The Genome of Caenorhabditis bovis

Article

Current Biology

The Genome of *Caenorhabditis bovis*

Highlights

- Isolated *Caenorhabditis bovis* from ear of a cow at livestock market in rural Kenya
- Sequenced genome using the Oxford Nanopore MinION in nearby field laboratory
- *C. bovis* genome is 40% smaller than *C. elegans* and encodes 7,000 fewer genes
- Most closely related to a species isolated from carrion

Authors

Lewis Stevens, Stefan Rooke, Laura C. Falzon, ..., Joseph Ogola, Eric M. Fèvre, Mark Blaxter

Correspondence

lewis.stevens@ed.ac.uk

In Brief

Stevens et al. isolate *Caenorhabditis bovis* from the ear of a cow in Kenya and sequence its genome in a nearby field laboratory using the Oxford Nanopore MinION. They compare the *C. bovis* genome with that of *C. elegans*, identify genes that may relate to its unusual ecology, and determine how *C. bovis* is related to other *Caenorhabditis* species.

Stevens et al., 2020, Current Biology *30*, 1–9
March 23, 2020
https://doi.org/10.1016/j.cub.2020.01.074

CellPress

Please cite this article in press as: Stevens et al., The Genome of *Caenorhabditis bovis*, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.01.074

Current Biology
Article

The Genome of *Caenorhabditis bovis*

Lewis Stevens,[1,10,*] Stefan Rooke,[2] Laura C. Falzon,[3,4] Eunice M. Machuka,[5] Kelvin Momanyi,[4] Maurice K. Murungi,[4] Samuel M. Njoroge,[4,6] Christian O. Odinga,[4] Allan Ogendo,[7] Joseph Ogola,[8] Eric M. Fèvre,[3,4] and Mark Blaxter[1,9]
[1]Institute of Evolutionary Biology, Ashworth Laboratories, School of Biological Sciences, University of Edinburgh, Edinburgh EH9 3JT, UK
[2]Usher Institute, College of Medicine and Veterinary Medicine, University of Edinburgh, Edinburgh EH9 3JT, UK
[3]Institute of Infection and Global Health, University of Liverpool, 8 West Derby Street, Liverpool L69 7BE, UK
[4]International Livestock Research Institute, Old Naivasha Road, PO Box 30709 00100, Nairobi, Kenya
[5]Biosciences, Eastern and Central Africa, International Livestock Research Institute (BecA-ILRI) Hub, Old Naivasha Road, PO Box 30709 00100, Nairobi, Kenya
[6]Centre for Microbiology Research, Kenya Medical Research Institute, KNH Grounds, PO Box 54840 00200, Nairobi, Kenya
[7]Veterinary Department, Busia County Government, PO Box Private Bag 50400, Busia, Kenya
[8]Veterinary Department, Bungoma County Government, PO Box 2489 50200, Bungoma, Kenya
[9]Present address: Tree of Life, Wellcome Sanger Institute, Cambridge CB10 1SA, UK
[10]Lead Contact
*Correspondence: lewis.stevens@ed.ac.uk
https://doi.org/10.1016/j.cub.2020.01.074

SUMMARY

The free-living nematode *Caenorhabditis elegans* is a key laboratory model for metazoan biology. *C. elegans* has also become a model for parasitic nematodes despite being only distantly related to most parasitic species. All of the ~65 *Caenorhabditis* species currently in culture are free-living, with most having been isolated from decaying plant or fungal matter. *Caenorhabditis bovis* is a particularly unusual species that has been isolated several times from the inflamed ears of Zebu cattle in Eastern Africa, where it is associated with the disease bovine parasitic otitis. *C. bovis* is therefore of particular interest to researchers interested in the evolution of nematode parasitism. However, as *C. bovis* is not in laboratory culture, it remains little studied. Here, by sampling livestock markets and slaughterhouses in Western Kenya, we successfully reisolated *C. bovis* from the ear of adult female Zebu. We sequenced the genome of *C. bovis* using the Oxford Nanopore MinION platform in a nearby field laboratory and used the data to generate a chromosome-scale draft genome sequence. We exploited this draft genome sequence to reconstruct the phylogenetic relationships of *C. bovis* to other *Caenorhabditis* species and reveal the changes in genome size and content that have occurred during its evolution. We also identified expansions in several gene families that have been implicated in parasitism in other nematode species. The high-quality draft genome and our analyses thereof represent a significant advancement in our understanding of this unusual *Caenorhabditis* species.

INTRODUCTION

The free-living nematode *Caenorhabditis elegans* is used extensively as a model for animal development, genetics, and neurobiology. As the most well-studied species within the phylum Nematoda, *C. elegans* has also become a model for this extremely abundant and diverse group of animals, many of which are parasites [1, 2]. Attempts to understand the evolutionary origins and genetic basis of nematode parasitism often involve comparisons between parasitic nematode species and *C. elegans* [3, 4]. However, *C. elegans* is only distantly related to most parasitic species, which limits the efficacy of comparative studies [1]. Recent years have seen significant progress in our understanding of *Caenorhabditis* diversity, with over 30 new species discovered in the last decade [5–8]. However, all of the ~65 species currently in culture are free-living, with the vast majority having been isolated from rotting fruits and flowers [5–8].

Caenorhabditis bovis [9] is therefore particularly unusual for a *Caenorhabditis* species, having been isolated several times from the outer auditory canals of Zebu cattle in Eastern Africa [10] and recently from Gyr cattle in South America [11]. *C. bovis* is believed to be the causative agent of bovine parasitic otitis, a disease that causes a range of symptoms including inflammation, dark brown discharge from the affected ear, and dullness [12]. In severe cases, bovine parasitic otitis can result in mortality [12]. As is typical for a *Caenorhabditis* species, *C. bovis* is believed to have a phoretic association with an invertebrate, with larvae of the Old World screwworm fly (*Chrysomya bezziana*) also being found in the ears of Zebu cattle [12, 13]. It is unclear to what extent bovine parasitic otitis is caused directly by *C. bovis* or by bacterial and/or fungal infections, and therefore to what extent *C. bovis* can be considered a parasite. Despite this, its close association with a vertebrate means that *C. bovis* is of particular interest to researchers interested in the evolution of nematode parasitism and in *Caenorhabditis* diversity. However, as *C. bovis* is not in laboratory culture, it remains little studied.

In collaboration with local veterinarians and scientists, we sampled cattle at livestock markets and slaughterhouses in

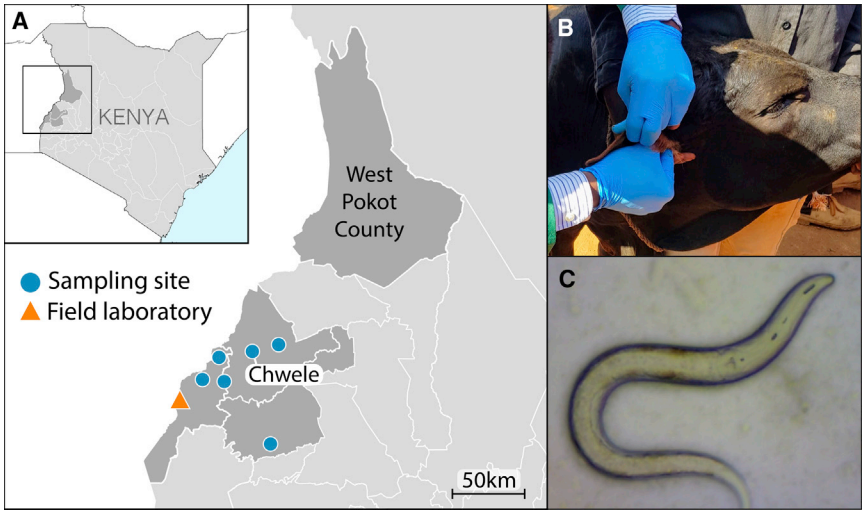

Figure 1. Cattle Sampling and Nematode Isolation

(A) Sampling locations in Western Kenya. We isolated *C. bovis* from an adult female Zebu sampled at a livestock market in Chwele. The animal was believed to have originated from West Pokot County. The location of the field laboratory in Busia is also shown. GPS coordinates and the number of animals sampled at each site can be found in Table S1.

(B) An animal being sampled using cotton wool soaked in physiological saline.

(C) Adult female *C. bovis* under a stereo microscope (*C. bovis* adults are ~1 mm in length [9]).

Western Kenya and successfully reisolated *C. bovis* from the ear of an adult female Zebu. We sequenced the genome of *C. bovis* in a nearby field laboratory using the Oxford Nanopore MinION platform and used the data to generate a high-quality, chromosome-scale draft genome sequence. We exploited this genome to determine the phylogenetic relationships of *C. bovis* to other species in the genus *Caenorhabditis*, including *C. elegans*, and reveal changes in genome size and content that have occurred during its evolution. We also reveal specific expansions in several gene families that may play a role in its unusual lifestyle. The high-quality draft genome and the analyses presented here represent a major step forward in our understanding of this unusual and understudied *Caenorhabditis* species.

RESULTS

Reisolation of *C. bovis*

We sampled a total of 44 cattle of various ages and breeds at livestock markets and slaughterhouses in three counties in Western Kenya (Figure 1A; Table S1). Sampling was performed by washing the outer auditory canal of each animal with cotton wool soaked in physiological saline (Figure 1B), which was subsequently inspected under a dissecting microscope. We identified only a single instance of bovine parasitic otitis. The affected animal was an adult female Zebu that was sampled at a livestock market in Chwele, Bungoma County (Figure 1A). The animal is believed to have originated from West Pokot County (Figure 1A). Although we noted no obvious clinical symptoms, the cotton wool sample had an unpleasant odor, consistent with previous reports of bovine parasitic otitis [12]. We isolated approximately 50 live nematode larvae from the sample, which were subsequently cultured on *E. coli*-seeded agar plates. The cultures thrived at 37°C on both nematode growth medium (NGM) and horse blood agar plates. Using a standard compound microscope, we identified adult nematodes as members of the genus *Caenorhabditis* based on their morphology (presence of a prominent pharyngeal bulb and filiform female tail). The morphology of the adult male tail (anteriorly closed fan, ray pattern with gap between GP2 and GP3, and a bent gubernaculum) was consistent with previous descriptions of *C. bovis* [9, 14].

A High-Quality, Chromosome-Scale *C. bovis* Reference Genome

We sought to generate a high-quality reference genome for *C. bovis*. We took advantage of the portability of the Oxford Nanopore MinION platform and sequenced the genome of *C. bovis* in a field laboratory in Busia, Western Kenya (Figure 1A). We generated 11.3 Gb of sequence data representing ~180-fold coverage of the *C. bovis* genome using two MinION v9.4 flow cells. Read length N50s were 11.4 kb and 4.3 kb, respectively, with the longest read spanning 242 kb (Table S2; Figure S1). We also sequenced the genome to ~210-fold coverage (13.3 Gb) using the Illumina MiSeq platform at the BecA-ILRI Hub in Nairobi, Kenya. We identified and discarded reads originating from contaminant organisms, including several bacterial species that are known mammalian pathogens, using taxon-annotated GC-coverage plots (Figure S2).

We assembled the *C. bovis* genome using the MinION long reads and corrected residual sequencing errors in the assembly using the Illumina short reads. The resulting assembly comprises 35 contigs spanning 62.7 Mb with a contig N50 of 7.6 Mb, with half of the assembly contained in just 4 contigs (Figures 2A and 2B; Table 1). The assembly is highly complete, with 94.2% of a conserved set of nematode genes being present and fully assembled. In contrast to other outcrossing species, whose genomes typically contain high levels of heterozygosity [16], we find that the genome of *C. bovis* contains surprisingly little heterozygosity. Using a variant calling approach, we estimate that 0.03% of sites in the *C. bovis* genome are heterozygous (1 heterozygous site every ~3,760 bp), with the k-mer distribution of the Illumina data indicating that the genome is essentially homozygous (Figure S3). Using protein sequences predicted from the genomes of related nematodes as homology evidence, we predicted 13,128 protein-coding genes in the *C. bovis* genome. We note that this number is considerably lower than the number of genes predicted in the genomes of other *Caenorhabditis* species [8]. However, the gene set contains 95.1% of a conserved set of nematode genes (Table 1), suggesting that the reduced count is not due to an incomplete gene set.

Chromosomal linkage groups are highly conserved in *Caenorhabditis* [17]. We defined 7,706 one-to-one orthologs between *C. bovis* and *C. elegans*, and exploited this conservation to assign 15 contigs (representing 99.4% of the *C. bovis* assembly) to the six *C. elegans* chromosomes (Figures 2A and S4).

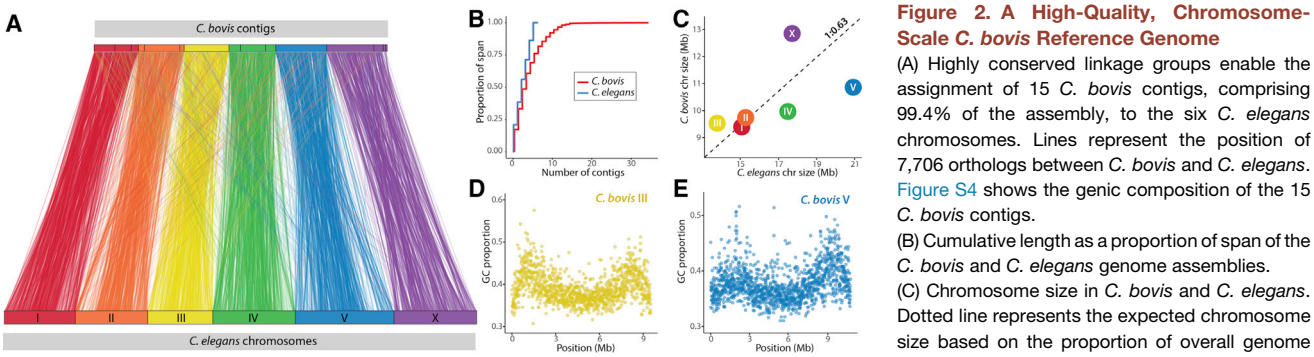

Figure 2. A High-Quality, Chromosome-Scale *C. bovis* Reference Genome

(A) Highly conserved linkage groups enable the assignment of 15 *C. bovis* contigs, comprising 99.4% of the assembly, to the six *C. elegans* chromosomes. Lines represent the position of 7,706 orthologs between *C. bovis* and *C. elegans*. Figure S4 shows the genic composition of the 15 *C. bovis* contigs.

(B) Cumulative length as a proportion of span of the *C. bovis* and *C. elegans* genome assemblies.

(C) Chromosome size in *C. bovis* and *C. elegans*. Dotted line represents the expected chromosome size based on the proportion of overall genome size between *C. elegans* and *C. bovis* (1:0.63).

(D and E) Patterns of variation in GC content (using an 8 kb sliding window) in *C. bovis* contigs 3 (chromosome III; D) and 1 (chromosome V; E), respectively, are consistent with the arms and centers organization present in the chromosomes of other *Caenorhabditis* species.

Chromosomes III and V are represented by single contigs, suggesting that these contigs represent complete *C. bovis* chromosomes. Both contigs also show patterns of variation in GC content characteristic of the arm and center organization present in the chromosomes of other *Caenorhabditis* species [17–19] (Figures 2D and 2E). The remaining chromosomes are each represented by 3–4 contigs (Figures 2A and S4).

The Position of *C. bovis* within *Caenorhabditis*

We sought to reconstruct the phylogenetic relationships of *C. bovis* to other species in the genus *Caenorhabditis*. We clustered over a million protein sequences predicted from the genomes of *C. bovis*, 32 other *Caenorhabditis* species [8, 18–23], and two outgroup taxa, *Diploscapter coronatus* [24] and *Diploscapter pachys* [25], into orthologous groups and selected 1,167 single-copy orthologs. Alignments of these orthologs were concatenated to form a supermatrix that was used to reconstruct the *Caenorhabditis* phylogeny using maximum likelihood. Our phylogenomic analysis resulted in a well-supported phylogeny (Figure 3) that was largely congruent with previously published phylogenies [5, 8, 22]. We recover *C. bovis* as sister to *Caenorhabditis plicata* with maximal support (bootstrap value of 100). The clade containing *C. bovis* and *C. plicata* is early-diverging within the genus *Caenorhabditis*, and the branches separating *C. bovis* and *C. plicata* are long, indicating that *C. bovis* is highly diverged from all other sequenced species, including *C. elegans*.

Comparison between the *C. bovis* and *C. elegans* Genomes

At 62.3 Mb, the *C. bovis* genome is the smallest *Caenorhabditis* genome published to date [8], and is nearly 40 Mb smaller than the *C. elegans* genome. All six *C. bovis* chromosomes are smaller than their *C. elegans* homologs (Figure 2C). Chromosome V is 49% smaller in *C. bovis* and contains fewer than half as many genes (2,302 and 4,992, respectively). Interestingly, the X chromosome is most conserved in size, being only 28% smaller in *C. bovis* and containing 20% fewer genes (2,260 and 2,782, respectively).

The majority of the overall difference in genome size (22.1 Mb or 58%) can be explained by a difference in protein-coding gene content, with the *C. bovis* genome containing 7,080 fewer predicted genes than *C. elegans*. The 13,128 *C. bovis* genes span 35.1 Mb, while the 20,208 *C. elegans* genes span 57.2 Mb, with genic DNA making up a similar proportion of each genome (56% and 57%, respectively). To understand what underlies the difference in gene number, we used the orthology clustering set described previously to compare the number of single-copy genes (those that do not cluster alongside another gene from the same species) and multi-copy genes (those that cluster alongside at least one other gene from the same species) in *C. bovis* and *C. elegans*. We find that the *C. bovis* gene set is substantially less redundant, with only 23% (3,095) of the gene set being classified as multi-copy, while 41% (8,351) of the *C. elegans* gene set is multi-copy. A particularly striking difference is in the number of G protein-coupled receptors (GPCRs), a large family of transmembrane proteins with chemosensory roles in *C. elegans* [26]. The *C. elegans* genome encodes 1,465 GPCRs, while the *C. bovis* genome contains just 326, accounting for 16% of the overall difference in gene number (Table S3). Several other large *C. elegans* gene families are also similarly underrepresented in the *C. bovis* genome, with differences in the number of nuclear hormone receptors (NHRs), major sperm proteins (MSPs), F-box proteins, and C-type lectins accounting for a further 10% of the difference in gene number (Table S3).

In addition to having fewer genes, the *C. bovis* genome contains a smaller proportion of repetitive DNA than the *C. elegans* genome (13% and 16%, respectively; Table S4), explaining a further 8.1 Mb (21%) of the difference in genome size. As is the case for *C. elegans*, repeats are underrepresented on the *C. bovis* X chromosome relative to the rest of the genome (5% versus 16%; Table S4). In *C. elegans*, repeats are distributed non-randomly within the five autosomes, with the chromosome arms being substantially more repeat rich than the centers (28% versus 9%; Figures 4A and 4B; [18]). In contrast, we find a more even distribution of repeats in the two fully assembled *C. bovis* chromosomes (III and V), with the center regions being marginally more repeat rich the arms (16% versus 14%, respectively, assuming the same proportional length of the arm and center domains as *C. elegans*; Figures 4A and 4B).

We compared gene structure in 7,706 genes that were single copy between *C. bovis* and *C. elegans*. Despite the genome being considerably smaller, *C. bovis* genes contain more introns than their *C. elegans* orthologs (8.6 and 6.6 introns per gene, respectively; Figure 4C; Table S5). This is consistent with previous analyses that have found that early-diverging

Table 1. Genome and Gene Set Metrics for *Caenorhabditis bovis* Assembly v1.0

	C. bovis v1	*C. elegans*
Span (Mb)	62.73	100.29
Number of contigs	35	7*
Contig N50 length (Mb)	7.56	17.49
Contig N50 number	4	3
Longest contig (Mb)	10.86	20.92
Repeat content (Mb)	8.19 (13.1%)	16.34 (16.3%)
BUSCO genome—complete (%)/fragmented (%)	94.2/4.6	98.7/0.7
Number of protein-coding genes	13,128	20,208
BUSCO gene set—complete (%)/fragmented (%)	95.2/3.2	98.7/0.9

Assembly and gene set completeness was assessed using BUSCO (version 3.0.2) with the "nematoda_odb9" dataset. *The *C. elegans* genome comprises 6 chromosomes and a 13 kb mitochondrial genome. WormBase ParaSite version WBPS12 of the *C. elegans* genome was used [15]. MinION sequencing statistics are shown in Figure S1 and Table S2. A taxon-annotated GC-coverage plot showing contigs from contaminating organisms that were removed from the *C. bovis* assembly is shown in Figure S2. Kmer spectra of the Illumina short-read data are shown in Figure S3.

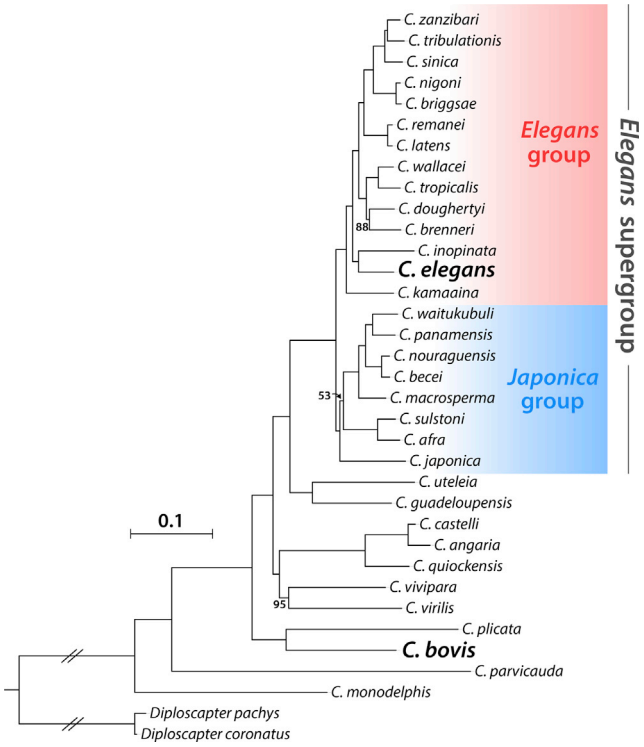

Figure 3. The Phylogenetic Position of *C. bovis* within *Caenorhabditis*
Phylogeny inferred using a supermatrix of 1,167 single-copy orthologs under the general time reversible substitution model with gamma-distributed rate variation among sites (GTR + Γ). *C. bovis* and *C. elegans* are highlighted in bold. The tree is rooted with the two *Diploscapter* species. Branch lengths are in substitutions per site; scale is shown. Bootstraps were 100 unless noted as branch annotations. Major clades as defined by [5] are highlighted.

Caenorhabditis species have retained more ancestral introns than their in-group relatives [22, 28]. However, *C. bovis* introns are, on average, less than half the size of *C. elegans* introns (157 bp and 319 bp, respectively; Table S5). Therefore, despite containing more introns, *C. bovis* genes contain on average less intronic DNA than their *C. elegans* orthologs (1,270 bp and 2,375 bp of intronic DNA per gene, respectively; Figure 4D; Table S5).

Expanded Gene Families in the *C. bovis* Genome

We sought to identify features of the *C. bovis* genome that may relate to its unusual ecology. Using the orthology clustering set described previously, we compared the *C. bovis* gene set to those of 32 other *Caenorhabditis* species. Despite having substantially fewer genes than other *Caenorhabditis* species, we identified several *C. bovis*-specific expansions in gene families that have independently been implicated in parasitism in other nematode species.

P-glycoproteins are members of the ATP-binding cassette (ABC) transporter family and are responsible for the removal of intracellular xenobiotics [29]. P-glycoproteins have been implicated in resistance to antihelminthic drugs in several parasitic nematode species [30–32]. We find evidence for two duplications of the ortholog of the *C. elegans* P-glycoprotein gene *pgp-11* in *C. bovis*, resulting in three distinct copies (Figure S5A). All other *Caenorhabditis* species, except for *C. monodelphis*, possess a single ortholog of *pgp-11* (Figure S5A). *C. elegans* strains that lack *pgp-11* function show increased susceptibility to ivermectin [33], a widely used antihelminthic drug, and genetic variation in the ortholog of *pgp-11* is associated with variation in ivermectin susceptibility in the horse parasite *Parascaris equorum* [34].

Fatty acid and retinol (FAR) proteins are responsible for the uptake and transport of lipids required for nematode metabolism and development [35]. FAR proteins have also been proposed to play a role in modulating host immune responses via the interference of lipid signaling pathways [36]. We find that the ortholog of the *C. elegans* FAR gene *far-8* has undergone two duplications in *C. bovis*, resulting in three distinct copies (Figure S5B). We also find evidence for the expansion of a family of proteins containing Kunitz-type serine protease inhibitor domains in *C. bovis* (Figure S5C). A Kunitz-type serine protease inhibitor secreted by the hookworm *Ancylostoma ceylanicum* has been shown to be capable of inhibiting mammalian host proteases [37]. The majority of species, including *C. elegans*, possess a single member of this family, while *C. bovis* possesses five. In addition, a family of galectin-domain-containing proteins appears to be restricted to *C. bovis*. Galectins are actively secreted by several parasitic nematode species and may interfere with mammalian host immune responses [38–40].

DISCUSSION

Here, we reisolated *C. bovis* from the ear of a female Zebu (*Bos taurus indicus*) in Western Kenya. We sequenced the genome of *C. bovis* using the Oxford Nanopore MinION platform in a nearby field laboratory and used the data to generate a high-quality, chromosome-scale reference genome. We exploited this genome sequence to reconstruct the phylogenetic relationships

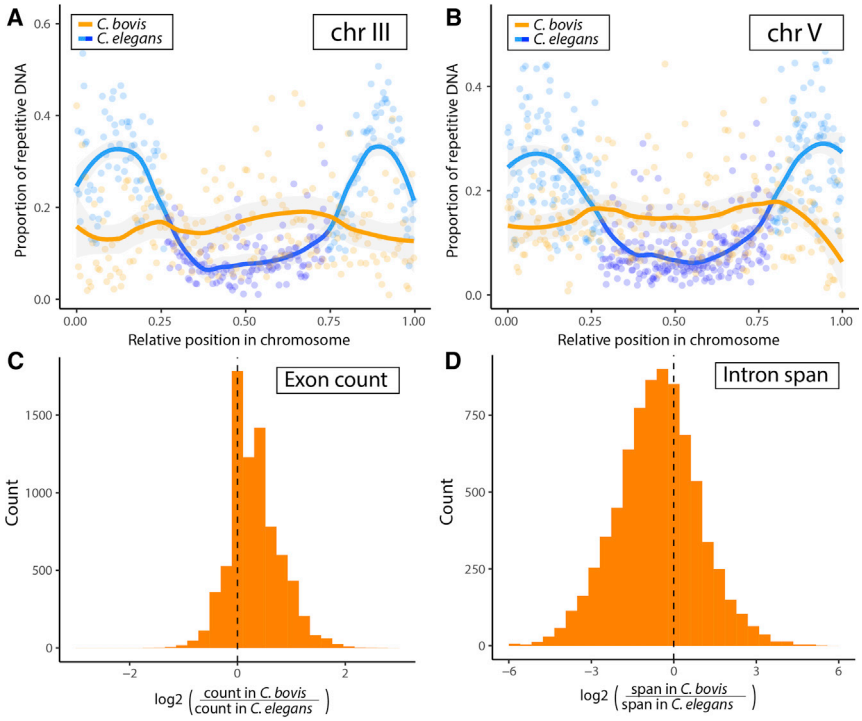

Figure 4. Comparison between the *C. bovis* and *C. elegans* Genomes

(A and B) Repeat densities across chromosome III (A) and V (B) in 50 kb windows in *C. bovis* and *C. elegans*. Lines represent loess smoothing functions fitted to the data for each species. Points and lines for *C. elegans* are colored by arms and centers domains (dark blue: centers, light blue: arms) as defined by [27]. Repeat content statistics for each chromosome are shown in Table S4.

(C and D) Histograms of the log2-transformed ratio of exon count (C) and intron span (D) in 7,706 genes in *C. bovis* compared to their orthologs in *C. elegans*. Untranslated regions (UTRs) are not annotated in *C. bovis* and so only coding exons and the intervening introns were considered in both species. Gene structure statistics are shown in Table S5. Counts of large gene families in *C. elegans* and *C. bovis* are shown in Table S3 and gene trees of expanded gene families in *C. bovis* are shown in Figure S5.

of *C. bovis* to other *Caenorhabditis* species, and identified expansions in gene families that may be associated with the unusual lifestyle of *C. bovis*.

The low level of heterozygosity in the *C. bovis* genome is surprising. Genomes of outcrossing *Caenorhabditis* species typically contain extremely high levels of heterozygosity [41, 42], which can complicate genome assembly [16]. To circumvent these issues, *Caenorhabditis* species are often deliberately inbred over multiple generations (e.g., by sibling mating) prior to sequencing [8]. While it is likely that our *C. bovis* cultures underwent some population bottlenecking during isolation and the subsequent two-week period of laboratory culture, if *C. bovis* has similar levels of heterozygosity to other outcrossing *Caenorhabditis* species, this alone is not sufficient to explain the low levels of heterozygosity we observe. Instead, it seems that the *C. bovis* population we sampled from is naturally highly inbred, suggesting that a very small number of nematodes are transported between hosts and that gene flow between demes is extremely rare. Resequencing other isolates would allow us to test if this is true of all *C. bovis* populations.

The placement of *C. bovis* as sister to *C. plicata* is intriguing. *C. plicata* has been isolated from carrion (once from a dead elephant in Kenya and once from a dead pine marten in Germany) and has a phoretic association with carrion beetles [10, 43, 44]. *C. plicata* is therefore the only *Caenorhabditis* species currently in culture that has been found in association with a vertebrate, with all others having been isolated from rotting plant or fungal matter [5–8]. In recent years, worldwide sampling has led to the discovery of many new *Caenorhabditis* species, but all efforts have been focused in habitats that resemble the decaying vegetable matter habitat identified as the home of *C. elegans* [5–7, 45, 46]. While there are anecdotal instances of other *Caenorhabditis* species being associated with vertebrates, including birds [47], dogs [48], and humans [49], no directed searches focusing on living or dead animal niches have been reported. It is therefore possible that there exists a largely undiscovered clade of vertebrate-associated *Caenorhabditis* species.

While the genome and the gene set of *C. bovis* are smaller than that of *C. elegans* and many other *Caenorhabditis* species, we identified several gene families that appear to have undergone expansion in *C. bovis*. Functional annotation of these expanded gene families revealed that several have been independently implicated in parasitism in other nematode species. While P-glycoproteins (and orthologs of *pgp-11* specifically) are associated with resistance to antihelminthic drugs such as ivermectin, what role the expansion of *pgp-11* plays in the biology of *C. bovis* remains unclear. Ivermectin has been found to be effective at killing *C. bovis* and in the treatment of bovine parasitic otitis in cattle in Tanzania [50]. In contrast, similar studies have found ivermectin and albendazole to be an ineffective treatment for bovine parasitic otitis in cattle in Brazil [51, 52]. Aside from P-glycoproteins, FAR proteins, galectins, and serpins are known to be actively secreted by several parasitic nematode species, and an immunomodulatory role has been proposed. It would be fascinating to explore the roles of these families (and many others) in the possible parasitic lifestyle of *C. bovis*. We note also that *C. bovis* appears to be adapted to life at 37°C in its bovine niche. This temperature is rapidly lethal to *C. elegans* [53, 54], and thus *C. bovis* must have adapted to be heat resistant.

We do not yet know enough about the biology of *C. bovis*, from either its genome or its limited biological literature, to classify it as a "true" parasite; *C. bovis* might instead be an opportunistic colonizer of niches created by other pathogens. Several other *Caenorhabditis* species associate with arthropods as phoretic hosts, and such phoresy is thought to serve as the major means of transport between scattered food patches [10]. *C. bovis* is

believed to be transported by dipterans [13], themselves associated with parasitism of bovine ears, and may have exploited the biology of these phoretic hosts to colonize a new niche, the bovine ear. Bacterial coinfection may be a prerequisite of colonization by *C. bovis*, may be exacerbated or encouraged by the presence of *C. bovis*, or may be initiated by *C. bovis* itself. Other Rhabditine species offer models for this last possibility: entomopathogenic species in the genus *Heterorhabditis* carry specific bacteria that play roles in killing their arthropod larval prey [55], and molluscicidal nematodes in the genus *Phasmarhabditis* induce bacterial sepsis in slugs and snail prey [56].

While the high-quality draft genome and the analyses presented here represent a major step forward in our understanding of this unusual and understudied *Caenorhabditis* species, it is only a beginning. Our ultimate goal is to establish long-term cultures and to apply the exquisite reverse genetic toolkits available for *Caenorhabditis* to understand the biology of this species. We would like the isolates to be available to any researcher via the *Caenorhabditis* Genetics Center (CGC), and we are currently seeking the appropriate permits for export from Kenya. We hope that these cultures, combined with the draft-genome sequence, will enable the interrogation of the biology of *C. bovis*, including the use of CRISPR-Cas9 technology to edit or disrupt loci that might be relevant for its unusual lifestyle. It is important to note, however, that we still know very little about *C. bovis in situ*, with details of its present-day prevalence, role in bovine parasitic otitis, and microbial associates remaining scarce. Therefore, any laboratory interrogation must happen alongside further study of *C. bovis* in Eastern Africa, in collaboration with local institutes and scientists.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- LEAD CONTACT AND MATERIALS AVAILABILITY
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Ethics Statement
  - Sampling, nematode isolation and culture
- METHOD DETAILS
  - DNA extraction
  - Oxford Nanopore MinION sequencing
  - Illumina MiSeq sequencing
  - Genome assembly
  - Gene prediction
  - Estimation of heterozygosity
  - Assignment of *C. bovis* contigs to chromosomes
  - Orthology inference and phylogenomics
  - Gene content and structure analyses
  - Repeat content analyses
- QUANTIFICATION AND STATISTICAL ANALYSIS
- DATA AND CODE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.cub.2020.01.074.

ACKNOWLEDGMENTS

We thank the University of Edinburgh's Davis Fund for the generous support of this project. We thank members of ILRI ZooLink team for their assistance and helpful discussions. We thank Karin Kiontke and David Fitch for assistance with identification of isolated nematodes. We also thank the three anonymous reviewers for their helpful comments on an earlier version of this manuscript. This work was supported by the Biotechnology and Biological Sciences Research Council, the Department for International Development, the Economic & Social Research Council, the Medical Research Council, the Natural Environment Research Council, and the Defense Science & Technology Laboratory, under the Zoonoses and Emerging Livestock Systems (ZELS) program, grant reference BB/L019019/1. It also received support from the CGIAR Research Program on Agriculture for Nutrition and Health (A4NH), led by the International Food Policy Research Institute (IFPRI). We also acknowledge the CGIAR Fund Donors (https://www.cgiar.org/funders/). L.S. is funded by a Baillie Gifford Studentship. S.R. is funded by MRC Grant Ref. MR/N013166/1.

AUTHOR CONTRIBUTIONS

L.S., M.B., and E.M.F. conceptualized the study. L.S., L.C.F., M.K.M., K.M., C.O.O., A.O., and J.O. performed field sampling. L.S. performed nematode isolation. L.S., S.R., and S.M.N. performed nematode culture. L.S. and S.R. performed DNA extraction and MinION sequencing. E.M.M. performed MiSeq sequencing. L.S. performed genome assembly, phylogenetic analysis, and comparative genomic analysis. L.S. and M.B. wrote the manuscript. All authors commented on and approved the final version of the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: October 23, 2019
Revised: December 10, 2019
Accepted: January 23, 2020
Published: February 27, 2020

REFERENCES

1. Blaxter, M., and Koutsovoulos, G. (2015). The evolution of parasitism in Nematoda. Parasitology *142* (*Suppl 1*), S26–S39.
2. Blaxter, M.L., De Ley, P., Garey, J.R., Liu, L.X., Scheldeman, P., Vierstraete, A., Vanfleteren, J.R., Mackey, L.Y., Dorris, M., Frisse, L.M., et al. (1998). A molecular evolutionary framework for the phylum Nematoda. Nature *392*, 71–75.
3. Bürglin, T.R., Lobos, E., and Blaxter, M.L. (1998). Caenorhabditis elegans as a model for parasitic nematodes. Int. J. Parasitol. *28*, 395–411.
4. Gilleard, J.S. (2004). The use of Caenorhabditis elegans in parasitic nematode research. Parasitology *128* (*Suppl 1*), S49–S70.
5. Kiontke, K.C., Félix, M.-A., Ailion, M., Rockman, M.V., Braendle, C., Pénigault, J.-B., and Fitch, D.H.A. (2011). A phylogeny and molecular barcodes for Caenorhabditis, with numerous new species from rotting fruits. BMC Evol. Biol. *11*, 339.
6. Félix, M.-A., Braendle, C., and Cutter, A.D. (2014). A streamlined system for species diagnosis in Caenorhabditis (Nematoda: Rhabditidae) with name designations for 15 distinct biological species. PLoS ONE *9*, e94723.
7. Ferrari, C., Salle, R., Callemeyn-Torre, N., Jovelin, R., Cutter, A.D., and Braendle, C. (2017). Ephemeral-habitat colonization and neotropical species richness of Caenorhabditis nematodes. BMC Ecol. *17*, 43.
8. Stevens, L., Félix, M.-A., Beltran, T., Braendle, C., Caurcel, C., Fausett, S., Fitch, D., Frézal, L., Gosse, C., Kaur, T., et al. (2019). Comparative genomics of 10 new *Caenorhabditis* species. Evol Lett *3*, 217–236.

9. Kreis, H.A. (1964). Ein neuer Nematode aus dem äusseren Gehörgang von Zeburindern in Ostafrika, Rhabditis bovis n. sp.(Rhabditidoidea; Rhabditidae). Schweiz. Arch. Tierh. *106*, 372–378.
10. Kiontke, K., and Sudhaus, W. (2006). Ecology of Caenorhabditis species. WormBook, 1–14.
11. Cardona, J., González, M., and Álvarez, J. (2010). Otitis bovina por Rhabditis bovis en Córdoba, Colombia. Reporte de dos casos. Rev. Mvz Cordoba *15*, https://doi.org/10.21897/rmvz.311.
12. Msolla, P., Semuguruka, W.D., Kasuku, A.A., and Shoo, M.K. (1993). Clinical observations on bovine parasitic otitis in Tanzania. Trop. Anim. Health Prod. *25*, 15–18.
13. Msolla, P., Kimera, I.S., Kassuku, A.A., and Semuguruka, W.D. (1989). The role of flies, manure and soil in the epidemiology of bovine parasitic otitis. Proc. 7th Tanzania Vet. Assoc. Sci. Conf. *7*, 101–109.
14. Sudhaus, W., and Kiontke, K. (1996). Phylogeny of Rhabditis subgenus Caenorhabditis (Rhabditidae, Nematoda)*. J. Zoological Syst. Evol. Res. *34*, 217–233.
15. Howe, K.L., Bolt, B.J., Shafie, M., Kersey, P., and Berriman, M. (2017). WormBase ParaSite - a comprehensive resource for helminth genomics. Mol. Biochem. Parasitol. *215*, 2–10.
16. Barrière, A., Yang, S.-P., Pekarek, E., Thomas, C.G., Haag, E.S., and Ruvinsky, I. (2009). Detecting heterozygosity in shotgun genome assemblies: Lessons from obligately outcrossing nematodes. Genome Res. *19*, 470–480.
17. Hillier, L.W., Miller, R.D., Baird, S.E., Chinwalla, A., Fulton, L.A., Koboldt, D.C., and Waterston, R.H. (2007). Comparison of C. elegans and C. briggsae genome sequences reveals extensive conservation of chromosome organization and synteny. PLoS Biol. *5*, e167.
18. C. elegans Sequencing Consortium. (1998). Genome sequence of the nematode C. elegans: a platform for investigating biology. Science *282*, 2012–2018.
19. Yin, D., Schwarz, E.M., Thomas, C.G., Felde, R.L., Korf, I.F., Cutter, A.D., Schartner, C.M., Ralston, E.J., Meyer, B.J., and Haag, E.S. (2018). Rapid genome shrinkage in a self-fertile nematode reveals sperm competition proteins. Science *359*, 55–61.
20. Stein, L.D., Bao, Z., Blasiar, D., Blumenthal, T., Brent, M.R., Chen, N., Chinwalla, A., Clarke, L., Clee, C., Coghlan, A., et al. (2003). The genome sequence of Caenorhabditis briggsae: a platform for comparative genomics. PLoS Biol. *1*, E45.
21. Mortazavi, A., Schwarz, E.M., Williams, B., Schaeffer, L., Antoshechkin, I., Wold, B.J., and Sternberg, P.W. (2010). Scaffolding a Caenorhabditis nematode genome with RNA-seq. Genome Res. *20*, 1740–1747.
22. Slos, D., Sudhaus, W., Stevens, L., Bert, W., and Blaxter, M. (2017). Caenorhabditis monodelphis sp. n.: defining the stem morphology and genomics of the genus Caenorhabditis. BMC Zoology *2*, 4.
23. Kanzaki, N., Tsai, I.J., Tanaka, R., Hunt, V.L., Liu, D., Tsuyama, K., Maeda, Y., Namai, S., Kumagai, R., Tracey, A., et al. (2018). Biology and genome of a newly discovered sibling species of Caenorhabditis elegans. Nat. Commun. *9*, 3216.
24. Hiraki, H., Kagoshima, H., Kraus, C., Schiffer, P.H., Ueta, Y., Kroiher, M., Schierenberg, E., and Kohara, Y. (2017). Genome analysis of Diploscapter coronatus: insights into molecular peculiarities of a nematode with parthenogenetic reproduction. BMC Genomics *18*, 478.
25. Fradin, H., Kiontke, K., Zegar, C., Gutwein, M., Lucas, J., Kovtun, M., Corcoran, D.L., Baugh, L.R., Fitch, D.H.A., Piano, F., and Gunsalus, K.C. (2017). Genome architecture and evolution of a unichromosomal asexual nematode. Curr. Biol. *27*, 2928–2939.e6.
26. Robertson, H.M. (1998). Two large families of chemoreceptor genes in the nematodes Caenorhabditis elegans and Caenorhabditis briggsae reveal extensive gene duplication, diversification, movement, and intron loss. Genome Res. *8*, 449–463.
27. Rockman, M.V., and Kruglyak, L. (2009). Recombinational landscape and population genomics of Caenorhabditis elegans. PLoS Genet. *5*, e1000419.
28. Kiontke, K., Gavin, N.P., Raynes, Y., Roehrig, C., Piano, F., and Fitch, D.H.A. (2004). Caenorhabditis phylogeny predicts convergence of hermaphroditism and extensive intron loss. Proc. Natl. Acad. Sci. USA *101*, 9003–9008.
29. Sheps, J.A., Ralph, S., Zhao, Z., Baillie, D.L., and Ling, V. (2004). The ABC transporter gene family of Caenorhabditis elegans has implications for the evolutionary dynamics of multidrug resistance in eukaryotes. Genome Biol. *5*, R15.
30. Xu, M., Molento, M., Blackhall, W., Ribeiro, P., Beech, R., and Prichard, R. (1998). Ivermectin resistance in nematodes may be caused by alteration of P-glycoprotein homolog. Mol. Biochem. Parasitol. *91*, 327–335.
31. Bourguinat, C., Ardelli, B.F., Pion, S.D.S., Kamgno, J., Gardon, J., Duke, B.O.L., Boussinesq, M., and Prichard, R.K. (2008). P-glycoprotein-like protein, a possible genetic marker for ivermectin resistance selection in Onchocerca volvulus. Mol. Biochem. Parasitol. *158*, 101–111.
32. Bartley, D.J., McAllister, H., Bartley, Y., Dupuy, J., Ménez, C., Alvinerie, M., Jackson, F., and Lespine, A. (2009). P-glycoprotein interfering agents potentiate ivermectin susceptibility in ivermectin sensitive and resistant isolates of Teladorsagia circumcincta and Haemonchus contortus. Parasitology *136*, 1081–1088.
33. Janssen, I.J.I., Krücken, J., Demeler, J., and von Samson-Himmelstjerna, G. (2013). Caenorhabditis elegans: modest increase of susceptibility to ivermectin in individual P-glycoprotein loss-of-function strains. Exp. Parasitol. *134*, 171–177.
34. Janssen, I.J.I., Krücken, J., Demeler, J., Basiaga, M., Kornaś, S., and von Samson-Himmelstjerna, G. (2013). Genetic variants and increased expression of Parascaris equorum P-glycoprotein-11 in populations with decreased ivermectin susceptibility. PLoS ONE *8*, e61635.
35. Garofalo, A., Rowlinson, M.-C., Amambua, N.A., Hughes, J.M., Kelly, S.M., Price, N.C., Cooper, A., Watson, D.G., Kennedy, M.W., and Bradley, J.E. (2003). The FAR protein family of the nematode Caenorhabditis elegans. Differential lipid binding properties, structural characteristics, and developmental regulation. J. Biol. Chem. *278*, 8065–8074.
36. Bradley, J.E., Nirmalan, N., Kläger, S.L., Faulkner, H., and Kennedy, M.W. (2001). River blindness: a role for parasite retinoid-binding proteins in the generation of pathology? Trends Parasitol. *17*, 471–475.
37. Milstone, A.M., Harrison, L.M., Bungiro, R.D., Kuzmic, P., and Cappello, M. (2000). A broad spectrum Kunitz type serine protease inhibitor secreted by the hookworm Ancylostoma ceylanicum. J. Biol. Chem. *275*, 29391–29399.
38. Turner, D.G., Wildblood, L.A., Inglis, N.F., and Jones, D.G. (2008). Characterization of a galectin-like activity from the parasitic nematode, Haemonchus contortus, which modulates ovine eosinophil migration in vitro. Vet. Immunol. Immunopathol. *122*, 138–145.
39. Kim, J.-Y., Cho, M.K., Choi, S.H., Lee, K.H., Ahn, S.C., Kim, D.-H., and Yu, H.S. (2010). Inhibition of dextran sulfate sodium (DSS)-induced intestinal inflammation via enhanced IL-10 and TGF-β production by galectin-9 homologues isolated from intestinal parasites. Mol. Biochem. Parasitol. *174*, 53–61.
40. Wang, W., Wang, S., Zhang, H., Yuan, C., Yan, R., Song, X., Xu, L., and Li, X. (2014). Galectin Hco-gal-m from Haemonchus contortus modulates goat monocytes and T cell function in different patterns. Parasit. Vectors *7*, 342.
41. Cutter, A.D., Baird, S.E., and Charlesworth, D. (2006). High nucleotide polymorphism and rapid decay of linkage disequilibrium in wild populations of Caenorhabditis remanei. Genetics *174*, 901–913.
42. Dey, A., Chan, C.K.W., Thomas, C.G., and Cutter, A.D. (2013). Molecular hyperdiversity defines populations of the nematode Caenorhabditis brenneri. Proc. Natl. Acad. Sci. USA *110*, 11056–11060.
43. Volk, J., and Others. (1950). Nematodes associated with earthworms and carrion beetles. Zool. Jahrb., Abt. Syst. Okol. Geogr. Tiere *79*, 1–70.
44. Sudhaus, W. (1974). Zur Systematik, Verbreitung, Ökologie und Biologie neuer und wenig bekannter Rhabditiden (Nematoda).

45. Félix, M.-A., and Braendle, C. (2010). The natural history of Caenorhabditis elegans. Curr. Biol. *20*, R965–R969.
46. Félix, M.-A., and Duveau, F. (2012). Population dynamics and habitat sharing of natural populations of Caenorhabditis elegans and C. briggsae. BMC Biol. *10*, 59.
47. Schmidt, G., and Kuntz, R.E. (1972). Caenorhabditis avicola sp. n.(Rhabditidae) found in a bird from Taiwan. Proc. Helminthol. Soc. Wash. *39*, 189–191.
48. Kreis, H.A., and Faust, E.C. (1933). Two new species of Rhabditis (Rhabditis macrocerca and R. clavopapillata) associated with dogs and monkeys in experimental Strongyloides studies. Trans. Am. Microsc. Soc. *52*, 162–172.
49. Scheiber, S.H. (1880). Ein Fall von mikroskopisch kleinen Rundwürmern—Rhabditis genitalis—im Urin einer Kranken. Virchows Arch. *82*, 161–175.
50. Msolla, P., Falmer-Hansen, J., Musemakweli, J., and Monrad, J. (1985). Treatment of bovine parasitic otitis using ivermectin. Trop. Anim. Health Prod. *17*, 166–168.
51. Verocai, G.G., Fernandes, J.I., Correia, T.R., Melo, R.M., Alves, P.A.M., Scott, F.B., and Grisi, L. (2009). Inefficacy of albendazole sulphoxide and ivermectin for the treatment of bovine parasitic otitis caused by rhabditiform nematodes. Pesqui. Vet. Bras. *29*, 910–912.
52. Ferraz, C.M., Sobral, S.A., Senna, C.C., Junior, O.F., Moreira, T.F., Tobias, F.L., de Freitas Soares, F.E., Geniêr, H.L.A., Vilela, V.L.R., Lima, J.A.C., et al. (2019). Combined use of ivermectin, dimethyl sulfoxide, mineral oil and nematophagous fungi to control Rhabditis spp. Vet. Parasitol. *275*, 108924.
53. Snutch, T.P., and Baillie, D.L. (1983). Alterations in the pattern of gene expression following heat shock in the nematode Caenorhabditis elegans. Can. J. Biochem. Cell Biol. *61*, 480–487.
54. Jones, D., and Candido, E.P. (1999). Feeding is inhibited by sublethal concentrations of toxicants and by heat stress in the nematode Caenorhabditis elegans: relationship to the cellular stress response. J. Exp. Zool. *284*, 147–157.
55. Forst, S., Dowds, B., Boemare, N., and Stackebrandt, E. (1997). Xenorhabdus and Photorhabdus spp.: bugs that kill bugs. Annu. Rev. Microbiol. *51*, 47–72.
56. Tan, L., and Grewal, P.S. (2002). Endotoxin activity of Moraxella osloensis against the grey garden slug, Deroceras reticulatum. Appl. Environ. Microbiol. *68*, 3943–3947.
57. Ruan, J., and Li, H. (2019). Fast and accurate long-read assembly with wtdbg2. Nat. Methods *17*, 155–158.
58. Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L. (2009). BLAST+: architecture and applications. BMC Bioinformatics *10*, 421.
59. Pundir, S., Martin, M.J., and O'Donovan, C. (2017). UniProt Protein Knowledgebase. Methods Mol. Biol. *1558*, 41–55.
60. Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics *34*, 3094–3100.
61. Laetsch, D.R., and Blaxter, M.L. (2017). BlobTools: interrogation of genome assemblies. F1000Res. *6*, 1287.
62. Vaser, R., Sović, I., Nagarajan, N., and Šikić, M. (2017). Fast and accurate de novo genome assembly from long uncorrected reads. Genome Res. *27*, 737–746.
63. Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv, arXiv:1303.3997. http://arxiv.org/abs/1303.3997.
64. Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A., Zeng, Q., Wortman, J., Young, S.K., and Earl, A.M. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS ONE *9*, e112963.
65. Smit, A., and Hubley, R. (2010). RepeatModeler Open-1.0 (ISB). http://www.repeatmasker.org/RepeatModeler/.
66. Smit, A.F.A., Hubley, R., and Green, P. (1996). RepeatMasker (ISB). http://www.repeatmasker.org/.
67. Hoff, K.J., Lange, S., Lomsadze, A., Borodovsky, M., and Stanke, M. (2016). BRAKER1: unsupervised RNA-seq-based genome annotation with GeneMark-ET and AUGUSTUS. Bioinformatics *32*, 767–769.
68. Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., and Zdobnov, E.M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics *31*, 3210–3212.
69. Marçais, G., and Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics *27*, 764–770.
70. Vurture, G.W., Sedlazeck, F.J., Nattestad, M., Underwood, C.J., Fang, H., Gurtowski, J., and Schatz, M.C. (2017). GenomeScope: fast reference-free genome profiling from short reads. Bioinformatics *33*, 2202–2204.
71. Broad Institute (2019). Picard (Broad Institute). https://broadinstitute.github.io/picard/.
72. Garrison, E., and Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. arXiv, arXiv:1207.3907. http://arxiv.org/abs/1207.3907.
73. Danecek, P., Schiffels, S., and Durbin, R. (2014). Multiallelic calling model in bcftools (-m). https://samtools.github.io/bcftools/call-m.pdf.
74. Emms, D.M., and Kelly, S. (2015). OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol. *16*, 157.
75. Nguyen, L.-T., Schmidt, H.A., von Haeseler, A., and Minh, B.Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. *32*, 268–274.
76. Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. *30*, 772–780.
77. Kocot, K.M., Citarella, M.R., Moroz, L.L., and Halanych, K.M. (2013). PhyloTreePruner: a phylogenetic tree-based approach for selection of orthologous sequences for phylogenomics. Evol. Bioinform. Online *9*, 429–435.
78. Laetsch, D.R., and Blaxter, M.L. (2017). KinFin: software for taxon-aware analysis of clustered protein sequences. G3 (Bethesda) *7*, 3349–3357.
79. Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., et al. (2014). InterProScan 5: genome-scale protein function classification. Bioinformatics *30*, 1236–1240.
80. Haas, B. (2007). TransposonPSI: an application of PSI-Blast to mine (retro-) transposon ORF homologies (Broad Institute).
81. Ellinghaus, D., Kurtz, S., and Willhoeft, U. (2008). LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. BMC Bioinformatics *9*, 18.
82. Rognes, T., Flouri, T., Nichols, B., Quince, C., and Mahé, F. (2016). VSEARCH: a versatile open source tool for metagenomics. PeerJ *4*, e2584.
83. Buchfink, B., Xie, C., and Huson, D.H. (2015). Fast and sensitive protein alignment using DIAMOND. Nat. Methods *12*, 59–60.
84. Huerta-Cepas, J., Szklarczyk, D., Forslund, K., Cook, H., Heller, D., Walter, M.C., Rattei, T., Mende, D.R., Sunagawa, S., Kuhn, M., et al. (2016). eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. Nucleic Acids Res. *44* (D1), D286–D293.
85. Hoang, D.T., Chernomor, O., von Haeseler, A., Minh, B.Q., and Vinh, L.S. (2018). UFBoot2: improving the ultrafast bootstrap approximation. Mol. Biol. Evol. *35*, 518–522.
86. Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., and Jermiin, L.S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. Nat. Methods *14*, 587–589.
87. Capella-Gutiérrez, S., Silla-Martínez, J.M., and Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics *25*, 1972–1973.

88. Letunic, I., and Bork, P. (2016). Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res. *44* (W1), W242-5.

89. Bateman, A., Coin, L., Durbin, R., Finn, R.D., Hollich, V., Griffiths-Jones, S., Khanna, A., Marshall, M., Moxon, S., Sonnhammer, E.L.L., et al. (2004). The Pfam protein families database. Nucleic Acids Res. *32*, D138–D141.

90. Schwarz, E.M. (2005). Genomic classification of protein-coding gene families. WormBook, 1–23.

91. Berriman, M., Coghlan, A., and Tsai, I.J. (2018). Creation of a comprehensive repeat library for a newly sequenced parasitic worm genome. Protoc. Ex. https://doi.org/10.1038/protex.2018.054.

92. Llorens, C., Futami, R., Covelli, L., Domínguez-Escribá, L., Viu, J.M., Tamarit, D., Aguilar-Rodríguez, J., Vicente-Ripolles, M., Fuster, G., Bernet, G.P., et al. (2011). The Gypsy Database (GyDB) of mobile genetic elements: release 2.0. Nucleic Acids Res. *39*, D70–D74.

93. Steinbiss, S., Willhoeft, U., Gremme, G., and Kurtz, S. (2009). Fine-grained annotation and classification of de novo predicted LTR retrotransposons. Nucleic Acids Res. *37*, 7002–7013.

94. R Core Team. (2018). R: A language and environment for statistical computing (R Foundation for Statistical Computing).

95. Wickham, H. (2009). Ggplot2: Elegant Graphics for Data Analysis, Second Edition (Springer Publishing Company).

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, Peptides, and Recombinant Proteins		
Cholesterol	Sigma	CAS#57-88-5
Potassium phosphate, monobasic	Fisher	CAS#7778-77-0
diPotassium hydrogen phosphate trihydrate	Millipore	CAS#16788-57-1
Bactotrypone	Millipore	CAS#91079-40-2
OmniPur Sodium Chloride	Millipore	CAS#7647-14-5
Columbia Agar	Millipore	Cat#27688-500G
Cell Lysis Solution	QIAGEN	Cat#158906
RNase Cocktail Enzyme Mix	Invitrogen	Cat#AM2286
Protein Precipitation Solution	QIAGEN	Cat#158910
Qubit BR/HS assay	ThermoScientific	Cat#Q32854/ Q32850
AMPure XP beads	Beckman Coulter	Cat#A63880
Ligation Sequencing Kit	Oxford Nanopore	Kit#SQK-LSK109
Ultra II Ligation Master Mix	New England Biolabs	Cat#E7595S
300 bp MiSeq reagent kit v3	Illumina	Cat#MS-102-3003
Deposited Data		
Raw sequence data	European Nucleotide Archive (ENA)	ENA: PRJEB34497
Genome assembly and annotation files	European Nucleotide Archive (ENA)	ENA: PRJEB34497
Software and Algorithms		
Guppy	N/A	https://community.nanoporetech.com
wtdbg2	[57]	https://github.com/ruanjue/wtdbg2
NCBI-BLAST+	[58]	ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/LATEST/
Diamond	[59]	https://github.com/bbuchfink/diamond
minimap2	[60]	https://github.com/lh3/minimap2
blobtools	[61]	https://blobtools.readme.io/docs
Racon	[62]	https://github.com/isovic/racon
Medaka	N/A	https://github.com/nanoporetech/medaka
BWA-MEM	[63]	https://github.com/lh3/bwa
Pilon	[64]	https://github.com/broadinstitute/pilon
RepeatModeler	[65]	http://www.repeatmasker.org/RepeatModeler/
RepeatMasker	[66]	http://www.repeatmasker.org/RMDownload.html
BRAKER	[67]	https://github.com/Gaius-Augustus/BRAKER
BUSCO	[68]	https://busco.ezlab.org/
JellyFish	[69]	https://www.cbcb.umd.edu/software/jellyfish/
GenomeScope	[70]	http://qb.cshl.edu/genomescope/
PicardTools	[71]	https://broadinstitute.github.io/picard/
freebayes	[72]	https://github.com/ekg/freebayes
bcftools	[73]	http://samtools.github.io/bcftools/bcftools.html
OrthoFinder	[74]	https://github.com/davidemms/OrthoFinder
IQ-TREE	[75]	http://www.iqtree.org/
MAFFT	[76]	https://mafft.cbrc.jp/alignment/software/
PhyloTreePruner	[77]	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3825643/

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
KinFin	[78]	https://github.com/DRL/kinfin
InterProScan	[79]	https://www.ebi.ac.uk/interpro/search/
TransposonPSI	[80]	http://transposonpsi.sourceforge.net/
GenomeTools	[81]	http://genometools.org/
RepeatClassifier	[66]	http://www.repeatmasker.org/RMDownload.html
VSEARCH	[82]	https://github.com/torognes/vsearch

LEAD CONTACT AND MATERIALS AVAILABILITY

There are restrictions to the availability of our cryopreserved *C. bovis* cultures as we do not yet have a material transfer agreement (MTA) in place to export these animals from Kenya to the UK. Further information about our cultures should be directed to the Lead Contact, Lewis Stevens (lewis.stevens@ed.ac.uk).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Ethics Statement

This study was approved by the Institutional Research Ethics Committee (IREC Reference No. 2017-08) and the Institutional Animal Care and Use Committee (IACUC Reference No. 2017-04 and 2017-04.1) at the International Livestock Research Institute, review bodies approved by the Kenyan National Commission for Science, Technology and Innovation. Approval to conduct the work was also obtained from the Department of Veterinary Services and the relevant offices of these Ministries at the county government level. All recruited animal owners gave written, informed consent prior to their inclusion in the study.

Sampling, nematode isolation and culture

Sampling was carried out as part of an existing surveillance program of zoonotic disease in humans at hospitals and livestock animals at livestock markets and slaughterhouses in three counties of Western Kenya (Figure 1; Table S1). A total of 44 cattle, including a range of local breeds and ages, were sampled. We restrained each animal manually and washed the external auditory canal using cotton wool soaked in physiological saline. Cotton wool samples were stored in 50 mL tubes and transported to the laboratory in a refrigerated box. We inspected 1-2 mL of saline from each sample under a dissecting microscope within 4 h of collection. Nematodes were isolated from the saline using a pipette and placed onto nematode growth medium (NGM) (1 g NaCl, 2 g Bactotryptone, 1.5 g KH2PO4, 0.25 g K2HPO4, 4mg cholesterol, 10 g agar, 500 mL deionized water) or blood agar (50 mL horse blood, 41 g Columbia blood agar base, 1L of deionized water) plates seeded with an environmentally-sourced *E. coli* strain. Plates were incubated at 37°C. The morphology of adult nematodes was examined using a standard compound microscope and compared to the previous morphological descriptions of *C. bovis* [9, 14].

METHOD DETAILS

DNA extraction

We harvested nematodes by washing each plate with phosphate-buffered saline (PBS) supplemented with 0.01% Tween20. The nematodes were washed three times with clean PBS and subsequently centrifuged to form a pellet. Pellets were stored at −40°C until extraction. We added 600 μL of Cell Lysis Solution (QIAGEN) and 20 μL of proteinase K (20 μg/μL) to each frozen pellet and incubated for four h at 56°C. 5 μL of RNase Cocktail Enzyme Mix (Invitrogen) was subsequently added and incubated at 37°C for one h. We added 200 μl of Protein Precipitation Solution (QIAGEN) and centrifuged at 15,000 rpm for 3 min. The supernatant was collected in a new tube and 600 μL of isopropanol added to precipitate the DNA. We centrifuged each tube at 15,000 rpm for 3 min and discarded the supernatant. The resulting DNA pellets were washed twice with 70% ethanol and briefly allowed to dry before being resuspended in 100 μL of elution buffer (10 mM Tris-Cl). DNA concentration was assessed using Qubit (Thermo Scientific).

Oxford Nanopore MinION sequencing

We sheared the DNA prior to sequencing by passing approximately 2 μg in a volume of 100 μl through either 26G or 29G insulin needle 5-10 times. Small fragments were removed by purifying DNA with 0.5x concentration Agencourt AMPure XP beads. We followed the "one-pot" ligation protocol for preparing Oxford Nanopore SQK-LSK108 libraries (https://www.protocols.io/view/one-pot-ligation-protocol-for-oxford-nanopore-libr-k9acz2e) but with the following modifications: we added 5 μl of SQK-LSK109 adaptor mix (AMX) instead of 20 μl of SQK-LSK108 AMX; we added 20 μl of NEB Ultra II Ligation Master Mix instead of 40 μl; we replaced the SQK-LSK108 adaptor binding beads (ABB) with either the SQK-LSK109 long fragment buffer (LFB) or short fragment

buffer (SFB). Thereafter, we followed the standard manufacturer's instructions for preparing and loading SQK-LSK109 libraries. Libraries were loaded on to two R9.4 flow cell and run for ~48 h using MinKNOW version 18.12.9. Raw data metrics are presented in Table S2.

Illumina MiSeq sequencing

We prepared one Nextera DNA Flex library as per manufacturer's instructions using ~100 ng of input DNA. The library fragment size was assessed using the Agilent TapeStation and library concentration was determined using Qubit dsDNA HS Assay Kit (Thermo Scientific, USA). The library was then sequenced using the Illumina MiSeq platform with a paired-end 300 bp MiSeq reagent kit v3 (Illumina, USA) at the BecA-ILRI Hub in Nairobi, Kenya.

Genome assembly

Software versions and relevant parameters are available in the Zenodo repository. We base called the MinION FAST5 data using the high accuracy model in Guppy (available at https://community.nanoporetech.com). We generated a preliminary assembly using wtdbg2 [57] and identified contaminants using taxon-annotated, GC-coverage plots (Figure S2) as implemented in blobtools [61]. Reads were mapped to the preliminary assembly using minimap2 [60] and the likely taxonomic origin of each contig was determined by searching NCBI nucleotide 'nt' or UniProt Reference Proteomes [59] using NCBI-BLAST+ [58] or DIAMOND [83], respectively. Reads originating from contaminant organisms were discarded. We generated the final assembly using wtdbg2. Sequencing errors were initially corrected by aligning the MinION reads to the assembly using minimap2 and performing four iterations of Racon [62] followed by a single iteration of Medaka (available at https://github.com/nanoporetech/medaka). Any remaining errors were corrected by aligning the Illumina MiSeq reads to the assembly using BWA-MEM [63] and performing two iterations of Racon followed by two iterations of Pilon [64].

Gene prediction

Prior to gene prediction, repeat sequences were identified *de novo* using RepeatModeler [65] and subsequently masked using RepeatMasker [66]. Protein-coding genes were predicted using BRAKER [67], using proteins sequences from the nematode-specific EggNOG database (which comprises sequences from *C. elegans*, *C. briggsae*, *C. remanei*, *C. japonica*, *Pristionchus pacificus* and *Trichinella spiralis*) [84] as homology evidence. Genome assembly and gene set completeness were assessed using BUSCO with the 'nematoda_odb9' database [68].

Estimation of heterozygosity

We estimated heterozygosity in the *C. bovis* genome using two approaches. We used Jellyfish [69] to count kmers (k = 19) in adaptor-trimmed and contaminant-free Illumina MiSeq reads and used the GenomeScope website [70] to estimate heterozygosity. To specifically call heterozygous sites in the *C. bovis* genome, we aligned the Illumina MiSeq reads to the *C. bovis* assembly using BWA-MEM and removed possible PCR duplicates from the resulting BAM file using PicardTools [71]. We performed variant calling using freebayes [72] and used bcftools [73] to remove variants sites that were dependent on strand or the position of the aligned read. We then estimated heterozygosity by dividing the total number of biallelic single nucleotide polymorphisms (SNPs) by the total number of sites (only those sites with a read depth $\geq$ 8 and $\leq$ 250, which represented 99.3% of the genome, were considered).

Assignment of *C. bovis* contigs to chromosomes

To assign *C. bovis* contigs to chromosomes, we identified one-to-one orthologs between *C. bovis* and *C. elegans* using a reciprocal best BLAST hit approach. Both proteomes were filtered so that they contained only the longest isoform per gene and searched against each other using blastp. Protein pairs which had reciprocal best BLAST hits with e-values < 1e-25 and a query coverage > 75% were declared as one-to-one orthologs. *C. bovis* contigs containing 10 or more *C. elegans* orthologs were assigned to the chromosome containing the majority of the *C. elegans* orthologs.

Orthology inference and phylogenomics

Accession details for all data used in this analysis are available in the Zenodo repository. We selected the protein sequence of the longest isoform of each protein-coding gene in *C. bovis*, 32 other species of *Caenorhabditis,* and the two outgroup taxa, *Diploscapter coronatus* and *Diploscapter pachys.* OrthoFinder [74] was used to cluster all protein sequences into putatively orthologous groups (OGs) using the default inflation value of 1.5. OGs containing loci which were present in at least 75% of species and which were, on average, single copy (mean count per species < 1.3) were selected. We aligned each selected OG using MAFFT [76] and generated a maximum likelihood tree along with 1000 ultrafast bootstraps [85] using IQ-TREE [75], allowing the best-fitting substitution model to be selected automatically [86]. Each tree was screened by PhyloTreePruner [77], collapsing nodes with bootstrap support < 90, and any OGs containing paralogs were discarded. If two representative sequences were present for any species (i.e., "in-paralogues") after this paralog screening step, only the longest of the two sequences was retained. We then realigned the remaining OGs using MAFFT and trimmed spuriously aligned regions using trimAl [87]. The trimmed alignments were subsequently concatenated to form a supermatrix using catfasta2phyml (available at https://github.com/nylander/catfasta2phyml). We inferred the species tree using IQ-TREE with the general time reversible model (GTR) with gamma-distributed rate variation among sites. The resulting tree was visualized using the iTOL web server [88]).

Gene content and structure analyses

To understand the large difference in protein-coding gene number between *C. bovis* and *C. elegans*, we used the orthology clustering set described previously to determine the level of redundancy in each gene set. For each species, we counted the number of loci in orthogroups containing two or more representatives from that species (multi-copy) and the number of loci in orthogroups containing a single representative from that species (single-copy). We also searched the longest isoform of each protein-coding gene from both species against the Pfam [89] database using InterProScan [79]. We then counted the number of loci in each species that were annotated as being GPCRs, NHRs, MSPs, F-box proteins, or C-type lectins. These gene families are known to constitute a substantial fraction of the *C. elegans* gene set [90].

We sought to identify gene families that have undergone expansion in the relatively small *C. bovis* gene set. We provided the orthology clustering set described previously to KinFin [78] to compare counts between *C. bovis* and all other species. We also searched the longest isoform of each protein-coding gene for all species against Pfam using InterProScan and provided the output to KinFin to annotate each orthogroup with a putative function. We screened the expanded gene families for functions that had previously been implicated in parasitism in other nematode species. To further affirm expansion in *C. bovis,* we generated gene trees for each orthogroup of interested using IQ-TREE as previously described.

To compare gene structure in *C. bovis* and *C. elegans*, we identified one-to-one orthologs in the orthology clustering set and extracted the exon counts and intron spans from the GFF annotation files of each species. As untranslated regions (UTRs) are not annotated in *C. bovis*, only coding exons and intervening introns were considered for both species. We calculated log2-transformed ratios of exon counts and intron spans for each gene pair using a Python script (available at https://github.com/lstevens17/cbovis_manuscript).

Repeat content analyses

To generate comprehensive repeat libraries and annotations for both *C. bovis* and *C. elegans*, we followed the approach of [91]. Briefly, we used TransposonPSI [80] to identify transposon sequences in both species, retaining those that were at least 50 bp in length. We also identified long terminal repeat (LTR) transposons in each species using LTRharvest [81]. We searched the resulting library with protein HMMs from Pfam and GyDB [92] using LTRdigest [93] and discarded any sequence that did not contain a transposable element domain. We also identified repetitive sequences *de novo* in both species using RepeatModeler. We then combined the resulting repeat libraries, classified each sequence using RepeatClassifier, and clustered the libraries at an identity of $\geq$ 80% using VSEARCH [82] to create a non-redundant repeat library for each species. We removed sequences from these repeat libraries that had significant homology to any member of the *C. elegans* gene set using TBLASTN. The resulting non-redundant and filtered repeat libraries were then provided to RepeatMasker which generated the final repeat annotations for each species along with genome file with repeat sequences masked with N's. We used a Python script (available at https://github.com/lstevens17/cbovis_manuscript) to compute repeat densities in 50 kb windows across chromosomes III and V in both *C. bovis* and *C. elegans*.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analyses were conducted using R (v3.5.1) [94] and Python 2.7. Gene structure ratios were log2-transformed using the math Python module. Loess smoothing curves were fitted to repeat densities using the ggplot2 R package (v2.3.1) [95].

DATA AND CODE AVAILABILITY

Raw sequence data and the genome assembly and annotation files have been deposited in the relevant INSDC databases under the accession NCBI:PRJEB34497. The assembly and gene set are also available to browse, query, and download at http://caenorhabditis.org/. Data files associated with this study have been deposited in Zenodo under the accession 10.5281/zenodo.3571457. Scripts and intermediate files associated with this study are available in the following GitHub directory https://github.com/lstevens17/cbovis_manuscript.