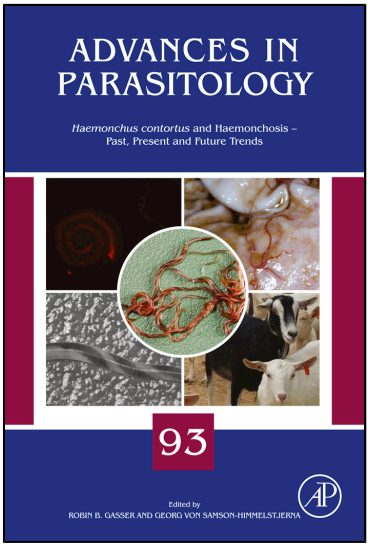

From Laing, R., Martinelli, A., Tracey, A., Holroyd, N., Gilleard, J.S., Cotton, J.A., 2016. *Haemonchus contortus*: Genome Structure, Organization and Comparative Genomics. In: Gasser, R.B., von Samson-Himmelstjerna, G., (Eds.), *Haemonchus contortus* and Haemonchosis – Past, Present and Future Trends, pp. 569–598.
ISBN: 9780128103951

Academic Press

CHAPTER THIRTEEN

Haemonchus contortus: Genome Structure, Organization and Comparative Genomics

R. Laing*, A. Martinelli§, A. Tracey§, N. Holroyd§, J.S. Gilleard¶,1, J.A. Cotton§

*University of Glasgow, Glasgow, Scotland, United Kingdom
§Wellcome Trust Sanger Institute, Cambridge, United Kingdom
¶University of Calgary, Calgary, AB, Canada
1Corresponding author: E-mail: jsgillea@ucalgary.ca

Contents

Abstract

One of the first genome sequencing projects for a parasitic nematode was that for *Haemonchus contortus*. The open access data from the Wellcome Trust Sanger Institute provided a valuable early resource for the research community, particularly for the identification of specific genes and genetic markers. Later, a second sequencing project was initiated by the University of Melbourne, and the two draft genome sequences for *H. contortus* were published back-to-back in 2013. There is a pressing need for long-range genomic information for genetic mapping, population genetics and functional

Advances in Parasitology, Volume 93
ISSN 0065-308X
http://dx.doi.org/10.1016/bs.apar.2016.02.016

Haemonchus contortus and Haemonchosis – Past, Present and Future Trends, First Edition, 2016, 569–598

genomic studies, so we are continuing to improve the Wellcome Trust Sanger Institute assembly to provide a finished reference genome for *H. contortus*. This review describes this process, compares the *H. contortus* genome assemblies with draft genomes from other members of the strongylid group and discusses future directions for parasite genomics using the *H. contortus* model.

1. INTRODUCTION

1.1 Overview and historical context

The first metazoan organism to have a full genome sequence published was the free-living nematode *Caenorhabditis elegans* in 1998; even now, the genome of this model organism is still one of the few truly 'complete' metazoan genomes (*C. elegans* Sequencing Consortium, 1998). The associated database WormBase (https://www.wormbase.org) was one of the first to integrate genomic, genetic and phenotypic data and published literature and to have ongoing curation. Although the *C. elegans* Sequencing Consortium played a central role in the development of whole genome sequencing approaches, until recently, parasitic nematodes have been largely left out of these endeavours. Indeed, whilst research groups working on bacterial and protozoan pathogens have rapidly advanced their genomic resources, the progress in producing high-quality reference genomes for the major parasitic helminths has been slower. There are a number of reasons for this, including the large size and complexity of helminth genomes relative to other classes of pathogens and the lower priority given to parasitic helminths by the major genome centres due to the relatively small size of the research communities. However, over the last 5–6 years, advances in massively parallel sequencing technologies have dramatically reduced raw sequencing costs which, together with related developments in informatics, has resulted in significant progress in sequencing helminth genomes (Ghedin et al., 2007; Jex et al., 2011; Mitreva et al., 2011; Protasio et al., 2012; Tsai et al., 2013). For example, a recent initiative at the Wellcome Trust Sanger Institute aims to provide draft reference genomes for 50 parasitic helminths of major importance to human and veterinary medicine and agricultural production https://www.sanger.ac.uk/research/initiatives/globalhealth/research/helminthgenomes/.

One of the first parasitic nematode genome sequencing projects to be initiated at one of the major genome centres was that for *Haemonchus contortus* (http://www.sanger.ac.uk/resources/downloads/helminths/haemonchus contortus.html). This project was started in 2004 using the standard

approach available at the time; a combination of capillary sequencing and physical mapping with Bacterial Artificial Chromosome (BAC) or cosmid/fosmid libraries. MHco3.ISE was chosen as the reference strain for this project. This strain is a version of the ISE isolate that has been passaged and maintained at the Moredun Research Institute since 2002 (Redman et al., 2008b). It was chosen because it is susceptible to all the major anthelmintic drugs and had been previously inbred by multiple rounds of serial passage infections with progeny from a single female worm (Roos et al., 2004). The draft genome sequence produced at this time provided a valuable, early resource for researchers in the field, particularly for the identification of specific individual genes and genetic markers (Kaminsky et al., 2008; McCavera et al., 2007; Redman et al., 2008b). This strain was further inbred using a single pair mating approach to produce the more highly inbred strain MHco3(ISE).N1 that was used to produce DNA template for subsequent sequencing in the genome project (Sargison, Redman, Naghra, Cotton and Gilleard, unpublished data). However, the relatively large genome size (~ 300 Mb) and high levels of sequence polymorphism, even in this partially inbred strain, made completion of a high-quality reference genome challenging and prohibitively expensive at that time. It was not until the availability and application of 'second-generation' sequencing approaches that significant progress on the assembly and annotation of a good quality 'draft' reference genome was made. Subsequently, another genome project was initiated to sequence the *H. contortus* McMaster strain which was derived from an Australian isolate. In 2013, the draft genome sequences of both the MHco3.ISE and McMaster strains were published in complementary papers (Laing et al., 2013; Schwarz et al., 2013). These draft genomes were of a quality such that >90% of genes are believed to be at least partially represented and mostly in scaffolds that contain multiple genes. The MHco3.ISE reference genome was 346.2 Mb in length, with a scaffold N50 of 83,287 bp and with 21,799 predicted protein-coding genes (Laing et al., 2013). The McMaster reference genome was 320 Mb in length, with a scaffold N50 of 56,300 bp and with 23,610 predicted protein-coding genes (Schwarz et al., 2013). Although these can be considered to be good quality draft sequences, and so valuable resources, they are still far from 'finished' genomes.

1.2 The need for a high-quality 'finished' reference genome

The rapid development of next-generation sequencing technologies has made large-scale sequencing faster and cheaper than ever before and has

resulted in a large number of parasitic helminth genome projects (Kumar et al., 2012). The strategy for most of these projects has been to use short-read technologies to sequence genomes at great depth and then undertake shotgun assembly, to avoid the labour-intensive process of physical mapping of large-insert genomic clones. Although mate pair libraries are often used to aid the assembly of contigs into larger scaffolds, most of the current helminth genome projects are producing draft genomes that are still highly fragmented and do not have funding to progress these to chromosome-level assemblies (Fig. 1). Although these draft genomes are extremely valuable for identifying 'genes of interest' and for some comparative and evolutionary studies, chromosome-level assemblies will greatly improve the power of future functional genomic and genetic studies.

Consequently, one of the major challenges facing the research community working on parasitic helminths at present is how to ensure that we continue to improve these genome assemblies and annotations for those parasites of high medical, veterinary or agricultural importance, particularly those species that can serve as model experimental systems. Unfortunately, there is a disproportionate cost and effort of finishing a draft reference genome and a poor return, in terms of peer-review publication, acts as a major disincentive. *Haemonchus contortus* is a good example of this, in that it is the best available model system for studying the biology of strongylid nematodes, for anthelmintic drug discovery and for anthelmintic resistance research (Burns et al., 2015; Kaminsky et al., 2008). For this latter research area, in particular, *H. contortus* provides an opportunity to apply both classical genetic mapping and population genomic studies to identify anthelmintic resistance loci (Gilleard and Beech, 2007; Redman et al., 2012). However, this work is extremely difficult with the current draft genomes and so we are continuing to improve the MHco3.ISE assembly to provide a finished reference genome for *H. contortus*.

Limitations of draft genome assemblies include:

- Ambiguity in genome size and content and an incomplete gene list for functional genomic studies
- An inability to identify haplotypic sequences and copy number variation
- Insufficient long-range information to allow genetic mapping by genetic crosses
- Insufficient long-range information to identify signatures of selection through population genetic/genomic approaches
- Incomplete and ambiguous membership of large and complex gene families
- Incomplete information on genome organization and regulation
- Inability to investigate potential long range genetic interactions

Figure 1 Limitations of draft genome assemblies.

2. FROM DRAFT ASSEMBLY TO REFERENCE GENOME

2.1 Progress so far

To generate the draft genomes (Laing et al., 2013; Schwarz et al., 2013), a combination of short and long insert Illumina libraries were generated, with the addition of 454 libraries for the MHco3.ISE assembly (see Mardis, 2008 for review of these sequencing technologies). Standard de novo assembly software (Velvet: Zerbino and Birney, 2008; Celera: Myers et al., 2000) was used together with HaploMerger (Huang et al., 2012) to reduce the amount of haplotypic expansion. Illumina and 454 are both short-read sequencing technologies, with particularly Illumina sequencing being highly cost-effective, in terms of generating large amounts of sequence data, and highly accurate, in terms of the correct identification of sequenced bases. Assembly involves finding overlaps between sequence reads, so that longer strings of contiguous known bases (contigs) can be reconstructed. Two aspects make this process difficult and have led to an enormous diversity of algorithms and software packages for genome assembly (eg, Bradnam et al., 2013; Earl et al., 2011; Nagarajan and Pop, 2013; Schatz et al., 2010 for reviews and comparison). First, the sheer number of reads produced by modern sequencing technologies means that algorithms and their software implementation need to be extremely efficient in finding overlaps between reads (eg, Simpson and Durbin, 2010). More fundamentally, repetitive strings occurring in DNA or the presence of polymorphism can make the reconstruction of the genome ambiguous and lead to fragmented assemblies. Sequencing bases from both ends of a range of different sizes of DNA fragments can show that certain contigs belong close to each other, with some unknown sequence in-between being represented by a gap in between two contigs. Strings of contigs joined in this way are known as scaffolds (Fig. 2). The presence of such gaps, and of sets of scaffolds where the order and orientation between them is unknown, is the hallmark of a draft assembly, and generally short-read technologies can only produce draft assemblies for all but the very smallest and simplest genomes. Each of the *Haemonchus* genome assemblies published in 2013 (Laing et al., 2013; Schwarz et al., 2013) was in over 10,000 scaffolds.

Since publication of the draft sequences, we have continued to improve the MHco3.ISE genome with the aim of assembling the six chromosomes that are predicted by karyotype analysis (Redman et al., 2008a), and so provide a true reference-quality genome assembly (Chain et al., 2009)

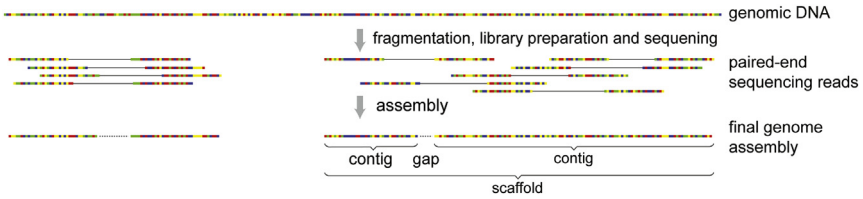

Figure 2 Sequence reads and assembly. A paired-end sequencing library is generated by randomly fragmenting a genome and some additional molecular biology steps before sequencing stretches of nucleotides from both ends of each fragment. The assembly process identifies stretches of overlapping nucleotides between reads to form contigs. Two contigs that contain reads from the same pair are united into the same scaffold, with unsequenced or ambiguous bases in-between leading to a gap between adjacent contigs.

for *H. contortus*. One key approach is to generate additional data that allow long-range scaffolding and contiguation of the existing assembly. We have generated long-read sequence data using the PacBio sequencing platform (Korlach et al., 2010) and used contigs from a de novo assembly of these data to bridge gaps and identify joins between scaffolds in the previous assembly. Longer-range scaffolding is improved by optical mapping — a technology in which individual DNA fragments are immobilized and digested with restriction enzymes or labelled at sequence-specific nicks in the DNA molecules (Lam et al., 2012; Samad et al., 1995). The distribution of these cuts or nicks is visualized microscopically to produce an ordered map of the relative location of short-sequence motifs along a molecule. Combining data from many thousands of molecules allows the reconstruction of very long maps, up to whole chromosomes for smaller, less-repetitive genomes. Aligning sequence data against these maps or using the maps to identify adjacent sequence contigs in other ways (eg, Dong et al., 2013) can help build long scaffolds with extensive sequence gaps that can then be filled using other methods.

A second component of the genome improvement work has been a programme of manual curation and genome improvement, initially based on the inspection of short reads mapped to the entire genome to allow correction of the assembly where necessary, and latterly also making use of longer PacBio reads. This is achieved using the genome viewing and editing tool Gap5 (Bonfield and Whitwham, 2010). Using this approach, it is possible to identify places throughout the genome where the sequence is too divergent, or where pairs of reads do not lie in the correct relative orientation or spacing, and these features give clues to places where the sequence is

incorrect, and how to correct the problems. For example, if many of the reads have their 'mate' — the other read of the pair — elsewhere in the assembly, it might indicate that the existing sequence needs to be broken, and an additional contig needs to be inserted. This read pair information, together with unusual patterns in coverage, can also identify regions where haplotypes have been too dissimilar to assemble and the contigs end prematurely — an issue that seems to have particularly plagued the automatic assemblies of *H. contortus*. Recently, significant progress in improving the contiguation of the genome assembly has been achieved by taking contigs de novo assembled from PacBio data, validating and assembling them with other data and then using these sequences — representing longer stretches of a single haplotype — to replace gaps or misassembled regions of the assembly.

This 'manual' improvement process is somewhat automated — Gap5 itself contains many features to help automatically find potential joins and identify potentially misassembled regions, and other software has also been developed to aid genome improvement. Algorithmic approaches to identifying problematic regions eg, REAPR (Hunt et al., 2013) are used to help locate possible misassemblies for manual improvement. Small errors can be automatically corrected (Otto et al., 2010), and automated approaches to close gaps (Tsai et al., 2010) and improve scaffolding using data from existing assemblies (Assefa et al., 2009) are all available. These approaches have already provided a much improved genome assembly, referred to here as MHco3.ISE.2015.

Table 1 shows a comparison of assembly statistics for the two published *H. contortus* genomes and the current *H. contortus* MHco3.ISE.2015 genome. While the published *H. contortus* genomes have comparable assembly metrics to other draft strongylid genomes (Table 2, Section 3),

Table 1 Assembly statistics for *Haemonchus contortus*

	***H. contortus* MHco3.ISE**	***H. contortus* McMaster**	***H. contortus* MHco3.ISE.2015**
Assembly size (Mb)	370	320	340
Number of scaffolds	26,044	14,419	5051
N50 count	1151	1684	16
N50 (kb)	83	56	7446
N90 count	5518	6085	552
N90 (kb)	11	13	32
Largest scaffold (kb)	947	345	25,917

Table 2 Genome statistics for *Haemonchus contortus* and other clade V nematodes

	H. contortus MHco3.ISE (MHco3.ISE.2015)	*H. contortus* McMaster	*Necator americanus*	*Ancylostoma ceylanicum*	*Oesophagostomum dentatum*	*Dictyocaulus viviparus*	*Angiostrongylus cantonensis*[a]	*Caenorhabditis elegans* WS248
Assembly size (Mb)	370 (340)	320	244	313	443	169	261	100
Number of scaffolds	26,044 (5051)	14,419	11,713	1737	64,258	17,715[b]	17,280	6
Scaffold N50 (kb)	83 (7446)	56	213	668	26	22.6[b]	42	17,493
GC content (%)	43	42	40	43	33	34.5	41	35
Repetitive sequence (%)	29	13	24	41	31	Not reported	25	16
Number of genes	21,869	23,610	19,151	30,738	25,291	14,306	17,482	20,388
Gene density (per Mb)	59	74	79	98	57	85	67	200
Mean gene footprint (bp)	6524	6167	4289	4560	2171	Not reported	Not reported	3037
Mean number introns per gene	8.8	6	5.4	11	4.3	Not reported	Not reported	5.1
Mean intron size (bp)	601	832	642	319	352	Not reported	Not reported	320

Figures are all as reported in the genome paper for each species cited in the text, except gene density figures calculated from number of genes and assembly size where necessary.
[a]*Angiostrongylus cantonensis* figures are for the Yong et al. (2015) assembly.
[b]*Dictyocaulus viviparus* assembly statistics are for contigs, not scaffolds.

the MHco3.ISE.2015 assembly is significantly better. Fifty percent of the genome is now assembled into 16 scaffolds (N50 = 7.4 Mb), and 80% of the genome is assembled into 66 scaffolds. The largest scaffold is 25.7 Mb in length. To visually display the scale of improvement, synteny plots against the *C. elegans* reference genome annotation (WS235) were generated using the 20 largest scaffolds for the three *H. contortus* assemblies (Fig. 3). The scaffolds and chromosomes are shown to scale, and for MHco3.ISE.2015, these 20 scaffolds comprise more than half of the genome. As expected, MHco3.ISE.2015 and the published MHco3.ISE assembly are highly collinear, although a number of corrected misassemblies interrupt this pattern. A more detailed investigation of synteny between the two *H. contortus* isolates is consistent with the general pattern of colinearity of the two assemblies: most McMaster scaffolds that show any orthology to MHco3.ISE.2015 scaffold 5 show orthology only to this scaffold (Fig. 4). Scaffold 5, which is approximately 6 Mb in size and syntenic with *C. elegans* chromosome I was chosen as an arbitrary representative scaffold close to the N50 of the improved assembly.

2.2 Future and ongoing work

The presence of extensive haplotypic sequence diversity in *Haemonchus* − even in the inbred MHco3.ISE isolate − means manual intervention is currently essential in improving the genome assembly, and much of the work described in Section 2.1 is still in progress. If the combination of approaches being used fails to produce a finished assembly of *H. contortus* chromosomes, other strategies may be required. One possibility may be to exploit the significant degree of synteny with *C. elegans* for an assisted chromosomal assembly, if necessary. Work to generate a genetic linkage map for *H. contortus* is also underway, consisting of a cross between the MHco3.ISE isolate and a genetically divergent multidrug-resistant isolate MHco18.UGA2004 originally isolated from a sheep at the University of Georgia in 2004 (Williamson et al., 2011). The genetic linkage map will aid final assembly and provide the first quantitative data on the rates of meiotic crossing over and gene conversion in this group of nematodes, which may be important in interpreting data from the forward genetic and population genetic studies currently underway or being planned in *Haemonchus*. Once genome improvement work is complete, the 'finished' genome will be coupled to a verified de novo annotation, and all data will be released via WormBase ParaSite http://parasite.wormbase.org/index.html and the

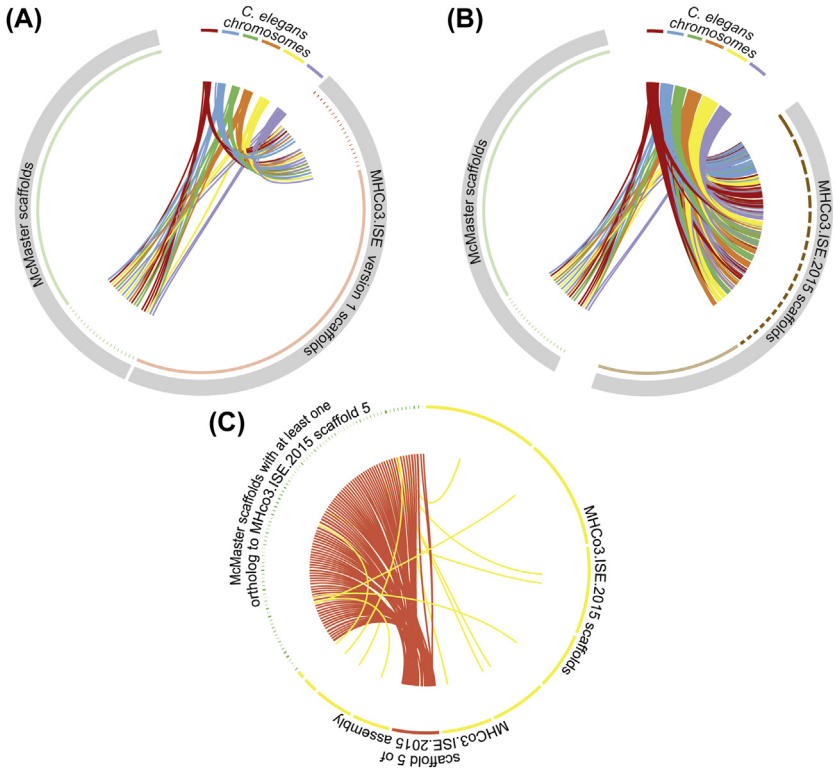

Figure 3 (A) Synteny between the 20 largest scaffolds in the published *H. contortus* McMaster and MHco3.ISE assemblies and the WormBase *C. elegans* genome (from WS235). The longest McMaster and MHco3.ISE scaffolds are marked, with the long contiguous bars representing the rest of the genome for each draft assembly to the same scale, showing how much of the genome is assembled only into small fragments. Lines link genomic locations of single-copy orthologs shared between the three genomes, shaded by chromosome location of the *C. elegans* ortholog. (B) Synteny between 20 largest scaffolds in the *H. contortus* McMaster and improved MHco3.ISE.2015 assemblies and *C. elegans*. Large McMaster scaffolds and MHco3.ISE.2015 scaffolds are marked, with the long contiguous bars representing the rest of the genome for each assembly. Lines link genomic locations of single-copy orthologs shared between the three genomes, shaded by chromosome location of the *C. elegans* ortholog. (C) Synteny between McMaster scaffolds and scaffold 5 of the MHco3.ISE.2015 assembly. McMaster scaffolds are shaded (only scaffolds with at least one single-copy ortholog on MHco3.ISE.2015 scaffold 5 are shown), MHco3.ISE.2015 scaffold 5 is shaded and outlines and other MHco3.ISE.2015 scaffolds with orthologs to any of the McMaster scaffolds shown are outlined. Single-copy orthology from McMaster scaffolds to MHco3.ISE.2015 scaffold 5 are linked by red (black in print versions) lines, all other links are yellow (grey in print versions).

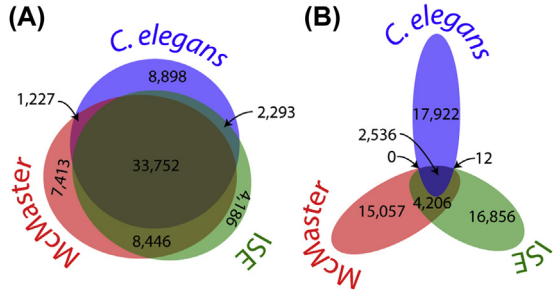

Figure 4 Overlap of predicted proteomes between two *Haemonchus contortus* genomes and *Caenorhabditis elegans*. In (A) genes are considered shared between genomes if they are members of a gene family present in both, while in (B) genes are considered shared between genomes only if they are one-to-one orthologs.

Wellcome Trust Sanger Institute ftp site ftp://ftp.sanger.ac.uk/pub/pathogens/Haemonchus/contortus.

3. GENOME STRUCTURE

Clade V nematodes share a conserved genetic core, which facilitates powerful comparative and functional genomics using a *C. elegans* model (Gilleard, 2013). However, nematode genomes are notoriously dynamic in structure and organization (Blaxter, 2003), and each species may be as readily defined by its unique adaptations to a particular niche as by its common ancestry. Since the publication of the *H. contortus* genomes in 2013, draft assemblies have been published for five more members of the strongylid group, the hookworms *Necator americanus* (Tang et al., 2014) and *Ancylostoma ceylanicum* (Schwarz et al., 2015), the nodule worm *Oesophagostomum dentatum* (Tyagi et al., 2015a), the bovine lungworm *Dictyocaulus viviparus* (Koutsovoulos et al., 2014) and two different assemblies for the zoonotic pathogen rat lungworm, *Angiostrongylus cantonensis* (Morassutti et al., 2013; Yong et al., 2015). We therefore have the opportunity to explore aspects of the *H. contortus* genome that are broadly conserved in clade V nematodes, that are a general feature of the strongylid group, and that are unique to this species. Vertebrate parasites are hypothesized to have evolved from a free-living ancestor though 'preadaptation' in insect hosts, so the publication of the genome of *Heterohabditis bacteriophora* (Bai et al., 2013), an insect parasite in a sister group to the strongylids, may provide insight to their evolution as vertebrate parasites. Further, we have the

opportunity to compare independent genome assemblies of two geographically isolated *H. contortus* strains.

3.1 Genome size, repetitive sequence and Guanine and Cytosine (GC) content

Strongylid genomes are significantly expanded relative to *C. elegans*; with the exception of *D. viviparus*, all are at least double the size (Table 2). A large proportion of this expansion appears to be repetitive sequences: 24% of the *N. americanus* and *A. cantonensis* genome assemblies are repetitive as is an incredible 41% of the *A. ceylanicum* genome. Strikingly, the *H. contortus* MHco3.ISE, *A. ceylanicum* and *O. dentatum* genomes all contain over 100 Mb of repeats (equivalent to the entire the *C. elegans* genome). The disparity in genome size between *H. contortus* MHco3.ISE and McMaster assemblies is largely due to a difference in repeat content, which is most likely of technical note rather than biological interest, as repeats are difficult to represent correctly in draft genome assemblies. Similarly, the presence of haplotypic sequence will also contribute to differences in sizes of draft assemblies. GC content is slightly higher in *H. contortus* and the hookworms than *O. dentatum* and *C. elegans* but this difference does not appear to correlate with a bias in base composition of the repetitive sequence, as the GC content is comparable in both MHco3.ISE and McMaster assemblies.

3.2 Chromosomal synteny

As shown in Section 2 (Fig. 2), the degree of chromosomal synteny between *H. contortus* and *C. elegans* is striking. However, gene order is generally poorly conserved, other than in small regions of microsynteny (Laing et al., 2011). This is consistent with comparison of the *C. elegans* and *Caenorhabditis briggsae* genomes, which found high rates of intrachromosomal rearrangements (Coghlan and Wolfe, 2002; Guiliano et al., 2002; Stein et al., 2003). Such rearrangements are suggested to occur more commonly than interchromosomal rearrangements because they require fewer DNA breakpoints and because conformation of the nuclear scaffold could maintain associations between local regions. Stein et al. (2003) also found that rearrangements were more common in the autosomes than in the X chromosome, and more frequent in the chromosomal arms than centres. The finished genome will reveal whether a similar pattern is true for *H. contortus*.

4. GENOME CONTENT

4.1 Gene number, size and structure

The apparent differences in genome content between *H. contortus* isolates (Table 2) are expected to be largely due to technical differences in genome assembly and annotation, although there may be some between-isolate and between-individual biological variation. Both assemblies were relatively fragmented at the time of publication and are known to contain some split genes (ie, single genes spanning unassembled contigs annotated as multiple genes) and some haplotypic sequences. Coverage of the coding portion of the genome is comprehensive in both *H. contortus* assemblies, with 91–92% of core eukaryotic genes (Parra et al., 2007) represented, so the number of predicted genes is expected to be an upper limit and may decrease slightly as assembly improves. The MHco3.ISE genome encodes a smaller number of genes, with longer transcripts (of a comparable size to *C. elegans*), so it will be used for between-species comparisons of gene structure in this section.

The *H. contortus* genome encodes a remarkably similar number of genes to *C. elegans*, despite the striking difference in genome size. Other strongylid genomes encode more widely varying numbers of genes (eg, 30% difference between hookworms; although this may be at least partly due to different methods of gene prediction). Gene number does not appear to correlate with genome size in the Strongylida, but may do so within genera (this is the case in members of the *Caenorhabditis* genus; Fierst et al., 2015). It is notable that both lungworm genome assemblies (*D. viviparus* and *A. cantonensis*) are smaller than those for other strongylids and have fewer gene models, although both of these assemblies are relatively fragmented, so it is unclear whether this represents a biological difference between strongylid groups or a technical artefact. However, relative to *C. elegans*, gene density is consistently lower (less than half) in all strongylids.

While the average transcript length is similar in *H. contortus* and *C. elegans*, the average gene length is more than double in the parasite. Comparison of a subset of one-to-one orthologs suggests that this is due to longer introns and a higher number of introns per gene. Although less so than in *H. contortus*, the average gene length in the hookworm genomes is significantly increased relative to *C. elegans*, again due to an expansion of intronic sequences; *A. ceylanicum* has more introns per gene, while *N. americanus* has larger introns. Notably, *O. dentatum* has a smaller average gene length than *C. elegans*,

resulting from a smaller number of similarly sized introns. While differences in gene content, such as expansion of particular genes and gene families with long introns, may affect these genome-wide averages — and these metrics in draft genomes are sensitive to technical variation — there is clearly a trend for intronic expansion in the hookworms as well as *H. contortus*.

The significance of this intronic expansion is unknown. Although a greater number of introns might promote protein diversity through alternative splicing, it is not clear what advantage large introns would confer on *H. contortus*, particularly if they result in greater time and energy costs for transcription (Castillo-Davis et al., 2002). Introns do play an integral role in the regulation of gene expression in other organisms and often contain small regulatory RNAs or binding sites for functional elements (Hube and Francastel, 2015), but the link between gene expression and intron size or structure in *Haemonchus* (or any other parasitic nematode) has not been explored. However, a large fraction of transposable elements (TEs) and repetitive DNA are also found in introns (Palazzo and Gregory, 2014) and, thus, intron expansion could equally reflect relatively unconstrained evolution.

4.2 Orthology and gene family evolution

Nematodes share a core set of highly conserved genes, enriched for roles in developmental processes and signalling pathways. However, despite their closer phylogenetic relationship and parasitic lifestyles, the number of one-to-one orthologs shared by *H. contortus* and the strongylid parasites is comparable with those shared with *C. elegans*. For example, Schwarz et al. (2015) report that *H. contortus* shares 5268 strict orthologs with *A. ceylanicum* and 4576 with *C. elegans*, although the precise numbers vary with the dataset and analysis method used. These differences may partly reflect the draft nature of the parasite genome assemblies, confounding the detection of single-copy orthologs that can be more easily identified in *C. elegans*, but it is also indicative of a significant degree of gene family diversity among members of the Strongylida. This diversity is highlighted by the marked difference in gene number between the two hookworm genomes and is consistent with reports of large-scale duplication events, lineage-specific amplifications and gene losses that appear to be widespread elsewhere in the phylum Nematoda (Markov et al., 2015). Such differences are even apparent between the annotated gene models for the two different *H. contortus* isolates (Fig. 4) — while most genes can be assigned to gene families shared by the two gene sets (and with *C. elegans*), few of them are

single-copy orthologs, suggesting that many gene families are differently annotated in the two assemblies. At the current state of completion of the two published assemblies, it is difficult to assess how much of the variation in gene content might be due to genuine genetic differences between isolates, to chance assembly of different haplotypic alleles in the two assemblies or to differences in the gene-finding approaches used.

Recent analysis of metabolic pathways and enzyme abundance in free-living and parasitic nematodes across the phylum did not find any completely conserved parasitism pathways, suggesting substantial divergence and adaptation to particular lifestyles and niches (Tyagi et al., 2015b). Clearly, similar life traits can be underpinned by different genomic adaptations, and it appears that such 'convergent' rather than orthologous evolution (Zarowiecki and Berriman, 2015) may be common in nematodes, particularly when comparing related groups of species that are adapted to live in different hosts, such as *H. contortus* and the hookworms.

Parasitic lifestyles theoretically afford the opportunity for genome simplification, due to metabolic provision by the host. In nematodes without free-living life stages, such as *Trichinella spiralis* and *Brugia malayi*, as well as in cestodes and trematodes, host metabolic provision appears to have resulted in a considerable loss of biosynthetic pathways (Tyagi et al., 2015b; Zarowiecki and Berriman, 2015). However, this does not appear to be a feature of the *H. contortus* genome. While the complement of genes involved in amino acid and carbohydrate metabolism appears to differ between *H. contortus* and *C. elegans* (see Laing et al., 2013), no large-scale losses of biosynthetic pathways are apparent. As proposed by Blaxter (2003), nematodes with both free-living and parasitic life stages probably cannot simplify their genomes. Further, as they require additional capabilities for host infection, migration, nutrient salvage and immune avoidance, their genomes may need to be more complex than their pre-parasitic ancestors.

For example, the *H. contortus* proteome is significantly enriched for proteases and aminopeptidases, which are required for tissue invasion, haemoglobin digestion, anticoagulation and modulation of host immune response. Large expansions are seen in the cathepsin protease families, involving both the cathepsin D aspartic proteases (83 genes) and the cathepsin B cysteine proteases (63 genes). Phylogenetic analysis with other clade V nematodes shows that these expansions form large monophyletic groups, including six distinct clusters of cathepsin D proteases and three distinct clusters of cathepsin B proteases (Laing et al., 2013). The magnitude

of the expansion in *H. contortus* is striking, even relative to the hookworms, and may reflect a requirement to digest particularly large volumes of host blood. However, the substantial number of cathepsin B cysteine (113) and cathepsin D aspartic (54) proteases in the non-blood-feeding strongylid *O. dentatum* highlights their diversity of function and suggests a role for other selective forces in driving protease gene family evolution, such as host immune selection.

Other gene families that have expanded in strongylid nematodes are less well characterized, such as the ASPs (activation-associated secreted proteins, also known as SCP/TAPs proteins). These encode cysteine-rich products of unknown function, but their expression is upregulated on host infection and, in the hookworms, on exposure to serum and 37°C (Hawdon et al., 1996, 1999). The *C. elegans* genome encodes 35 ASPs, but the family is significantly expanded in the parasites with 161 ASPs in *H. contortus*, 137 in *N. americanus*, 284 in *O. dentatum* and a remarkable 432 in *A. ceylanicum* (Schwarz et al., 2015; Tyagi et al., 2015a). Schwarz et al. (2015) describe two other intriguing gene families: one group of 24 strongylid-specific genes termed SL4Ps, and another (SCVPs) which is greatly expanded in strongylids; both sets of genes are predicted to encode secreted proteins of unknown function.

However, the *H. contortus* genome also encodes a number of gene families that are notably conserved relative to *C. elegans*. These include various chemoreceptors and detoxification enzyme systems. For example, the *C. elegans* genome is significantly enriched for cytochrome P450s (CYPs), a large group of enzymes that metabolize endogenous and exogenous toxins. This gene family has been characterized in *H. contortus*, due to its potential role in anthelmintic metabolism (Laing et al., 2015). Although the majority of CYPs with endogenous roles appear to have one-to-one orthologs, *H. contortus* lacks the characteristic large clusters of tandem gene duplications seen in CYP families with exogenous roles in *C. elegans*. Comparative analysis with the CYP family in other strongylid nematodes suggests a similar trend (Fig. 5), although small clusters of strongylid-specific CYPs are apparent. These include a small group of parasite CYPs (four in *H. contortus* and *O. dentatum*, six in *A. ceylanicum* and two in *N. americanus*) branching with the *C. elegans* CYP34 and CYP35 families, members of which have a putative role in metabolizing anthelmintic compounds (Laing et al., 2010). The necromenic clade V insect parasite *Pristionchus pacificus* has an extremely expanded family of detoxification enzymes, including 198 CYPs, which had been hypothesized to represent a preadaptation to

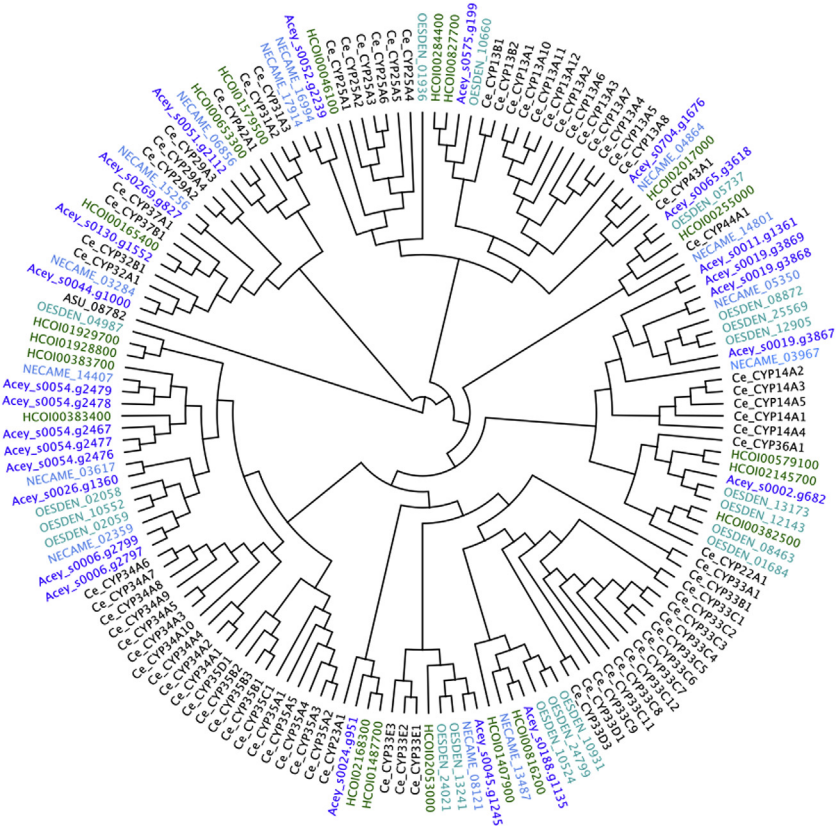

Figure 5 The cytochrome P450 family in *Caenorhabditis elegans*, *Haemonchus contortus*, *Necator americanus*, *Ancylostoma ceylanicum* and *Oesophagostomum dentatum*. Neighbour-joining tree of predicted polypeptide sequences encoding Interpro P450 domain (IPR001128), with *Ascaris suum* CYP as an out-group.

parasitism (Dieterich et al., 2008). However, in the CYP family at least, the relatively small numbers in *H. contortus*, the strongylid parasites of humans and the more closely related insect parasite *H. bacteriophora* suggest that this is not the case.

Major aims of sequencing the *H. contortus* genome were to better understand anthelmintic resistance and to develop novel methods of parasite control. Most of our understanding of drug targets and modes of action have come from studies in *C. elegans*, but important differences in gene families encoding anthelmintic targets were identified in the *H. contortus* genome. Notably, *H. contortus* lacks an ortholog of the major target of ivermectin in *C. elegans*, *glc-1*, which encodes a glutamate-gated chloride channel

(GluCl) subunit (Dent et al., 2000; Ghosh et al., 2012). However, the *H. contortus* genome encodes two alternative subunits, *glc-5* and *glc-6*, which are absent from *C. elegans*, in addition to orthologs of *glc-2*, *glc-3*, *glc-4*, *avr-14* and *avr-15*. Interestingly, heterologous expression of *H. contortus glc-6* can rescue the ivermectin resistance phenotype of a *C. elegans* GluCl mutant strain, suggesting that this gene not only encodes a GluCl that responds to the drug but can replace the function of nonorthologous *C. elegans* genes (Glendinning et al., 2011). This appears to be a remarkably divergent gene family in strongylid nematodes, given its role in nervous system function, with only 3 family members reported in *N. americanus* (homologues of *glc-2*, *glc-3* and *avr-14*) and 10 family members described in *A. ceylanicum* (including homologues of *glc-2*, *avr-14* and *avr-15*). However, there are also clear between-species differences in the nicotinic acetylcholine receptor family (Neveu et al., 2010), which are targets of levamisole and monepantel, and in the β-tubulin family (Saunders et al., 2013), which includes the major target of the benzimidazoles.

The P-glycoproteins (PGPs) are members of the ATP-binding cassette (ABC) transporter family and have been implicated in resistance in *H. contortus* and related parasitic nematodes to various classes of anthelmintic (reviewed by Ardelli, 2013; Kotze et al., 2014). The *H. contortus pgp* family differs significantly from *C. elegans*; there are 10 *pgp* genes in the parasite relative to 14 in *C. elegans*, and species-specific gene duplications appear to have occurred frequently (Laing et al., 2013). For example, gene duplications in *C. elegans* (*pgp-3* and *pgp-4*; *pgp-12*, *pgp-13* and *pgp-14*) correspond to single genes in *H. contortus*, while two paralogous copies of *C. elegans pgp-9* are present in the parasite genome. *Caenorhabditis elegans pgp-5*, *pgp-6*, *pgp-7* and *pgp-8* are clustered in the genome and have no orthologs in *H. contortus*, while the parasite genome encodes *two genes*, *Hco-pgp-16* and *Hco-pgp-17*, which are absent from *C. elegans* (Fig. 6).

While the *pgp* families in *A. ceylanicum*, *N. americanus* and *O. dentatum* are less well characterized than in *H. contortus*, preliminary analysis reveals some clear trends. First, these *pgp* genes tend to branch with individual members of the *C. elegans pgp* family, but form distinct strongylid clusters, including *H. contortus pgp*s. In *A. ceylanicum* and *N. americanus*, single-copy orthologs could be detected for almost all *H. contortus pgp*s, including *Hco-pgp-16* and *Hco-pgp-17*, which are absent from *C. elegans*. The ABC family genes of *O. dentatum* were very fragmented and only two partial length *pgp*s were sufficiently complete for analysis, but these fit the same pattern, branching with *H. contortus Hco-pgp-3* and *Hco-pgp-10*. As is the case for

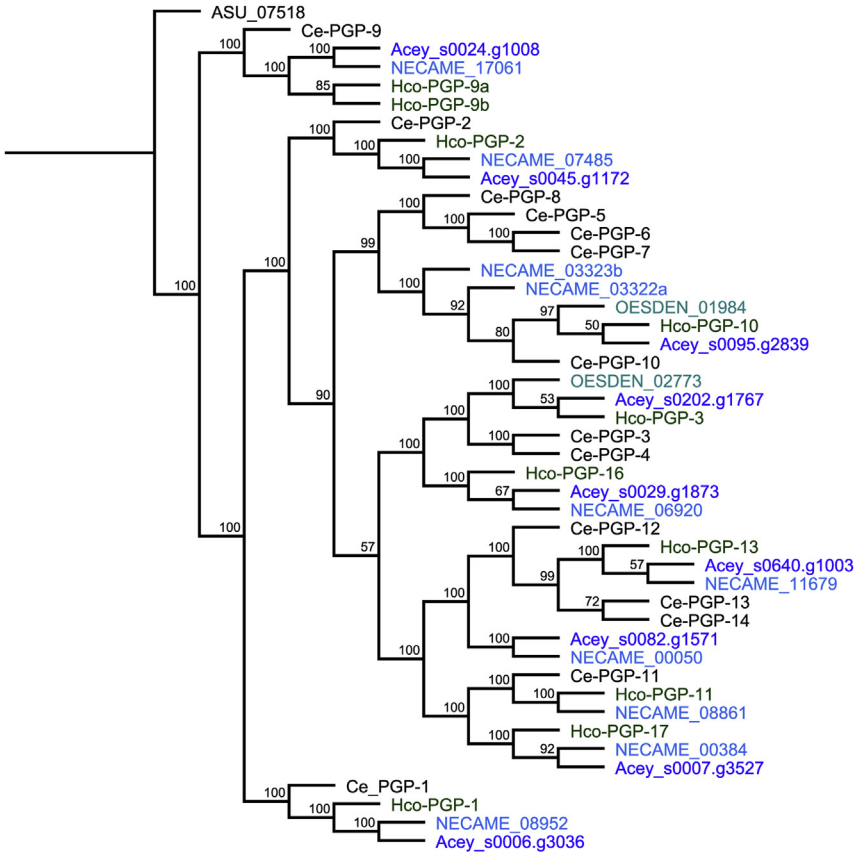

Figure 6 The P-glycoprotein family in *Caenorhabditis elegans*, *Haemonchus contortus*, *Necator americanus*, *Ancylostoma ceylanicum* and *Oesophagostomum dentatum*. Neighbour-joining tree of predicted polypeptide sequences, with *Ascaris suum* PGP as an outgroup.

H. contortus, no *A. ceylanicum*, *N. americanus* or *O. dentatum* orthologs were detected for the *C. elegans pgp-5*, *pgp-6*, *pgp-7* and *pgp-8* cluster. The *H. contortus* tandem duplication of *Hco-pgp-9* was not detected in the other parasite draft genomes, but *N. americanus* has a tandem duplication of *pgp-10*. The *A. ceylanicum* and *N. americanus* genomes encode a *pgp* gene that is not present in *H. contortus* or *C. elegans*, but branches with *pgp-12*, *pgp-13* and *pgp-14*.

4.3 Operons and *trans*-splicing

Operons are clusters of 2–8 genes, which are co-transcribed from the same promoter. Around 17% of *C. elegans* genes are in operons (Allen et al.,

2011), and 23% of them appear to be conserved in *H. contortus* (Laing et al., 2013). A further 10% appear to be partially conserved, ie, at least two orthologs are present in the correct order and orientation, but one or more of the genes appear in a different order or in the opposite orientation in *H. contortus* compared with *C. elegans*, or are missing. This is similar to findings in *N. americanus*, where 28% of *C. elegans* operons appear to be conserved (Tang et al., 2014).

In *C. elegans*, functional constraints are thought to maintain operonic intergenic distance at around 100 bp, although a small number of operons with much larger intergenic distances exist (Morton and Blumenthal, 2011). However, the average intergenic distance in putative *H. contortus* operons is nearly 10-fold greater (Laing et al., 2013). For this analysis, nine putative *H. contortus* operons with intergenic distances of >10 kb were excluded on the assumption that they represented conservation of microsynteny only, but a more recent comparison with putative operons in *N. americanus* (average intergenic distance of 9141 bp; Tang et al., 2014) suggests they should not be ruled out on this basis.

Polycistronic mRNAs transcribed from operons are resolved by 5′ *trans*-splicing with spliced leader sequences (SL1 and SL2); usually, the first gene is *trans*-spliced with SL1, and the downstream genes are *trans*-spliced with SL2. In *H. contortus*, the first genes in all putative operons are SL1 *trans*-spliced, but less than half of the downstream genes appear to be SL2 *trans*-spliced; of these genes, the majority are also SL1 *trans*-spliced. In *C. elegans*, operons with downstream SL1 *trans*-splicing and longer intergenic distances often contain an extra promoter in the intercistronic region (Blumenthal et al., 2015), but the degree to which this is true of *H. contortus* operons is currently unknown.

4.4 Repetitive sequence and mobile elements

As described in Section 3, strongylid genomes contain large amounts of repetitive sequence, which greatly influences genome size. Repetitive sequences include TEs, such as retrotransposons and DNA transposons, and various inverted, tandem and dispersed repeats. The *H. contortus* genome contains 2–2.5% retrotransposons and 2.1–5% DNA transposons (McMaster and MHco3.ISE, respectively) and transcriptomic data suggest that active transposition is occurring. Many TE insertions will have no impact on fitness and may accumulate unnoticed, but the large size and fecundity of *H. contortus* populations should promote rapid selection for any TE insertions that are beneficial and against any that are deleterious. In light of the major

expansions seen in various gene families, and the propensity of *H. contortus* to rapidly adapt to different hosts and environments (Gilleard, 2013), it is tempting to speculate that active transposition could be a significant driver of genetic diversity in this species.

Numerous tandem repeats are found throughout the *H. contortus* genome. Although their function and derivation are largely unknown, their high mutation rate makes them particularly useful for resolving relationships among different populations. Microsatellites are repetitive sequences composed of short tandem repeats and have provided extremely valuable tools for population-genetics studies and fingerprinting of laboratory isolates (Redman et al., 2008b). In *H. contortus*, microsatellites are often found to be associated with one or more copies of a downstream repeat element, HcRep, which is transcribed, but is of unknown function (Callaghan and Beh, 1994; Hoekstra et al., 1997; Laing et al., 2011).

5. FUTURE DIRECTIONS

In recent years, there has been an exponential increase in the availability of draft nematode genomes. For example, the Helminth Genomes Initiative is a joint venture by the Wellcome Trust Sanger Institute and the Genome Institute at Washington University, aiming to generate genome sequences for the parasitic helminths that have the greatest impact on human health and are responsible for diseases of veterinary and agricultural importance. The intention is to rapidly produce draft data for 50 helminths, to complement efforts to generate high-quality reference genomes for exemplars of major parasite groups and for the most important pathogens. These data are all available on WormBase ParaSite. Meanwhile, there have been major advances in the functional annotation of genomes, with comprehensive transcriptome data sets now available for many helminth species (http://nematode.net). Together, the increasing richness of both genomic and transcriptomic resources for strongylid nematodes provides an unprecedented opportunity to better understand the evolution and genetic basis of important traits such as pathogenicity and drug resistance in *H. contortus* and related parasites.

The falling price and increasing throughput of sequencing technology has driven the availability of draft genomes for an increasing number of strongylid species, and is also opening up the prospects of sequencing of multiple individuals and undertaking genomic analysis of *H. contortus* populations. There is some evidence that *H. contortus* shows particularly high

levels of genetic diversity, with a high density of SNP and small indel variants (Redman et al., 2008a), presumably driven by high population sizes. Genome diversity data, combined with the improved genome assembly, should shed light on this, and on other aspects of the diversity, epidemiology and genetics of *H. contortus*. This aspect is further explored in Chapter "Genetic Diversity and Population Structure of *Haemonchus contortus*" by Gilleard and Redman (2016), of this volume.

In both *H. contortus* draft genome papers, and the majority of parasitic nematode genomes published since, the focus lies almost entirely on the protein-coding fraction of the genome. However, there is increasing interest in the important roles performed by nonprotein-coding sequences (Gerstein et al., 2010), including those that are transcribed into functional RNAs (such as tRNAs, rRNAs, small noncoding RNAs and long noncoding RNAs) and regulatory sequences such as transcription start sites and transcription factor binding sites. For example, microRNAs are small noncoding RNAs, which regulate post-transcriptional gene expression. Nearly 200 microRNAs have been identified in the *H. contortus* genome, many of which appear to be unique to this species and may reflect adaptations to different environments and lifestyles (Winter et al., 2012). Their differential expression between the L3 and adult life stages suggests involvement in developmental gene regulation, a function that is well characterized in *C. elegans* (Karp et al., 2011), and a recent study highlights the upregulation of a single *H. contortus* miRNA in two ivermectin-resistant laboratory isolates and two ivermectin-resistant backcross isolates (Gillan and Devaney, in preparation).

The wealth of genomic, transcriptomic and epigenomic data now available for *C. elegans* via modENCODE (Brown and Celniker, 2015) provides a compelling argument for a broader view of functional content in the *H. contortus* genome and in related parasites with reference genomes. Comparative analysis with the *C. elegans* data may identify many conserved aspects of genome organization and function, but it should also be possible to apply similar experimental approaches (eg, CHIP-seq for genome-wide profiling of histone modification or of proteins interacting with DNA (Landt et al., 2012)), albeit on a smaller scale, to study parasites directly. One epigenetic process for which *C. elegans* data cannot provide a model is DNA methylation, as *C. elegans* apparently completely lacks 5-methyl cytosine (Simpson et al., 1986). However, this methylation mark present in the clade I parasite *T. spiralis* varies through the life cycle (Gao et al., 2012) and is directed by small RNA molecules (Sarkies et al., 2015). The presence of genes encoding the cytosine methylation machinery appears

to vary across the nematode phylum (Gutierrez and Sommer, 2004; Wasmuth et al., 2008), but the pathway is conserved in another (distantly related) parasite of clade I, *Romanomermis culcivorax* (see Schiffer et al., 2013). The discovery of adenine methylation in *C. elegans* (see Greer et al., 2015) raises the possibility of an alternative epigenetic mechanism in parasitic nematodes, but the role of DNA methylation and its distribution in other nematodes is currently unknown.

The available genomes of *H. contortus* are already enabling a wide range of exciting molecular research on this species, and a 'finished' reference *H. contortus* genome is now imminent. Although species-specific adaptations must be considered, *H. contortus* is well placed as a parasite model for members of the order Strongylida. Therefore, the availability of a reference genome, combined with draft sequences for related species of human and veterinary importance, will allow researchers to undertake more detailed genetic and genomic analyses than have been possible previously. This comparative approach is likely to be crucial for our understanding of important traits, such as anthelmintic resistance, which, in light of the size and complexity of strongylid genomes, may require whole genome and genetic mapping approaches to solve.

ACKNOWLEDGEMENTS

JAC, AM, NH and AT are funded by the Wellcome Trust through their core support of the Wellcome Trust Sanger Institute (grant 098051). Improvement of the *H. contortus* genome is funded by Wellcome Trust grant 098051 and a BBRSC Strategic Lola grant (BB/M003949/1). RL acknowledges funding from the same BBSRC Strategic Lola and from the Scottish Government through the Strategic Partnership for Animal Science Excellence (SPASE). JG acknowledges support from the Canadian Institutes of Health Research (CIHR) (grant 230927). The authors thank Robin Beech for manual curation of the *H. contortus* P-glycoprotein family; Guillaume Sallé, Stephen Doyle, Hayley Bennett and Collette Britton for comments on this manuscript and Matthew Berriman for his continued support of the *H. contortus* genome project.

REFERENCES

Allen, M.A., Hillier, L.W., Waterston, R.H., Blumenthal, T., 2011. A global analysis of *C. elegans* trans-splicing. Genome Res. 21, 255.

Ardelli, B.F., 2013. Transport proteins of the ABC systems superfamily and their role in drug action and resistance in nematodes. Parasitol. Int. 62, 639.

Assefa, S., Keane, T.M., Otto, T.D., Newbold, C., Berriman, M., 2009. ABACAS: algorithm-based automatic contiguation of assembled sequences. Bioinformatics 25, 1968.

Bai, X., Adams, B.J., Ciche, T.A., Clifton, S., Gaugler, R., Kim, K.S., Spieth, J., Sternberg, P.W., Wilson, R.K., Grewal, P.S., 2013. A lover and a fighter: the genome sequence of an entomopathogenic nematode *Heterorhabditis bacteriophora*. PLoS One 8, e69618.

Blaxter, M.L., 2003. Nematoda: genes, genomes and the evolution of parasitism. Adv. Parasitol. 54, 101.

Blumenthal, T., Davis, P., Garrido-Lecca, A., 2015. Operon and non-operon gene clusters in the *C. elegans* genome. WormBook 1.

Bonfield, J.K., Whitwham, A., 2010. Gap5—editing the billion fragment sequence assembly. Bioinformatics 26, 1699.

Bradnam, K.R., Fass, J.N., Alexandrov, A., Baranay, P., Bechner, M., Birol, I., Boisvert, S., Chapman, J.A., Chapuis, G., Chikhi, R., Chitsaz, H., Chou, W.C., Corbeil, J., Del Fabbro, C., Docking, T.R., Durbin, R., Earl, D., Emrich, S., Fedotov, P., Fonseca, N.A., Ganapathy, G., Gibbs, R.A., Gnerre, S., Godzaridis, E., Goldstein, S., Haimel, M., Hall, G., Haussler, D., Hiatt, J.B., Ho, I.Y., Howard, J., Hunt, M., Jackman, S.D., Jaffe, D.B., Jarvis, E.D., Jiang, H., Kazakov, S., Kersey, P.J., Kitzman, J.O., Knight, J.R., Koren, S., Lam, T.W., Lavenier, D., Laviolette, F., Li, Y., Li, Z., Liu, B., Liu, Y., Luo, R., Maccallum, I., Macmanes, M.D., Maillet, N., Melnikov, S., Naquin, D., Ning, Z., Otto, T.D., Paten, B., Paulo, O.S., Phillippy, A.M., Pina-Martins, F., Place, M., Przybylski, D., Qin, X., Qu, C., Ribeiro, F.J., Richards, S., Rokhsar, D.S., Ruby, J.G., Scalabrin, S., Schatz, M.C., Schwartz, D.C., Sergushichev, A., Sharpe, T., Shaw, T.I., Shendure, J., Shi, Y., Simpson, J.T., Song, H., Tsarev, F., Vezzi, F., Vicedomini, R., Vieira, B.M., Wang, J., Worley, K.C., Yin, S., Yiu, S.M., Yuan, J., Zhang, G., Zhang, H., Zhou, S., Korf, I.F., 2013. Assemblathon 2: evaluating de novo methods of genome assembly in three vertebrate species. Gigascience 2, 10.

Brown, J.B., Celniker, S.E., 2015. Lessons from modENCODE. Annu. Rev. Genomics Hum. Genet. 16.

Burns, A.R., Luciani, G.M., Musso, G., Bagg, R., Yeo, M., Zhang, Y., Rajendran, L., Glavin, J., Hunter, R., Redman, E., Stasiuk, S., Schertzberg, M., Angus McQuibban, G., Caffrey, C.R., Cutler, S.R., Tyers, M., Giaever, G., Nislow, C., Fraser, A.G., MacRae, C.A., Gilleard, J., Roy, P.J., 2015. *Caenorhabditis elegans* is a useful model for anthelmintic discovery. Nat. Commun. 6, 7485.

Callaghan, M.J., Beh, K.J., 1994. Characterization of a tandemly repetitive DNA sequence from *Haemonchus contortus*. Int. J. Parasitol. 24, 137.

Castillo-Davis, C.I., Mekhedov, S.L., Hartl, D.L., Koonin, E.V., Kondrashov, F.A., 2002. Selection for short introns in highly expressed genes. Nat. Genet. 31, 415.

Chain, P.S., Grafham, D.V., Fulton, R.S., Fitzgerald, M.G., Hostetler, J., Muzny, D., Ali, J., Birren, B., Bruce, D.C., Buhay, C., Cole, J.R., Ding, Y., Dugan, S., Field, D., Garrity, G.M., Gibbs, R., Graves, T., Han, C.S., Harrison, S.H., Highlander, S., Hugenholtz, P., Khouri, H.M., Kodira, C.D., Kolker, E., Kyrpides, N.C., Lang, D., Lapidus, A., Malfatti, S.A., Markowitz, V., Metha, T., Nelson, K.E., Parkhill, J., Pitluck, S., Qin, X., Read, T.D., Schmutz, J., Sozhamannan, S., Sterk, P., Strausberg, R.L., Sutton, G., Thomson, N.R., Tiedje, J.M., Weinstock, G., Wollam, A., Genomic Standards Consortium Human Microbiome Project Jumpstart, C, Detter, J.C., 2009. Genomics. Genome project standards in a new era of sequencing. Science 326, 236.

Coghlan, A., Wolfe, K.H., 2002. Fourfold faster rate of genome rearrangement in nematodes than in *Drosophila*. Genome Res. 12, 857.

C. elegans Sequencing Consortium, 1998. Genome sequence of the nematode *C. elegans*: a platform for investigating biology. Science 282, 2012.

Dent, J.A., Smith, M.M., Vassilatis, D.K., Avery, L., 2000. The genetics of ivermectin resistance in *Caenorhabditis elegans*. Proc. Natl. Acad. Sci. U.S.A. 97, 2674.

Dieterich, C., Clifton, S.W., Schuster, L.N., Chinwalla, A., Delehaunty, K., Dinkelacker, I., Fulton, L., Fulton, R., Godfrey, J., Minx, P., Mitreva, M., Roeseler, W., Tian, H., Witte, H., Yang, S.P., Wilson, R.K., Sommer, R.J., 2008. The *Pristionchus pacificus*

genome provides a unique perspective on nematode lifestyle and parasitism. Nat. Genet. 40, 1193.

Dong, Y., Xie, M., Jiang, Y., Xiao, N., Du, X., Zhang, W., Tosser-Klopp, G., Wang, J., Yang, S., Liang, J., Chen, W., Chen, J., Zeng, P., Hou, Y., Bian, C., Pan, S., Li, Y., Liu, X., Wang, W., Servin, B., Sayre, B., Zhu, B., Sweeney, D., Moore, R., Nie, W., Shen, Y., Zhao, R., Zhang, G., Li, J., Faraut, T., Womack, J., Zhang, Y., Kijas, J., Cockett, N., Xu, X., Zhao, S., Wang, J., Wang, W., 2013. Sequencing and automated whole-genome optical mapping of the genome of a domestic goat (*Capra hircus*). Nat. Biotechnol. 31, 135.

Earl, D., Bradnam, K., St John, J., Darling, A., Lin, D., Fass, J., Yu, H.O., Buffalo, V., Zerbino, D.R., Diekhans, M., Nguyen, N., Ariyaratne, P.N., Sung, W.K., Ning, Z., Haimel, M., Simpson, J.T., Fonseca, N.A., Birol, I., Docking, T.R., Ho, I.Y., Rokhsar, D.S., Chikhi, R., Lavenier, D., Chapuis, G., Naquin, D., Maillet, N., Schatz, M.C., Kelley, D.R., Phillippy, A.M., Koren, S., Yang, S.P., Wu, W., Chou, W.C., Srivastava, A., Shaw, T.I., Ruby, J.G., Skewes-Cox, P., Betegon, M., Dimon, M.T., Solovyev, V., Seledtsov, I., Kosarev, P., Vorobyev, D., Ramirez-Gonzalez, R., Leggett, R., MacLean, D., Xia, F., Luo, R., Li, Z., Xie, Y., Liu, B., Gnerre, S., MacCallum, I., Przybylski, D., Ribeiro, F.J., Yin, S., Sharpe, T., Hall, G., Kersey, P.J., Durbin, R., Jackman, S.D., Chapman, J.A., Huang, X., DeRisi, J.L., Caccamo, M., Li, Y., Jaffe, D.B., Green, R.E., Haussler, D., Korf, I., Paten, B., 2011. Assemblathon 1: a competitive assessment of de novo short read assembly methods. Genome Res. 21, 2224.

Fierst, J.L., Willis, J.H., Thomas, C.G., Wang, W., Reynolds, R.M., Ahearne, T.E., Cutter, A.D., Phillips, P.C., 2015. Reproductive mode and the evolution of genome size and structure in *Caenorhabditis* nematodes. PLoS Genet. 11, e1005323.

Gao, F., Liu, X., Wu, X.P., Wang, X.L., Gong, D., Lu, H., Xia, Y., Song, Y., Wang, J., Du, J., Liu, S., Han, X., Tang, Y., Yang, H., Jin, Q., Zhang, X., Liu, M., 2012. Differential DNA methylation in discrete developmental stages of the parasitic nematode *Trichinella spiralis*. Genome Biol. 13, R100.

Gerstein, M.B., Lu, Z.J., Van Nostrand, E.L., Cheng, C., Arshinoff, B.I., Liu, T., Yip, K.Y., Robilotto, R., Rechtsteiner, A., Ikegami, K., Alves, P., Chateigner, A., Perry, M., Morris, M., Auerbach, R.K., Feng, X., Leng, J., Vielle, A., Niu, W., Rhrissorrakrai, K., Agarwal, A., Alexander, R.P., Barber, G., Brdlik, C.M., Brennan, J., Brouillet, J.J., Carr, A., Cheung, M.S., Clawson, H., Contrino, S., Dannenberg, L.O., Dernburg, A.F., Desai, A., Dick, L., Dose, A.C., Du, J., Egelhofer, T., Ercan, S., Euskirchen, G., Ewing, B., Feingold, E.A., Gassmann, R., Good, P.J., Green, P., Gullier, F., Gutwein, M., Guyer, M.S., Habegger, L., Han, T., Henikoff, J.G., Henz, S.R., Hinrichs, A., Holster, H., Hyman, T., Iniguez, A.L., Janette, J., Jensen, M., Kato, M., Kent, W.J., Kephart, E., Khivansara, V., Khurana, E., Kim, J.K., Kolasinska-Zwierz, P., Lai, E.C., Latorre, I., Leahey, A., Lewis, S., Lloyd, P., Lochovsky, L., Lowdon, R.F., Lubling, Y., Lyne, R., MacCoss, M., Mackowiak, S.D., Mangone, M., McKay, S., Mecenas, D., Merrihew, G., Miller 3rd, D.M., Muroyama, A., Murray, J.I., Ooi, S.L., Pham, H., Phippen, T., Preston, E.A., Rajewsky, N., Ratsch, G., Rosenbaum, H., Rozowsky, J., Rutherford, K., Ruzanov, P., Sarov, M., Sasidharan, R., Sboner, A., Scheid, P., Segal, E., Shin, H., Shou, C., Slack, F.J., 2010. Integrative analysis of the *Caenorhabditis elegans* genome by the modENCODE project. Science 330, 1775.

Ghedin, E., Wang, S., Spiro, D., Caler, E., Zhao, Q., Crabtree, J., Allen, J.E., Delcher, A.L., Guiliano, D.B., Miranda-Saavedra, D., Angiuoli, S.V., Creasy, T., Amedeo, P., Haas, B., El-Sayed, N.M., Wortman, J.R., Feldblyum, T., Tallon, L., Schatz, M., Shumway, M., Koo, H., Salzberg, S.L., Schobel, S., Pertea, M., Pop, M., White, O., Barton, G.J., Carlow, C.K., Crawford, M.J., Daub, J., Dimmic, M.W., Estes, C.F., Foster, J.M.,

Ganatra, M., Gregory, W.F., Johnson, N.M., Jin, J., Komuniecki, R., Korf, I., Kumar, S., Laney, S., Li, B.W., Li, W., Lindblom, T.H., Lustigman, S., Ma, D., Maina, C.V., Martin, D.M., McCarter, J.P., McReynolds, L., Mitreva, M., Nutman, T.B., Parkinson, J., Peregrin-Alvarez, J.M., Poole, C., Ren, Q., Saunders, L., Sluder, A.E., Smith, K., Stanke, M., Unnasch, T.R., Ware, J., Wei, A.D., Weil, G., Williams, D.J., Zhang, Y., Williams, S.A., Fraser-Liggett, C., Slatko, B., Blaxter, M.L., Scott, A.L., 2007. Draft genome of the filarial nematode parasite *Brugia malayi*. Science 317, 1756.

Ghosh, R., Andersen, E.C., Shapiro, J.A., Gerke, J.P., Kruglyak, L., 2012. Natural variation in a chloride channel subunit confers avermectin resistance in *C. elegans*. Science 335, 574.

Gilleard, J., Redman, E., 2016. Genetic diversity and population structure of *Haemonchus contortus*. In: Gasser, R., Samson-Himmelstjerna, G.V. (Eds.), Haemonchus contortus and Haemonchosis Past, Present and Future Trends. vol. 93, pp. 31–68,.

Gilleard, J.S., Beech, R.N., 2007. Population genetics of anthelmintic resistance in parasitic nematodes. Parasitology 134, 1133.

Gilleard, J.S., 2013. *Haemonchus contortus* as a paradigm and model to study anthelmintic drug resistance. Parasitology 140, 1506.

Glendinning, S.K., Buckingham, S.D., Sattelle, D.B., Wonnacott, S., Wolstenholme, A.J., 2011. Glutamate-gated chloride channels of *Haemonchus contortus* restore drug sensitivity to ivermectin resistant *Caenorhabditis elegans*. PLoS One 6, e22390.

Greer, E.L., Blanco, M.A., Gu, L., Sendinc, E., Liu, J., Aristizabal-Corrales, D., Hsu, C.H., Aravind, L., He, C., Shi, Y., 2015. DNA methylation on N6-adenine in *C. elegans*. Cell 161, 868.

Guiliano, D.B., Hall, N., Jones, S.J., Clark, L.N., Corton, C.H., Barrell, B.G., Blaxter, M.L., 2002. Conservation of long-range synteny and microsynteny between the genomes of two distantly related nematodes. Genome Biol. 3. RESEARCH0057.

Gutierrez, A., Sommer, R.J., 2004. Evolution of dnmt-2 and mbd-2-like genes in the free-living nematodes *Pristionchus pacificus*, *Caenorhabditis elegans* and *Caenorhabditis briggsae*. Nucleic Acids Res. 32, 6388.

Hawdon, J.M., Jones, B.F., Hoffman, D.R., Hotez, P.J., 1996. Cloning and characterization of *Ancylostoma*-secreted protein. A novel protein associated with the transition to parasitism by infective hookworm larvae. J. Biol. Chem. 271, 6672.

Hawdon, J.M., Narasimhan, S., Hotez, P.J., 1999. *Ancylostoma* secreted protein 2: cloning and characterization of a second member of a family of nematode secreted proteins from *Ancylostoma caninum*. Mol. Biochem. Parasitol. 99, 149.

Hoekstra, R., Criado-Fornelio, A., Fakkeldij, J., Bergman, J., Roos, M.H., 1997. Microsatellites of the parasitic nematode *Haemonchus contortus*: polymorphism and linkage with a direct repeat. Mol. Biochem. Parasitol. 89, 97.

Huang, S., Chen, Z., Huang, G., Yu, T., Yang, P., Li, J., Fu, Y., Yuan, S., Chen, S., Xu, A., 2012. HaploMerger: reconstructing allelic relationships for polymorphic diploid genome assemblies. Genome Res. 22, 1581.

Hube, F., Francastel, C., 2015. Mammalian introns: when the junk generates molecular diversity. Int. J. Mol. Sci. 16, 4429.

Hunt, M., Kikuchi, T., Sanders, M., Newbold, C., Berriman, M., Otto, T.D., 2013. REAPR: a universal tool for genome assembly evaluation. Genome Biol. 14, R47.

Jex, A.R., Liu, S., Li, B., Young, N.D., Hall, R.S., Li, Y., Yang, L., Zeng, N., Xu, X., Xiong, Z., Chen, F., Wu, X., Zhang, G., Fang, X., Kang, Y., Anderson, G.A., Harris, T.W., Campbell, B.E., Vlaminck, J., Wang, T., Cantacessi, C., Schwarz, E.M., Ranganathan, S., Geldhof, P., Nejsum, P., Sternberg, P.W., Yang, H., Wang, J., Wang, J., Gasser, R.B., 2011. *Ascaris suum* draft genome. Nature 479, 529.

Kaminsky, R., Ducray, P., Jung, M., Clover, R., Rufener, L., Bouvier, J., Weber, S.S., Wenger, A., Wieland-Berghausen, S., Goebel, T., Gauvry, N., Pautrat, F.,

Skripsky, T., Froelich, O., Komoin-Oka, C., Westlund, B., Sluder, A., Maser, P., 2008. A new class of anthelmintics effective against drug-resistant nematodes. Nature 452, 176.

Karp, X., Hammell, M., Ow, M.C., Ambros, V., 2011. Effect of life history on microRNA expression during *C. elegans* development. RNA 17, 639.

Korlach, J., Bjornson, K.P., Chaudhuri, B.P., Cicero, R.L., Flusberg, B.A., Gray, J.J., Holden, D., Saxena, R., Wegener, J., Turner, S.W., 2010. Real-time DNA sequencing from single polymerase molecules. Methods Enzymol. 472, 431.

Kotze, A.C., Hunt, P.W., Skuce, P., von Samson-Himmelstjerna, G., Martin, R.J., Sager, H., Krucken, J., Hodgkinson, J., Lespine, A., Jex, A.R., Gilleard, J.S., Beech, R.N., Wolstenholme, A.J., Demeler, J., Robertson, A.P., Charvet, C.L., Neveu, C., Kaminsky, R., Rufener, L., Alberich, M., Menez, C., Prichard, R.K., 2014. Recent advances in candidate-gene and whole-genome approaches to the discovery of anthelmintic resistance markers and the description of drug/receptor interactions. Int. J. Parasitol. Drugs Drug Resist. 4, 164.

Koutsovoulos, G., Makepeace, B., Tanya, V.N., Blaxter, M., 2014. Palaeosymbiosis revealed by genomic fossils of Wolbachia in a strongyloidean nematode. PLoS Genet. 10, e1004397.

Kumar, S., Koutsovoulos, G., Kaur, G., Blaxter, M., 2012. Toward 959 nematode genomes. Worm 1, 42.

Laing, S.T., Ivens, A., Laing, R., Ravikumar, S., Butler, V., Woods, D.J., Gilleard, J.S., 2010. Characterization of the xenobiotic response of *Caenorhabditis elegans* to the anthelmintic drug albendazole and the identification of novel drug glucoside metabolites. Biochem. J. 432, 505.

Laing, R., Hunt, M., Protasio, A.V., Saunders, G., Mungall, K., Laing, S., Jackson, F., Quail, M., Beech, R., Berriman, M., Gilleard, J.S., 2011. Annotation of two large contiguous regions from the *Haemonchus contortus* genome using RNA-seq and comparative analysis with *Caenorhabditis elegans*. PLoS One 6, e23216.

Laing, R., Kikuchi, T., Martinelli, A., Tsai, I.J., Beech, R.N., Redman, E., Holroyd, N., Bartley, D.J., Beasley, H., Britton, C., Curran, D., Devaney, E., Gilabert, A., Hunt, M., Jackson, F., Johnston, S.L., Kryukov, I., Li, K., Morrison, A.A., Reid, A.J., Sargison, N., Saunders, G.I., Wasmuth, J.D., Wolstenholme, A., Berriman, M., Gilleard, J.S., Cotton, J.A., 2013. The genome and transcriptome of *Haemonchus contortus*, a key model parasite for drug and vaccine discovery. Genome Biol. 14, R88.

Laing, R., Bartley, D.J., Morrison, A.A., Rezansoff, A., Martinelli, A., Laing, S.T., Gilleard, J.S., 2015. The cytochrome P450 family in the parasitic nematode *Haemonchus contortus*. Int. J. Parasitol. 45, 243.

Lam, E.T., Hastie, A., Lin, C., Ehrlich, D., Das, S.K., Austin, M.D., Deshpande, P., Cao, H., Nagarajan, N., Xiao, M., Kwok, P.Y., 2012. Genome mapping on nanochannel arrays for structural variation analysis and sequence assembly. Nat. Biotechnol. 30, 771.

Landt, S.G., Marinov, G.K., Kundaje, A., Kheradpour, P., Pauli, F., Batzoglou, S., Bernstein, B.E., Bickel, P., Brown, J.B., Cayting, P., Chen, Y., DeSalvo, G., Epstein, C., Fisher-Aylor, K.I., Euskirchen, G., Gerstein, M., Gertz, J., Hartemink, A.J., Hoffman, M.M., Iyer, V.R., Jung, Y.L., Karmakar, S., Kellis, M., Kharchenko, P.V., Li, Q., Liu, T., Liu, X.S., Ma, L., Milosavljevic, A., Myers, R.M., Park, P.J., Pazin, M.J., Perry, M.D., Raha, D., Reddy, T.E., Rozowsky, J., Shoresh, N., Sidow, A., Slattery, M., Stamatoyannopoulos, J.A., Tolstorukov, M.Y., White, K.P., Xi, S., Farnham, P.J., Lieb, J.D., Wold, B.J., Snyder, M., 2012. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. Genome Res. 22, 1813.

Mardis, E.R., 2008. Next-generation DNA sequencing methods. Annu. Rev. Genomics Hum. Genet. 9, 387.

Markov, G.V., Baskaran, P., Sommer, R.J., 2015. The same or not the same: lineage-specific gene expansions and homology relationships in multigene families in nematodes. J. Mol. Evol. 80, 18.

McCavera, S., Walsh, T.K., Wolstenholme, A.J., 2007. Nematode ligand-gated chloride channels: an appraisal of their involvement in macrocyclic lactone resistance and prospects for developing molecular markers. Parasitology 134, 1111.

Mitreva, M., Jasmer, D.P., Zarlenga, D.S., Wang, Z., Abubucker, S., Martin, J., Taylor, C.M., Yin, Y., Fulton, L., Minx, P., Yang, S.P., Warren, W.C., Fulton, R.S., Bhonagiri, V., Zhang, X., Hallsworth-Pepin, K., Clifton, S.W., McCarter, J.P., Appleton, J., Mardis, E.R., Wilson, R.K., 2011. The draft genome of the parasitic nematode *Trichinella spiralis*. Nat. Genet. 43, 228.

Morassutti, A.L., Perelygin, A., DE Carvalho, M.O., Lemos, L.N., Pinto, P.M., Frace, M., Wilkins, P.P., Graeff-Teixeira, C., DA Silva, A.J., 2013. High throughput sequencing of the *Angiostrongylus cantonensis* genome: a parasite spreading worldwide. Parasitology 140, 1304.

Morton, J.J., Blumenthal, T., 2011. Identification of transcription start sites of trans-spliced genes: uncovering unusual operon arrangements. RNA 17, 327.

Myers, E.W., Sutton, G.G., Delcher, A.L., Dew, I.M., Fasulo, D.P., Flanigan, M.J., Kravitz, S.A., Mobarry, C.M., Reinert, K.H., Remington, K.A., Anson, E.L., Bolanos, R.A., Chou, H.H., Jordan, C.M., Halpern, A.L., Lonardi, S., Beasley, E.M., Brandon, R.C., Chen, L., Dunn, P.J., Lai, Z., Liang, Y., Nusskern, D.R., Zhan, M., Zhang, Q., Zheng, X., Rubin, G.M., Adams, M.D., Venter, J.C., 2000. A whole-genome assembly of *Drosophila*. Science 287, 2196.

Nagarajan, N., Pop, M., 2013. Sequence assembly demystified. Nat. Rev. Genet. 14, 157.

Neveu, C., Charvet, C.L., Fauvin, A., Cortet, J., Beech, R.N., Cabaret, J., 2010. Genetic diversity of levamisole receptor subunits in parasitic nematode species and abbreviated transcripts associated with resistance. Pharmacogenet. Genomics 20, 414.

Otto, T.D., Sanders, M., Berriman, M., Newbold, C., 2010. Iterative Correction of Reference Nucleotides (iCORN) using second generation sequencing technology. Bioinformatics 26, 1704.

Palazzo, A.F., Gregory, T.R., 2014. The case for junk DNA. PLoS Genet. 10, e1004351.

Parra, G., Bradnam, K., Korf, I., 2007. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. Bioinformatics 23, 1061.

Protasio, A.V., Tsai, I.J., Babbage, A., Nichol, S., Hunt, M., Aslett, M.A., De Silva, N., Velarde, G.S., Anderson, T.J., Clark, R.C., Davidson, C., Dillon, G.P., Holroyd, N.E., LoVerde, P.T., Lloyd, C., McQuillan, J., Oliveira, G., Otto, T.D., Parker-Manuel, S.J., Quail, M.A., Wilson, R.A., Zerlotini, A., Dunne, D.W., Berriman, M., 2012. A systematically improved high quality genome and transcriptome of the human blood fluke Schistosoma mansoni. PLoS Negl. Trop. Dis. 6, e1455.

Redman, E., Grillo, V., Saunders, G., Packard, E., Jackson, F., Berriman, M., Gilleard, J.S., 2008a. Genetics of mating and sex determination in the parasitic nematode *Haemonchus contortus*. Genetics 180, 1877.

Redman, E., Packard, E., Grillo, V., Smith, J., Jackson, F., Gilleard, J.S., 2008b. Microsatellite analysis reveals marked genetic differentiation between *Haemonchus contortus* laboratory isolates and provides a rapid system of genetic fingerprinting. Int. J. Parasitol. 38, 111.

Redman, E., Sargison, N., Whitelaw, F., Jackson, F., Morrison, A., Bartley, D.J., Gilleard, J.S., 2012. Introgression of ivermectin resistance genes into a susceptible *Haemonchus contortus* strain by multiple backcrossing. PLoS Pathog. 8, e1002534.

Roos, M.H., Otsen, M., Hoekstra, R., Veenstra, J.G., Lenstra, J.A., 2004. Genetic analysis of inbreeding of two strains of the parasitic nematode *Haemonchus contortus*. Int. J. Parasitol. 34, 109.

Samad, A., Huff, E.F., Cai, W., Schwartz, D.C., 1995. Optical mapping: a novel, single-molecule approach to genomic analysis. Genome Res. 5, 1.

Sarkies, P., Selkirk, M.E., Jones, J.T., Blok, V., Boothby, T., Goldstein, B., Hanelt, B., Ardila-Garcia, A., Fast, N.M., Schiffer, P.M., Kraus, C., Taylor, M.J., Koutsovoulos, G.,

Blaxter, M.L., Miska, E.A., 2015. Ancient and novel small RNA pathways compensate for the loss of piRNAs in multiple independent nematode lineages. PLoS Biol. 13, e1002061.

Saunders, G.I., Wasmuth, J.D., Beech, R., Laing, R., Hunt, M., Naghra, H., Cotton, J.A., Berriman, M., Britton, C., Gilleard, J.S., 2013. Characterization and comparative analysis of the complete *Haemonchus contortus* beta-tubulin gene family and implications for benzimidazole resistance in strongylid nematodes. Int. J. Parasitol. 43, 465.

Schatz, M.C., Delcher, A.L., Salzberg, S.L., 2010. Assembly of large genomes using second-generation sequencing. Genome Res. 20, 1165.

Schiffer, P.H., Kroiher, M., Kraus, C., Koutsovoulos, G.D., Kumar, S., Camps, J.I., Nsah, N.A., Stappert, D., Morris, K., Heger, P., Altmuller, J., Frommolt, P., Nurnberg, P., Thomas, W.K., Blaxter, M.L., Schierenberg, E., 2013. The genome of Romanomermis culicivorax: revealing fundamental changes in the core developmental genetic toolkit in Nematoda. BMC Genomics 14, 923.

Schwarz, E.M., Korhonen, P.K., Campbell, B.E., Young, N.D., Jex, A.R., Jabbar, A., Hall, R.S., Mondal, A., Howe, A.C., Pell, J., Hofmann, A., Boag, P.R., Zhu, X.Q., Gregory, T., Loukas, A., Williams, B.A., Antoshechkin, I., Brown, C., Sternberg, P.W., Gasser, R.B., 2013. The genome and developmental transcriptome of the strongylid nematode *Haemonchus contortus*. Genome Biol. 14, R89.

Schwarz, E.M., Hu, Y., Antoshechkin, I., Miller, M.M., Sternberg, P.W., Aroian, R.V., 2015. The genome and transcriptome of the zoonotic hookworm *Ancylostoma ceylanicum* identify infection-specific gene families. Nat. Genet. 47, 416.

Simpson, J.T., Durbin, R., 2010. Efficient construction of an assembly string graph using the FM-index. Bioinformatics 26, i367.

Simpson, V.J., Johnson, T.E., Hammen, R.F., 1986. *Caenorhabditis elegans* DNA does not contain 5-methylcytosine at any time during development or aging. Nucleic Acids Res. 14, 6711.

Stein, L.D., Bao, Z., Blasiar, D., Blumenthal, T., Brent, M.R., Chen, N., Chinwalla, A., Clarke, L., Clee, C., Coghlan, A., Coulson, A., D'Eustachio, P., Fitch, D.H., Fulton, L.A., Fulton, R.E., Griffiths-Jones, S., Harris, T.W., Hillier, L.W., Kamath, R., Kuwabara, P.E., Mardis, E.R., Marra, M.A., Miner, T.L., Minx, P., Mullikin, J.C., Plumb, R.W., Rogers, J., Schein, J.E., Sohrmann, M., Spieth, J., Stajich, J.E., Wei, C., Willey, D., Wilson, R.K., Durbin, R., Waterston, R.H., 2003. The genome sequence of *Caenorhabditis briggsae*: a platform for comparative genomics. PLoS Biol. 1, E45.

Tang, Y.T., Gao, X., Rosa, B.A., Abubucker, S., Hallsworth-Pepin, K., Martin, J., Tyagi, R., Heizer, E., Zhang, X., Bhonagiri-Palsikar, V., Minx, P., Warren, W.C., Wang, Q., Zhan, B., Hotez, P.J., Sternberg, P.W., Dougall, A., Gaze, S.T., Mulvenna, J., Sotillo, J., Ranganathan, S., Rabelo, E.M., Wilson, R.K., Felgner, P.L., Bethony, J., Hawdon, J.M., Gasser, R.B., Loukas, A., Mitreva, M., 2014. Genome of the human hookworm *Necator americanus*. Nat. Genet. 46, 261.

Tsai, I.J., Otto, T.D., Berriman, M., 2010. Improving draft assemblies by iterative mapping and assembly of short reads to eliminate gaps. Genome Biol. 11, R41.

Tsai, I.J., Zarowiecki, M., Holroyd, N., Garciarrubio, A., Sanchez-Flores, A., Brooks, K.L., Tracey, A., Bobes, R.J., Fragoso, G., Sciutto, E., Aslett, M., Beasley, H., Bennett, H.M., Cai, J., Camicia, F., Clark, R., Cucher, M., De Silva, N., Day, T.A., Deplazes, P., Estrada, K., Fernandez, C., Holland, P.W., Hou, J., Hu, S., Huckvale, T., Hung, S.S., Kamenetzky, L., Keane, J.A., Kiss, F., Koziol, U., Lambert, O., Liu, K., Luo, X., Luo, Y., Macchiaroli, N., Nichol, S., Paps, J., Parkinson, J., Pouchkina-Stantcheva, N., Riddiford, N., Rosenzvit, M., Salinas, G., Wasmuth, J.D., Zamanian, M., Zheng, Y., Taenia solium Genome Consortium, Cai, X., Soberon, X., Olson, P.D., Laclette, J.P., Brehm, K., Berriman, M., 2013. The genomes of four tapeworm species reveal adaptations to parasitism. Nature 496, 57.

Tyagi, R., Joachim, A., Ruttkowski, B., Rosa, B.A., Martin, J.C., Hallsworth-Pepin, K., Zhang, X., Ozersky, P., Wilson, R.K., Ranganathan, S., Sternberg, P.W., Gasser, R.B., Mitreva, M., 2015a. Cracking the nodule worm code advances knowledge of parasite biology and biotechnology to tackle major diseases of livestock. Biotechnol. Adv. 33.

Tyagi, R., Rosa, B.A., Lewis, W.G., Mitreva, M., 2015b. Pan-phylum comparison of nematode metabolic potential. PLoS Negl. Trop. Dis. 9, e0003788.

Wasmuth, J., Schmid, R., Hedley, A., Blaxter, M., 2008. On the extent and origins of genic novelty in the phylum Nematoda. PLoS Negl. Trop. Dis. 2, e258.

Williamson, S.M., Storey, B., Howell, S., Harper, K.M., Kaplan, R.M., Wolstenholme, A.J., 2011. Candidate anthelmintic resistance-associated gene expression and sequence polymorphisms in a triple-resistant field isolate of *Haemonchus contortus*. Mol. Biochem. Parasitol. 180, 99.

Winter, A.D., Weir, W., Hunt, M., Berriman, M., Gilleard, J.S., Devaney, E., Britton, C., 2012. Diversity in parasitic nematode genomes: the microRNAs of *Brugia pahangi* and *Haemonchus contortus* are largely novel. BMC Genomics 13, 4.

Yong, H.S., Eamsobhana, P., Lim, P.E., Razali, R., Aziz, F.A., Rosli, N.S., Poole-Johnson, J., Anwar, A., 2015. Draft genome of neurotropic nematode parasite *Angiostrongylus cantonensis*, causative agent of human eosinophilic meningitis. Acta Trop. 148, 51.

Zarowiecki, M., Berriman, M., 2015. What helminth genomes have taught us about parasite evolution. Parasitology 142 (Suppl. 1), S85.

Zerbino, D.R., Birney, E., 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 18, 821.