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. Nematodes are numerically abundant, attaining millions of individuals per square meter. 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—relate to those used by other animals? Although there are undoubtedly nematode-specific components to C. elegans the 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. 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.

For example, if a gene such as a particular nuclear hormone receptor subtype 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, 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 phylogenetic separation of the nematodes from other groups. Current best estimates of the time of divergence range from 1200 million to 600 million years ago. There are about 35 animal groups whose body plans are distinct enough to warrant elevation to phylum status. After 130 years of phylogeny, 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.

Morphological phylogenies have usually indicated that the pseudocoelemate 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). 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 proposed that the nematodes, along with four other pseudocoelemate phyla (nematomorphs, priapulids, kinorhynchs, 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. 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. 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. 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. The Ecdysozoa hypothesis is not universally accepted, as it...
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 strongyloid 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 other species [and particularly other species of nematode (16)], these hypotheses will be tested more rigorously.

C. elegans: Sequence to Biology

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 pseudocelomic 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 introverted organ. (C) The phylogeny proposed by Aguinaldo et al. (14), with the nematodes and arthropod 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 (exon. In addition, about 20% of genes are organized as operons, many genes (about 80%) are trans-spliced to a common spliced leader.

**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 strongylids (which include human hookworms) can be grouped as a single clade (clade V). The morphologically rather uniform rhabditids are apparently very diverse genetically. A second group of terrestrial free-living nematodes, the cephalobins, are similarly linked with plant-parasitic (tylenchid), fungal-feeding (aphelenchid), and animal-parasitic (strongylid) 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).
nents, 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⁺-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 strongylid 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 strongylid 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 VCS 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 tt, 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 tetrasaccharide, 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 (43); 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 β-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 ribosomal 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* sequences 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 intra-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 sequencing 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 metazoans and offer new tools to control and eradicate nematode pathogens.

**References and Notes**

10. Expanded versions of the phylogenetic trees in Fig. 1, giving the names of all the phyla, are available on the *World Wide Web* at http://www.ed.ac.uk/~mbx/science/fig1.html. and at www.sciencemag.org/feature/data/985136.shl.
16. An expanded version of Fig. 2, giving full names of taxa analyzed, is available at http://www.ed.ac.uk/~mbx/fig2fig2.html. and at www.sciencemag.org/feature/data/985136.shl.
18. H. Hulter et al., *personal communication.*
22. C. elegans predicted genes in *WormPep* (release 14) were processed using the Pfam domain 1.6 system [E. L. L. Sonnehammer and D. Kahn, *Protein Sci.* 3, 482 (1994)] and the resulting domains (groups of >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 version of a *WormPep* data set encompassing about one-third of the complete data set identified many of these domains [E. L. L. Sonnehammer and R. Durbin, *Genomics* 46, 200 (1997)]. A general overview of the data set and annotation is available at www.sciencemag.org/content//data/c-elegans.html.

**C. ELEGANS: SEQUENCE TO BIOLOGY**

**Special Section**
Specal Section

C. elegans: Sequence to Biology


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 strongyloid 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 for more information.


42. D. V. Pastrana et al., Infect. Immun., in press.

43. D. Guiliano and M. Blaxter, unpublished data.


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.