Review | TRENDS in Genetics Vol.21 No.10 October 2005

Full text provided by www.sciencedirect.com
SCIENCE DIRECT

Comparative genomics of nematodes

Makedonka Mitreva[1], Mark L. Blaxter[2], David M. Bird[3] and James P. McCarter[1,4]

[1]Genome Sequencing Center, Department of Genetics, Washington University School of Medicine, St Louis, MO 63108, USA
[2]Institute of Evolutionary Biology, School of Biological Sciences, University of Edinburgh, Edinburgh, UK, EH9 3JT
[3]Center for the Biology of Nematode Parasitism, North Carolina State University, Raleigh, NC 27695, USA
[4]Divergence Inc., St Louis, MO 63141, USA

Recent transcriptome and genome projects have dramatically expanded the biological data available across the phylum Nematoda. Here we summarize analyses of these sequences, which have revealed multiple unexpected results. Despite a uniform body plan, nematodes are more diverse at the molecular level than was previously recognized, with many species- and group-specific novel genes. In the genus *Caenorhabditis*, changes in chromosome arrangement, particularly local inversions, are also rapid, with breakpoints occurring at 50-fold the rate in vertebrates. Tylenchid plant parasitic nematode genomes contain several genes closely related to genes in bacteria, implicating horizontal gene transfer events in the origins of plant parasitism. Functional genomics techniques are also moving from *Caenorhabditis elegans* to application throughout the phylum. Soon, eight more draft nematode genome sequences will be available. This unique resource will underpin both molecular understanding of these most abundant metazoan organisms and aid in the examination of the dynamics of genome evolution in animals.

Introduction

To extend the usefulness of completed, high-quality, model organism genome sequences, many projects have also sampled the genomes of related species to various levels of completion from draft genome coverage to shallow surveys. One of the unknowns of this strategy is the extent to which model organism genomes will be representative of the genomes of related species. The nematode *Caenorhabditis elegans* was the first multicellular organism for which a complete genome sequence was generated [1] and it remains the only metazoan (animal) for which the sequence of every nucleotide (i.e 100 278 047 nt) has been finished to a high level of confidence [2]. Recent sampling across the phylum Nematoda (Box 1) is now enabling investigators to address questions of phylum-level conservation and diversity. What are the advantages and disadvantages of using *C. elegans* genome as a model for other nematodes? What is the mode and tempo of genome evolution across the Nematoda, and how do they compare with other phyla?

Corresponding author: Mitreva, M. (mmitreva@watson.wustl.edu).
Available online 15 August 2005

Sequencing the nematodes

Not including *C. elegans*, recent sequencing efforts have generated 560 000 expressed sequence tags (ESTs) and genome survey sequences (GSSs) from >38 species of nematodes representing all clades except Enoplia (sufficient biological material is unavailable) [3] (Figure 1 and Box 2). A high-quality draft-genome sequence of *Caenorhabditis briggsae* has been generated and an extensive comparative analysis with *C. elegans* completed [4]. A draft genome sequence of *Brugia malayi* has also been produced [5] and GenBank submission is expected in late 2005. For information on additional genome projects, see Table 1.

Genomic disparity across the phylum Nematoda

Rapid gene discovery and cross-species comparisons in Nematoda have been enabled by generating and analyzing ESTs, single-pass sequence reads from randomly selected cDNA clones. In 2000 there were only 24 000 ESTs from nematodes other than *C. elegans* in public databases, but by December 2004 ~350 000 had been deposited, mainly from two large-scale projects [3,6]. cDNA libraries have been made from a variety of life-cycle stages and dissected tissues. Some libraries have taken advantage of the conserved nematode *trans*-spliced leader sequence SL1, found at the 5′ end of many nematode transcripts, to amplify full-length cDNAs [7,8]. More than a dozen single-species analyses of these EST datasets have been published, focusing on parasites of humans (e.g. Refs [9,10]), animals (e.g. Ref. [11]) and plants (e.g. Ref. [12]). The first meta-analysis of the genomic biology of the phylum Nematoda used >250 000 ESTs originating from 30 species, clustered into 93 000 genes and grouped into 60 000 gene families [13]. This data collection was used to estimate the degree to which 'genespace' (the diversity of distinct

Box 1. The phylum Nematoda

Nematodes, or roundworms, are an ancient and diverse group of organisms and the most abundant of all metazoans [79]. They exploit multiple ecological niches and include free-living terrestrial and marine microbivores, meiofaunal predators, herbivores, and animal and plant parasites (Figure 2). Parasitic nematodes are important pathogens of humans, infecting several billion people, and causing damage to domesticated animals and crops. Parasitism has arisen multiple times during nematode evolution, and all major clades include parasites. Molecular phylogenetics defines three major nematode classes, which can be further divided into five clades (Figure 1): Dorylaimia (Clade I in Ref. [80]), Enoplia (Clade II) and Chromadorea (including Spirurina – Clade III; Tylenchina – Clade IV; and Rhabditina – Clade V). *C. elegans* is a member of Rhabditina.

 0168-9525/$ - see front matter © 2005 Elsevier Ltd. All rights reserved. doi:10.1016/j.tig.2005.08.003

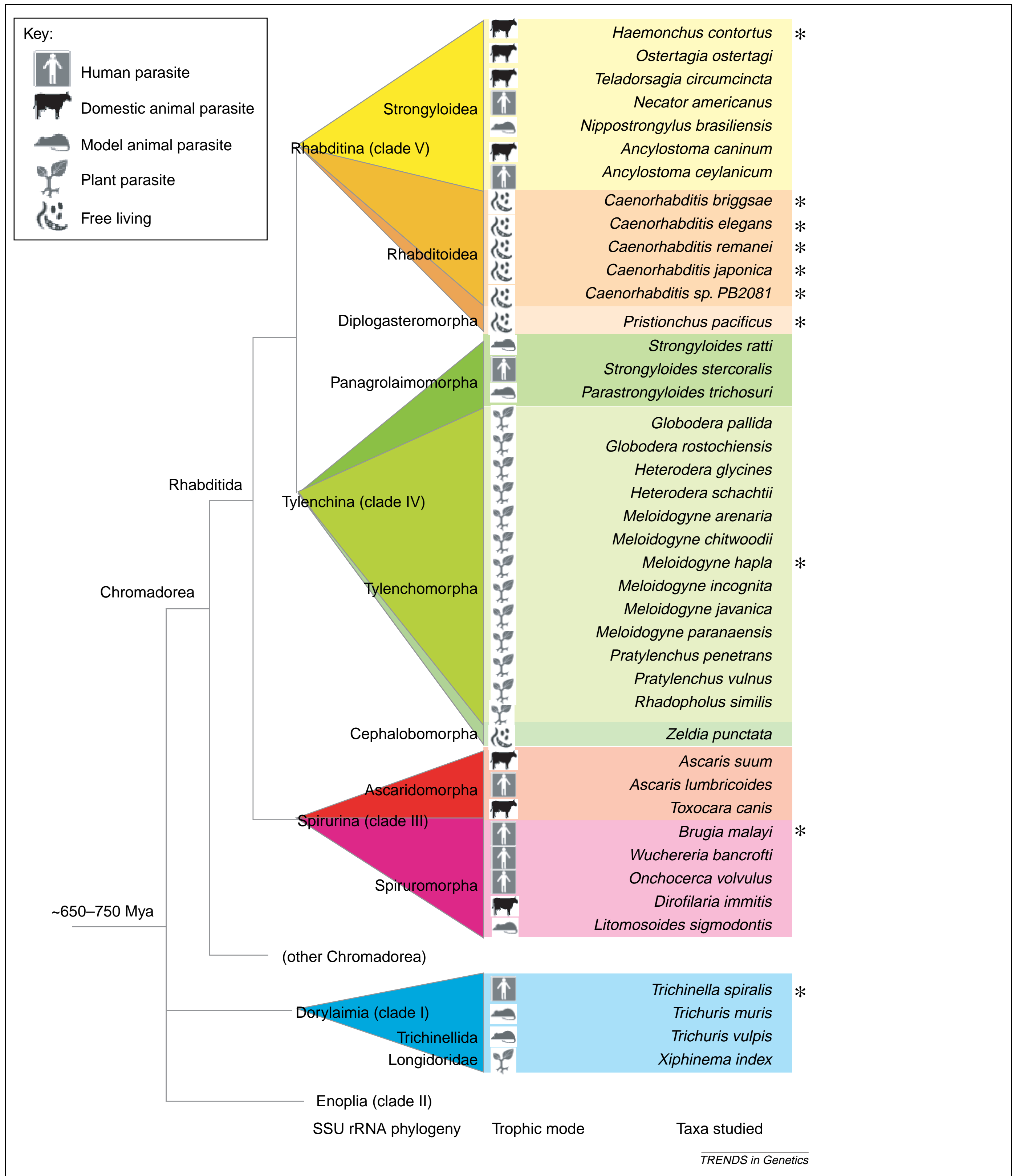

Figure 1. Genome information across the phylum Nematoda. All species with either significant numbers of ESTs in public databases (>100) or genome projects are arranged phylogenetically based on small subunit (18S ribosomal RNA) (SSU) rRNA phylogeny [4]. Species with genome projects completed or underway are indicated by asterisks. Adapted with permission from Ref. [4].

genes) within nematodes has been sampled. Analysis of bacterial genespace had shown that continued addition of complete genomes yielded diminishing returns of novelty, suggesting that a large percentage of bacterial genespace was already sampled [14]. However, more recent ecosystem sampling of marine microbes has revealed the vast genetic complexity present in such environments. Sequencing of Sargasso Sea microbes yielded 148 previously

Box 2. Comparing full and partial genomes

For the foreseeable future, eukaryotic molecular biologists will be analyzing sequence information from a limited number of complete genomes [1], an increasing number of draft genomes [4] and an even larger number of partial genomes from which only limited sampling has occurred [13]. Figure I provides a graphical illustration of currently available sequence data comparing the *C. elegans* proteome with the collections of sequences from other non-*Caenorhabditis* nematodes and displaying the highest scoring match for each dataset. Evolutionary change is not uniform across genomes, with varying rates of change for different genes and lineages resulting in a distribution. Overall, the observed distribution is in keeping with the known SSU rRNA phylogeny (see Figure 1), higher pairwise similarity BLAST bitscores (normalized scores independent of the scoring matrix chosen) are found relative to partial genomes from the more closely related Rhabditina (Clade V) and Tylenchina (Clade IV) than to the more distantly related Spirurina (Clade III) and Enoplia (Clade I) (Figure Ia,b). Thus, in the Clade I and III combined versus Clade V comparison (Figure Ib), 9558 proteins had higher scoring matches in Clade V, 224 proteins had equal matches in both groups, and 6489 had better matches to Clades I and III. Such an approach is useful in visualizing candidate genes that could have homologs or orthologs available for study in multiple clades or where gene loss might have occurred in one lineage. Other techniques such as SimiTri plots enable the visualization of query sequences versus three target sequence collections simultaneously [38]. However, conclusions about gene distribution depend on statistical analysis taking into account the partial nature of the sequence collections and the resulting caveats such as variation in contig length and matches to weaker homologs as opposed to true orthologs not yet sequenced. Ideally, complete predicted proteomes based on multiple lines of evidence would be available. Yet even in the case of a complete genome, such as *C. elegans*, generating such a high-quality predicted proteome has been an iterative process with significant revision occurring over several years [2], a circumstance that is unlikely to be matched for the numerous other nematode genomes.

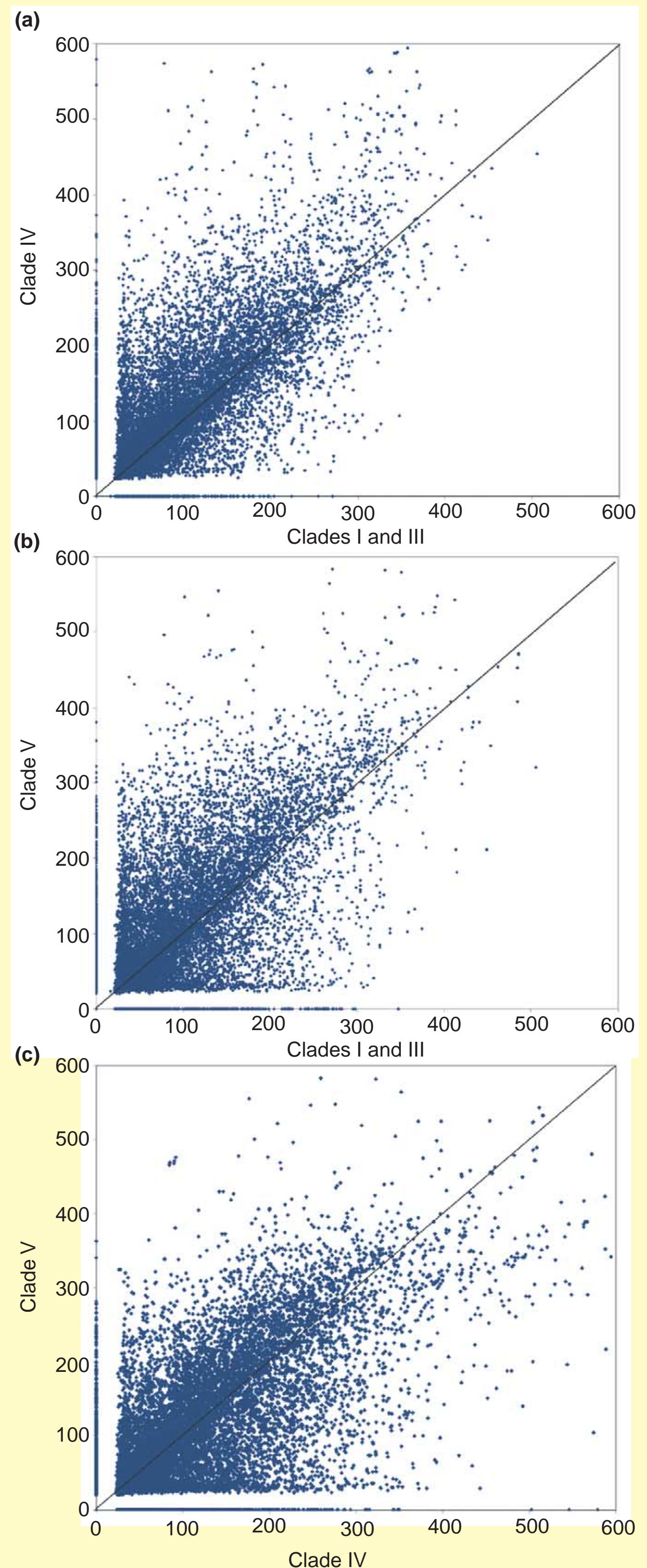

Figure I. The similarity between *Caenorhabditis elegans* proteins and putative genes products from other nematodes. TBLASTN-based similarity comparisons of the predicted *C. elegans* proteome (as defined in the Wormpep release 132–22 288 peptides: www.sanger.ac.uk/Projects/C_elegans/WORMBASE/current/wormpep.shtml) with available EST and contig consensus sequences from major nematode clades: Clades I and III combined, 98 391 sequences; Clade IV, 109 085 sequences; Clade V (excluding *Caenorhabditis* spp.), 49 771 sequences. For each graph, the relative bitscore of the top match of each *C. elegans* protein to each clade or group of clades is displayed. *C. elegans* was compared with: **(a)** Clades I and III versus IV (16 241 comparisons return nonzero values for both clades); **(b)** Clades I and III versus V (16 271 comparisons return nonzero values for both clades); **(c)** Clade IV versus V (16 475 comparisons return nonzero values for both clades). For the three graphs together, a total of 154 data points have bitscores >650 and are beyond the scale (0.31%). The species included were: Clade I: *Trichinella spiralis*, *Trichuris vulpis*, *Trichuris muris*, *Xiphinema index*; Clade III: *Ascaris suum, Ascaris lumbricoides*, *Brugia malayi*, *Dirofilaria immitis*, *Onchocerca volvulus*, *Toxocara canis*; Clade IV: *Globodera rostochiensis*, *Heterodera glycines*, *Heterodera schachtii*, *Meloidogyne arenaria*, *Meloidogyne chitwoodi*, *Meloidogyne hapla*, *Meloidogyne incognita*, *Meloidogyne javanica*, *Meloidogyne paranaensis*, *Pratylenchus penetrans*, *Pratylenchus vulnus*, *Parastrongyloides trichosuri*, *Strongyloides ratti*, *Strongyloides stercoralis*, *Zeldia* puncta; Clade V: *Ancylostoma caninum*, *Ancylostoma ceylanicum*, *Haemonchus contortus*, *Necator americanus*, *Nippostrongylus brasiliensis*, *Ostertagia ostertagi*, *Teladorsagia circumcincta*, *Pristionchus pacificus*.

Table 1. Genome sizes and chromosome numbers of nematode taxa for which a genome project is underway[a]

Species	Clade	Trophic ecology	Type of genome project	Number of chromosomes(n)[b]	Size of genome (Mb)	Funding
Caenorhabditis elegans	V	Bacteriovore	Full genome sequence complete [1]	6	100.2	NHGRI and MRC
Caenorhabditis briggsae	V	Bacteriovore	Whole genome draft complete [4]	6	105	NHGRI and Wellcome Trust
Caenorhabditis remanei	V	Bacteriovore	Whole genome draft complete	6	~131	NHGRI
Caenorhabditis japonica	V	Bacteriovore	Whole genome draft planned	6	–	NHGRI
Caenorhabditis sp. c.f. PB2801	V	Bacteriovore	Whole genome draft planned	6	–	NHGRI
Pristionchus pacificus	V	Omnivore	Whole genome draft in progress	6	~110	NHGRI
Brugia malayi	III	Vertebrate parasite	Whole genome shotgun complete [5]	5	~100	NIAID
Haemonchus contortus	V	Vertebrate parasite	Whole genome draft in progress	6	~55	Wellcome Trust
Meloidogyne hapla	IV	Plant parasite	Pooled BAC sequencing planned	16	62.5	NSF and USDA
Trichinella spiralis	I	Vertebrate parasite	Whole genome draft in progress	?[c]	270	NHGRI

[a]Abbreviations: NIAD, National Institute of Allergy and Infectious Diseases (USA); NHGRI, National Human Genome Research Institute (USA); NSF, National Science Foundation (USA); MRC, Medical Research Council (UK); USDA, United States Department of Agriculture.
[b]Sex determination in rhabditines is by an XX-XO mechanism; thus males have one less chromosome than females. In some taxa, such as the filarial nematode *Brugia malayi*, XY sex chromosomes have been described [81], and in some tylenchine plant parasites sex determination is driven by environmental cues (see Ref. [82]).
[c]Other trichosomatids have $n=4$.

unknown bacterial phylotypes and 1.2 million genes originating from at least 1800 genomic species [15].

In nematodes, despite the availability of the genomes of two *Caenorhabditis* species, genespace appears far from thoroughly sampled because the addition of each new species to the analysis yielded a linear increase in new gene discovery. With a BLAST bit score cut-off of 50, 30–70% of genes from each species had no non-nematode homolog, and ~60% of the genes had a homolog elsewhere within the Nematoda. Because ESTs often only partially cover the cognate mRNA, even after clustering, this could have been a result of comparisons using fragments of genes, but even when only sequences of >400 bp were considered, an average of 45% of the genes of each species had no non-nematode homolog and 23% of the genes were unique to the species sampled. Therefore, despite a deceptively uniform body plan (Figure 2), nematodes seem to be more diverse at the molecular level than was previously recognized. The set of ~20 000 genes and ~12 000 gene families represented by *C. elegans* [1] provides a starting point for exploring this diversity, and captures many of the conserved gene families shared with other eukaryotes, but it represents only a small portion of the expanding total nematode genespace. Because sequencing has been performed in only a few dozen of the estimated one million or more nematode species, the vast majority of nematode genespace remains unsampled.

Surprisingly, many genes were identified from the ESTs that have clear orthologs in other animals but are missing from the complete genome of *C. elegans* (and the draft genome of *C. briggsae*). These genes were probably lost in the lineage leading to the caenorhabditids, emphasizing the importance of gene loss in genome evolution. A striking example is in the Hox gene complex, where *C. elegans* has many fewer genes than arthropods or vertebrates. This paucity of genes (six different genes corresponding to four of the eight paralogous genes found in *Drosophila melanogaster*) was initially suggested to be the result of the 'primitive' origin of nematodes. Because Hox genes are involved in anterior–posterior patterning of the body axis, this coincided neatly with the perception of nematodes as morphologically simple animals. However, a survey of other nematodes revealed the presence of orthologs of at least seven of the expected families, and thus that *C. elegans* has lost Hox genes during evolution [16,17]. A unique gene fusion of Hox paralogs in the filarial nematode *Brugia malayi* (Spirurina) suggests that these genes were lost piecemeal.

What sorts of features characterize the new genes that have evolved in the Nematoda? In the rodent parasite *Nippostrongylus brasiliensis*, novel sequences were nearly tenfold enriched for genes containing predicted signal peptides relative to more conserved genes with homologies beyond Nematoda [18]. One speculation is that the complexities of host–parasite interaction, including the host immune response, are driving accelerated molecular evolutionary change. Comparing gene expression between parasitic and free-living nematodes suggests that novel genes could be associated with dispersal and infective life-cycle stages (Figure 3). Genes expressed in survival or dispersal stages, including the *C. elegans* dauer L3 and the infective L3 in *Strongyloides stercoralis*, were more likely to be novel compared with genes expressed in adult or larval growth stages, and showed a weaker signature of shared expression [19]. Similar analysis on additional species, stages and tissues will refine this view of how patterns of gene expression have changed during nematode evolution.

Caenorhabditis shows a remarkable rate of intra-chromosomal rearrangement

Nematode genomes, like those of other eukaryotes, are organized as multiple chromosomes, except in the case of

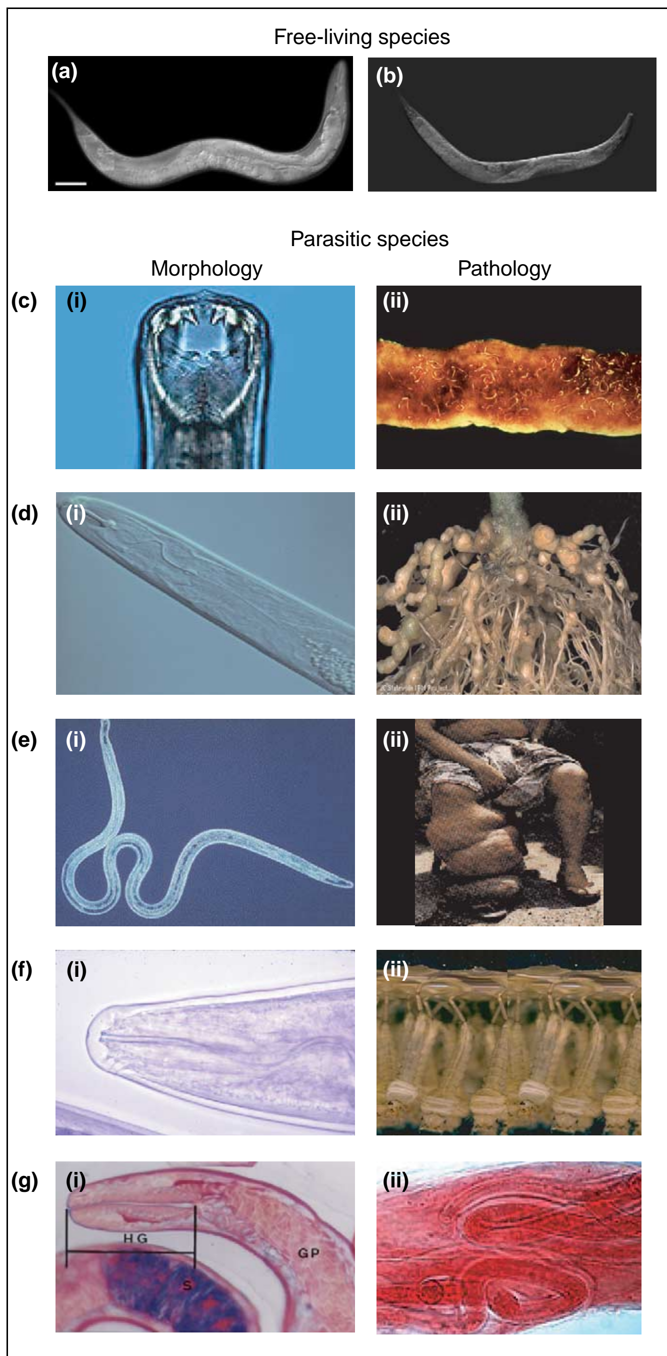

Figure 2. Nematode morphology and pathology. **(a)** *Caenorhabditis elegans*, Clade V, free-living soil nematode, adults are 1.5 mm long. Photo is courtesy of Ian D. Chin-Sang (Queen's University, Kingston, ON, Canada). **(b)** *Pristionchus pacificus*, Clade V, free-living species, adults are 1 mm long. Photo courtesy of Ralf Sommer (Max-Planck Institute, Germany). **(c)** *Ancylostoma* species, Clade V, vertebrate hookworm parasite, adults are 10 mm in length: (i) anterior end of adult *Ancylostoma caninum;* (ii) *A. caninum* in a strip of small intestine from a dog (photo courtesy of Eisenback and Zunke; http://www.mactode.com/Pages/Nemapix.html). **(d)** *Meloidogyne* species, Clade IVb, plant parasite, larvae are 400 μm, adult female diameter 600 μm: (i) anterior part of L2 larva; (ii) infected tomato roots (photo courtesy of the University of California IPM Project; www.ipm.ucdavis.edu/PMG/r52200111.html). **(e)** Filarial species, Clade III, human parasite, females are 5 cm, males are 18 cm: (i) *Brugia malayi* L3 (photo courtesy of Sinclair Stammers, Imperial College, London); (ii) lymphatic filarial infection (photo taken from Peters and Gilles, 1991). **(f)** *Romanomermis culicivorax*, Clade I, entomopathogenic parasite, body size 5–25 mm: (i) anterior end of adult; (ii) coiled within mosquito (photos courtesy of Edward G. Platzer, University of California, Riverside, CA). **(g)** *Trichinella spiralis*, Clade I, vertebrate parasite, females are 3 mm, males are 1.5 mm: (i) anterior end of adult (photo courtesy of Yuzo Takahashi, Gifu University, Gifu, Japan); (ii) larvae encysted in vertebrate muscle cell (photo courtesy of Eisenback and Zunke; http://www.mactode.com/Pages/Nemapix.html).

Parascaris univalens (Clade III); nematode chromosome number (n) ranges up to >40, but in general is between four and eight [20]. In the group of nematodes closest to *C. elegans*, the Rhabditina, n is usually six (Table 1). Nematode chromosomes are acrocentric (with distributed centromere function) and have little in the way of structure discernible by conventional banding techniques: most nematode chromosomes are visualized as small dots. Exceptions to this trend are found in the ascaridids (Clade III), where there can be large genomes (up to 2.5 Gb) with few chromosomes. *Parascaris univalens* (c.f. *equorum* var. 5) has a single large chromosome (∼2 Gb) in its germ-line cells (*Ascaris suum* has $n=24$). However, *Parascaris* and *Ascaris* chromosomes in somatic cells undergo chromatin diminution, a process whereby large germ-line chromosomes are specifically cleaved in particular regions to yield many thousands of minichromosomes [21]. Chromatin diminution involves loss of intergenic DNA and some genes [22,23], and addition of new telomeres to the chromosome fragments, and is associated with rapid chromosomal evolution [24]. Chromatin diminution has been described from other phyla, but is apparently limited in nematodes to a subset of ascaridids.

Comparative vertebrate genomics can define conserved synteny blocks across the whole vertebrate radiation, and has permitted the prediction of the ancestral chromosome set for amniotes (reptiles, birds and mammals) [25]. With only two, closely related, completed nematode genomes to analyze, it is perhaps premature to discuss chromosomal evolution in the Nematoda, but it is already apparent that the dynamics of chromosomal change are different from that known in vertebrates. Comparison of the genomes of *C. elegans* and *C. briggsae* suggests that >4000 chromosome rearrangement events have occurred since they last shared a common ancestor an estimated 80–110 million years ago (Mya) [4,26]. These events include a preponderance of local inversions and transpositions, but relatively few between-chromosome translocations. The rate of chromosomal evolution (breakpoints per megabase per million years) in *Caenorhabditis* exceeds that of the previous record holders (drosophilid flies) by fivefold, and that of vertebrates by fiftyfold [27]. This high rate of within-chromosome rearrangement is in contrast to the relatively stable chromosome number of six, suggesting there might be an important mechanism stabilizing chromosome number in the Rhabditina.

Congruent with intra-chromosomal rearrangement far exceeding interchromosomal translocations, analysis of conserved synteny between *B. malayi* and *C. elegans* revealed a preponderance of within-chromosome linkage, but little local synteny (homologs of genes separated ∼60 kb in *B. malayi* were separated by a mean of >4 Mb in *C. elegans*) [28]. *B. malayi* has five chromosome pairs (Table 1) [28], and comparisons of genes linked in *B. malayi* but on different chromosomes in *C. elegans* suggests that the changed chromosome number is the result of a fusion between genomic segments corresponding to *C. elegans* chromosomes I and III. Because the related filarioid *Setaria* is reported as having $n=6$ [20], the fusion could have occurred in the onchocercine lineage (Clade III).

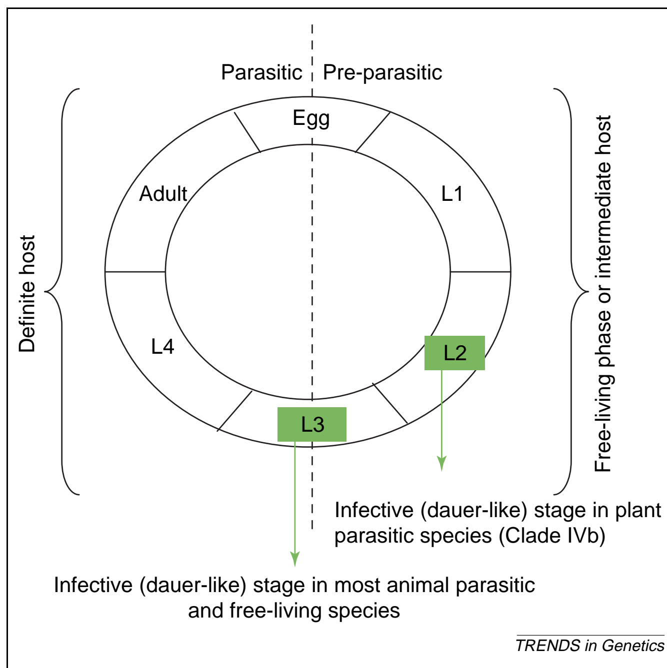

Figure 3. The life cycle of a parasitic nematode. The nematode has four larval stages after hatching that are punctuated by molts. The adult nematode resides in the definitive host, but enters this host as an L3 or an L2 stage. The arrested dauer stage in *Caenorhabditis elegans* is an L3. In some nematode species (such as *Ascaris* spp. and *Tylenchina* spp.) the L2 hatches from the egg after undergoing a molt; or the L1 infective muscle stage larvae (in *Trichinella spiralis*) infect a paratenic host, but the definitive host life cycle is essentially L3 driven. Abbreviation: L, larval stage.

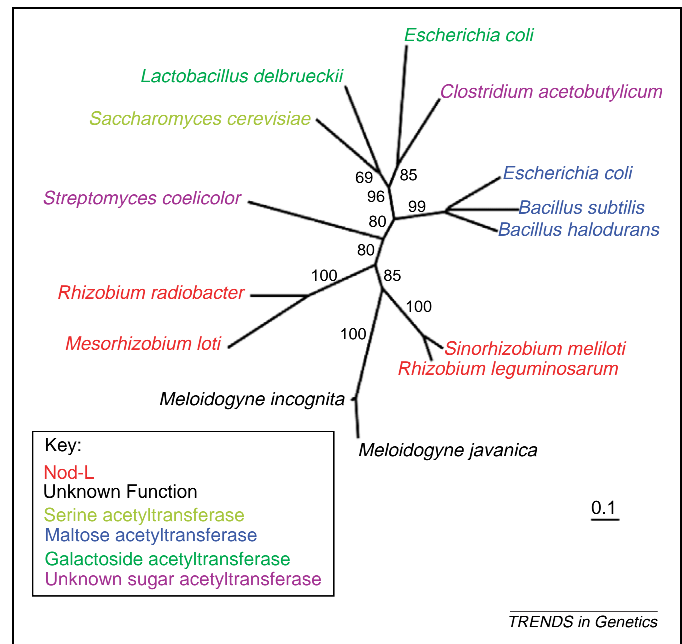

Figure 4. Phylogram of NodL-like proteins from plant parasitic nematodes and bacteria [36]. NodL-like proteins were identified in *Meloidogyne* species using BLAST searches and aligned to a set of homologs from prokaryotes (color-coded according to function on the phylogram). This unrooted tree was generated by neighbor-joining using protein distance measures, with percent support from 1000 nonparametric bootstrap replicates indicated [78]. The scale bar represents 0.1 amino acid replacements per site.

Origins of parasitism and evidence for horizontal gene transfer into plant parasitic nematode genomes

With the multiple origins of parasitism in Nematoda, different strategies and molecular innovations are likely to underlie adaptations in different lineages. Mechanisms that could affect evolution to parasitism include gene duplication and diversification, alterations in genes controlling metabolic and developmental functions, adaptation of pre-existing genes to encode new functions and acquisition of genes from other species [horizontal gene transfer (HGT)]. Although HGT is a widely accepted route for accelerated evolution in prokaryotes [29–31] few substantiated cases have been documented in eukaryotic species, where it remains a controversial subject [32–34].

Accumulating evidence supports a bacterial origin for some genes in plant parasitic nematodes, including cyst nematodes, root-knot nematodes and migratory endoparasites [35]. Among those proposed are genes encoding cellulases and pectinases, enabling the endogenous production of enzymes that can degrade two major components of plant cell walls and could have major roles in parasitism. Most of these genes were identified on the basis of biochemical or immunological criteria, with claims of HGT being supported by phylogenetic incongruency. Additional sequence information from other eukaryotes will help to clarify on a case-by-case basis whether certain candidate genes actually arose by HGT or are derived by descent from an ancient ancestor [34].

A larger set of candidate HGT genes with potential roles in parasitism in root-knot nematodes (*Meloidogyne* spp.) was identified using a phylogenetic incongruency filter and all bacterial sequences available from GenBank. Twelve genes were identified as having been potentially horizontally acquired, including all previously identified HGT candidates for which ESTs were present. Of the newly identified HGT candidates, four have highest similarity to genes in rhizobia – nitrogen-fixing soil bacteria that nodulate plant roots (Figure 4) [36]. Plant parasitic nematodes, free-living ancestral nematodes and rhizobia share the same soil environment, and recent evidence shows that root-knot nematodes signal plants through the same machinery as rhizobia [37]. The remaining eight candidates have putative functions that might be directly related to the ability of these nematodes to parasitize plants, indicating that horizontal gene transfer could be an important route of evolution from free-living nematodes to the parasitic lifestyle. Interestingly, no striking cases of potential HGT from prokaryotes have yet been identified in either free-living nematodes (some of which feed on bacteria) or animal parasitic nematodes, although candidate transfers from a fungal genome into *C. elegans* have been identified [38].

Applying functional genomics tools across the phylum

Beyond the acquisition of sequence data, understanding the biology of nematodes requires determination of the function of gene products in distantly related nematode species on a genome-wide scale. Functional genomics approaches, including RNA interference (RNAi), microarrays and proteomics, are generating genome-wide data, greatly accelerating the rate of functional characterization of genes that had not previously been studied biochemically or genetically.

Introduction of double-stranded RNA (dsRNA) homologous to a target gene can result in a null or hypomorphic

phenotype through post-transcriptional gene silencing. This phenomenon, RNAi, was first demonstrated in *C. elegans* [39], and has subsequently been successfully applied in many eukaryote systems. Although the mechanism responsible for the RNAi response is evolutionarily conserved, the 'spreading' of RNAi to give a systemic response to the localized introduction of dsRNA could be limited to a few species. In *C. elegans*, an RNAi response can be robustly achieved by microinjection [39], soaking [40], or feeding [41] of dsRNA. However, for most *Caenorhabditis* species, including *C. briggsae*, a systemic RNAi response can only be achieved if the dsRNA is delivered by microinjection. It has been suggested that two specific transport machineries are needed for a systemic RNAi response: one for transport of the external dsRNA across epithelia into the organism and a second for further transport to other cells [42].

Initially, RNAi was applied in *C. elegans* to screen individual genes, and then adapted for high-throughput screening of gene sets from whole chromosomes [43,44], cDNA collections [45] and ultimately genome-wide clone sets [46–48]. RNAi information is now available for 19 707 *C. elegans* genes (96% of molecular loci), with 4402 genes (22% of those surveyed) displaying visible phenotypic defects in at least one screen [46–48]. Although such high-throughput screens are incredibly valuable, false negatives are still likely to occur because RNAi is less effective in certain tissues (e.g. the nervous system), only laboratory growth conditions are tested and only a subset of visible phenotypes are accessed. Additional screens have found more phenotypes by using genetic backgrounds sensitized for RNAi such as *rrf-3* [49] or devising methods for detecting more subtle phenotypes such as defects in fat storage [50]. Applying RNAi to parasitic nematodes poses significant additional challenges beyond those faced in a model system like *C. elegans*. The complexity of obligate parasitic life cycles, with movement into and out of the host, makes both the delivery of dsRNA and the assessment of phenotype difficult. Culture systems have been developed making it possible to maintain some parasitic species outside their hosts for significant periods of time. RNAi has been documented in five plant [51–54], one animal [55] and two human parasitic nematodes [56,57]. The dsRNAs were delivered by soaking different developmental stages of these parasitic species (adult female in *B. malayi*, L4 in *N. brasiliensis*, L3 in *Onchocerca volvulus*, L2 in *Heterodera glycines* and *Globodera pallida*, *Globodera rostochiensis*, *Meloidogyne incognita* and the egg in *Meloidogyne artellia*) (Figure 3). In each parasitic species, a few genes have so far been targeted and different efficiencies observed. For example, in *N. brasiliensis* RNAi knockdown of acetylcholinesterase genes suppressed acetylcholinesterase E (AchE) secretory protein levels by >90%, an effect that persisted through six days in culture. In *H. glycines* and *G. pallida*, targeting cysteine proteinase and C-type lectin genes affected subsequent *in planta* development, and targeting the major sperm protein confirmed that the RNAi effect can persist for several days. In *O. volvulus*, RNAi knockdown of cathepsin L and cathepsin Z-like genes reduced the rate of molting from L3 to L4 by 92% and 86%, respectively. In *B. malayi* RNAi targeting two housekeeping genes (the drug target β-tubulin and the RNA polymerase II large subunit) specifically reduced transcript levels and resulted in death of adult female nematodes in culture. RNAi in *B. malayi*, targeting a maternally expressed gene that has a role only in the embryo led to specific transcript reduction and visible defects in 50% of released larvae. Despite these successes in the use of RNAi to investigate gene function in parasitic nematodes, it will be a challenge to develop these methodologies for high-throughput screens. Differing susceptibilities to systemic RNAi between nematodes could also add complexity as additional species are tested.

As an alternative, it might be possible to extrapolate from *C. elegans* RNAi phenotypes to understanding which orthologous genes might have crucial roles in other nematodes, including parasites where faster screening is not yet possible. Using sequence data from parasitic nematodes, *C. elegans* orthologs and homologs with RNAi phenotypes can be identified for many genes of interest. In several species (*M. incognita*, *S. stercoralis*), there is a strong correlation between sequence conservation of a parasite gene and its *C. elegans* homolog and the presence of an RNAi phenotype in *C. elegans* [19,58].

Patterns of gene expression can, in the absence of other functional information, suggest tissue- or stage-specific roles for particular genes of interest. Gene expression can be measured by hybridization of tissue- or stage-specific mRNA to representative microarrays, by the use of serial analysis of gene expression (SAGE), or by assessing EST abundance from specific libraries. For the model nematode *C. elegans*, all three approaches have been used to identify the transcriptomes associated with larval to adult development, aging, dauer larvae, and male and female gametogenesis (e.g. Ref. [59]). Transgenic *C. elegans* expressing a biochemically tagged poly(A)-binding protein under the control of a tissue-specific promoter as a method of isolating RNA from specific cells [60] has permitted the definition of tissue-specific transcriptomes. For other nematode species detailed expression analysis is just beginning. Many of the EST projects involve sampling from stage- or tissue-specific libraries, yielding a rough-grained view of relative expression levels, and these have been used to identify stage-specifically expressed genes [61–63]. Nisbet and Gasser [64] performed a microarray study of gender-specific expression in *Trichostrongylus vitrinus*, a parasitic rhabditine of small ruminants, identifying genes with gender-specific expression patterns and thus with potential roles in gametogenesis, embryogenesis and reproduction. Similar studies investigating gender-specific genes in the human filarial parasite *B. malayi* are underway (B.W. Li and G.J. Weil, unpublished data). cDNA microarray analysis of *Ascaris suum* fourth-stage larvae revealed differential expression of a number of genes during expulsion from the host intestine [65]. A microarray study of *Ancylostoma caninum* using clones from a recent EST project [66] has identified genes differentially expressed between infective and parasitic larval stages [67].

High-throughput methods for separation, quantitation and identification of hundreds of proteins from biological

samples (proteomics) can also help to define the function and expression profiles of all proteins encoded within a given genome. Two-dimensional proteome maps of mixed-stage wild-type *C. elegans* populations under different conditions were published in the late 1990s (e.g. Ref. [68]). More recently temperature-dependent and stage-specific changes in the proteome [69] or germ-line development-related proteins [70] have been determined. Protein interaction maps for *C. elegans*, derived from high-throughput yeast two-hybrid screens, are also being generated [71]. Proteomics tools have been used in parasitic nematodes, mainly to analyze excretory or secretory products (ESPs) that are believed to have roles in formation of the host–parasite complex and in induction of changes in the host cells [72]. Separation of purified secreted proteins from the tylenchine plant parasite *Meloidogyne incognita* by two-dimensional (2D) electrophoresis visualized 40 proteins, of which the seven most abundant were identified by micro-sequencing [73]. More than 100 ESPs from *Haemonchus contortus*, a rhabditine parasite that infects small ruminants, were identified by mass spectrometry, following 2D electrophoresis [74]. These ESPs were also tested for recognition by hyperimmune sera, and the novel immune-recognized ESPs might assist future development of parasite vaccines or therapies for autoimmune diseases [75]. Proteomics approaches have been applied to investigate ESPs in the mammalian skeletal muscle parasite *Trichinella spiralis*, identifying two members of a family of nematode-specific proteins [76]. Proteomics tools have also been used to investigate protein expression patterns in parasitic nematodes exposed to different environmental conditions [77].

Concluding remarks

In less than a decade, the generation of one complete, two draft and >30 partial genomes from nematodes has revolutionized the understanding of these organisms at the molecular level. Comparative genome analyses reveal that despite a uniform body plan, nematodes have more molecular diversity than previously recognized, with much genespace yet to be sampled. The sequence of a single species, such as the model nematode *C. elegans*, provides a crucial starting point, but cannot represent the molecular diversity of Nematoda. Phylogenetically informed genome sequencing is therefore essential for advances in genomics, evolution and infectious disease biology. Areas of particularly rapid evolution in nematodes appear to include intra-chromosomal rearrangements and modifications to secretory proteins, at least in certain parasites. Other parasites might rely on virulence factors acquired by horizontal gene transfer for infection. There will soon be eight more draft nematode genome sequences available. The continued acquisition of genome data and their application through functional tools such as RNAi knockdown, microarrays and proteomics, will be important for expanding understanding of nematode genomic organization, evolution and infectious disease biology.

Acknowledgements

This work was supported by grants from NIH-NIAID and The Wellcome Trust. We thank John Martin and Elizabeth Scholl for technical support. J.P.M. is an employee and equity holder of Divergence Inc.; this research was not company funded.

References

1 The *C. elegans* Sequencing Consortium. (1998) Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* 282, 2012–2018
2 Chen, N. *et al.* (2005) WormBase: a comprehensive data resource for *Caenorhabditis* biology and genomics. *Nucleic Acids Res*. 33 Database Issue, D383–389
3 Parkinson, J. *et al*. (2003) 400 000 nematode ESTs on the Net. *Trends Parasitol.* 19, 283–286
4 Stein, L.D. *et al*. (2003) The genome sequence of *Caenorhabditis briggsae*: a platform for comparative genomics. *PLoS Biol.* 1, E45
5 Ghedin, E. *et al*. (2004) First sequenced genome of a parasitic nematode. *Trends Parasitol.* 20, 151–153
6 McCarter, J.P. *et al*. (2003) Nematode gene sequences: update for December 2003. *J. Nematol.* 35, 465–469
7 Krause, M. and Hirsh, D. (1987) A trans-spliced leader sequence on actin mRNA in *C. elegans*. *Cell* 49, 753–761
8 Mitreva, M. *et al*. (2004) A survey of SL1-spliced transcripts from the root-lesion nematode *Pratylenchus penetrans*. *Mol. Genet. Genomics* 272, 138–148
9 Daub, J. *et al*. (2000) A survey of genes expressed in adults of the human hookworm, *Necator americanus*. *Parasitology* 120, 171–184
10 Blaxter, M. *et al*. (2002) The *Brugia malayi* genome project: expressed sequence tags and gene discovery. *Trans. R. Soc. Trop. Med. Hyg.* 96, 7–17
11 Tetteh, K.K. *et al*. (1999) Identification of abundantly expressed novel and conserved genes from the infective larval stage of *Toxocara canis* by an expressed sequence tag strategy. *Infect. Immun.* 67, 4771–4779
12 Dautova, M. *et al*. (2001) Single pass cDNA sequencing – a powerful tool to analyse gene expression in preparasitic juveniles of the southern root-knot nematode *Meloidogyne incognita*. *Nematology* 3, 129–139
13 Parkinson, J. *et al*. (2004) A transcriptomic analysis of the phylum Nematoda. *Nat. Genet.* 36, 1259–1267
14 Tatusov, R.L. *et al*. (2001) The COG database: new developments in phylogenetic classification of proteins from complete genomes. *Nucleic Acids Res.* 29, 22–28
15 Venter, J.C. *et al*. (2004) Environmental genome shotgun sequencing of the Sargasso Sea. *Science* 304, 66–74
16 Aboobaker, A. and Blaxter, M. (2003) Hox gene evolution in nematodes: novelty conserved. *Curr. Opin. Genet. Dev.* 13, 593–598
17 Aboobaker, A. and Blaxter, M. (2003) Hox gene loss during dynamic evolution of the nematode cluster. *Curr. Biol.* 13, 37–40
18 Harcus, Y.M. *et al*. (2004) Signal sequence analysis of expressed sequence tags from the nematode *Nippostrongylus brasiliensis* and the evolution of secreted proteins in parasites. *Genome Biol.* 5, R39
19 Mitreva, M. *et al*. (2004) Comparative genomics of gene expression in the parasitic and free-living nematodes *Strongyloides stercoralis* and *Caenorhabditis elegans*. *Genome Res.* 14, 209–220
20 Walton, A.C. (1959) Some parasites and their chromosomes. *J. Parasitol.* 45, 1–20
21 Muller, F. and Tobler, H. (2000) Chromatin diminution in the parasitic nematodes *Ascaris suum* and *Parascaris univalens*. *Int. J. Parasitol.* 30, 391–399
22 Etter, A. *et al*. (1991) Eliminated chromatin of *Ascaris* contains a gene that encodes a putative ribosomal protein. *Proc. Natl. Acad. Sci. U. S. A.* 88, 1593–1596
23 Teschke, C. *et al*. (1991) The highly variable pentameric repeats of the AT-rich germline limited DNA in *Parascaris univalens* are the telomeric repeats of somatic chromosomes. *Nucleic Acids Res.* 19, 2677–2684
24 Bachmann-Waldmann, C. *et al*. (2004) Chromatin diminution leads to rapid evolutionary changes in the organization of the germ line genomes of the parasitic nematodes *A. suum* and *P. univalens*. *Mol. Biochem. Parasitol.* 134, 53–64
25 International Chicken Genome Sequencing Consortium. (2004) Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* 432, 695–716
26 Coghlan, A. and Wolfe, K.H. (2002) Fourfold faster rate of genome rearrangement in nematodes than in *Drosophila*. *Genome Res.* 12, 857–867

27 Blaxter, M. (2003) Comparative genomics: two worms are better than one. *Nature* 426, 395–396

28 Whitton, C. *et al*. (2004) A genome sequence survey of the filarial nematode *Brugia malayi*: repeats, gene discovery, and comparative genomics. *Mol. Biochem. Parasitol.* 137, 215–227

29 Jain, R. *et al*. (1999) Horizontal gene transfer among genomes: the complexity hypothesis. *Proc. Natl. Acad. Sci. U. S. A.* 96, 3801–3806

30 Lawrence, J.G. (1999) Gene transfer, speciation, and the evolution of bacterial genomes. *Curr. Opin. Microbiol.* 2, 519–523

31 Ochman, H. *et al*. (2000) Lateral gene transfer and the nature of bacterial innovation. *Nature* 405, 299–304

32 Brinkman, F.S.L. *et al*. (2002) Evidence that plant-like genes in *Chlamydia* species reflect an ancestral relationship between chlamydiaceae, cyanobacteria, and the choloroplast. *Genome Res.* 12, 1159–1167

33 Stanhope, M.J. *et al*. (2001) Phylogenetic analyses do not support horizontal gene transfers from bacteria to vertebrates. *Nature* 411, 940–944

34 Davison, A. and Blaxter, M. (2005) Ancient origin of glycosyl hydrolase family 9 cellulase genes. *Mol. Biol. Evol.* 22, 1273–1284

35 Uehara, T. *et al*. (2001) PCR-based cloning of two beta-1,4-endoglucanases from the root-lesion nematode *Pratylenchus penetrans*. *Nematology* 3, 335–341

36 Scholl, E.H. *et al*. (2003) Horizontally transferred genes in plant-parasitic nematodes: a high-throughput genomic approach. *Genome Biol.* 4, R39

37 Weerasinghe, R.R. *et al*. (2005) Root-knot nematodes and bacterial Nod factors elicit common signal transduction events in *Lotus japonicus* root hair cells. *Proc. Natl. Acad. Sci. U. S. A.* 102, 3147–3152

38 Parkinson, J. and Blaxter, M.L. (2003) SimiTri – visualizing similarity relationships for groups of sequences. *Bioinformatics* 19, 390–395

39 Fire, A. *et al*. (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* 391, 806–811

40 Tabara, H. *et al*. (1998) RNAi in *C. elegans*: soaking in the genome sequence. *Science* 282, 430–431

41 Timmons, L. and Fire, A. (1998) Specific interference by ingested dsRNA. *Nature* 395, 854

42 Feinberg, E.H. and Hunter, C.P. (2003) Transport of dsRNA into cells by the transmembrane protein SID-1. *Science* 301, 1545–1547

43 Fraser, A.G. *et al*. (2000) Functional genomic analysis of *C. elegans* chromosome I by systematic RNA interference. *Nature* 408, 325–330

44 Gonczy, P. *et al*. (2000) Functional genomic analysis of cell division in *C. elegans* using RNAi of genes on chromosome III. *Nature* 408, 331–336

45 Maeda, I. *et al*. (2001) Large-scale analysis of gene function in *Caenorhabditis elegans* by high-throughput RNAi. *Curr. Biol.* 11, 171–176

46 Kamath, R.S. *et al*. (2003) Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. *Nature* 421, 231–237

47 Rual, J.F. *et al*. (2004) Toward improving *Caenorhabditis elegans* phenome mapping with an ORFeome-based RNAi library. *Genome Res.* 14, 2162–2168

48 Sonnichsen, B. *et al*. (2005) Full-genome RNAi profiling of early embryogenesis in *Caenorhabditis elegans*. *Nature* 434, 462–469

49 Simmer, F. *et al*. (2003) Genome-wide RNAi of *C. elegans* using the hypersensitive rrf-3 strain reveals novel gene functions. *PLoS Biol.* 1, E12

50 Ashrafi, K. *et al*. (2003) Genome-wide RNAi analysis of *Caenorhabditis elegans* fat regulatory genes. *Nature* 421, 268–272

51 Urwin, P.E. *et al*. (2002) Ingestion of double-stranded RNA by preparasitic juvenile cyst nematodes leads to RNA interference. *Mol lant-Microbe Interacts* 15, 747–752

52 Fanelli, E. *et al*. (2005) Analysis of chitin synthase function in a plant parasitic nematode, *Meloidogyne artellia*. *Gene* 349, 87–95

53 Rosso, M.N. *et al*. (2005) Application of RNA interference to root-knot nematode genes encoding esophageal gland proteins. *Molecular Plant Microbe Interactions* 18, 615–620

54 Chen, Q. *et al*. (2005) Functional analysis of pathogenicity proteins of the potato cyst nematode *Globodera rostochiensis* using RNAi. *Molecular Plant Microbe Interactions* 18, 621–625

55 Hussein, A.S. *et al*. (2002) Suppression of selected acetylcholinesterase expression in *Nippostrongylus brasiliensis* by RNA interference. *Mol. Biochem. Parasitol.* 122, 91–94

56 Aboobaker, A.A. and Blaxter, M.L. (2003) Use of RNA interference to investigate gene function in the human filarial nematode parasite *Brugia malayi*. *Mol. Biochem. Parasitol.* 129, 41–51

57 Lustigman, S. *et al*. (2004) RNA interference targeting cathepsin L and Z-like cysteine proteases of *Onchocerca volvulus* confirmed their essential function during L3 molting. *Mol. Biochem. Parasitol.* 138, 165–170

58 McCarter, J. *et al*. (2003) Analysis and functional classification of transcripts from the nematode *Meloidogyne incognita*. *Genome Biol.* R26, 1–19

59 Kim, S.K. *et al*. (2001) A gene expression map for *Caenorhabditis elegans*. *Science* 293, 2087–2092

60 Roy, P.J. *et al*. (2002) Chromosomal clustering of muscle-expressed genes in *Caenorhabditis elegans*. *Nature* 418, 975–979

61 Gregory, W.F. *et al*. (1997) Differentially expressed, abundant trans-spliced cDNAs from larval *Brugia malayi*. *Mol. Biochem. Parasitol.* 87, 85–95

62 Li, B.W. *et al*. (2004) Quantitative analysis of gender-regulated transcripts in the filarial nematode *Brugia malayi* by real-time RT-PCR. *Mol. Biochem. Parasitol.* 137, 329–337

63 Zang, X. *et al*. (1999) A novel serpin expressed by blood-borne microfilariae of the parasitic nematode *Brugia malayi* inhibits human neutrophil serine proteinases. *Blood* 94, 1418–1428

64 Nisbet, A.J. and Gasser, R.B. (2004) Profiling of gender-specific gene expression for *Trichostrongylus vitrinus* (Nematoda: Strongylida) by microarray analysis of expressed sequence tag libraries constructed by suppressive-subtractive hybridisation. *Int. J. Parasitol.* 34, 633–643

65 Morimoto, M. *et al*. (2003) *Ascaris suum*: cDNA microarray analysis of 4th stage larvae (L4) during self-cure from the intestine. *Exp. Parasitol.* 104, 113–121

66 Mitreva, M. *et al*. (2005) Investigating hookworm genomes by comparative analysis of two *Ancylostoma* species. *BMC Genomics* 6, 58

67 Moser, J.M. *et al*. (2005) Gene expression profiles associated with the transition to parasitism in *Ancylostoma caninum* larvae. *Mol. Biochem. Parasitol.* 143, 39–48

68 Bini, L. *et al*. (1997) Two-dimensional gel electrophoresis of *Caenorhabditis elegans* homogenates and identification of protein spots by microsequencing. *Electrophoresis* 18, 557–562

69 Madi, A. *et al*. (2003) Mass spectrometric proteome analysis for profiling temperature-dependent changes of protein expression in wild-type *Caenorhabditis elegans*. *Proteomics* 3, 1526–1534

70 Bantscheff, M. *et al*. (2004) Differential proteome analysis and mass spectrometric characterization of germ line development-related proteins of *Caenorhabditis elegans*. *Proteomics* 4, 2283–2295

71 Li, S. *et al*. (2004) A map of the interactome network of the metazoan *C. elegans*. *Science* 303, 540–543

72 Kwan-Lim, G.E. *et al*. (1989) Secreted antigens of filarial nematodes: a survey and characterization of *in vitro* excreted/secreted products of adult *Brugia malayi*. *Parasite Immunol.* 11, 629–654

73 Jaubert, S. *et al*. (2002) Direct identification of stylet secreted proteins from root-knot nematodes by a proteomic approach. *Mol. Biochem. Parasitol.* 121, 205–211

74 Yatsuda, A.P. *et al*. (2003) Comprehensive analysis of the secreted proteins of the parasite *Haemonchus contortus* reveals extensive sequence variation and differential immune recognition. *J. Biol. Chem.* 278, 16941–16951

75 Summers, R.W. *et al*. (2005) *Trichuris suis* therapy in Crohn's disease. *Gut* 54, 87–90

76 Gare, D. *et al*. (2004) Developmental regulation and secretion of nematode-specific cysteine-glycine domain proteins in *Trichinella spiralis*. *Mol. Biochem. Parasitol.* 134, 257–266

77 Islam, M.K. *et al*. (2004) The proteome expression patterns in adult *Ascaris suum* under exposure to aerobic/anaerobic environments analyzed by two-dimensional electrophoresis. *Parasitol. Res.* 93, 96–101

78 Felsenstein, J. (1985) Confidence limits on phylogenies: an approach using the bootstrap. *Evolution Int. J. Org. Evolution* 39, 783–791

79 Platt, H.M. (1994) Foreword. In *The Phylogenetic Systematics of Free-living Nematodes* (Lorenzen, S., ed.), pp. i–ii, The Ray Society

80 Blaxter, M.L. *et al*. (1998) A molecular evolutionary framework for the phylum Nematoda. *Nature* 392, 71–75

81 Sakaguchi, Y. *et al*. (1983) Karyotypes of *Brugia pahangi* and *Brugia malayi* (Nematoda: Filarioidea). *J. Parasitol.* 69, 1090–1093

82 Siddiqi, M.R., ed. (2000) *Tylenchida: Parasites of Plants and Insects*, Commonwealth Institute of Parasitology