BMC Genomics

RESEARCH

Open Access

Chromosome-length genome assembly of *Teladorsagia circumcincta* – a globally important helminth parasite in livestock

Shamshad Ul Hassan[1,2], Eng Guan Chua[2], Erwin A. Paz[1,2], Chin Yen Tay[2], Johan C. Greeff[3], Dieter G. Palmer[3], Olga Dudchenko[4,5], Erez Lieberman Aiden[4,5,6,7], Graeme B. Martin[1] and Parwinder Kaur[1*]

Abstract

Background Gastrointestinal (GIT) helminthiasis is a global problem that affects livestock health, especially in small ruminants. One of the major helminth parasites of sheep and goats, *Teladorsagia circumcincta*, infects the abomasum and causes production losses, reductions in weight gain, diarrhoea and, in some cases, death in young animals. Control strategies have relied heavily on the use of anthelmintic medication but, unfortunately, *T. circumcincta* has developed resistance, as have many helminths. Vaccination offers a sustainable and practical solution, but there is no commercially available vaccine to prevent Teladorsagiosis. The discovery of new strategies for controlling *T. circumcincta*, such as novel vaccine targets and drug candidates, would be greatly accelerated by the availability of better quality, chromosome-length, genome assembly because it would allow the identification of key genetic determinants of the pathophysiology of infection and host-parasite interaction. The available draft genome assembly of *T. circumcincta* (GCA_002352805.1) is highly fragmented and thus impedes large-scale investigations of population and functional genomics.

Results We have constructed a high-quality reference genome, with chromosome-length scaffolds, by purging alternative haplotypes from the existing draft genome assembly and scaffolding the result using chromosome conformation, capture-based, in situ Hi-C technique. The improved (Hi-C) assembly resulted in six chromosome-length scaffolds with length ranging from 66.6 Mbp to 49.6 Mbp, 35% fewer sequences and reduction in size. Substantial improvements were also achieved in both the values for N50 (57.1 Mbp) and L50 (5 Mbp). A higher and comparable level of genome and proteome completeness was achieved for Hi-C assembly on BUSCO parameters. The Hi-C assembly had a greater synteny and number of orthologs with a closely related nematode, *Haemonchus contortus*.

Conclusion This improved genomic resource is suitable as a foundation for the identification of potential targets for vaccine and drug development.

Keywords Genome assembly, Brown stomach worm, *Teladorsagia circumcincta*, Hi-C sequencing, 3D-DNA, Bioinformatics

*Correspondence:
Parwinder Kaur
parwinder.kaur@uwa.edu.au

Full list of author information is available at the end of the article

Background

Roundworms (phylum Platyhelminthes; class Nematoda) include some economically important species that infect livestock globally and incur huge annual losses in production [1, 2]. For example, *Teladorsagia circumcincta*, also known as the brown stomach worm, infects small ruminants including sheep [3] and is one of the major problematic helminth species in the southwestern part of Australia. This region has a Mediterranean-type climate with winter rainfall that favours the propagation of the larval stages of *T. circumcincta* on pasture [4].

The life cycle of *T. circumcincta* continues when third-stage (L_3) larvae on pasture are ingested by grazing sheep, exsheath and invade the mucosa of the abomasum where they develop into the fourth stage (L_4). Immature worms emerge from the mucosa into the gastric lumen where they develop into adult males and females and become sexually mature. The infection leads to functional disruption of the gastric mucosa, oedema of abomasal folds and sloughing of the mucosal lining, resulting in increased production of mucus, decreased production of acid, increased serum levels of pepsinogen and, possibly, protein deficiency (hypoalbuminemia). The host can suffer anorexia, dehydration, weight loss and diarrhoea, collectively leading to significant economic losses [2]. The helminth eggs leave the host in faecal material to re-contaminate the pasture and complete the life cycle, thus leading to recurrent infections [1].

For decades, control of the infection has relied on the extensive use of anthelmintic medications that were originally able to control the helminths, including *T. circumcincta.* Unfortunately, this practice has led to widespread development of resistance to some of the most effective anthelmintics on the market, including monepantel [5, 6]. Among the alternative, sustainable options are vaccination, but for *T. circumcincta*, a vaccine is not commercially available [7]. All issues considered; therefore, we need to be able to identify new targets for vaccine and drug development and elucidate the mechanisms that lead to anthelmintic resistance. Clearly, a good starting point in this quest would be a high-quality reference genome assembly.

Advances in high-throughput sequencing technologies over the past two decades have triggered a massive output of genomic data. The improvements in the technology provide an opportunity to revisit the original sequencing and genome assembling attempts. The original sequencing attempt that resulted in a highly fragmented genome thus offers a real opportunity to develop a high-quality genomic resource for *T. circumcincta*, potentially allowing major gains in our basic understanding of the physiology, evolutionary biology, pathogenesis of infection, host immune response, and the mechanisms that underpin the anthelmintic resistance [8, 9].

Table 1 Quality assessments of the original and Hi-C integrated genome assemblies of *T. circumcincta*

Parameters	Original assembly	Purged Hi-C assembly[a]	Chromosomes only[b]
Number of scaffolds	81,730	52,860	6
Total size of scaffolds (bp)	700,607,159	614,147,289	363,201,533
Longest scaffold (bp)	1,473,920	66,586,561	66,586,561
N50 scaffold length (bp) count	47,089	57,133,369	62,733,602
L50 scaffold count	3152	5	3
Number of contigs	213,313	175,861	76,733
Longest contig (bp)	98,345	98,345	98,345
N50 contig length (bp)	3624	4009	5700
L50 contig count	40,623	32,178	14,421
BUSCO assessment[c]			
Complete (single-copy + duplicated)	2099 (67%)	2112 (67.5%)	1840 (58.8%)
Complete and single-copy	1835 (58.6%)	1978 (63.2%)	1821 (58.2%)
Complete and duplicated	264 (8.4%)	134 (4.3%)	19 (0.6%)
Fragmented	350 (11.2%)	358 (11.4%)	269 (8.6%)
Missing	682 (21.8%)	661 (21.1%)	1022 (32.6%)

[a]Includes all available scaffolds; [b]Includes only six chromosome-length scaffolds; [c]BUSCO assessment was performed using the nematode odb10 dataset which contains 3131 orthologs.

In the present study, we aimed to improve the current *T. circumcincta* draft genome to a chromosome-length assembly, using chromosome conformation capture technique, or in situ Hi-C [10], and thus increase the value of the genome resource by annotating and analysing it for genome-wide synteny and orthologs.

Results

Genome contiguity and completeness

The original draft genome assembly (GCA_002352805.1) was highly fragmented with 81,730 scaffolds, with N50 of 47,089 bp and L50 of 3152, and a total size of 700 Mbp (Table 1). Following the purging of alternative haplotypes and the integration of Hi-C sequencing data, the new Hi-C assembly contained 52,860 scaffolds approximately 35% fewer sequences than the original draft. Notably, of these, six were chromosome-length as shown in Fig. 1, with lengths ranging from 66.6 Mbp to 49.6 Mbp. Substantial improvements were achieved in both the values for N50 (57.1 Mbp) and L50 (5 Mbp). The longest scaffold had increased markedly in length, from approximately 1.4 Mbp in the original assembly to nearly 66.6 Mbp in the Hi-C assembly, while the estimated genome size was reduced from 700 Mbp to 614 Mbp, probably due to improved identification and separation of haplotypes.

Next, BUSCO (with nematode odb10 data) was used to assess and compare the genome completeness levels of both assemblies. After adding scaffolds (n=353)

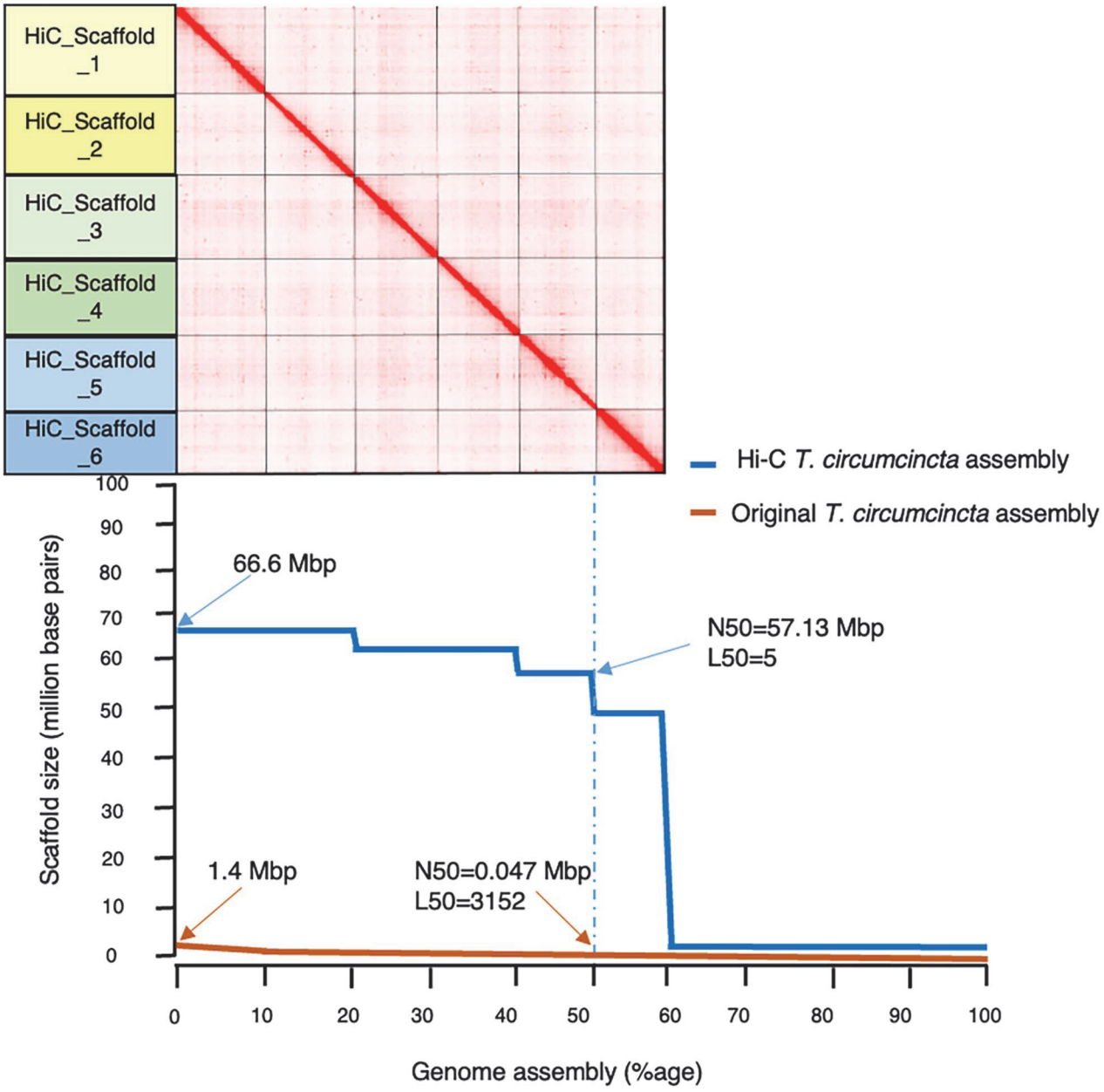

Fig. 1 Comparison of the Hi-C and draft genome assemblies for genome contiguity and completeness. Top: Hi-C matrix of the spatial clustering of Hi-C reads to six chromosome-length scaffolds in Hi-C assembly. The interactive contact map is available at https://www.dnazoo.org/assemblies/Teladorsagia_circumcincta. Bottom: comparison of the scaffold lengths of Hi-C and draft genome assemblies (values for N50 and L50 are indicated for both assemblies)

to the Hi-C assembly from the draft assembly that contained missing BUSCOs, we detected a higher level of genome completeness in the Hi-C assembly, with 67.5% (2112/3131) of BUSCO genes identified compared to 67% (2099/3131) in the original assembly (Table 1). More importantly, the Hi-C assembly contained 143 more single-copy and 130 fewer duplicated BUSCO genes, than the original assembly, indicating a significant reduction in the number of duplicated sequences. We then examined the genome completeness of only the six chromosome-length scaffolds, achieving an overall completeness score of 58.8% compared to 67.5% in the entire Hi-C-assembly. The sequences for the missing BUSCOs were retrieved manually from https://www.orthodb.org/ and 1269 scaffolds containing missing BUSCOs were added to the six chromosome-length scaffolds and the completeness score rose to 67.1%, very similar to the Hi-C assembly containing 52,860 scaffolds.

Table 2 Comparison of genome annotations in the purged, Hi-C integrated and original genome assemblies of *T. circumcincta*

Parameters	Purged Hi-C assembly	Original assembly
mRNAs (n)	30,055	39,896
Genes (n)	28,082	37,276
Exons (n)	239,113	281,766
CDS (n)	239,106	281,759
Introns (n)	197,880	231,167
Start Codon (n)	26,422	32,685
Stop codon (n)	27,090	33,400
BUSCO assessment*		
Complete (single-copy + duplicated)	2402 (76.7%)	2409 (76.9%)
Complete and single-copy	1834 (58.6%)	1816 (58.0%)
Complete and duplicated	568 (18.1%)	593 (18.9%)
Fragmented	239 (7.6%)	250 (8%)
Missing	490 (15.7%)	472 (15.1%)

*BUSCO assessment was performed using the nematode odb10 dataset which contains 3131 orthologs.

Genome and functional annotations

The genome annotation results generated from the Braker2 pipeline are outlined in Table 2. The annotated Hi-C assembly had fewer genes (28,082) and mRNA transcripts (30,055), compared to the original draft (37,276 genes; 39,896 mRNA transcripts), but the BUSCO assessment scores of both protein sequence sets were highly comparable. In the Hi-C assembly, the overall genome completeness level was 76.7%, slightly less than that of the original assembly (76.9%). However, it is important to note that, in comparison to the original assembly, the Hi-C assembly contained more single-copy (58% vs. 58.6%), fewer duplicates (18.9% vs. 18.1%) and fewer fragmented (8% vs. 7.6%) orthologs, demonstrating the improvement in genome accuracy and fragmentation.

The complete functional annotation outcomes are available in Additional File 1. Overall, based on the protein sequences extracted from the annotated Hi-C assembly, nearly half of the predicted Gene Ontology (GO) terms (49.18%; 12,265 terms), were classified under the molecular function category, followed by the cellular component (31.68%; 8,133 terms) and biological processes (19.14%; 4,915 terms). As depicted in Fig. 2 some of the most frequent GO biological process terms were 'translation', 'intracellular signal transduction', 'carbohydrate metabolic process', 'regulation of transcription' and 'intracellular protein transport'. The most frequent GO terms in the cellular component category included 'integral component of membrane', "nucleus', 'cytoplasm', 'extracellular region' and 'plasma membrane'. In the molecular function group, bindings to nucleic acids and both ATP and GTP, as well as metal ions, including zinc and calcium, were the most common GO terms predicted.

Genome synteny analysis

Both versions of the *T. circumcincta* assembly were compared with *H. contortus* using pairwise synteny analysis because *H. contortus* has a near-complete genome assembly [11] and, more importantly, phylogenetic analysis shows that it is closely related to *T. circumcincta* [12]. The synteny between the *H. contortus* genome and the original assembly for *T. circumcincta* was relatively poor (Fig. 3A) and greatly improved with the Hi-C assembly (Fig. 3B). It is important to note the strikingly high level of synteny between all six chromosome-length scaffolds in the Hi-C assembly and the six chromosomal sequences of *H. contortus.* Further, synteny analysis allowed identification, for the first time, of the X-chromosome in *T. circumcincta*, with Hi-C scaffold 6 evident as the counterpart of the X-chromosome of *H. contortus*. Interestingly, no syntenic links could be drawn between any unplaced scaffolds in the Hi-C assembly and *H. contortus* genome sequences, perhaps because the parameters were too stringent during the alignment process and when bundling the syntenic links in Circos.

Orthology analysis

Using OrthoVenn2, the protein sequences from annotated *T. circumcincta* Hi-C assembly were also compared with those from *H. contortus*, as well as with two other more distant parasitic nematode species, *Burgia malayi* and *Trichinella spiralis*. Of 12,504 ortholog clusters, 3,214 were shared by all four species (Fig. 4a and b). As expected, the closely related helminths, *T. circumcincta* and *H. contortus*, shared the most orthologs (7,332 clusters), whereas *T. circumcincta* shared only 3,318 orthologs with *B. malayi* and 3,291 orthologs with *T. spiralis*. Using Orthofinder, we also compared the number of orthologs shared between *H. contortus* and the original and Hi-C assemblies of *T. circumcincta*. As shown in Fig. 4c, the Hi-C assembly shared significantly more orthologs (6948) with *H. contortus* than the original draft (5313).

Discussion and conclusion

The present project aims to improve the current genome reference for *T. circumcincta*, a helminth nematode that is important for small ruminant livestock [8]. By purging alternative haplotypes and using in situ Hi-C to order, orient, correct and anchor draft sequences to chromosomes [10, 13], we have been able to improve the draft genome and create the first chromosome-length assembly for *T. circumcincta*.

The Hi-C assembly is more contiguous and complete than the previously available draft, and, at 614 Mbp, 13%

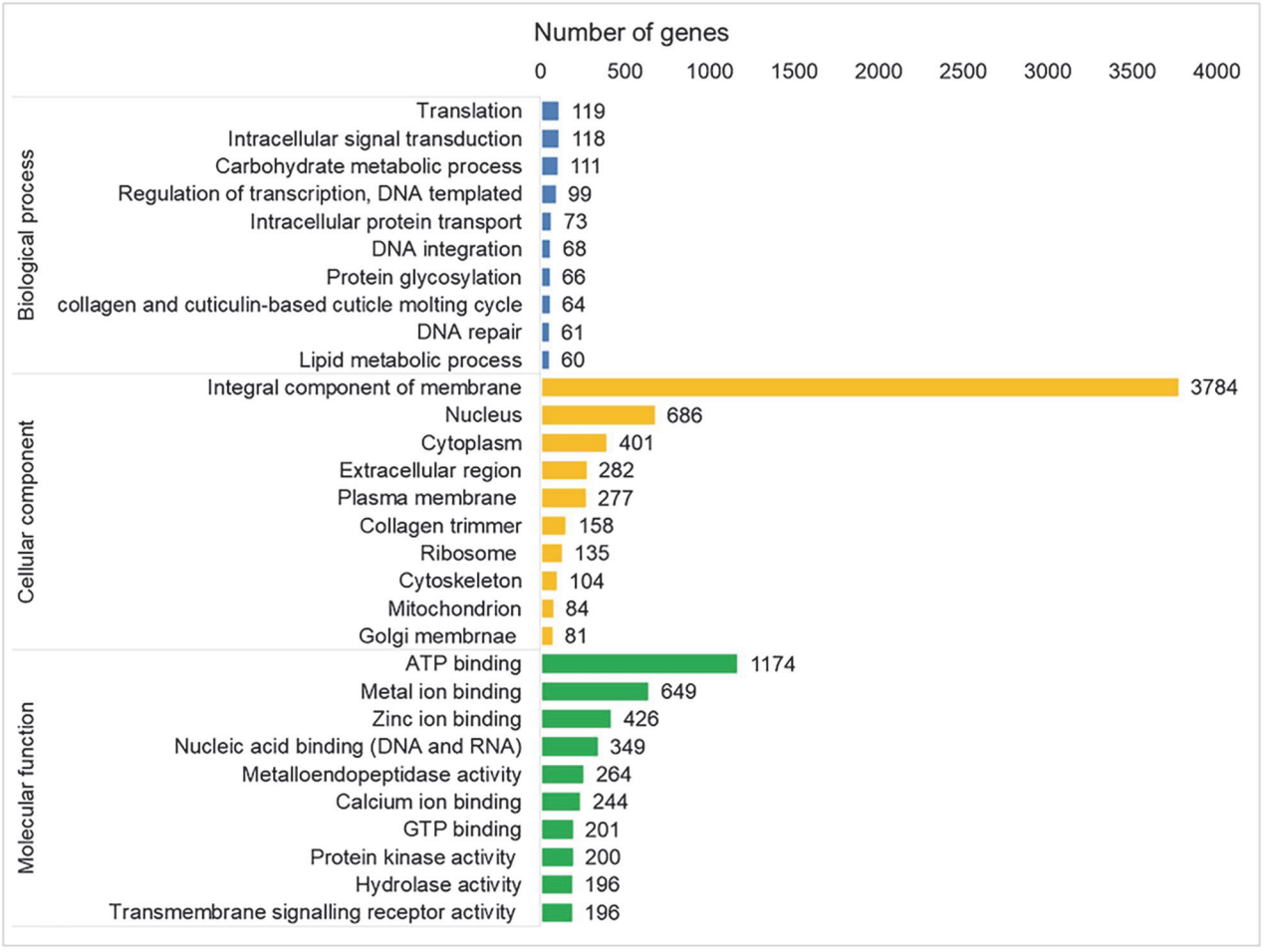

Fig. 2 Bar plots depicting the 10 most abundant Gene Ontology (GO) terms in the Hi-C assembly, for biological processes, cellular components and molecular functions

smaller than the original assembly. This reduction in size makes the revised genome of *T. circumcincta* more consistent with that of *H. contortus*, another helminth nematode of the same clade, where the genome size has recently been reduced from 465 Mbp to 283 Mbp [11]. The karyotype (2n=12) of the *T. circumcincta* genome, identified for the first time in the present analysis, is also consistent with that of *H. contortus* [11], as well as that of *C. elegans*, a model organism that is a free-living nematode [14]. Furthermore, the synteny analysis between the chromosome-length assemblies of *T. circumcincta* and *H. contortus* suggest that chromosomes are syntenic [12] but, while genes are conserved between the two species, the gene order is not, and different regions are linked to different chromosomes [11]. For example, Hi-C Scaffold 6 is syntenic to Chromosome-X on *H. contortus*, whereas Hi-C Scaffold 1 is syntenic to Chromosome 5, Hi-C Scaffold 2 is syntenic to Chromosome 4, and Hi-C Scaffold 3 is syntenic to Chromosome 3.

After genome annotation, there were fewer genes in the Hi-C *T. circumcincta* assembly because haplotypes had been removed and contiguity increased, compared to the original *T. circumcincta* assembly [15]. Although the number of predicted proteins was reduced in the Hi-C assembly, completeness and accuracy were identical for both assemblies, suggesting that, during Hi-C assembly, the rearrangements and reductions in fragmentation increased the number of curated gene models [15]. The single-copy orthologs (SCOs) were also compared across four helminth species from different clades – *T. circumcincta, H. contortus, B. malayi* and *T. spiralis.* As *T. circumcincta* and *H. contortus* belong to the same clade-Va, they share more SCOs (7332) with each other than they share with the other species showing that clade variation can affect the number of shared SCOs within helminths as *T. circumcincta* shares 3318 SCOs with *B. malayi* (clade-III) and 3291 SCOs with *T. spiralis* (clade-I). This variation in shared SCOs is an outcome of speciation and differences among life cycle stages of each helminth

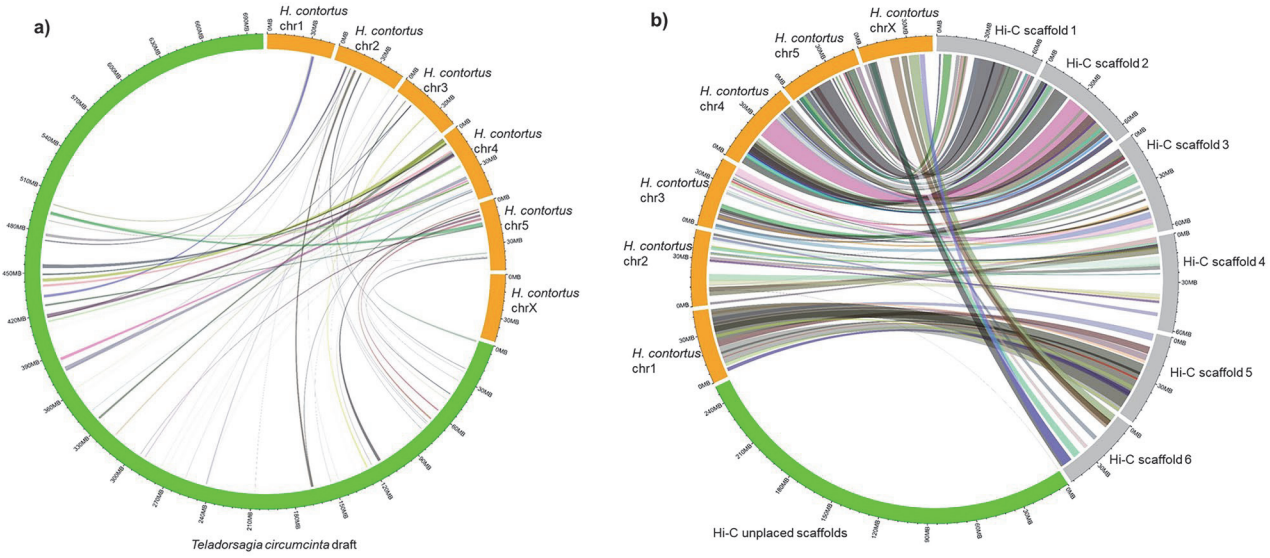

Fig. 3 Syntenic relationships between *Haemonchus contortus* genome (orange) and **(a)** the original genome assembly (green) for *T. circumcincta*; and **(b)** the Hi-C genome assembly for *T. circumcincta* (chromosome-length scaffolds in grey; unplaced scaffolds in green). Syntenic links were bundled using the following parameters: --max_gap = 1,000,000 --min_bundle_size = 10,000 min_bundle_membership = 5

– for example, *T. spiralis* with a broad host range, lives in muscle and small intestine [16], whereas infective larvae of *T. circumcincta* and *H. contortus* are found on pastures and infect the abomasum [17], and *B. malayi* requires the mosquito as an intermediate host and infects lymph nodes [18].

Our improved Hi-C assembly still contains several unplaced scaffolds. The analysis of completeness and accuracy of the six Hi-C scaffolds (~59% BUSCO; Table 1) suggests that most of the genetic information is retained in the chromosome-length scaffolds. A total of 1275 scaffolds (six chromosome-length scaffolds plus 1269 unplaced scaffolds), has the completeness level like that for the total scaffolds in the Hi-C assembly (52,860), indicating redundancy in the unplaced scaffolds.

In conclusion, our chromosome-length scaffold assembly and annotation have advanced the genomics of the economically important small ruminant nematode parasite, *T. circumcincta* (isolated from Western Australia). The availability of a better reference genome, with greater comprehension of the genetic architecture of Teladorsagiosis, will help phylogenomic analysis of helminths of various clades [19], and help understand the parasite biology and host-parasite interactions. Ultimately, this information should lead to new options for vaccine and drug targets and, most importantly, pave the way to sustainable solutions for gastrointestinal parasitism [20]. Finally, the inclusion of long-read sequencing (from PacBio or Oxford Nanopore) should help resolve the unplaced scaffolds in the current version of the genome assembly [21, 22].

Materials and methods

Helminth collection and identification

Helminths were collected from the abomasum (predilection site for *T. circumcincta)* of sheep obtained from the Western Australian Meat Marketing Company (WAMMCO). The sheep had been naturally infected with *T. circumcincta*, an important helminth in the southwest of Western Australia. The abomasal contents were carefully scraped onto a sieve (mesh size 150 μm) and washed thoroughly and placed in a petri-dish from which individual helminths were removed with the aid of a dissecting microscope. Helminth species were identified based on morphological characteristics (Fig. 5) using differential contrast and compound microscopy. Males were identified by the shape and length of spicules which are up to 450 μm in length; females were identified by the presence of a vulvar flap, annular rings and their body length (10–12 mm; about twice that of males) [3]. Eggs can also be seen in females near the vulvar flap from where they are laid. The worms were then thoroughly washed with physiological saline and stored at −80 °C until processing. Extracted DNA (see below) was subjected to PCR using helminth specific ITS2 primers, as previously described [23]. Helminth's identity was confirmed by Sanger sequencing of the PCR product followed by a blastn search against the NCBI database.

DNA extraction

Briefly, the helminths (100 mature male and female *Teladorsagia circumcincta* in equal ratios) were mechanically homogenized using a sterile micro-pestle in a microcentrifuge tube containing 200 μL of Tris-EDTA

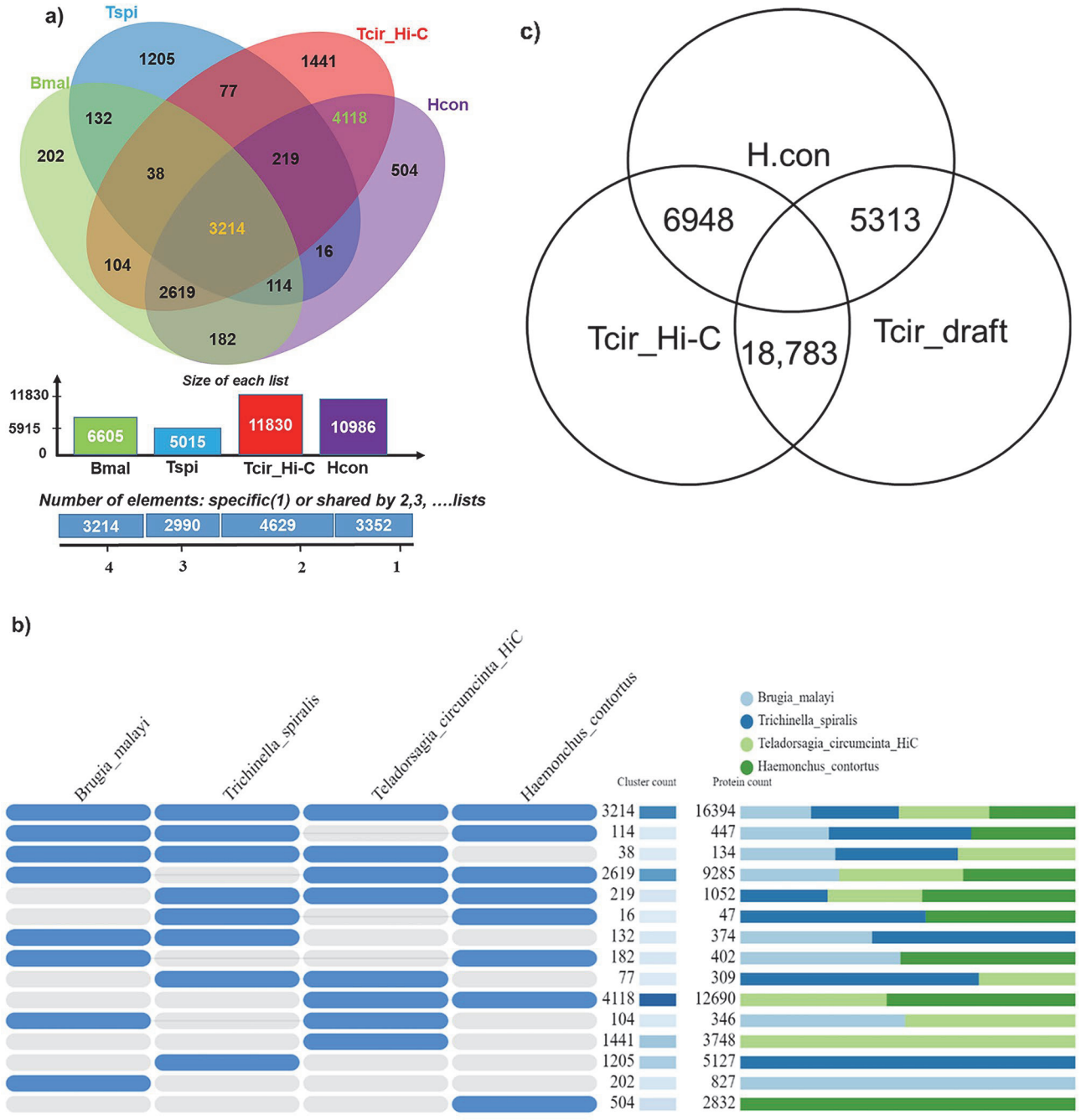

Fig. 4 Orthologs shared among helminth species. **(a)** Venn diagram showing comparisons and distribution of orthologous clusters shared among *Burgia malayi* (Bmal, clade-III nematode), *Trichinella spiralis* (Tspi, clade-I nematode), *H. contortus* (Hcon, clade-Va nematode), *T. circumcincta* Hi-C assembly (Tcir_Hi-C, clade-Va nematode) . The species formed 14,185 clusters of which 12,504 were orthologous (contained in at least two species) and 1,681 were single-copy gene clusters. **(b)** Table showing the pattern of occurrence of shared orthologues among Bmal, Tspi, Hcon and Tcir_Hi-C. **(c)** Venn-diagram indicating one-to-one OrthologuesStats inferred from Orthofinder by comparing proteomes of *H. contortus* with *T. circumcincta* Hi-C and *T. circumcincta* draft

buffer, 1% (v/v) β-mercaptoethanol, 200 mg proteinase K, 10 mg/ml RNAase, 0.5 M EDTA and 10% (v/v) sodium dodecyl sulphate. The cell lysate was then incubated at 65 °C for 2 h. After incubation, an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1) was added and the mixture was centrifuged at 10,000 *g* for 5 min. The supernatant was collected into a sterile microcentrifuge tube and resuspended with an equal volume of chloroform:isoamyl alcohol (24:1). After centrifugation, the supernatant was again collected into a sterile microcentrifuge tube, this time with ice-cold ethanol (95% v/v) to precipitate the DNA. The DNA pellet was washed

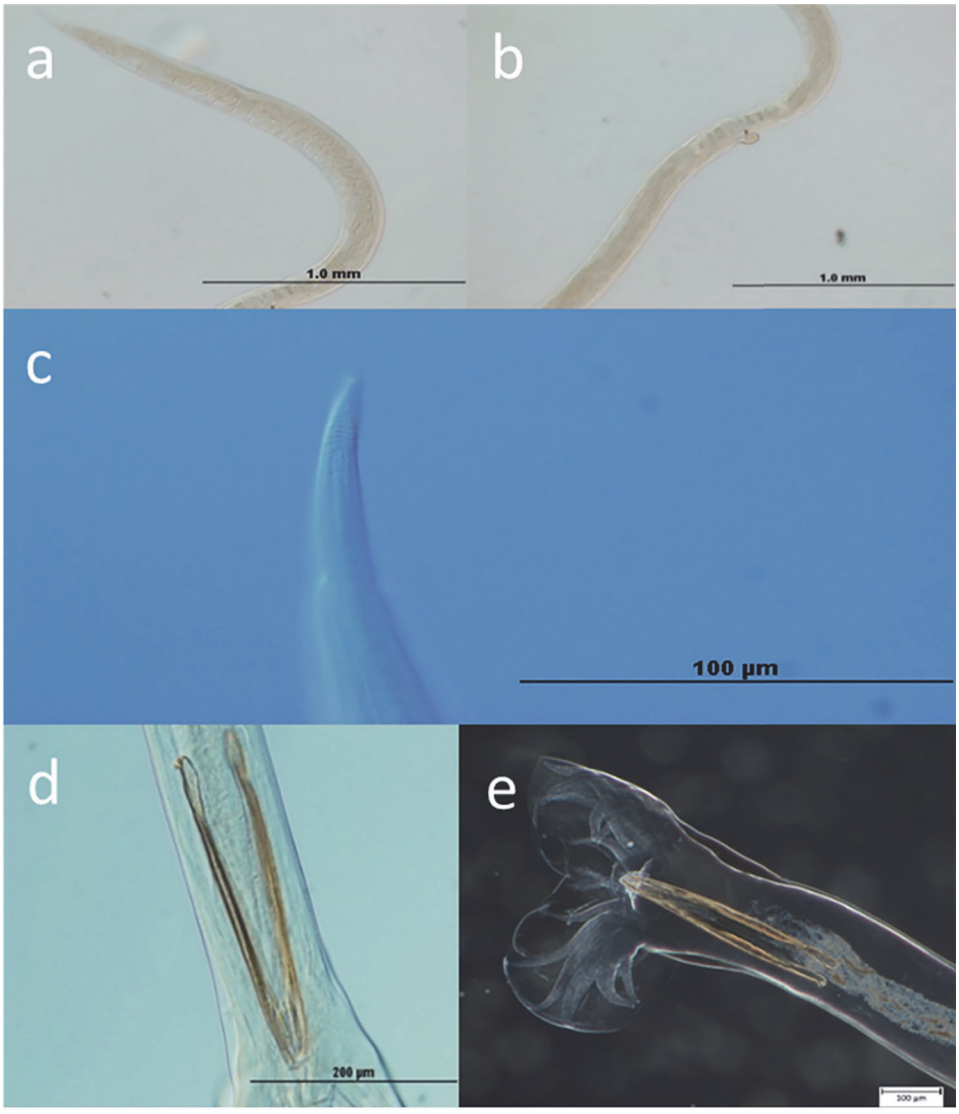

Fig. 5 Morphological identification of *T. circumcincta*. (a) Eggs towards the posterior end of the female; (b) Vulvar flap towards the posterior end of the female; (c) Annular rings towards the posterior end of the female; (d) and (e) Spicules towards the posterior, a specific characteristic of the male of this species

with ethanol (70% v/v) before being resuspended in 50 µL DEPC water. The integrity of the extracted DNA was assessed by electrophoresis on 1% (w/v) agarose gel. The quality and quantity of the DNA were assessed using a NanoDrop 2000 spectrophotometer (Thermofisher, USA) and a Qubit 2.0 fluorometer (Thermofisher, USA).

PCR amplification of the helminth specific ITS2 region

The ITS2 primer sequences were 5'-CTTAATGATCTC-GCCTAGACG-3' (forward) and 5'-TTTCATCGATAC-GCGAATCG-3' (reverse). A 50 µL reaction mixture (reaction buffer 10 µL; forward and reverse primer 2 µL each; DNA polymerase 1 µL; DNA sample 3 µL; water 32 µL) was run through 35 cycles of PCR with MyTaq HS DNA (Bioline, Canada), using following conditions: initial denaturation at 95 °C for 1 min followed by 35 amplification cycles, each comprising denaturation at 95 °C for 15 s, annealing at 54 °C for 30 s, and extension at 72 °C for 10 s.

Hi-C sequencing, chromosome-length scaffolding and quality assessment

In situ Hi-C sequencing was performed as described previously [10] using 100 adult *T. circumcincta*, including both males and females. We constructed one in situ library which was then sequenced using the Illumina NovaSeq 6000 platform. The generated Hi-C reads were used to anchor, order, orient, and correct misjoins in the existing draft genome assembly (GCA_002352805.1) using the 3D de novo assembly (3D-DNA) pipeline [24].

Before scaffolding with Hi-C reads, the draft assembly was run through purge haplotigs software [25]. The resulting assembly was then polished using the Juicebox Assembly Tools [13]. The resulting contact map was visualized using Juicebox visualization software [13]. QUAST (v5.0.2) was used to assess the assembly metrics [26]. Benchmark for Universal Single Copy Orthologues (BUSCO, v5.1.2) was used in *genome mode* to determine the genome completeness [27]. In this analysis, the sequences for missing BUSCOs in the Hi-C assembly were retrieved manually from https://www.orthodb.org/ (orthoDB v10) and blasted against the draft genome to obtain the relevant scaffolds which were then addedto the Hi-C assembly. The list of added scaffolds can be found in Additional File 2.

Genome and functional annotations

The original (GCA_002352805.1) and Hi-C integrated draft genome assemblies were annotated using Braker2 v2.1.6 [28]. First, each genome was softmasked using RepeatMasker v4.1.1 [29] with custom repeat library built upon itself by RepeatModeler v2.0.1 [29]. The Braker2 was run with the --etpmode parameter enabled to train GeneMark-ETP [30] with RNA-Seq data and protein hints. The GeneMark-ETP predictions were then used for training AUGUSTUS, following which genes with hints were predicted by AUGUSTUS [30–34]. Five sets of *T. circumcincta* RNA-Seq data (sequence read accession numbers SRX1507697, SRX1507698, SRX2485888, SRX2485887, SRX2485886) derived from two previous studies [8, 35], were downloaded from the NCBI Database and aligned to both the original draft and our improved Hi-C version of genome assemblies, using STAR (v2.7.6a) with default parameters [36, 37]. The *Caenorhabditis elegans* proteome from the UniProt Database served as protein hints when running Braker2. BUSCO was run in *protein mode* to assess the annotation results. After genome annotation, functional analysis was performed using the web-based Gene Ontology Functional Enrichment Annotation Tool (GO FEAT) [38].

Genome synteny and orthology analyses

Genome-wide synteny was analysed using Cactus v1.3.0 and halSynteny [39] to compare the Hi-C integrated *T. circumcincta* genome assembly with the original GCA_002352805.1 genome assembly, and the genome of a closely related helminth species, *Haemonchus contortus* (GCA_000469685.2). A hierarchical alignment (hal) output file was generated using the Cactus package, and a PSL output file with syntenic links was generated using the halSynteny function within Cactus, using the following parameters: --minBlockSize 10,000 --maxAnchorDistance 1,000,000. The syntenic links were bundled using Circos tools v0.69-8 in Galaxy platform v7 [40, 41] and then visualized using shinyCircos [42]. The single copy orthologs in both the original and Hi-C integrated *T. circumcincta* genome assemblies, as well as the draft assembly of *Haemonchus contortus*, were inferred using Orthofinder [43]. OrthoVenn2 [44] was also used to compare the orthologs between four nematode species: *Burgia malayi*; *Trichinella spiralis*; *H. contortus; T. circumcincta* [12].

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-023-09172-0.

Additional File 1

Additional File 2

Acknowledgements

S.U.H. was supported by a joint PhD scholarship from the University of Agriculture Faisalabad (Pakistan) and the University of Western Australia (reference number PS-2(11) FDP/17/8071). Hi-C data were created in collaboration with the DNA Zoo Consortium (www.dnazoo.org). DNA Zoo sequencing effort is supported by Illumina, Inc., IBM, and the Pawsey Supercomputing Center. P.K. is supported by the University of Western Australia. Special thanks to Ashling Charles from the team at DNA Zoo Australia for routine data processing support. We also acknowledge the resources provided by the Department of Primary Industries and Regional Development (DPIRD) and Western Australian Meat Marketing Company (WAMMCO).

Author Contribution

Conceptualization: S.U.H., P.K. and C.Y.T. Computational analysis and data interpretation: S.U.H., P.K. and E.G.C. Investigation; S.U.H., P.K., O.D., E.L.A., J.C.G., E.A.P., C.Y.T., and D.G.P. Writing (original draft); S.U.H. Writing (review and editing); G.B.M. and E.G.C. All authors read and approved the final manuscript.

Funding

S.U.H. received a joint scholarship from the University of Agriculture Faisalabad (Pakistan) and the University of Western Australia for PhD studies (reference number PS-2(11) FDP/17/8071). E.L.A. was supported by the Welch Foundation (Q-1866), a McNair Medical Institute Scholar Award, an NIH Encyclopedia of DNA Elements Mapping Center Award (UM1HG009375), a US-Israel Binational Science Foundation Award (2019276), the Behavioral Plasticity Research Institute (NSF DBI-2021795), NSF Physics Frontiers Center Award (NSF PHY-2019745), and an NIH CEGS (RM1HG011016-01A1).

Data Availability

The interactive Hi-C contact map for the genome assembly is available at www.dnazoo.org. The genome assembly and intermediate files can be accessed here; https://www.dropbox.com/sh/czjlxso80stoqts/AAA0wnAO0qttk8i3--rHOPFba?dl=0.

Declarations

Ethics approval and consent to participate

Not applicable. No live animals were used.

Consent for publication

Not applicable.

Competing interests

The author(s) declare no competing interests.

Author details

[1]UWA School of Agriculture and Environment, The University of Western Australia, 6009 Crawley, WA, Australia

[2]Helicobacter Research Laboratory, The Marshall Centre for Infectious Disease Research and Training, School of Biomedical Sciences, University of Western Australia, Perth, WA, Australia
[3]Department of Primary Industries and Regional Development, Western Australia 3 Baron Hay Court, South Perth 6151, WA, Australia
[4]The Center for Genome Architecture, Department of Molecular and Human Genetics, Baylor College of Medicine, 77030 Houston, TX, USA
[5]Center for Theoretical Biological Physics, Rice University, 77005 Houston, TX, USA
[6]Shanghai Institute for Advanced Immunochemical Studies, ShanghaiTech, Pudong, China
[7]Broad Institute of MIT and Harvard, Cambridge, MA, USA

Received: 27 September 2022 / Accepted: 8 February 2023
Published online: 15 February 2023

References

1. Stear M, Bishop S, Henderson N, Scott I. A key mechanism of pathogenesis in sheep infected with the nematode *Teladorsagia circumcincta*, Animal Health Research Reviews. 4 (2003) 45–52. doi: https://doi.org/10.1079/ahrr200351. PMID: 12885208.
2. Craig TM, CHAPTER 22 - Helminth Parasites of the Ruminant Gastrointestinal Tract, in: D.E. Anderson, D.M. Rings, editors, Food Animal Practice (Fifth Edition), Saunders WB. Saint Louis, 2009: pp. 78–91. https://doi.org/10.1016/B978-141603591-6.10022-3.
3. Roeber F, Jex AR, Gasser RB. Chapter Four - Next-Generation Molecular-Diagnostic Tools for gastrointestinal nematodes of Livestock, with an emphasis on small ruminants: a turning point? In: Rollinson D, editor. Advances in parasitology. Academic Press; 2013. pp. 267–333. https://doi.org/10.1016/B978-0-12-407705-8.00004-5.
4. O'Connor LJ, Walkden-Brown SW, Kahn LP. Ecology of the free-living stages of major trichostrongylid parasites of sheep. Vet Parasitol. 2006;142:1–15. https://doi.org/10.1016/j.vetpar.2006.08.035.
5. Turnbull F, Devaney E, Morrison AA, Laing R, Bartley DJ. Genotypic characterisation of monepantel resistance in historical and newly derived field strains of *Teladorsagia circumcincta*. Int J Parasitology: Drugs Drug Resist. 2019;11:59–69. https://doi.org/10.1016/j.ijpddr.2019.10.002.
6. Kaplan RM, Vidyashankar AN. An inconvenient truth: Global worming and anthelmintic resistance, Veterinary Parasitology. 186 (2012)70–78. https://doi.org/10.1016/j.vetpar.2011.11.048.
7. Nisbet AJ, McNeilly TN, Wildblood LA, Morrison AA, Bartley DJ, Bartley Y, Longhi C, McKendrick IJ, Palarea-Albaladejo J, Matthews JB. Successful immunization against a parasitic nematode by vaccination with recombinant proteins. Vaccine. 2013;31:4017–23. https://doi.org/10.1016/j.vaccine.2013.05.026.
8. Choi Y-J, Bisset SA, Doyle SR, Hallsworth-Pepin K, Martin J, Grant WN, Mitreva M. Genomic introgression mapping of field-derived multiple-anthelmintic resistance in *Teladorsagia circumcincta*. PLOS Genet. 2017;13:e1006857. https://doi.org/10.1371/journal.pgen.1006857.
9. Greenwood JM, Ezquerra AL, Behrens S, Branca A, Mallet L. Current analysis of host–parasite interactions with a focus on next generation sequencing data,Zoology.119 (2016)298–306. https://doi.org/10.1016/j.zool.2016.06.010.
10. Rao SSP, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES, Aiden EL. A 3D map of the Human Genome at Kilobase Resolution reveals principles of chromatin looping. Cell. 2014;159:1665–80. https://doi.org/10.1016/j.cell.2014.11.021.
11. Doyle SR, Tracey A, Laing R, Holroyd N, Bartley D, Bazant W, Beasley H, Beech R, Britton C, Brooks K, Chaudhry U, Maitland K, Martinelli A, Noonan JD, Paulini M, Quail MA, Redman E, Rodgers FH, Sallé G, Shabbir MZ, Sankaranarayanan G, Wit J, Howe KL, Sargison N, Devaney E, Berriman M, Gilleard JS, Cotton JA. Genomic and transcriptomic variation defines the chromosome-scale assembly of *Haemonchus contortus*, a model gastrointestinal worm. Commun Biology. 2020;3:656. https://doi.org/10.1038/s42003-020-01377-3.
12. Coghlan A, Tyagi R, Cotton JA, Holroyd N, Rosa BA, Tsai IJ, Laetsch DR, Beech RN, Day TA, Hallsworth-Pepin K, Ke H-M, Kuo T-H, Lee TJ, Martin J, Maizels RM, Mutowo P, Ozersky P, Parkinson J, Reid AJ, Rawlings ND, Ribeiro DM, Swapna LS, Stanley E, Taylor DW, Wheeler NJ, Zamanian M, Zhang X, Allan F, Allen JE, Asano K, Babayan SA, Bah G, Beasley H, Bennett HM, Bisset SA, Castillo E, Cook J, Cooper PJ, Cruz-Bustos T, Cuéllar C, Devaney E, Doyle SR, Eberhard ML, Emery A, Eom KS, Gilleard JS, Gordon D, Harcus Y, Harsha B, Hawdon JM, Hill DE, Hodgkinson J, Horák P, Howe KL, Huckvale T, Kalbe M, Kaur G, Kikuchi T, Koutsovoulos G, Kumar S, Leach AR, Lomax J, Makepeace B, Matthews JB, Muro A, O'Boyle NM, Olson PD, Osuna A, Partono F, Pfarr K, Rinaldi G, Foronda P, Rollinson D, Samblas MG, Sato H, Schnyder M, Scholz T, Shafie M, Tanya VN, Toledo R, Tracey A, Urban JF, Wang L-C, Zarlenga D, Blaxter ML, Mitreva M, Berriman M. International Helminth Genomes Consortium, Comparative genomics of the major parasitic worms, Nature Genetics. 51 (2019) 163–174. https://doi.org/10.1038/s41588-018-0262-1.
13. Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, Shamim MS, Machol I, Lander ES, Aiden AP, Aiden EL. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. Science. 2017;356:92–5. https://doi.org/10.1126/science.aal3327.
14. Roelens B, Schvarzstein M, Villeneuve AM. Manipulation of Karyotype in *Caenorhabditis elegans* reveals multiple inputs driving pairwise chromosome Synapsis during Meiosis. Genetics. 2015;201:1363–79. https://doi.org/10.1534/genetics.115.182279.
15. Thrash A, Hoffmann F, Perkins A. Toward a more holistic method of genome assembly assessment. BMC Bioinf. 2020;21:249. https://doi.org/10.1186/s12859-020-3382-4.
16. Gottstein B, Pozio E, Nöckler K. Epidemiology, diagnosis, treatment, and control of trichinellosis. Clin Microbiolgy Reviews. 2009;22:127–45. https://doi.org/10.1128/CMR.00026-08.
17. Zajac AM. Gastrointestinal nematodes of small ruminants: life cycle, anthelmintics, and diagnosis, Veterinary Clinics: Food Animal Practice. 22 (2006) 529–541. doi: https://doi.org/10.1016/j.cvfa.2006.07.006. PMID: 17071351.
18. Paily KP, Hoti SL, Das PK. A review of the complexity of biology of lymphatic filarial parasites. J Parasitic Dis. 2009;33:3–12. https://doi.org/10.1007/s12639-009-0005-4.
19. Viney M. The genomic basis of nematode parasitism. Brief Funct Genomics. 2017;17:8–14. https://doi.org/10.1093/bfgp/elx010.
20. Viney M. How can we understand the genomic basis of Nematode Parasitism? Trends in Parasitology. 2017;33:444–52. https://doi.org/10.1016/j.pt.2017.01.014.
21. Young ND, Stroehlein AJ, Kinkar L, Wang T, Sohn W-M, Chang BCH, Kaur P, Weisz D, Dudchenko O, Aiden EL, Korhonen PK, Gasser RB. High-quality reference genome for *Clonorchis sinensis*. Genomics. 2021;113:1605–15. https://doi.org/10.1016/j.ygeno.2021.03.001.
22. Nath S, Shaw DE, White MA. Improved contiguity of the threespine stickleback genome using long-read sequencing, G3 (Bethesda). 11 (2021)jkab007. https://doi.org/10.1093/g3journal/jkab007.
23. Learmount J, Conyers C, Hird H, Morgan C, Craig BH, von Samson-Himmelstjerna G, Taylor M. Development and validation of real-time PCR methods for diagnosis of *Teladorsagia circumcincta* and *haemonchus contortus* in sheep. Vet Parasitol. 2009;166:268–74. https://doi.org/10.1016/j.vetpar.2009.08.017.
24. Durand NC, Shamim MS, Machol I, Rao SSP, Huntley MH, Lander ES, Aiden EL. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. Cell Syst. 2016;3:95–8. https://doi.org/10.1016/j.cels.2016.07.002.
25. Roach MJ, Schmidt SA, Borneman AR. Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. BMC Bioinf. 2018;19:460. https://doi.org/10.1186/s12859-018-2485-7.
26. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. Bioinformatics. 2013;29:1072–5. https://doi.org/10.1093/bioinformatics/btt086.
27. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015;31:3210–2. https://doi.org/10.1093/bioinformatics/btv351.
28. Brůna T, Hoff KJ, Lomsadze A, Stanke M, Borodovsky M. BRAKER2: Automatic Eukaryotic Genome Annotation with GeneMark-EP + and AUGUSTUS Supported by a Protein Database, BioRxiv. (2020) 2020.08.10.245134. https://doi.org/10.1101/2020.08.10.245134.
29. Smit AFA, Hubley R. & P. Green RepeatMasker at http://repeatmasker.org
30. Brůna T, Lomsadze A, Borodovsky M. GeneMark-EP+: eukaryotic gene prediction with self-training in the space of genes and proteins. NAR Genomics and Bioinformatics. 2020;2. https://doi.org/10.1093/nargab/lqaa026.
31. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. Nat Methods. 2015;12:59–60. https://doi.org/10.1038/nmeth.3176.
32. Lomsadze A, Ter-Hovhannisyan V, Chernoff YO, Borodovsky M. Gene identification in novel eukaryotic genomes by self-training algorithm. Nucleic Acids Res. 2005;33:6494–506. https://doi.org/10.1093/nar/gki937.

33. Iwata H, Gotoh O. Benchmarking spliced alignment programs including Spaln2, an extended version of Spaln that incorporates additional species-specific features. Nucleic Acids Res. 2012;40:e161. https://doi.org/10.1093/nar/gks708.
34. Gotoh O, Morita M, Nelson DR. Assessment and refinement of eukaryotic gene structure prediction with gene-structure-aware multiple protein sequence alignment. BMC Bioinf. 2014;15:189. https://doi.org/10.1186/1471-2105-15-189.
35. McNeilly TN, Frew D, Burgess STG, Wright H, Bartley DJ, Bartley Y, Nisbet AJ. Niche-specific gene expression in a parasitic nematode; increased expression of immunomodulators in *Teladorsagia circumcincta* larvae derived from host mucosa. Sci Rep. 2017;7:7214. https://doi.org/10.1038/s41598-017-07092-0.
36. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29:15–21. https://doi.org/10.1093/bioinformatics/bts635.
37. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. The sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25:2078–9. https://doi.org/10.1093/bioinformatics/btp352.
38. Araujo FA, Barh D, Silva A, Guimarães L, Ramos RTJ. GO FEAT: a rapid web-based functional annotation tool for genomic and transcriptomic data, Scientific Reports. 8 (2018)1794. https://doi.org/10.1038/s41598-018-20211-9.
39. Paten B, Earl D, Nguyen N, Diekhans M, Zerbino D, Haussler D. Cactus: algorithms for genome multiple sequence alignment. Genome Res. 2011;21:1512–28. https://doi.org/10.1101/gr.123356.111.
40. Rasche H, Hiltemann S, Circos G. User-friendly Circos plots within the Galaxy platform, GigaScience.9(2020). https://doi.org/10.1093/gigascience/giaa065.
41. Krzywinski M, Schein J, Birol İ, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. Circos: an information aesthetic for comparative genomics. Genome Res. 2009;19:1639–45. https://doi.org/10.1101/gr.092759.109.
42. Yu Y, Ouyang Y, Yao W. shinyCircos: an R/Shiny application for interactive creation of Circos plot. Bioinformatics. 2018;34:1229–31. https://doi.org/10.1093/bioinformatics/btx763.
43. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biology. 2019;20:238. https://doi.org/10.1186/s13059-019-1832-y.
44. Xu L, Dong Z, Fang L, Luo Y, Wei Z, Guo H, Zhang G, Gu YQ, Coleman-Derr D, Xia Q, Wang Y. OrthoVenn2: a web server for whole-genome comparison and annotation of orthologous clusters across multiple species. Nucleic Acids Res. 2019;47:W52–8. https://doi.org/10.1093/nar/gkz333.

Publisher's Note