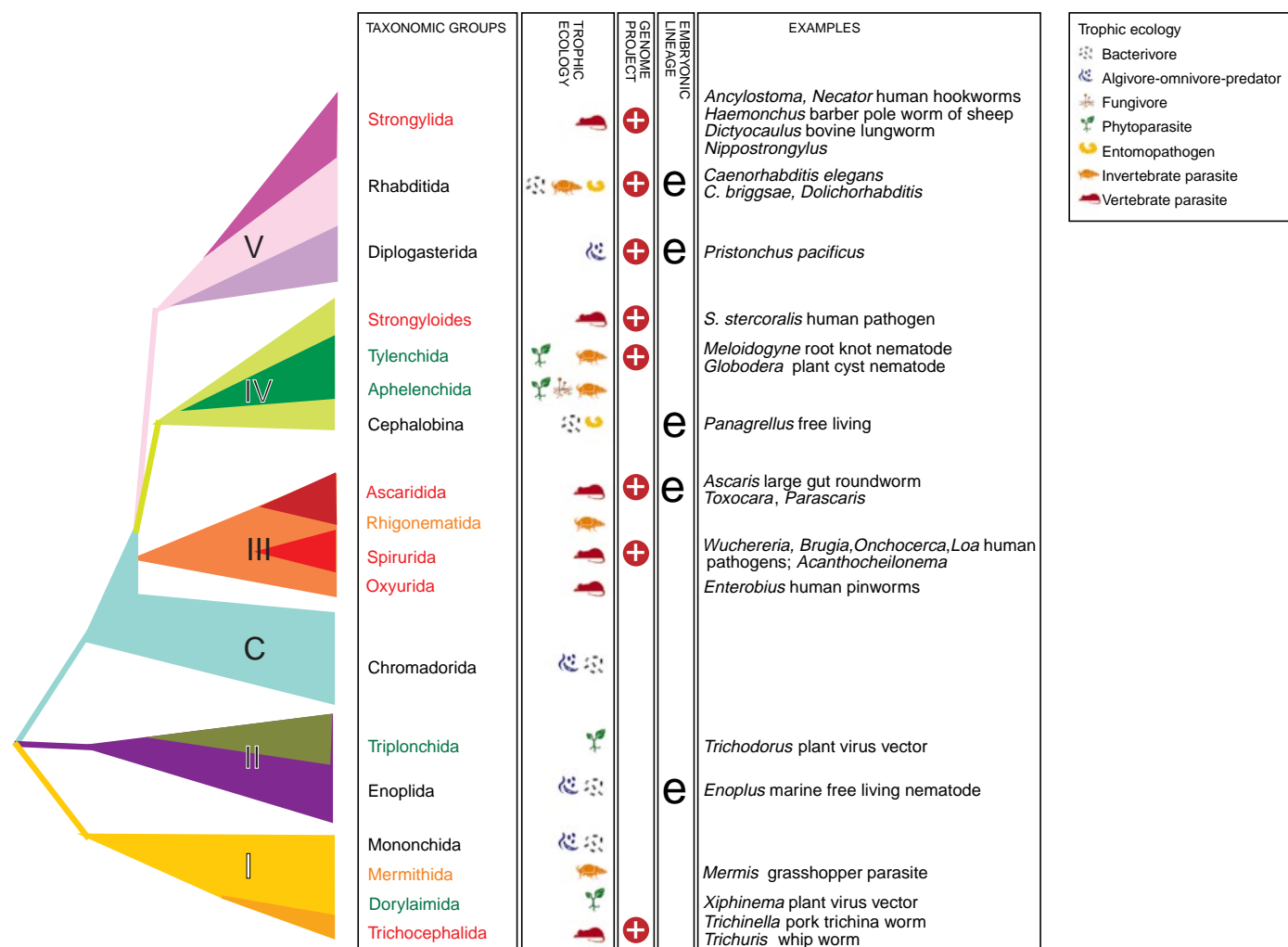
phyla whose interrelationships are not clearly resolved. (B) The phylogeny proposed by Nielsen (7), wherein nematodes are recognized as protostomes and are grouped with other phyla having an anterior introvert organ. (C) The phylogeny proposed by Aguinaldo *et al.* (14), with the nematodes and arthropods joined in a clade of molting animals.

classes: genes that do have homologs in other organisms that have not yet been sequenced (group 1) or that evolve at such a rate or in such a manner as to make the homology undetectable (group 2), genes that are specific to the nematodes (group 3), and genes that are unique to *C. elegans* and its closest relatives (group 4). Group 3 will be of most interest to parasitologists and pharmacologists because it will include the genes particular to building and running the nematode body plan. Within groups 1 and 2 will be genes that have been multiplied to form families or adapted to distinct functions in nematodes compared to other groups.

*Caenorhabditis elegans* differs from other organisms not only in its basic body plan, but also in many facets of metabolism and molecular biology. One such feature of the *C. elegans* genome is that many genes (about 80%) are trans-spliced to a common spliced leader exon. In addition, about 20% of genes are organized as operons, cotranscribed sets of two or more genes (26). This operonic structure has been demonstrated in one other species closely related to *C.*

*elegans* (*Dolichorhabditis*) (27). The significance of the operonic organization of genes is not clear in general, though some instances of genes with related function being cotranscribed have been noted. In that it differs from cis-splicing, the trans-splicing machinery may rely on novel or diverged proteins. Other sources of difference include facets of intermediate biochemistry (for example, nematodes have a functional glyoxalate cycle and can synthesize polyunsaturated fatty acids de novo) and the biosynthesis of the cuticle.

Our domain analysis of the *C. elegans* predicted protein data set suggests that there are ~400 distinct domains that appear to be unique to nematodes (28). These *C. elegans*- or nematode-specific domains include large and small protein segments, and families with more than 50 members, many of which are predicted to be extracellular (24). One source of functional information about these nematode-specific proteins is the large body of work on parasitic nematodes. For animal parasites, the cuticle and its surface are major players in the host-parasite interface. Immune attack is directed against surface compo-



**Fig. 2.** The phylum Nematoda: a cartoon illustrating the molecular phylogenetic analysis of nematode diversity (16). Sequences were abstracted from published reports and analyzed as described (18, 45). *Caenorhabditis elegans* is a rhabditid nematode, part of a diverse assemblage of microbivorous soil-dwelling species. These were traditionally classified in a distinct order from other free-living species (the diplogasterids, such as *Pristionchus pacificus*) and parasitic orders. Molecular phylogenetic analysis with ribosomal small subunit RNA genes (and other genes) strongly suggests that the rhabditids, the diplogasterids, and the animal-parasitic strongyloids (which include human hookworms) can be grouped as a single clade (clade V). The morphologically rather uniform rhabditids are apparently very diverse ge-

netically. A second group of terrestrial free-living nematodes, the cephalobes, are similarly linked with plant-parasitic (tylenchid), fungal-feeding (aphelenchid), and animal-parasitic (strongyloid) groups (clade IV). Several major human parasites (including *Ascaris* and the filarial nematodes) are shown to be very closely related (clade III). These three clades (traditionally given the name Secernentea) arise from a group of microbivorous aquatic/water film nematodes (the Chromadorida, clade C). Two other major clades can be discerned. Clade II includes plant-parasitic (Triplonchida) as well as free-living (Enoplida) members. Clade I links parasites of insects (Mermithida), plants (Dorylaimida), and animals (Trichocephalida) with free-living omnivores (Mononchida).

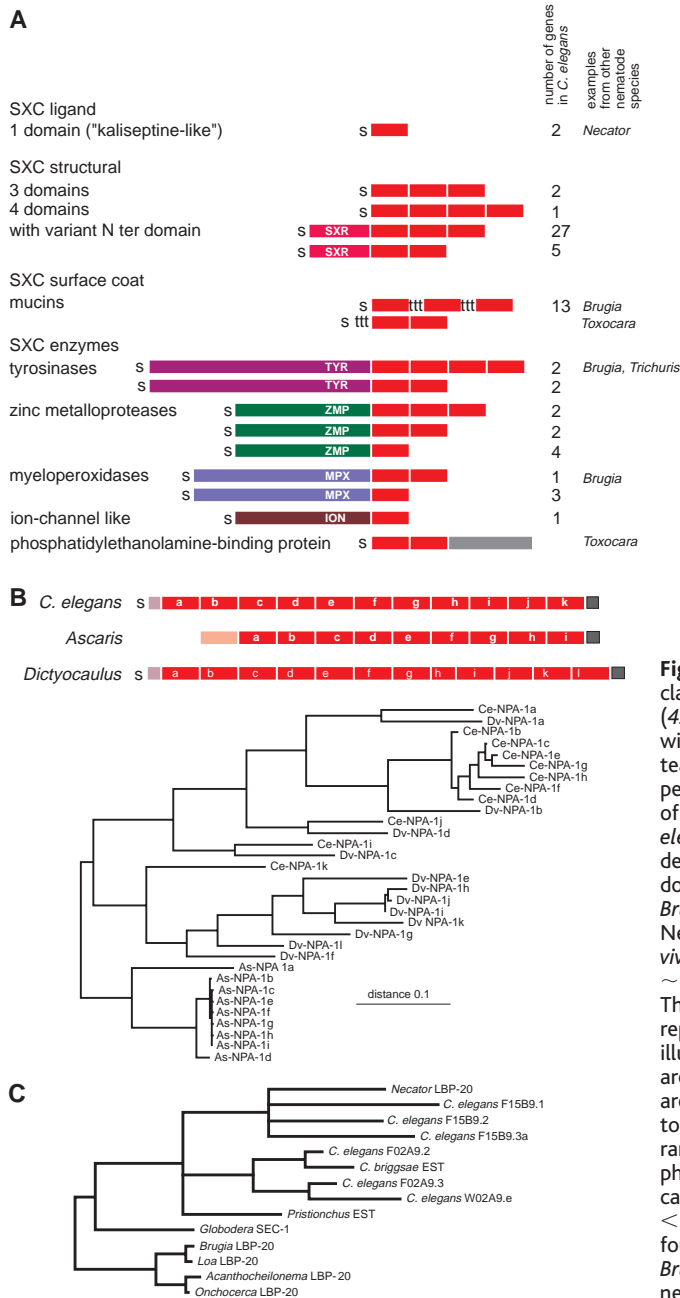
ments, and surface-located enzymes and other effectors mediate immune resistance, host manipulation, and nutritional uptake (29). The identification and cloning of animal-parasite surface proteins has been a major theme in molecular parasitology, and this program has identified proteins and domains with novel structures and functions.

One such domain is the SXC (six-cysteine) domain first identified in surface coat components of the parasitic ascaridid *Toxocara canis* (30). The SXC domain is short (36 to 42 amino acids), with six conserved cysteines (believed to be disulfide-bonded) and a number of other conserved residues. We have found 75 genes in *C. elegans* that contain 184 SXC motifs (Fig. 3A) (31). These include genes with only SXC motifs (up to four), mucin-like genes with SXC motifs separated by serine- or threonine-rich segments, and genes where a recognizable enzymatic domain is flanked by SXC motifs. The enzymes identified include tyrosinases, myeloperoxidases, and astacin-like zinc metalloproteases. The mucin-

like and SXC-only genes tend to be clustered as families in the genome. SXC domains have also been identified in other nematodes: in *Ascaris*, *Brugia*, *Trichuris muris* (a mouse-parasitic relative of human whipworm), and *Necator* (the human hookworm) (32). The SXC motif is likely to be a domain involved in protein-protein interaction, possibly specific to extracellular matrices such as the nematode cuticle. The SXC domain may also act as a signaling ligand (like the epidermal growth factor domain). Two non-nematode peptides with SXC-like features are known from sea anemone toxins, where they act as voltage-sensitive K<sup>+</sup>-channel blockers. In hookworms and in *C. elegans* similar secreted, single SXC-domain genes are present that may be diffusible ligands for as yet unknown receptors (33).

Two other nematode-specific gene families were first identified in parasitic nematodes as antigens in infection. These have subsequently been shown to be lipid-binding proteins, which may play roles in nutrient scavenging from the host or transport of lipid within the nematode. The first is an allergen identified in *Ascaris* and also found in strongyloid and filarial nematodes, where it is surface-located. It is the major allergen of *Ascaris* and is an important determinant of disease reactions in humans. It has been called the nematode polyprotein allergen (NPA), as it is first synthesized as a large peptide, which is cleaved into 15-kD monomers. They are predicted to fold as four α-helix bundles, and therefore to bind lipid buried within a hydrophobic core (34). In some species, such as *Ascaris*, the repeat unit is relatively monomorphic in sequence, whereas in others [such as the strongyloid lungworm *Dictyocaulus viviparus* (35)] each repeat is significantly different. The relationship of the differences in sequence to lipid binding specificity, if any, is unknown. Our analysis of the complete genome sequence revealed that *C. elegans* also has a NPA homolog (spread over cosmids VC5 and F27B10), which has variable repeat units like *Dictyocaulus* (Fig. 3B). Because of the diversity of sequence, it is unlikely that this gene would have been found by conventional means, but it can now be used to examine the organismal biology of the protein, the significance of repeat variation, and the regulation of its processing.

An unrelated small lipid-binding protein, LBP-20, also predicted to fold as four α helices, was first described from the surface of the human river blindness parasite *Onchocerca volvulus* (36). This 20-kD antigen has homologs in other filarial nematodes, and there is growing



**Fig. 3.** Nematode-specific proteins first identified in parasites. (A) The different classes of SXC-containing proteins found in *C. elegans* and other nematodes (45). The SXC domain is indicated by the red boxes. Other domains associated with SXC domains are S, signal peptide; ION, ion channel-like; MP, metalloprotease/astacin domain; TYR, tyrosinase domain; SXR, SXC-related domain; PX, peroxidase domain; and ttt, threonine- and/or serine-rich domain. To the right of each gene type is given the number of different genes in each class in the *C. elegans* genome, and other nematode species where this gene family has been demonstrated. The phosphatidylethanolamine-binding protein with two SXC domains at its COOH-terminus has only been found in *Toxocara* (30); the *Brugia*, *Onchocerca*, and *C. elegans* homologs do not have SXC domains. (B) Nematode polyprotein allergens. The NPA homologs of *C. elegans*, *Dictyocaulus viviparus*, and *Ascaris suum* are compared. Each gene encodes a polyprotein with ~15-kD domains separated by tetrabasic, subtilisin-like protease cleavage sites. The *Ascaris* sequence is derived from partial cDNAs encompassing only nine repeats. Repeat h of *Dictyocaulus* is truncated. Below the cartoon is a tree illustrating the diversity of repeat sequences in the NPAs. The *Ascaris* repeats are very similar to each other, whereas the *C. elegans* and *Dictyocaulus* repeats are more divergent (35). (C) LBP-20 homologs from many nematodes compared to the *C. elegans* gene family. LBP-20 homologs were identified from a wide range of nematode species (36). The aligned sequences were subjected to phylogenetic analysis by neighbor-joining algorithm, and the statistical significance of the resulting trees was tested by bootstrap analysis (45); nodes with <50% bootstrap support are collapsed. The six *C. elegans* representatives are found as two pairs (one head-to-head, one head-to-tail) and two single copies. *Brugia*, *Loa*, *Onchocerca*, and *Acanthocheilonema* are animal-parasitic filarial nematodes. *Globodera* is a plant parasite. *Necator* is a gut parasite.

interest in its potential as a vaccine component and as a marker of immune status in onchocerciasis. The *C. elegans* genome project has identified six homologs of this protein, and others have been sequenced from *C. briggsae*, *Pristionchus pacificus*, the plant parasite *Globodera pallida*, and *Necator* (36). Fortuitously, one of the *C. elegans* homologs was also identified in a promoter-trapping screen designed to define expression patterns for random genes using a  $\beta$ -galactosidase marker gene in transgenic *C. elegans* (37). This *C. elegans* gene is expressed in the somatic musculature, whereas the parasitic homologs are synthesized in the hypodermis and are secreted to the surface. Perhaps other members of the LBP-20 family are hypodermal in *C. elegans*. Could LBP-20 be used to trick nematodes into assimilating toxic lipid analogs ignored by their hosts?

### Comparative Nematode Genomics

An efficient way of identifying a large number of expressed genes is through the expressed sequence tag (EST) strategy (38). EST projects have now been carried out on a number of other nematodes, including *C. briggsae* and the free-living diplogasterid model *Pristionchus pacificus*. The World Health Organization has sponsored the Filarial Genome Project, which has generated 16,500 ESTs from the human parasite *Brugia malayi* (22, 39). Smaller EST data sets have been generated from *Onchocerca*, *Strongyloides stercoralis* (a human gut parasite), *N. americanus*, *Ascaris*, *Trichuris*, *Toxocara*, and *Nippostrongylus brasiliensis* (a model rodent gut strongylid) (see Fig. 2). When compared with the *C. elegans* genome, these data sets can be used to refine and confirm *C. elegans* gene predictions, identify conserved residues, examine the evolutionary histories of the nematode genes, and define potentially nematode-specific genes. As expected from the ribosomal RNA phylogenetic studies (Fig. 2), the rhabditid and strongylid EST data sets show highest overall similarity to *C. elegans*, whereas the *Trichuris* data set is least similar. Surprisingly, in the *Trichuris* data set, more than 50% of the genes are novel (or pioneer) despite having the complete *C. elegans* gene set for comparison. This hints at genetic and functional diversity within the nematodes, which sampling from one species would not have revealed.

To complement the *C. elegans* sequence, substantial portions (>5%) of the sequence of the genome of the closely related *C. briggsae* have also been determined. Comparison of segments sequenced from both species reveals that, in general, gene order has been closely conserved, and synteny cloning is feasible (40). The *C. briggsae* genome appears to be slightly smaller than that of *C. elegans*, as both intergenic and intronic regions are shorter. The major differences seen are attributable to the insertion of transposable elements and the rearrangement of relatively large DNA segments. Comparison of the *C. briggsae* and *C. elegans* sequences serves to confirm intron-exon predictions (in that the level of conservation of DNA sequence is much higher within exons) and highlights potential control regions. As first demonstrated for the *hsp-70* genes, comparison of upstream regions between these two species is a powerful way of identifying promoter elements: Conserved segments prove to have promoter activity (41).

It is also informative to examine genome structure and gene order in distantly related nematodes. As part of the Filarial Genome Project, a map of the *Brugia* genome is being constructed (22). Although full chromosomal comparisons are not yet possible, sequence of a 65-kb segment surrounding a gene of interest [a macrophage migration inhibition factor homolog (42)] has revealed conservation of local gene order and synteny between *C. elegans* and *Brugia* (43). Even with the limited sequence data available, some contrasts are already evident. Introns in *C. elegans* can be separated into two classes: common short introns (37 to 80 bases) and rarer long ones (>150 bases) (44). *Brugia* does not appear to have this preponderance of short introns (most are >300 bases).

The *C. briggsae* and *Brugia* data suggest that comparative se-

quencing of selected extensive genomic regions will reveal unexpected features of nematode sequence, gene evolution, and genome evolution that cannot be accessed through the static picture of a single genome. When integrated with the emerging synthesis of sequence with biology in *C. elegans*, these comparative data will both enhance our understanding of the biology of all metazoa and offer new tools to control and eradicate nematode pathogens.

### References and Notes

- J. Lambshhead, *Oceanis* **19**, 5 (1993); H. M. Platt, in *The Phylogenetic Systematics of Free-Living Nematodes*, S. Lorenzen, Ed. (The Ray Society, London, 1994); G. Boucher and J. D. Lambshhead, *Conserv. Biol.* **9**, 1594 (1994); J. H. Lawton et al., *Nature* **391**, 72 (1998).
- T. A. Platonova and V. V. Gal'tsova, *Nematodes and Their Role in the Meiobenthos* (Nakua, Leningrad, 1976).
- D. Riddle, T. Blumenthal, B. Meyer, J. Priess, Eds., *C. elegans II* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1997).
- G. Ruvkun and O. Hobert, *Science* **282**, 2033 (1998).
- A. Sidow and W. K. Thomas, *Curr. Biol.* **4**, 596 (1994).
- H. Philippe, A. Chenuil, A. Adoutte, *Dev. Suppl.* **15** (1994); R. A. Raff, *The Shape of Life: Genes, Development and the Evolution of Animal Form* (Univ. of Chicago Press, Chicago, 1996); R. A. Raff, C. R. Marshall, J. M. Turbeville, *Annu. Rev. Ecol. Syst.* **25**, 351 (1994); J. R. Vanfleteren et al., *Mol. Phylogenet. Evol.* **3**, 92 (1994); R. F. Doolittle, D.-F. Feng, S. Tsang, G. Cho, E. Little, *Science* **271**, 470 (1996); G. A. Wray, J. S. Levinton, L. H. Shapiro, *ibid.* **274**, 568 (1996); D.-F. Feng, G. Cho, R. F. Doolittle, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 13028 (1997).
- C. Nielsen, *Animal Evolution. Interrelationships of the Living Phyla* (Oxford Univ. Press, Oxford, 1995).
- E. Haeckel, *Generelle Morphologie der Organismen* (Georg Reimer, Berlin, 1866).
- S. Conway Morris, *Dev. Suppl.*, **1** (1994).
- Expanded versions of the phylogenetic trees in Fig. 1, giving the names of all the phyla, are available at <http://www.ed.ac.uk/~mbx/science/Figure1.html> and at [www.sciencemag.org/feature/data/985136.shl](http://www.sciencemag.org/feature/data/985136.shl).
- R. C. Brusca and G. J. Brusca, *Invertebrates* (Sinauer, Sunderland, MA, 1990).
- V. V. Malakhov, *Nematodes. Structure, Development, Classification and Phylogeny* (Smithsonian Institution Press, Washington, DC, 1994).
- K. M. Halanych et al., *Science* **267**, 1641 (1995).
- A. M. A. Aguinaldo et al., *Nature* **387**, 489 (1997).
- A. R. Mushegian, J. R. Garey, J. Martin, L. X. Liu, *Genome Res.* **8**, 590 (1998).
- An expanded version of Fig. 2, giving full names of taxa analyzed, is available at <http://www.ed.ac.uk/~mbx/figure2.html> and at [www.sciencemag.org/feature/data/985136.shl](http://www.sciencemag.org/feature/data/985136.shl).
- M. L. Blaxter and D. M. Bird, in (3), pp. 851–878.
- S. Kampfer, C. Sturmbauer, C. J. Ott, *Invertebr. Biol.* **117**, 29 (1998); M. L. Blaxter et al., *Nature* **392**, 71 (1998); V. V. Aleshin, O. S. Kedrova, I. A. Milyutina, N. S. Vladychenskaya, N. B. Petrov, *Russ. J. Nematol.* **6**, 175 (1998); D. H. A. Fitch, B. Bugaj-gaweda, S. W. Emmons, *Mol. Biol. Evol.* **12**, 346 (1995); D. H. A. Fitch and W. K. Thomas, in (3), pp. 815–850.
- B. Goldstein, L. M. Frisse, W. K. Thomas, *Curr. Biol.* **8**, 157 (1997); D. A. Voronov, Y. V. Panchin, S. E. Spiridonov, *Nature* **395**, 28 (1998).
- R. J. Martin, A. P. Robertson, H. Bjorn, *Parasitology* **114**, S111 (1997).
- R. Prichard, *Vet. Parasitol.* **54**, 259 (1994).
- The Filarial Genome Project, *Parasitol. Today*, in press; T. A. Moore et al., *Mol. Biochem. Parasitol.* **79**, 243 (1996); see also the Filarial Genome Project web site at <http://helios.bto.ed.ac.uk/mbx/fgn/filgen1.html>.
- C. elegans* Genome Sequencing Consortium, *Science* **282**, 2012 (1998).
- H. Hulter et al., personal communication.
- S. A. Chervitz et al., *Science* **282**, 2022; N. D. Clarke and J. M. Berg, *ibid.*, p. 2018.
- M. Krause and D. Hirsh, *Cell* **49**, 753 (1987); X.-Y. Huang and D. Hirsh, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 8640 (1989); M. L. Blaxter and L. X. Liu, *Int. J. Parasitol.* **26**, 1025 (1996).
- T. Blumenthal and K. Steward, in (3), pp. 115–117.
- C. elegans* predicted genes in WormPep (release 14) were processed using the Pfam domain 1.6 system [E. L. L. Sonnhammer and D. Kahn, *Protein Sci.* **3**, 482 (1994)] and the resulting domains (groups of  $\geq 2$  protein segments with significant similarity to each other) were compared to SwissProt 35 and SPTREMBL. All domains with significant similarities to non-nematode proteins were eliminated, leaving 409 apparently nematode-specific domains containing from 58 to 2 members. The analysis was performed by S. J. Jones. A previous analysis of a WormPep data set encompassing about one-third of the complete data set identified many of these domains [E. L. L. Sonnhammer and R. Durbin, *Genomics* **46**, 200 (1997)]. A general overview of the data set and annotation is available at [www.sciencemag.org/feature/data/c-elegans.shl](http://www.sciencemag.org/feature/data/c-elegans.shl).
- D. Gems and R. M. Maizels, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 1665 (1996); D. G. Gems et al., *J. Biol. Chem.* **270**, 18517 (1995).
- M. L. Blaxter, A. P. Page, W. Rudin, R. M. Maizels, *Parasitol. Today* **8**, 243 (1992); R. M. Maizels, D. A. P. Bundy, M. E. Selkirk, D. F. Smith, R. M. Anderson, *Nature* **365**, 797 (1993).
- See <http://www.ed.ac.uk/~mbx/science/sxc.html> and [www.sciencemag.org/feature/data/985136.shl](http://www.sciencemag.org/feature/data/985136.shl) for more information about *C. elegans* SXC domain proteins.
- D. Gerrits, J. Daub, M. Blaxter, unpublished data.
- H. Schweitz et al., *J. Biol. Chem.* **270**, 25121 (1995).

34. L. A. McReynolds, M. W. Kennedy, M. E. Selkirk, *Parasitol. Today* **9**, 403 (1993); M. W. Kennedy *et al.*, *Biochemistry* **34**, 6700 (1995); M. W. Kennedy, J. E. Allen, A. S. Wright, A. B. McCrudden, A. Cooper, *Mol. Biochem. Parasitol.* **71**, 41 (1995).
35. H. J. Spence, J. Moore, A. Brass, M. W. Kennedy, *Mol. Biochem. Parasitol.* **57**, 339 (1993); C. Britton, J. Moore, J. S. Gilleard, M. W. Kennedy, *ibid.* **72**, 77 (1995); J. Moore and M. Blaxter, unpublished data. The NPA repeats were aligned and subjected to analysis using the neighbor-joining method in PAUP\*4d64 using the mean character distance setting. See <http://www.ed.ac.uk/~mbx/science/npa.html> for more information on the NPA homologs.
36. T. I. M. Tree *et al.*, *Mol. Biochem. Parasitol.* **69**, 185 (1995); M. W. Kennedy *et al.*, *J. Biol. Chem.* **272**, 29442 (1997). *C. elegans* LBP-20 homologs were identified in WormPep 14 and genomic sequence. LBP-20 homologs from other species were identified in GenBank and dbEST. Additional LBP-20 genes have been isolated from the filarial nematodes *Brugia*, *Loa*, and *Acanthocheilonema* (J. Allen and J. Bradley, personal communication; D. Guiliano, personal communication), and from the stronglylid *Necator* (J. Daub and M. Blaxter, unpublished data). See <http://www.ed.ac.uk/~mbx/science/lbp20.html> or [www.sciencemag.org/feature/data/985136.shl](http://www.sciencemag.org/feature/data/985136.shl) for more information.
37. I. A. Hope, *Development* **113**, 399 (1991); I. A. Hope *et al.*, *Trends Genet.* **12**, 370 (1996).
38. L. L. Fulton, L. Hillier, R. K. Wilson, in *Caenorhabditis elegans. Modern Biological Analysis of an Organism*, H. F. Epstein and D. C. Shakes, Eds. (Academic Press, San Diego, CA, 1996), pp. 571–582; W. R. McCombie *et al.*, *Nature Genet.* **1**, 124 (1992); R. Waterston *et al.*, *ibid.*, p. 114; Y. Kohara, *Tanpakushitsu Kakusan Koso* **41**, 715 (1996).
39. M. L. Blaxter *et al.*, *Mol. Biochem. Parasitol.* **77**, 77 (1996).
40. P. Kuwabara and S. Shah, *Nucleic Acids Res.* **22**, 159 (1994).
41. M. F. P. Heschl and D. L. Baillie, *J. Mol. Evol.* **31**, 3 (1990); J. S. Gilleard, J. D. Barry, I. L. Johnstone, *Mol. Cell. Biol.* **17**, 2301 (1997).
42. D. V. Pastrana *et al.*, *Infect. Immun.*, in press.
43. D. Guiliano and M. Blaxter, unpublished data.
44. C. Fields, *Nucleic Acids Res.* **18**, 1509 (1990).
45. D. L. Swofford, G. J. Olsen, P. J. Waddell, D. M. Hillis, in *Molecular Systematics*, D. M. Hillis, C. Moritz, B. K. Mable, Eds. (Sinauer, Sunderland, MA, 1996), pp. 407–514. The tree presented is a bootstrap consensus phylogram. All the nodes in this tree are supported >60%.
46. Supported by the Darwin Trust. I thank colleagues in the Filarial Genome Project, the Nematode Phylogeny consortium, and the parasitic nematology community for discussions and insights. S. Jones performed the Pfam analysis of nematode-specific domains. J. Hodgkin, J. Allen, D. Guiliano, and two reviewers offered welcome comments on the manuscript.

# POWERSURGE

NEW! *Science* Online's Content Alert Service  
 Knowledge is power. If you'd like more of both, there's only one source that delivers instant updates on breaking science news and research findings: *Science*'s Content Alert Service. This free enhancement to your *Science* Online subscription delivers e-mail summaries of the latest news and research articles published each Friday in *Science* – *instantly*. To sign up for the Content Alert service, go to *Science* Online – but make sure your surge protector is working first.

**Science**  
[www.sciencemag.org](http://www.sciencemag.org)

For more information about Content Alerts go to [www.sciencemag.org](http://www.sciencemag.org).  
 Click on Subscription button, then click on Content Alert button.