View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Explore Bristol Research

Viney, M. (2018). The genomic basis of nematode parasitism. *Briefings in Functional Genomics*, *17*(1), 8-14. [elx010]. https://doi.org/10.1093/bfgp/elx010

Peer reviewed version

Link to published version (if available):
10.1093/bfgp/elx010

Link to publication record in Explore Bristol Research
PDF-document

This is the author accepted manuscript (AAM). The final published version (version of record) is available online via Oxford University Press at https://academic.oup.com/bfg/article-lookup/doi/10.1093/bfgp/elx010. Please refer to any applicable terms of use of the publisher.

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:
http://www.bristol.ac.uk/pure/about/ebr-terms

OXFORD UNIVERSITY PRESS | Briefings in Functional Genomics

The Genomic Basis of Nematode Parasitism

Journal:	*Briefings in Functional Genomics*
Manuscript ID	Draft
Manuscript Type:	Review Paper
Date Submitted by the Author:	n/a
Complete List of Authors:	Viney, Mark; University of Bristol, School of Biological Sciences
Keywords:	nematode, parasitism

SCHOLARONE™ Manuscripts

For Review

The Genomic Basis of Nematode Parasitism

Mark Viney

School of Biological Sciences, University of Bristol, BS8 1TQ, UK.

Contact: Mark.Viney@bristol.ac.uk

Phone: 0117 394 1203

Key words. Nematode, Parasite, Genomics, Evolution

For Peer Review

Abstract

Nematodes are highly abundant animals, and many species have a parasitic lifestyle. Nematode parasites are important pathogens of humans and other animals, and there is considerable interest in understanding their molecular and genomic adaptations to nematode parasitism. This has been approached in three main ways: comparing the genomes of closely related parasitic and free-living taxa, comparing the gene expression of parasitic and free-living life-cycle stages of parasitic nematode species, and analysing the molecules that parasitic nematodes excrete and secrete. To date these studies show that many species of parasitic nematodes have genomes that have large gene families coding for proteases / peptidases, protease inhibitors, SCP/TAPS proteins and acetycholinesterases, and in many cases there is evidence that these appear to be used by parasitic stages inside hosts, and are often secreted. Many parasitic nematodes have taxa-specific gene families that also appeared to be involved in parasitism, emphasising that there is still much to be discovered about what it takes to be a parasitic nematode.

For Peer Review

There are many nematode parasites

Nematodes are the most abundant and species extant group of metazoans. Memorably, some 100 years ago Nathan Cobb said of nematodes: "*In short, if all the matter in the universe except the nematodes were swept away, our world would still be dimly recognizable … we should find its mountains, hills, vales, rivers, lakes, and oceans represented by a film of nematodes. The location of towns would be decipherable, since for … human beings there would be corresponding massing of certain nematodes. Trees would still stand in ghostly rows representing our streets and highways. The location of various plants and animals would still be decipherable…"* [1]. He made the point well, that nematodes are ubiquitous, if rather rarely seen. As Cobb also makes plain, they are also very common parasites, infecting all manner of multicellular animals and plants, in all terrestrial and aquatic environments, such that being parasitized by nematodes is a normal feature of life.

Parasitic nematodes are diverse in their biology and life histories, so it can be difficult to make generalisations. However, by way of introduction, let us consider the most common nematode parasite of humans, *Ascaris lumbricoides*. Adult male and female worms (females are about 30 cm long, 5 mm in diameter; males are a little smaller) live in the host gut where they reproduce. Females lay eggs – some 250,000 per day – and these pass out of the host in its faeces. *Ascaris* eggs have a resistant egg shell and so the eggs persist in the environment for considerable periods of time. Hosts become infected when they accidently ingest eggs. Once inside the host gut the eggs hatch to release a larva; this leaves the gut and migrates around the host body before returning to the gut where the worms then mature as adults, reproduce, and so the cycle continues. Other species of nematodes that parasitize vertebrates can live in other within-host sites, including blood vessels, the lymphatic system, and within the tissues. Parasitic nematodes have a myriad of different life cycles; for example, where larvae (rather than eggs) infect hosts; where worms are transmitted among hosts by arthropod vectors (in which case the parasite then lives in two different host species during its life), and where worms are transmitted when a host predates upon an infected host. Nematodes parasitize vertebrates, invertebrates and plants, though studies are biased towards those that parasite vertebrates – a bias that will persist in this article.

Particular challenges for parasitic nematodes, to which they are adapted, are surviving within the host and transmitting among hosts. For parasites of vertebrates an important

For Peer Review

challenge is surviving the host anti-parasite immune response. There are two general strategies that parasitic nematodes use. The first is to evade the host immune response either by using a molecular disguise, by protecting themselves from host immune system effector molecules, or by living within certain within-host niches that are less immunologically exposed. The second, not mutually exclusive approach, is immunomodulation of the host, which appears to be very widespread among nematodes [2]. To immunomodulate their hosts nematodes release molecules into the host, and these interact with the host immune system to alter the host's immune response to the parasite's advantage. Though these immunomodulation phenomena are now well known, the detailed molecular mechanisms underlying these processes are currently largely unknown.

Parasitic nematodes cause immense harm to humans and other animals. More than 1.5 billion people are infected with gastrointestinal nematode parasites, with this concentrated in the young, poor of the developing world [3, 4], and as such the World Health Organization recognises these infections as a neglected tropical disease [5]. The WHO estimates that in 2012, worldwide intestinal nematode infection caused a disease burden, measured by the Years Lost due to Disability (YLD), of 5 million YLD. This loss is greater than for other infectious diseases – 4 million for malaria, 4.5 million for HIV/AIDS [6].

Nematode parasitism of agricultural animals is also very common and its control is necessary to maintain productivity and profitability [7, 8]. The production losses caused by parasitic nematode infection translates into economic costs: in the UK the estimated annual cost of nematode infection of sheep is €99 million. Across the EU the annual sales of anthelmintic drugs to treat nematode infections total approximately €400 million [8], but it is likely that this is actually only a small fraction of the true costs [9].

Nematodes are diverse

Nematodes have a conservative morphology making traditional approaches to their taxonomy frustratingly hard, even for experts. Perhaps perversely, the nematodes are actually a highly diverse group of organisms. In molecular phylogenetic analyses nematodes often have considerably extended phylogenetic branches compared with other taxa, pointing to nematodes' elevated molecular evolutionary rates [10, 11]. One consequence of this is that many nematode species (and other taxonomic groupings) have genes and gene families that are specific to them [12].

For Peer Review

This diversity is also seen in nematode genomes. *C. elegans* was the first nematode whose genome was sequenced, revealing a 100 Mb genome containing about 20,000 genes. Since then the genomes of other nematodes have also be sequenced, particularly those of parasitic species. There are currently about 25 nematode genome sequences published, with about 100 projects on-going [13]. This work has already revealed substantial diversity in the size and content of nematode genomes. For example, sequenced nematode genomes now range from *c.* 20 Mb in *Pratylenchus coffeae* (a pest of bananas and other crops) coding for about 7,000 genes, to 370 Mb in *Haemonchus contortus* (a gastrointestinal parasite of sheep) coding for *c.* 20,000 genes [14, 15]. This variation in genome size is due to changes in gene number, as well as changes in the size and number of introns, and in the extent of non-coding, intervening sequence, including repetitive sequence [15–18]. It is also interesting to note that even within the *Caenorhabditis* genus that there are substantial differences in genome size and gene number. For example, genome size differs more than two-fold ranging from 79 Mb in *C. tropicalis* to 190 Mb in *C. brenneri*, while estimates of gene number range from the low of *C. elegans* to *c.* 35,000 in *C. sinica*. Despite these substantial differences among nematode species, signals of the conservation of gene order across large nematode distances can still be seen [15, 17, 19–21].

Phylogenetic analysis of nematodes has shown that they have independently evolved parasitism on up to 18 different occasions [22, 23]. Consequently, many parasitic nematodes have close relatives that have a free-living lifestyle. Species that parasitize different hosts (vertebrates, invertebrates, plants) are also often in the same nematode clades, perhaps suggesting that there is some deep conservation of adaptations to a parasitic lifestyle independent of the type of host being parasitized.

Approaches to discovering the genetic and genomic basis of parasitism in nematodes

Parasitism is biologically fascinating, and so there is very considerable interest in understanding the adaptations that underlie the parasitic lifestyle. This basic biological interest is, of course, supplemented by a desire to understand the biology of these parasites so as to find ways to control them and the harm that they cause.

Prior to the advent of genomics a whole series of physiological and molecular studies were aimed at investigating aspects of parasitic nematodes' biology which, *a priori,* were likely

For Peer Review

to be different in parasitic nematodes, compared with free-living species. So, for example, the metabolism of many nematode parasites of the vertebrate gut was investigated, given the likely special gaseous, pH *etc.* conditions in this environment. Similarly, because parasitic nematodes have to survive the host immune response there was considerable interest directed to understanding the structure and composition of parasitic nematodes' surface cuticle and its interaction with the host immune response.

In pursuing the aim of understanding the genetic basis of nematode parasitism there is an underlying rationale that all such studies use, and it is useful to make this rationale explicit. It is this: parasitic nematodes have evolved from free-living ancestors and as this has happened parasites will have adapted existing traits of their free-living nematode ancestor, and evolved new traits, which together underlie the parasitic lifestyle. Therefore, comparing free-living nematodes with parasitic nematodes can be used to uncover the molecular basis of nematode parasitism. This is the first type of genomic study used to understand nematode parasites. A second approach, but one using the same fundamental rationale, is to compare, within a species, parasitic life cycle stages with free-living life cycle stages. Finally, the third approach that is used seeks to discover which molecules parasitic stages likely excrete / secrete (*i.e.* ES material), with the rationale that many of these molecules will interact with the host in prosecuting a parasitic lifestyle.

1. Comparing the genomes of parasitic and free-living nematodes

This is potentially an enormously powerful approach to understand the genomic basis of nematode parasitism, but care must be taken in making the correct comparisons given the multiple, independent origins of nematode parasitism. Specifically, nematodes occur in five phylogenetic clades [22], with many clades containing both parasitic and free-living taxa. The most appropriate and powerful phylogenomic comparisons are those that compare taxa within the clades, rather than between clades. This is because parasitic *vs.* free-living comparisons across different clades confounds differences in lifestyles with the different evolutionary history of those clades. In contrast, comparisons between taxa with different lifestyles within the same clade can reasonably be used to infer the genomic bases of those differences in lifestyle. Given the very extensive genomic diversity among the nematodes, it is arguable that we currently have too few nematode genomes to make strong inferences about common genomic themes underlying parasitism across different nematode clades [13, 24]. Instead, more phylogenetically local, focused comparisons are probably more appropriate.

For Peer Review

Because *C. elegans* was the first nematode genome to be sequenced, by necessity many studies of the genomes of parasitic nematodes have only been able to use *C. elegans* available as a comparator species. However, important parasitic species are in the same clade – for example, the hookworms *Necator* spp. and *Ancylostoma* spp., parasites of livestock such as *Haemonchus* spp., together with common laboratory models such as *Heligmosomoides polygyrus* and *Nippostrongylus brasiliensis* [22].

One particularly useful genomic comparison between parasitic and free-living species involves six species within one sub-clade of nematodes. Specifically, this compared four obligate con-generic parasitic species (*Strongyloides* spp.), one facultative parasitic species (*Parastrongyloides trichosuri*), and one free-living species (*Rhabditophanes* sp.) all of whom are closely phylogenetically related (**Figure 1**). Comparison of these species' genomes was used to infer where genes and gene families arose, and where genes and gene families were lost, during the evolutionary history of the species and, crucially, as parasitism evolved (**Figure 1**). This analysis identified over 1,000 gene families that were evolutionarily acquired as the parasitic genera *Parastrongyloides* and *Strongyloides* evolved. Of these gene families, the two largest code for astacin-like zinc-metallopeptidases and SCP/TAPS proteins (known too by many other names, including CAP-domain, ASP and VAL proteins [25]). Specifically, the parasitic species *S. ratti* and *S. stercoralis* have 184 and 237 astacin-like metallopeptidase coding genes, compared with 36 in the free-living relative *Rhabditophanes* (*C. elegans* has 40). For SCP/TAPS proteins, there were 89 and 113 coding genes in *S. ratti* and *S. stercoralis*, respectively, compared with 12 in *Rhabditophanes* (*C. elegans* has 36) [18].

The observed comparative expansion of astacin-like metallopeptidase coding genes in these parasitic taxa is consistent with similarly large gene families coding for peptidases and proteases in other parasitic nematode species [12]. For example, the *Ascaris suum* genome has 456 peptidase coding genes, with those coding for metallopeptidases and serine proteases predominating. The *Toxocara canis* genome has many similar genes too – 165, 107 and 60 (total 332) coding for metallo- cysteine and serine peptidases, respectively, together representing 89 % of all the predicated peptidase coding genes in this species [26]. In the whipworm *Trichuris suis* there are large numbers of genes coding for two classes of peptidases (116 for class S1 and 42 for class S8) [27]; *Dictyocaulus viviparus* has 478 protease coding genes [28]. Together, these results begin to suggest a

For Peer Review

common theme: that genomes of a range of parasitic nematodes have very many genes coding for proteases or peptidases [24], though the specific class of proteases / peptidase coded for appears to not be particularly consistent among different taxa.

The possible role for such an array of proteases and peptidases in nematode parasites is not known, but there are two main ideas for the role that they might play. Firstly, they might be used by the parasite to digest components of the host, at least in part to acquire food. *Strongyloides* spp., as many other parasitic nematode species, undergoes a within-host migration before finally settling in the gut. This migration requires the breakdown of host tissue, which presumably these parasite-produced enzymes achieve. Furthermore, *Strongyloides* actually lives within and continuously burrows through the mucosa of the host's small intestine [29]. Here the parasite is presumably digesting host tissue, which is then used as a food source by the parasite. In *Strongyloides* it is likely that the metallopeptidases are secreted by the parasite to achieve this. Other species of parasitic nematodes feed on the host in other ways, and the ability to digest host proteins would therefore seem to be a key component of living a parasitic lifestyle. The second potential role of proteases in the parasitic lifestyle concerns immunomodulation of their hosts. Here the idea is that these proteases act against, and disable, protein components of the host immune response, which thereby facilitates the parasites' survival.

Turning to the second group of genes that are comparatively expanded in *Strongyloides*, compared with free-living species, the SCP/TAPS proteins. The expansion of this gene family in *Strongyloides* also chimes with analyses of genomes of other species [12]. For example, the hookworms *Ancylostoma ceylanicum* and *Necator americanus* have 432 and 137 of these genes [30, 31], and *Haemonchus contortus* 161, compared with 36 in the free-living species *C. elegans* (and as a further comparison, there are 33 in *Pristionchus pacifiicus*, which lives in close association with beetles, but does not appear to be a parasite [32]). This gene family is also reported to be expanded in another clade V nematode, *Dictyocaulus viviparus* [28].

There has been substantial interest in the SCP/TAPS proteins in parasitic nematodes, in large part because these molecules appear to be abundant in nematode parasites [25], with some of these proteins being immunodominant. Despite this notoriety, what these molecules actually do in parasitic nematodes remains far from clear. *In vitro* evidence is consistent with some of these molecules interacting with components of the host immune

For Peer Review

responses [33]. In *Heligmosomoides polygyrus* and *Ancylostoma caninum* SCP/TAPS molecules are on (and closely below) the parasite surface, as well as associated with the parasite's gut and oesophagus. This is consistent with these molecules having a role in immunomodulating the host, but also consistent with interacting with the host more widely [25, 34]. More generally, related molecules occur in a wide variety of non-parasitic species, perhaps suggesting that the conserved features of these proteins are simply domains that can be usefully deployed in many settings [25].

Within the *Strongyloides* clade of nematodes, other gene families that expanded coincident with the evolution of parasitism include acetylcholinesterase-coding genes (33 and 34 in *S. ratti* and *S. stercoralis*, respectively, compared with 3 in *Rhabditophanes*; *C. elegans* has 4), receptor-type protein tyrosine phosphatase-coding genes (83 and 75 in *S. ratti* and *S. stercoralis*, respectively, compared with 21 in *Rhabditophanes*; *C. elegans* has 59). The role of parasitic nematode-produced acetycholinesterases in the parasitic lifestyle is also not fully understood. In some species of gastrointestinal parasitic nematodes these molecules are secreted into the host gut where they have been thought to play a role in modulating the activity of the gut – including muscular contraction of the gut, as well as secretion of mucus and other fluid – with this facilitating parasite survival [35]. However, more recently parasite acetycholinesterases have also been found to have an immunoregulatory role [36]. Careful assessment of this large family of genes in *Strongyloides* very strongly suggests that, when translated, many of these gene products will be enzymatically inactive [37]. One possibility, therefore, is that this comparatively expanded gene family is a functionless family, though this then begs the question of why such a large family of putatively non-functional proteins became expanded and are maintained in a genome, and why they appear to be secreted by the parasitic stages of the life cycle. An alternative explanation is that these genes are coding for proteins with other functionalities that are completely unrelated to the acetylcholinesterase-like enzymatic function.

Importantly, in many of these genomic comparisons between parasitic and free-living species, gene families whose products are (i) hitherto unknown and (ii) unique to the taxa in which they are described have been found [12]. Given the diversity among nematode taxa, the existence of taxa-specific gene families should be expected. Indeed, such taxa-specific gene families should be as likely to occur among free-living species, as well as

For Peer Review

parasitic. Because of the taxa-specific nature of these gene families, their possible role in parasitism is unclear, though this is clearly a research challenge for the future.

2. Comparing parasitic and free-living stages

This approach exploits the fact that many species of parasitic nematode have life cycles where some stages are parasitic and some are free-living (with these used for transmission among hosts). Comparing gene expression in parasitic and free-living life cycle stages might therefore identify the genes and proteins used specifically (or more) by the parasitic stages, which can then be used to infer what structures, metabolism, molecular processes *etc*. underlie the parasitic lifestyle of the species concerned.

Interpreting results from such comparisons can be problematic, for two related reasons. Firstly, these comparisons involve both different life cycle stages (for example larvae and adults) and different lifestyles (free-living *vs.* parasitic). Put formally, such comparisons therefore confound life cycle stage and lifestyle. This is a real problem because there can be considerable differences in the biology of different nematode life cycle stages, even when they are all living in the same environment. This can be seen using an example from the life cycle of the free-living nematode *C. elegans*. It has about 170 genes that code for collagen proteins, which are the principal component of its cuticle (as for all nematodes). Nematodes, including *C. elegans*, moult through four larval stages, before moulting into adult stages, and at each moult a new cuticle is formed. In *C. elegans* different groups of collagen coding genes are used by each life cycle stage. Therefore transcriptomic comparison of different *C. elegans* life cycle stages would reveal life cycle stage-specific differences in the transcription of collagen coding genes. If these different life cycle stages had different lifestyles (such as free-living and parasitic) then we might wrongly infer that collagen molecules were an adaptation to the parasitic lifestyle. A second, related problem of comparing transcriptomes or proteomes between different nematode life cycle stages is that the results obtained depend on what is compared with what.

Notwithstanding these caveats, comparative transcriptomic methods have been widely used, mainly focused on two within-host parasitic stages: (i) larval stages that infect hosts, and (ii) adult stages living and reproducing within hosts. The results are difficult to synthesise and summarise because of the diversity of comparisons made among many different parasitic taxa. In studies of infective larvae newly arrived within a host, increased transcription of genes whose products are involved in growth, metabolism, sensory

For Peer Review

processes *etc.* are observed (for example in *D. viviparous* and *H. contortus* [15, 28]), consistent with the parasites resuming growth and development once entry into a host has been achieved. The expression of genes coding for proteases, and protease inhibitors, is also often seen (for example in *Ancylostoma* spp. and *Ascaris suum* [21, 31, 38]) presumably reflecting the need of migratory larvae stages to digest host tissue as they migrate.

For parasitic adult stages, there is also a comparative increase in the expression of genes coding for a range of enzymes including proteases, peptidases, hydrolases, catalases (for example in *N. americanus, H. contortus, Toxocara canis* and *Ascaris suum* [15, 21, 26, 30]), and protease inhibitors. Also, consistent with adult stages growing and reproducing, there is a comparative increase in transcription of genes whose products are involved in DNA replication, cuticle formation, spermatogenesis.

A more nuanced analysis of the hookworm *Ancylostoma ceylanicum* compared parasitic stages at different times after infection, showing that in worms' first few days within a host SCP/TAPS proteins had comparatively enhanced transcription, but in more established worms there was a comparative increase in the transcription of genes whose products are likely involved in growth (for example genes coding for cuticle components, and proteins that are able to bind cytoskeleton proteins [31]). Later still in infection the transcription profile alters towards genes coding for protein tyrosine phosphatases, serine / threonine kinases, and C-lectins. These observations make the important, more general point that parasitic nematodes live a dynamic life where aspects of their biology change as the within-host environment alters. For example, hosts have changing physiological states (*e.g.* reproductive state, nutritional state, seasonal changes *etc.*) as well as different infections and co-infections [39] with consequent effects on the host immune response, all of which changes the within-host environment to which parasitic nematodes are exposed.

For *Strongyloides* nematodes, an unusual feature of their life cycle largely overcomes the problem of confounding lifestyle with life cycle stage. This is because the *Strongyloides* life cycle has two adult generations – one parasitic (which is female only) and one free-living (which is dioecious). In this case, comparison of parasitic adult female worms with free-living adult female worms compares lifestyle, without life cycle stage being a confounding factor. Comparison of the transcriptome of these free-living and parasitic adults of *S. ratti* showed that the parasitic females had significantly enhanced transcription of astacin-

For Peer Review

metallopeptidase and SCP/TAPS coding genes – the same gene families that are comparatively enlarged in these genomes [18]. In addition to this, genes coding for transthyretin-like proteins, prolyl endopeptidases, aspartic peptidases, acetylcholinesterases, trypsin inhibitors also had comparatively greater expression in parasitic females [18]. Importantly, up to a third of all genes differentially expressed between these stages were hitherto unknown, hypothetical protein-coding genes, including some gene families that appear to be unique to these parasitic taxa [18], with similar phenomena reported in other species too [40].

For some larger parasitic adult worms gene transcription in different regions of the worms' bodies have been analysed [41]. For example, the whipworms *Trichuris* have a structurally modified anterior end (called the stichosome), and this anterior end lies embedded within the hosts' mucosa, while the worms' posterior end lies free in the host gut lumen into which worm eggs are then easily liberated. Transcriptional analysis of the stichosome (compared with the rest of the adult worm) of *T. suis* shows comparatively greater expression of many genes coding for peptidases and for porins (along with many other genes), consistent with the idea that the stichosome function is to allow the worm to burrow into host tissue [27, 42]. The transcriptional profile of the gut of two parasitic nematodes (*H. contortus* and *T. canis*, [15, 26]) has also shown that genes coding for a range of protease / peptidases (and protease and peptidase inhibitors), and products involved in binding with and transporting a range of molecules, are comparatively upregulated, consistent with the *a priori* view of the function of the worm's gut.

In summary, studies that have sought to discover the genes specifically used by within-host parasitic stages have shown that many different gene families are used. Often these genes belong to families that appear to be comparatively expanded in parasitic species, strengthening the inference that these genes have a special role in parasitism. Beyond this, many of the other genes specifically expressed by these adult stages are consistent with growth and reproduction, which is clearly the principal biological function of within-host adult parasitic stages.

3. Secreted molecules

Identifying the molecules that nematode parasites secrete into their hosts aims to understand the interactive interface between parasites and their hosts. Two approaches are used: proteomic analysis of excretory / secretory (ES) molecules, and genomic

For Peer Review

analysis of genes whose products are predicted to be secreted.

There has been relatively limited proteomic analysis of parasitic nematodes. Of particular note are the analysis of the ES of *Heligmosomoides polygyrus*, showing that this contained more than 400 ES-specific proteins. SCP-TAPS proteins dominated the ES, but they also contained proteases, lysozymes, apyrases and acetycholinesterases [25]. In the hookworm *Ancylostoma caninum* its ES also contained similar groups of proteins, but lectins and galectins too [40]. Given that different species of parasitic nematodes have different within-host niches, and so different biology, the molecules they secrete to interact with their host are likely to vary in respect of their different biology. For *Haemonchus contortus*, proteomic analyses of its ES has focused on those proteins that are immunologically recognized by hosts [43]. A wide variety of molecules were identified including, once again, proteases. However, notably, the most immunogenic molecules in these ES was a transthyretin-like protein [43]. In *S. ratti* the secreted proteome of parasitic and free-living females, and infective larvae was compared, finding a range of proteins specifically secreted by the parasitic stage, including proteases, prolyl endopeptidases and acetylcholinesterases [44].

Proteomic analyses are the most direct way of characterising parasites' ES. However, many studies have used a transcriptomic approach where ES molecules are computationally predicted based on the presence of signal peptides and the absence of trans-membrane domains – a so-called secretome – though, of course, this approach does not only identify those proteins specifically secreted outside of the worm. While this is a widespread approach, there is a worrying empirical discordance between protein and transcript abundance, probably because of post-translational modifications which decouples the abundance of mRNA transcripts from the quantity of protein present [45–47].

Many of the results from analyses of predicted secretome overlap with results obtained from the two other approaches (above) used to investigate nematode parasitism. The predicted secretome can be large – in *Necator americanus* it is a third of the entire proteome. In this species, genes whose expression was comparatively greater in parasitic stages are more likely to be predicted secretory molecules (*N. americanus, S. ratti*) [18, 30].

For Peer Review

In many species, proteases and peptidases are predicted to be secreted (*Ascaris suum, Toxocara canis, Trichuris suis, S. ratti* [18, 21, 26, 27]); in *N. americanus* over half (325 of 592) of all proteases fall into this category, and 19 % of the 478 of *Dictyocaulus viviparus* [28]. Some SCP/TAPS proteins are also predicted to be secreted (in *Ascaris suum*, *Toxocara canis, S. ratti, D. viviparus*). But, species also differ in their predicted secretome. For example in *Ascaris suum*, o-linked glycosylated proteins are the single largest group of predicted secreted proteins, while other species have a diversity of molecules often implicated in immunomodulatory roles [21].

The genomic basis of nematode parasitism and prospects for the future

While impressive advances have been made in interrogating nematode genomes these studies are still essentially at an early stage, especially given the diversity that exists among the nematodes. Notwithstanding this, it appears that the genomes of many parasitic nematodes are abundantly resourced with some particular gene families – those coding for proteases / peptidases, protease inhibitors, SCP/TAPS proteins, acetycholinesterases *etc*. – and in many cases these genes appear to be used especially during the within-host parasitic stages, and are often secreted [18, 44]. The apparent preponderance of proteases / peptidases and protease inhibitors makes the, perhaps obvious, point that parasitic nematodes make their living by managing and manipulating host proteins and that, similarly, hosts seek to interact with parasite proteins (which parasites, in turn, aim to resist). The apparent widespread occurrence of SCP/TAPS proteins in parasitic nematodes is fascinating, but it remains frustrating that we have such a limited understanding of what these molecules might do.

One consistent, if somewhat opaque, finding in many of these studies is that there are often taxa-specific genes and gene families, and that these often have many of the hallmarks of being involved in the parasitic lifestyle. Discovering what the products of these genes do is clearly an important task to be able to fully understand the biology of nematode parasitism. This also makes the more general point that different groups of nematodes' genomes are specialised in different ways for a parasitic lifestyle, and in the molecules they use and secrete as they interact with the host. While this diversity of function might be daunting, it might also give very good purchase on separating out nematode-wide, from taxa-specific, aspects of nematode parasitism.

For Peer Review

Over the next few years it can be envisaged that further great strides will be made in interrogating the genomes of parasitic nematodes. However, beyond genomic analyses functional manipulations of genomes will be needed to rigorously test hypotheses concerning the function of particular genes and their products. Despite the enormous success of transgenesis, RNAi analysis and genome editing *etc.* in *C. elegans*, these methods have proved very much harder to establish with parasitic species. For example, it appears that parasitic nematodes are generally refractory to RNAi [48], whereas *C. elegans* is, by lucky chance, particularly susceptible (and indeed more susceptible than most other *Caenorhabditis* spp. [49]). Cleary, for parasitic nematodes the use of such approaches is considerably more tricky because of the need for these parasites to be maintained in (usually vertebrate) hosts. However, progress is being made, particularly with *Strongyloides* spp., where transgenesis is possible, and where the first evidence for the use of CRISPR has now been obtained [50]. Using these and other techniques we will eventually reveal how nematodes live their parasitic lives, and so we will answer a fundamentally fascinating question, but also be able to consider how to manipulate and control these organisms so as to alleviate human and animal suffering.

Acknowledgements

I would like to thank James Cotton for supplying data for the figure, Rebecca Cole and Vicky Hunt for useful discussions, and the Wellcome Trust for funding support.

Figure Legend

Figure 1. The phylogeny of *Strongyloides* and its relatives and the evolution of parasitism. A phylogeny of four species of *Strongyloides, Parastrongyloides trichosuri* and *Rhabditophanes* sp. *Rhabditophanes* is a free-living species. *Parastrongyloides* can have multiple free-living adult generations, but can also be parasitic, making it a facultative parasite. *Strongyloides* is an obligate parasitic species because, even though it has a free-living adult generation (as does *Parastrongyloides*), *Strongyloides'* life cycle requires a parasitic adult generation every generation. Where parasitism and obligate parasitism are inferred to arise is shown. The boxes show, in descending vertical order, the number of (i) gene families originating on each branch (+, in blue), (ii) gene families with at least one duplication event on each branch (+, in blue, in parentheses), and (iii) the number of gene

For Peer Review

families with at least one loss on each branch (-, in red). The tree branch lengths do not show the relative distance among the taxa. Data from [18].

References

1. Cobb N. Nematodes and their relationships. In: *Yearbook United States Department of Agriculture*. United States Department of Agriculture. pp.457-490. 1914.

2. Maizels RM, Balic A, Gomez-Escobar N, *et al.* Helminth parasites – masters of regulation. *Immunol Rev* 2004;**201**:89-116.

3. Savioli l, Albonico M. Soil-transmitted helminthiasis. *Nat Rev Microbiol* 2004;**2**:618-619.

4. Pullan RL, Smith JL, Jasrasaria R, *et al.* Global numbers of infection and disease burden of soil transmitted helminth infections in 2010. *Parasit Vectors* 2014;**7**:37.

5. Hotez PJ, Kamath A. Neglected tropical diseases in sub-saharan Africa: Review of their prevalence, distribution, and disease burden. *PLoS NTD* 2009;**3**:e412.

6. WHO Health Statistics and Information Systems, Estimates for 2000-2012, Disease Burden : who.int/healthinfo/global_burden_disease/estimates/en/index2.html

7. Nieuwhof GJ, Bishop SC. Costs of the major endemic diseases of sheep in Great Britain and the potential benefits of reduction in disease impact. *Anim Sci* 2005;**81**:23-29.

8. Selzer PM. In: *Antiparasitic and Antibacterial Drug Discovery – From Molecular Targets to Drug Candidates*. Wiley-Blackwell, 2009.

9. Charlier J, Höglund J, von Samson-Himmelstjerna G, *et al*. Gastrointestinal nematode infections in adult dairy cattle: Impact on production, diagnosis and control. *Vet Parasitol* 2009;**164**:70-79.

10. Aguinaldo AMA, Turbeville JM, Linford LS, *et al.* Evidence for a clade of nematodes, arthropods and other moulting animals. *Nature* 1997;**387**:489-493.

For Peer Review

11. De Ley PA. quick tour of nematode diversity and the backbone of nematode phylogeny (January 25, 2006), *WormBook*, ed. The *C. elegans* Research Community, WormBook, doi/10.1895/wormbook.1.41.1, http://www.wormbook.org.

12. Parkinson J, Mitreva M, Whitton C, *et al.* A transcriptomic analysis of the phylum Nematoda. *Nat Genet* 2004;**36**:259-1267.

13. Blaxter ML, Koutsovoulos G. The evolution of parasitism in Nematode. *Parasitol* 2015;**142**:S26-S39.

14. Burke M, Scholl EH, Bird DM, *et al.* The plant parasite *Pratylenchus coffeae* carries a minimal nematode genome. *Nematol* 2015;**17**:621-637.

15. Laing R, Kikuchi T, Martinelli A, *et al*. The genome and transcriptome of *Haemonchus contortus*, a key model parasite for drug and vaccine discovery. *Genome Biol* 2013;**14**:R88.

16. Dieterich C, Clifton SW, Schuster LN, *et al.* The *Pristionchus pacificus* genome provides a unique perspective on nematode lifestyle and parasitism. *Nat Genet* 2008;**40**:1193-1198.

17. Mitreva M, Jasmer DP, Zarlenga DS, *et al*. The draft genome of the parasitic nematode *Trichinella spiralis. Nat Genet* 2011;**43**:228-235.

18. Hunt VL, Tsai IJ, Coghlan A, *et al.* The genomic basis of parasitism in the *Strongyloides* clade of nematodes. *Nat Genet* 2016;**48**:299-307.

19. Guiliano DB, Hall N, Jones SJ, *et al*. Conservation of long-range synteny and microsynteny between the genomes of two distantly related nematodes. *Genome Biol* 2002;**3**:research0057.1–0057.14

20. Ghedin E, Wang S, Spiro D, *et al.* Draft genome of the filarial nematode parasite *Brugia malayi. Science* 2007;**317**:756-760.

For Peer Review

21. Jex AR, Liu S, Young ND, *et al*. *Ascaris* draft genome. *Nature* 2011;**479**:529-533.

22. Blaxter ML, DeLey P, Garey JR, *et al*. A molecular evolutionary framework for the phylum Nematoda. *Nature* 1998;**392**:71-75.

23. Weinstein SB, Kuris AM. Independent origins of parasitism in Animalia. *Biol Let* 2016;**12**:20160324.

24. Zarowiecki M, Berriman M. What helminth genomes have taught us about parasite evolution. *Parasitol* 2015;**142**:S85-S97.

25. Hewitson JP, Harcus Y, Murray J, *et al.* Proteomic analysis of secretory products from the model gastrointestinal nematode *Heligmosomoides polygyrus* reveals dominance of Venom Allergen-Like (VAL) proteins. *J Proteomics* 2011;**74**:1573-1594.

26. Zhu X-Q, Korhonen PK, Cai H, *et al.* Genetic blueprint of the zoonotic pathogen *Toxocara canis*. *Nat Commun* 2015;**6**:6145.

27. Jex AR, Nejsum P, Schwarz EM, *et al.* Genome and transcriptome of the porcine whipworm *Trichuris suis. Nat Genet* 2011;**46**:701-706.

28. McNulty SN, Strübe C, Rosa BA, *et al. Dictyocaulus viviparus* genome, variome and transcriptome elucidate lungworm biology and support future intervention. *Sci Rep* 2016;**6**:20316.

29. Viney ME, Kikuchi T. *Strongyloides ratti* and *S. venezuelensis* – rodent models of *Strongyloides* infection. *Parasitol* 2016;doi:10.1017/S0031182016000020.

30. Tang YT, Gao X, Rosa BA, *et al.* Genome of the human hookworm *Necator americanus*. *Nat Genet* 2014;**46**:261-269.

31. Schwarz EM, Hu Y, Antoshechkin I, *et al.* The genome and transcriptome of the zoonotic hookworm *Ancylostoma ceylanicum* identify infection-specific gene families. *Nat Genet* 2015;**47**:416-422.

For Peer Review

32. Sommer RJ. *Pristionchus pacificus* (August 14, 2006), *WormBook*, ed. The *C. elegans* Research Community, WormBook, doi/10.1895/wormbook.1.102.1, http://www.wormbook.org.

33. Del Valle A, Forbes BF, Harrison LM, *et al.* Isolation and molecular cloning of a secreted hookworm platelet inhibitor from adult *Ancylostoma caninum*. *Molec Biochem Parasitol* 2003;**129**:167-177.

34. Zhan B, Liu Y, Badamchian M, *et al*. Molecular characterisation of the *Ancylostoma*-secreted protein family from the adult stage of *Ancylostoma caninum*. *Int J Parasitol* 2003;**33**:897-907.

35. Selkirk ME, Lazari O, Matthews JB. Functional genomics of nematode acetylcholinesterases. *Parasitol* 2005;**131**:S3-S18.

36. Vaux R, Schnoeller C, Berkachy R, *et al*. Modulation of the immune response by nematode secreted acetylcholinesterase revealed by heterologous expression in *Trypanosoma musculi. PLoS Path* 2016;doi: 10.1371/journal.ppat.1005998

37. Hunt VL, Tsai IJ, Selkirk ME, *et al.* The genome of *Strongyloides* spp. gives insights into protein families with a putative role in nematode parasitism. *Parasitol* 2017;doi: 10.1017/S0031182016001554.

38. Mitreva M, McCarter JP, Arasu P, *et al.* Investigating hookworm genomes by comparative analysis of two *Ancylostoma* species. *BMC Genomics* 2005;**6**:58.

39. Viney ME, Graham AL. Patterns and processes in parasite co-infection in animals and humans. *Adv Parasitol* 2013;**82**:321-369.

40. Mulvenna J, Hamilton B, Nagaraj SH, *et al.* Proteomic analysis of the excretory/secretory component of the blood-feeding stage of the hookworm, *Ancylostoma caninum*. *Mol Cell Proteomics* 2009;**8**:109-121.

For Peer Review

41. Yin Y, Martin J, Abubucker S, *et al.* Intestinal transcriptomes of nematodes: comparison of the parasites *Ascaris suum* and *Haemonchus contortus* with the free-living *Caenorhabditis elegans*. *PLoS NTD* 2008;**2**:e269.

42. Foth BJ, Tsai IJ, Reid AJ, *et al.* Whipworm genome and dual-species transcriptome anlayses provide molecular insights into an intimate host-parasite interaction. *Nat Genet* 2014;**46**:693-700.

43. Yatsuda AP, Krijgsveld J, Cornelissen AWCA, *et al.* Comprehensive analysis of the secreted proteins of the parasite *Haemonchus contortus* reveals extensive sequence variation and differential immune recognition. *J Biol Chem* 2003;**9**:16941-1651.

44. Soblik H, Younis AE, Mitreva M, *et al.* Life cycle stage-resolved proteomic analysis of the excretome/secretome from *Strongyloides ratti* – identification of stage-specific proteases. *Mol Cell Proteomics* 2011;**10**:M111.010157.

45. Foss EJ, Radulovic D, Shaffer SA, *et al.* Genetic basis of proteome variation in yeast. *Nat Genet* 2007;**39**:1369-1375.

46. Ghazalpour A, Bennett B, Petyuk VA, *et al.* Comparative analysis of proteome and transcriptome variation in mouse. *PLoS Genet* 2011;**7**:e1001393.

47. Haider S, Pal R. Integrated analysis of transcriptomic and proteomic data. *Curr Genomics* 2013;**14**:91-110.

48. Selkirk Me, Huang SC, Knox DP, *et al.* The development of RNA interference (RNAi) in gastrointestinal nematodes. *Parasitol* 2012;**139**:605-612.

49. Winston WM, Sutherlin M, Wright AJ, *et al*. *Caenorhabditis elegans* SID-2 is required for environmental RNA interference. *Proc Natl Acad Sci USA* 2007;**104**:10565-10570.

50. Lok JB, Shao H, Massey HC, *et al*. Transgenesis in *Strongyloides* and related parasitic nematodes: historical perspectives, current functional genomic applications and progress towards gene disruption and editing. *Parasitol* 2016;doi: http://dx.doi.org/10.1017/S0031182016000391

For Peer Review

Author Biography

Mark Viney studies the biology of nematodes, particularly parasitic species, seeking to understand how the environment controls their development, and how they are adapted to their parasitic lifestyle. (nematode.bio.bris.ac.uk)

Summary Key Points

1. Nematodes are ubiquitous and abundant animals and many species are parasites.
2. Nematodes have evolved parasitism on multiple different occasions.
3. Phylogenetically appropriate comparisons of parasitic and free-living nematodes can reveal nematodes' genomic adaptations to parasitism.
4. Different parasitic nematodes have taxa-specific adaptations to parasitism, though the comparative expansion of genes and gene families coding for proteases and SCP/TAPS proteins appears to be a common theme among them.
5. Parasitic nematodes' secreted molecules are the key interface between host and parasite, and discovering these molecules will be instrumental in understanding how nematodes are adapted to their parasitic lifestyle.

For Peer Review

Figure 1

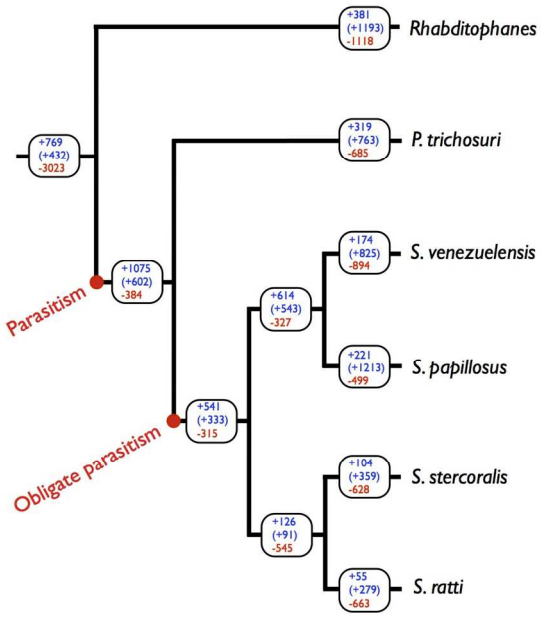

Figure 1. The phylogeny of Strongyloides and its relatives and the evolution of parasitism. A phylogeny of four species of Strongyloides, Parastrongyloides trichosuri and Rhabditophanes sp. Rhabditophanes is a free-living species. Parastrongyloides can have multiple free-living adult generations, but can also be parasitic, making it a facultative parasite. Strongyloides is an obligate parasitic species because, even though it has a free-living adult generation (as does Parastrongyloides), Strongyloides' life cycle requires a parasitic adult generation every generation. Where parasitism and obligate parasitism are inferred to arise is shown. The boxes show, in descending vertical order, the number of (i) gene families originating on each branch (+, in blue), (ii) gene families with at least one duplication event on each branch (+, in blue, in parentheses), and (iii) the number of gene families with at least one loss on each branch (-, in red). The tree branch lengths do not show the relative distance among the taxa. Data from [18].

297x209mm (150 x 150 DPI)