The Genome of Caenorhabditis bovis

Citation for published version:
Stevens, L, Rooke, S, Falzon, LC, Machuka, EM, Momanyi, K, Murungi, MK, Njoroge, SM, Odinga, CO, Ogendo, A, Ogola, J, Fèvre, EM & Blaxter, M 2020, 'The Genome of Caenorhabditis bovis', Current Biology. https://doi.org/10.1016/j.cub.2020.01.074

Digital Object Identifier (DOI):
10.1016/j.cub.2020.01.074

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Publisher's PDF, also known as Version of record

Published In:
Current Biology

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
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. Fève, 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.
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. Févre,3,4 and Mark Blaxter1,9

1Institute of Evolutionary Biology, Ashworth Laboratories, School of Biological Sciences, University of Edinburgh, Edinburgh EH9 3JT, UK
2Usher Institute, College of Medicine and Veterinary Medicine, University of Edinburgh, Edinburgh EH9 3JT, UK
3Institute of Infection and Global Health, University of Liverpool, 8 West Derby Street, Liverpool L69 7BE, UK
4International Livestock Research Institute, Old Naivasha Road, PO Box 30709 00100, Nairobi, Kenya
5Biosciences, Eastern and Central Africa, International Livestock Research Institute (BecA-ILRI) Hub, Old Naivasha Road, PO Box 30709 00100, Nairobi, Kenya
6Centre for Microbiology Research, Kenya Medical Research Institute, KNH Grounds, PO Box 54840 00200, Nairobi, Kenya
7Veterinary Department, Busia County Government, PO Box Private Bag 50400, Busia, Kenya
8Veterinary Department, Bungoma County Government, PO Box 2489 50200, Bungoma, Kenya
9Present address: Tree of Life, Wellcome Sanger Institute, Cambridge CB10 1SA, UK
10Lead 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 an 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...
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].

We sequenced the genome of an adult female Zebu that was 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.

**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]).

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).
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 separating C. bovis is highly diverged from all other sequenced species, 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), 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]).

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...
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).


ded 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 anthelmintic 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 ppg-11 function showed increased susceptibility to ivermectin [33], a widely used anthelmintic drug, and genetic variation in the ortholog of ppg-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 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
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 bottlenecks 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 anthelmintic 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 molluscidal 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 oitis, 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](https://doi.org/10.1016/j.cub.2020.01.074).

**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., Allon, 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., Strother, R., Calleymey-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., Faussett, S., Fitch, D., Frézal, L., Gosses, C., Kaur, T., et al. (2019). _Comparative genomics of 10 new _Caenorhabditis_ species_. Evol Lett 3, 217–236.
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 Rhabditidae. J. Parasitol. 20, 263–265.
49. Ferraz, C.M., Sobral, S.A., Senna, C.C., Junior, O.F., Moreira, T.F., Tobias, F.L., de Freitas Soares, F.E., Genieˆr, H.L.A., Vilela, V.L.R., Lima, J.A.C., et al. (2013). Combined use of invermectin, dimethyl sulfoxide, mineral oil and nematophagous fungi to control Rhabditis spp. Vet. Parasitol. 196, 1089–1024.
50. Msolla, P., Falmer-Hansen, J., Musemakweli, J., and Monrad, J. (1985). Alterations in the pattern of gene expression of nematodes induced by heat stress in the nematode Caenorhabditis elegans. Curr. Biol. 17, 3349–3357.
51. Verocai, G.G., Fernandes, J.J., Correia, T.R., Melo, R.M., Alves, P.A.M., and Scott, F.B., and Grisi, L. (2009). Inefficacy of albendazole sulphoxide and invermectin in the treatment of cattle parasitic ostitis 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., Genieˆr, H.L.A., Vilela, V.L.R., Lima, J.A.C., et al. (2019). Combined use of invermectin, 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 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 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. Pendir, 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., Sövíc, I., Nagarajan, N., and Sikić, 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., Abee, T., She, S., Priest, M., Abouelliel, A., Saktihikumar, S., Guomo, C.A., Zeng, G., Wortman, J., Young, S.K., and Earl, A.M. (2014). Phlox: 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/.
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-Ripollés, 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., Willhoft, 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](https://community.nanoporetech.com) |
| wtdbg2 | [57] | [https://github.com/ruanjue/wtdbg2](https://github.com/ruanjue/wtdbg2) |
| NCBI-BLAST+ | [58] | [ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/LATEST/](ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/LATEST/) |
| Diamond | [59] | [https://github.com/bbuchfink/diamond](https://github.com/bbuchfink/diamond) |
| minimap2 | [60] | [https://github.com/lh3/minimap2](https://github.com/lh3/minimap2) |
| blobtools | [61] | [https://blobtools.readme.io/docs](https://blobtools.readme.io/docs) |
| Racon | [62] | [https://github.com/isovic/racon](https://github.com/isovic/racon) |
| Medaka | N/A | [https://github.com/nanoporetech/medaka](https://github.com/nanoporetech/medaka) |
| BWA-MEM | [63] | [https://github.com/lh3/bwa](https://github.com/lh3/bwa) |
| Pilon | [64] | [https://github.com/broadinstitute/pilon](https://github.com/broadinstitute/pilon) |
| RepeatModeler | [65] | [http://www.repeatmasker.org/RepeatModeler/](http://www.repeatmasker.org/RepeatModeler/) |
| RepeatMasker | [66] | [http://www.repeatmasker.org/RMDownload.html](http://www.repeatmasker.org/RMDownload.html) |
| BRAKER | [67] | [https://github.com/Gaius-Augustus/BRAKER](https://github.com/Gaius-Augustus/BRAKER) |
| BUSCO | [68] | [https://mbio-ezlab.org/](https://mbio-ezlab.org/) |
| JellyFish | [69] | [https://www.cbcb.umd.edu/software/jellyfish/](https://www.cbcb.umd.edu/software/jellyfish/) |
| GenomeScope | [70] | [http://qb.cshl.edu/genomescope/](http://qb.cshl.edu/genomescope/) |
| PicardTools | [71] | [https://broadinstitute.github.io/picard/](https://broadinstitute.github.io/picard/) |
| freebayes | [72] | [https://github.com/ekg/freebayes](https://github.com/ekg/freebayes) |
| bcftools | [73] | [http://samtools.github.io/bcftools/bcftools.html](http://samtools.github.io/bcftools/bcftools.html) |
| OrthoFinder | [74] | [https://github.com/davidemms/OrthoFinder](https://github.com/davidemms/OrthoFinder) |
| IQ-TREE | [75] | [http://www.iqtree.org/](http://www.iqtree.org/) |
| MAFFT | [76] | [https://mafft.cbrc.jp/alignment/software/](https://mafft.cbrc.jp/alignment/software/) |
| PhyloTreePruner | [77] | [https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3825643/](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3825643/) |

(Continued on next page)
**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-k8ac22e) 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

---

**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 [82] | http://www.repeatmasker.org/RMDDownload.html |
| VSEARCH [82] | https://github.com/torognes/vsearch |

---

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
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 [81]. 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 [88].

**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 > 8 and ≤ 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 isofrom 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 isofrom 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 interest 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 \( R > 80\% \) 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.