DATA NOTE

The genome sequence of the Australian filarial nematode, *Cercopithifilaria johnstoni* [version 2; peer review: 3 approved]

Kirsty McCann¹, Warwick Grant¹*, Stephen R. Doyle²*

¹Department of Physiology, Anatomy & Microbiology, La Trobe University, Bundoora, Australia
²Parasites & Microbes Programme, Wellcome Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK

* Equal contributors

**Abstract**

We present a genome assembly and annotation of an individual female *Cercopithifilaria johnstoni*, a parasitic filarial nematode that is transmitted by hard ticks (Ixodidae) to infect a broad range of native Australian murid and marsupial hosts. The genome sequence is 76.9 Mbp in length, and although in draft form (N50 = 99 kbp, N50[n] = 232), is largely complete based on universally conserved orthologs (BUSCOs; genome = 94.9%, protein = 96.5%) and relative to other related filarial species. These data represent the first genomic resources for the genus *Cercopithifilaria*, a group of parasites with a broad host range, and form the basis for comparative analysis with the human-infective parasite, *Onchocerca volvulus*, both of which are responsible for similar eye and skin pathologies in their respective hosts.

**Keywords**
Cercopithifilaria johnstoni, filarial nematode, genome assembly, Illumina MiSeq

This article is included in the Wellcome Sanger Institute gateway.

**Open Peer Review**

Invited Reviewers

| Version | Invited Reviewers |
|---------|-------------------|
| version 1 | 1 | 2 | 3 |
| 12 Oct 2021 | report | report | report |
| version 2 | 1 | 2 | 3 |
| (revision) | ☑️ | ☑️ | ☑️ |
| 03 Dec 2021 | report | report | report |

1. James Wasmuth ID, University of Calgary, Calgary, Canada
   University of Calgary, Calgary, Canada

2. Jane Hodgkinson ID, University of Liverpool, Liverpool, UK

3. Neil Young ID, The University of Melbourne, Melbourne, Australia

Any reports and responses or comments on the article can be found at the end of the article.
Corresponding authors: Warwick Grant (W.Grant@latrobe.edu.au), Stephen R. Doyle (stephen.doyle@sanger.ac.uk)

Author roles: **McCann K**: Data Curation, Formal Analysis, Investigation, Writing – Review & Editing; **Grant W**: Conceptualization, Funding Acquisition, Investigation, Project Administration, Supervision, Writing – Review & Editing; **Doyle SR**: Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Investigation, Methodology, Project Administration, Software, Supervision, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This research was supported by a Bill and Melinda Gates Grand Challenges Exploration Grant (WG) [OPP1087697], internal research funding from La Trobe University (WG), and an Illumina MiSeq Grant (SRD). SRD is supported by a UKRI Future Leaders Fellowship [MR/T020733/1] and the Wellcome Trust through core funding to the Wellcome Sanger Institute [206194].

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2021 McCann K et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: McCann K, Grant W and Doyle SR. The genome sequence of the Australian filarial nematode, *Cercopithifilaria johnstoni* [version 2; peer review: 3 approved] Wellcome Open Research 2021, 6:259 https://doi.org/10.12688/wellcomeopenres.17258.2

First published: 12 Oct 2021, 6:259 https://doi.org/10.12688/wellcomeopenres.17258.1
Species taxonomy
Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Protostomia; Ecdysozoa; Nematoda; Chromadorea; Rhabditida; Spirurina; Spiriuromorphida; Filarioidea; Onchocercidae, Cercopithifilaria, Cercopithifilaria johnstoni (taxon ID: 2874296)

Introduction
Cercopithifilaria johnstoni (Mackerras, 1954) is a parasitic filarial nematode transmitted by ixodid ticks to infect a diverse range of native Australian mammalian hosts (Spratt & Haycock, 1988), including monotremes, marsupials, and native rodents. The ability to infect such a broad host range is unusual for a filarial parasite; however, it is yet to be determined if this reflects permissive infectivity and persistence in diverse hosts or cryptic species diversity among morphologically indistinguishable parasites. Over 30 years ago, investigation of C. johnstoni infection of native hosts and experimentally-infected laboratory rats (Rattus norvegicus) revealed that C. johnstoni could cause skin and ocular immunopathologies that appear to be analogous to those seen in humans infected with Onchocerca volvulus (Spratt & Haycock, 1988; Vuong et al., 1993), the causative agent of the neglected tropical disease onchocerciasis. This research prompted the hypothesis that C. johnstoni infection of R. norvegicus could provide an immunologically relevant and experimentally tractable laboratory model of onchocerciasis. Motivated by this hypothesis and progress in the development of C. johnstoni as a laboratory model, we have generated a draft genome assembly and annotation to understand the basic biology of the parasite. These genome data will facilitate the investigation of hypotheses relating to host specificity, provide a resource for comparative analysis between related filarial species, and in particular, be used to characterise the genetic determinants of disease pathology and their relevance to human onchocerciasis.

Genome sequence report
The genome was sequenced from DNA extracted from a single female parasite collected via post-mortem dissection of an Australian bush rat, R. fuscipes (Figure 1a). A total of 24,374,948 300 bp paired-end reads representing ~190-fold coverage of the genome were obtained by Illumina MiSeq sequencing. Trimmed reads (n = 22,065,411) were assembled, which, after contamination (Figure 2) and haplotype removal, resulted in an assembly with a total length of 76.9 Mbp in 2,091 scaffold sequences with a scaffold N50 of 99,003 bp and N50(n) of 232 (Table 1). Compared to other filarial nematodes with assembled genomes, the C. johnstoni assembly ranked 6th of 18 based on both genome contiguity (N50) and completeness (Genome BUSCOs); we note that three assemblies with better genome contiguity and completeness statistics - O. volvulus (Cotton et al., 2016), Brugia malayi (Foster et al., 2020), and Loa loa (prjna246086) (Tallon et al., 2014) - were all assembled using high-throughput sequencing together with one or more long molecule technologies, i.e., long-read PacBio sequencing and optical mapping, to improve contiguity whereas a further two assemblies - L. loa (prjna377577) (Desjardins et al., 2013) and O. flexuosa (prjna230512) - have incorporated long-range mate-pair sequencing libraries for scaffolding. The assembly includes a complete mitochondrial genome for C. johnstoni (contig ID: c_johnstoni mitochondrial_genome), which we used together with other complete mitochondrial genomes of filarial nematodes to demonstrate the phylogenetic placement of C. johnstoni (Figure 3). These data robustly recapitulate the known phylogeny of filarial nematodes and place C. johnstoni within a monophyletic clade with two rodent-infective parasites, Acanthocheilonema vitæae and Litomosoides sigmodontis. Annotation of the C. johnstoni genome identified 10,565 genes and 11,690 transcripts, broadly consistent with the number of reported annotation features for other filarial nematodes (Table 1; range = 8,140-16,203 for both gene and transcript features). Similar to the genome statistics described above, the annotation of the predicted proteome is also highly resolved, with 96.5% complete BUSCOs identified (Table 1). These data demonstrate the utility of using a large collection of diverse metazoan proteins to guide the annotation of a genome in the absence of species-specific data, for example, RNA-seq.

The immunopathology of O. volvulus infection is hypothesised to be driven by the recognition of immunoreactive proteins of Wolbachia (Saint André et al., 2002), a species of intracellular bacteria found in several filarial nematodes species (Figure 3; closed circles) where it is thought to play a symbiotic role in host metabolism and/or reproduction (Taylor et al., 2005). The similar pathologies caused by C. johnstoni infection of rats and O. volvulus infection of humans prompted us to examine the presence of Wolbachia in our C. johnstoni assembly. Analysis of raw sequencing reads revealed only 0.1% of C. johnstoni reads classified as bacterial, with only a single read matching Wolbachia in our custom Kraken database; for context, analysis of O. volvulus raw sequencing reads against the same database revealed, on average, 1.98% of reads were derived from Wolbachia (n = 32 O. volvulus whole-genome sequencing datasets (Choi et al., 2016); range = 0.08 - 13.26%; average library size = 34 million reads), which is consistent with previous estimates based on mapped reads to the O. volvulus nuclear and Wolbachia genomes (Armoo et al., 2017). Alignment of C. johnstoni protein-coding sequences to a diverse collection of Wolbachia reference genomes (Lefoulon et al., 2020) revealed 18 candidates; only two proteins,
Figure 1. (a) The bush rat, *Rattus fuscipes*, is one of several host species infected by and from which *Cercopithifilaria johnstoni* used in this study were collected (photo: K. McCann). (b) Sampling site (yellow point) from which bush rats were collected in the Mogo State Forest near Mogo, NSW, Australia.

Figure 2. Decontamination screen using BlobTools. The plot shows variation in GC (guanine+cytosine) content (x-axis), mapped read coverage (y-axis), and blast-classification (colours, see key above) of the assembly scaffolds, from which putative contaminants are commonly identified as outliers of the distributions.
## Table 1. Genome assembly statistics of *Cercopithifilaria johnstoni* and related Clade III filarial nematodes.

| Species/ID | Assembly length (bp) | Sequences (n) | N50 length (bp) | N50 (n) | Genes/transcripts (n) | Protein BUSCOs (%) | Genome BUSCOs (%) |
|------------|----------------------|---------------|-----------------|----------|----------------------|-----------------------|------------------|
| *C. johnstoni* (current study) | 76,938,880 | 2092 | 99,003 | 37 | 10565 | C:94.9 [S:94.2, D:0.7] | 2 |
| *A. viteae* (prjeb1697) | 77,350,906 | 6796 | 25,808 | 300 | 11690 | C:90.5 [S:88.9, D:1.6] | 2 |
| *B. malayi* (prjna10729) | 88,235,797 | 197 | 142,147 | 2 | 12,928 | C:97.6 [S:96.5, D:0.6] | 2 |
| *B. pahangi* (prjeb497) | 90,545,113 | 14029 | 65,666 | 300 | 16,904 | C:89.8 [S:88.7, D:1.1] | 2 |
| *B. timori* (prjeb4663) | 64,930,714 | 32963 | 4,919 | 874 | 16,203 | C:54.9 [S:54.6, D:0.3] | 2 |
| *D. immitis* (prjeb1797) | 88,309,529 | 16061 | 71,281 | 37 | 12,857 | C:92.0 [S:89.8, D:0.2] | 2 |
| *E. elaphi* (prjeb502) | 82,568,297 | 8078 | 25,590 | 377 | 10,410 | C:77.6 [S:77.3, D:0.3] | 2 |
| *L. sigmodontis* (prjeb3075) | 64,813,410 | 3165 | 8,078 | 2 | 10,246 | C:92.5 [S:90.6, D:1.9] | 2 |
| *L. loa* (prjna246086) | 96,405,338 | 2250 | 180,288 | 377 | 12,473 | C:97.6 [S:96.3, D:1.3] | 2 |
| *L. loa* (prjna37757) | 91,373,458 | 5773 | 114,388 | 377 | 14,908 | C:96.4 [S:95.7, D:0.7] | 2 |
| *O. flexuosa* (prjeb512) | 81,588,927 | 5250 | 25,590 | 377 | 14,908 | C:48.4 [S:48.2, D:0.2] | 2 |
| *O. ochengi* (prjna1465) | 94,405,338 | 91,373,458 | 2250 | 180,288 | 12,473 | C:96.3 [S:96.3, D:1.3] | 2 |
| *O. ochengi* (prjeb1465) | 91,660,559 | 8573 | 174,388 | 377 | 14,908 | C:96.3 [S:96.3, D:1.3] | 2 |
| *O. volvulus* (prjeb513) | 96,427,137 | 2583 | 25,485,961 | 2 | 12,186 | C:97.7 [S:97.4, D:0.3] | 2 |
| *S. digitata* (prjeb1697) | 78,770,088 | 1283 | 121,247 | 2 | 12,186 | C:97.7 [S:97.4, D:0.3] | 2 |
| *W. bancrofti* (prjeb536) | 76,991,470 | 1350 | 9,917 | 22 | 13,990 | C:97.7 [S:97.4, D:0.3] | 2 |
| *W. bancrofti* (prjna275548) | 90,325,107 | 5105 | 56,670 | 22 | 13,990 | C:97.7 [S:97.4, D:0.3] | 2 |

1 WormBase Parasite release 16 (*Howe et al., 2017*).
2 BUSCOs: C: complete, S: complete, single copy, D: fragmented, M: missing.

**Genome assembly statistics** of *Cercopithifilaria johnstoni* and related Clade III filarial nematodes. **Sequences (n)**: number of sequences. **Assembly length (bp)**: total assembly length in base pairs. **N50 length (bp)**: length of the largest contig. **N50 (n)**: number of sequences included in N50. **Genes/transcripts (n)**: number of genes/transcripts. **Protein BUSCOs (%)**: percent of protein-wide BUSCOs. **Genome BUSCOs (%)**: percent of genome-wide BUSCOs.
Figure 3. Phylogenetic placement of Cercopithifilaria johnstoni among related filarial nematodes. A maximum-likelihood tree is shown, generated from whole mitochondrial DNA alignments. Node labels represent bootstraps from 1000 replicates. The presence or absence of Wolbachia in each species is indicated by the closed and open circles; for Wolbachia-positive species, the Wolbachia supergroup is indicated (C, D, F) (Ferri et al., 2011; Gerth et al., 2014; McNulty et al., 2012).

Methods
Sample collection
As part of a larger program of fieldwork to investigate natural transmission of C. johnstoni in a wild, free-ranging population of Australian bush rats Rattus fuscipes (Figure 1a), 8 naturally infected bush rats were transferred from the site of collection in the Mogo State Forest, N.S.W., Australia (GPS coordinates: -35.7689484, 150.1027441; Figure 1b) to the La Trobe University Animal Research Facility in Bundoora, Vic., Australia (permits: AEC 13-23, NSW – Scientific Licence 5L 101280, VIC – Scientific Permit 10007169).

All efforts were made to ameliorate any suffering of animals through providing large cages and keeping their habitat and diet as close as possible to that of the wild. The study was also closely monitored by the facility veterinarian. The rats were housed singly in large plastic tubs approximately 0.5 m x 1 m square and 1 m deep, with a hinged mesh lid. The tubs were filled with leaf litter and contained small hollow logs for refuge. Rats were fed a mix of standard rat diet supplemented with meal worms. The adult parasite that was sequenced was recovered post-mortem from a single female rat who was euthanised by CO₂ asphyxia on advice of the facility veterinarian following a short illness of unknown origin.
DNA extraction, library preparation, and sequencing
A single adult female worm (approximately 7 cm in length) was cut into approximately 1 cm length pieces using a sterile scalpel blade before being placed in a lysis solution (lysis buffer and proteinase K solution) for 18 h. Genomic DNA from the worm lysate was extracted using an ISOLATE II Genomic DNA Kit (Bioline, Australia) following the manufacturer’s instructions, except for the following modification: the sample was eluted from the extraction column in 50 µl of extraction buffer, which was passed back through the extraction column a second time to collect additional DNA remaining on the column before further analysis.

Genomic DNA (500 ng in 50 µl) was sheared before sequencing library preparation using a Covaris S220 Focused-ultrasonicator with the following settings optimised for generating fragments approximately 400-600 bp: Peak incidence power = 175 W; Duty factor = 5%; cycles per burst = 200; treatment time = 55 s. A DNA sequencing library was prepared from 500 ng DNA using a NEBNext Ultra Library Prep Kit for Illumina, following the manufacturer’s instructions. The resulting library was run on a 2% agarose gel, from which a gel cut was made to extract the 500-700 bp fragment fraction, which was subsequently purified using a Promega Gel and PCR clean-up kit (Promega, Australia).

Before assembly, raw sequencing reads were first visualised for quality and inherent bias using FastQC version 0.11.9. Reads were adapted and quality trimmed using Trimmomatic version 0.32 (Bolger et al., 2014) (CROP:150 SLIDINGWINDOW:10:20 MINLEN:100), after which 22,065,411 paired-end reads were retained for assembly. Genome size was estimated from the trimmed reads using GenomeScope 2.0 (Ranallo-Benavidez et al., 2020), which predicted a length of 63.24 Mbp.

De novo genome assembly was performed using SPAdes version 3.10.1 (Prijibelski et al., 2020) using default parameters. The raw assembly was decontaminated, first using Redundans (Pryszcz & Gabaldón, 2016) to remove additional haplotypes present in the assembly, followed by BlobTools (Laetsch & Blaxter, 2017) to identify putative bacterial and host contamination present in the assembly (Figure 2). Only scaffolds containing hits to “Nematoda” or “no-hit” (the origin of these sequences is unclear but could potentially be novel nematode sequences) and with a mapped average read depth of 10 or greater were retained. The decontaminated assembly was further scaffolded using OPERA-LG (Gao et al., 2016) to encourage unique joins that could not be previously made due to alternative haplotypes present, followed by a second-round using Redundans to fill gaps. The iterative improvements to the assembly are documented in Table 2, demonstrating improved contiguity while maintaining and recovering conserved BUSCOs.

The mitochondrial genome was assembled independently of the nuclear genome. Briefly, mitochondrially-derived sequencing reads were identified by mapping all trimmed reads to mitochondrial genomes of Onchocerca volvulus (NC_001861.1), Acanthocheilonema viteae (HQ186249.1), Brugia malayi (NC_004298.1), Dirofilaria immitis (AJ537512.1), Litomosoides sigmodontis (AP017689.1), Loa loa (HQ186250.1), Onchocerca ochengi (KX181290.2), and Wuchereria bancrofti (HQ184469.1). Reads that mapped were then de novo assembled using Velvet version 1.2.10 (Zerbino & Birney, 2008) using default parameters, with kmer=99 identified as optimal

### Table 2. Iterative improvement of the Cercopithifilaria johnstoni genome assembly.

| Assembly statistics | Spades | Spades + Redundans | Spades + Redundans + Blobtools | Spades + Redundans + Blobtools + OPERA-LG | Spades + Redundans + Blobtools + OPERA-LG + gap filling (Redundans) |
|---------------------|--------|--------------------|-------------------------------|---------------------------------------------|-------------------------------------------------------------|
| Assembly size (bp)  | 79,062,707 | 77,312,925 | 77,015,453 | 77,032,887 | 76,924,992 |
| Sequences (n)       | 7,152 | 3,117 | 2,568 | 2,263 | 2,091 |
| N50 (bp)            | 88,758 | 91,012 | 91,596 | 99,003 | 99,003 |
| N50 (n)             | 263 | 253 | 252 | 232 | 232 |
| Average length (bp) | 11,054.63 | 24,803.63 | 29,990.44 | 34,040.16 | 36,788.61 |
| Largest scaffold (bp) | 588,165 | 588,165 | 588,165 | 588,165 | 588,166 |
| Ns (bp)             | 56,933 | 56,921 | 56,921 | 74,355 | 3,888 |
| Gaps (n)            | 299 | 298 | 298 | 603 | 414 |
| Genome BUSCOs (n=982)      | Spades     | Spades + Redundans | Spades + Redundans + Blobtools | Spades + Redundans + Blobtools + OPERA-LG | Spades + Redundans + Blobtools + OPERA-LG + gap filling (Redundans) |
|---------------------------|------------|--------------------|--------------------------------|---------------------------------------------|-------------------------------------------------------------------|
| Complete                  | 929 (94.6%)| 930 (94.7%)        | 930 (94.7%)                    | 930 (94.7%)                                 | 932 (94.9%)                                                       |
| Complete, single          | 922 (93.9%)| 923 (94%)          | 923 (94%)                      | 923 (94%)                                   | 925 (94.2%)                                                      |
| Complete, duplicate       | 7 (0.7%)   | 7 (0.7%)           | 7 (0.7%)                       | 7 (0.7%)                                    | 7 (0.7%)                                                         |
| Fragmented                | 40 (4.1%)  | 39 (4.0%)          | 39 (4.0%)                      | 40 (4.1%)                                   | 38 (3.9%)                                                        |
| Missing                   | 13 (1.3%)  | 13 (1.3%)          | 13 (1.3%)                      | 12 (1.2%)                                   | 12 (1.2%)                                                        |

using Velvet-optimiser version 2.2.5. Velvet was unsuccessful in producing a closed mtDNA genome, so an iterative mapping and joining approach was used to manually curate the assembly, resulting in a complete single contig of 13,716 bp. Validation of the assembly was performed by multiple sequence alignment to available filarial mtDNA genomes above using Mesquite version 3.04 (Maddison & Maddison, 2019) and visualised in progressiveMauve (20150213) (Darling et al., 2010).

**Genome annotation**

The mtDNA genome sequence was initially annotated using MITOS (Bernt et al., 2013). The *C. johnstoni* annotation was improved manually by comparing sequence alignments and GFF3 annotation files from *C. johnstoni* with the closely related filarial nematodes *L. loa*, *D. immitis*, *A. viteae*, *B. malayi*, *O. ochengi*, *O. volvulus*, *W. bancrofti*.

The nuclear genome assembly was annotated using Braker v2 (Brúna et al., 2021). As no RNA-seq data were available, we generated hints (predicted introns, start and stop codons) for Braker using the ProtHin pipeline; spliced alignments were generated by mapping proteins from OrthoDB Metazoan protein database, from which evidence (prothint_augustus.gff) was used as an input to Braker.

Annotation statistics were determined using GAG (Geib et al., 2018).

The final GFF containing both nuclear and mitochondrial genome annotations was converted to EMBL format for submission to ENA using EMBLmyGFF3 (Norling et al., 2018).

**Genome and annotation completeness**

Genome and annotation completeness was estimated using BUSCO (Benchmarking Universal Single-Copy Orthologues) version 4 (Seppey et al., 2019) with lineage set to nematode_odb9 and mode set to “genome” or “protein” for the assembly or protein-coding genes, respectively, using “Caenorhabditis” as a training species for gene identification. Comparative genome assembly statistics were generated using assembly-stats version 1.0.1. All genomic and proteomic data from available assemblies of related filarial nematode species were obtained from WormBase ParaSite release 16 (Howe et al., 2017).

**Phylogenetic analysis**

Phylogenetic placement of *C. johnstoni* was performed by comparing its assembled mitochondrial genome to publicly available mitochondrial genomes of filarial nematodes. Mitochondrial genomes from the following species were downloaded from NCBI: *A. viteae* (accession number: HQ186249.1), *B. malayi* (NC_004298.1), *B. pahangi* (CM022469.1), *B. timori* (AP017686.1), *Chandlerella quiscalii* (NC_014486.1), *D. immitis* (NC_005305.1), *D. repens* (NC_029975.1), *Gongylonema pulchrum* (NC_026687.1), *L. loa* (HQ186250.1), *L. sigmodontis* (AP017689.1), *Mansonella perstans* (MT361687.1), *O. flexuosa* (NC_016172.1), *O. ochengi* (NC_031891.2), *O. volvulus* (NC_001861.1), *Setaria digitata* (NC_014282.1), *Spirocerca lupi* (NC_021135.1), *Thelazia callipaeda* (NC_018363.1), and *W. bancrofti* (NC_016186.1).

Whole mitochondrial genome alignment was performed using MAFFT version 7.480 (Katoh & Standley, 2013) (--globalpair --maxiterate 16), from which a maximum likelihood phylogeny was estimated using IQ-TREE version 2.1.2 (Minh et al., 2020) under the GTR+F+R4 model and 1,000 bootstrap replicates to estimate support for bipartitions. The genomes of *S. digitata*, *T. callipaeda*, *G. pulchrum*, and *S. lupi* were used as outgroups. The resulting tree was visualised using the ggtree package (Yu et al., 2017) in R.

**Wolbachia analyses**

The presence of Wolbachia was assessed in three ways. First, raw sequencing reads were assessed using Kraken2 (--db: minikraken_20141208 raw sequencing reads were assessed using Kraken2 (--db: minikraken_20141208) against an in-house database consisting of the publicly available “minikraken_20141208” database (--db: minikraken_20141208). Wolbachia analyses were performed using Mafft version 7.480 (Katoh & Standley, 2013) (--globalpair --maxiterate 16), from which a maximum likelihood phylogeny was estimated using IQ-TREE version 2.1.2 (Minh et al., 2020) under the GTR+F+R4 model and 1,000 bootstrap replicates to estimate support for bipartitions. The genomes of *S. digitata*, *T. callipaeda*, *G. pulchrum*, and *S. lupi* were used as outgroups. The resulting tree was visualised using the ggtree package (Yu et al., 2017) in R.
and wTpre (NZ_CM003641). Second, all protein-coding sequences derived from the genome annotation were aligned against the above-mentioned Wolbachia genomes using exonerate 2.4.0 (Slater & Birney, 2005), from which hits were queried using BLASTP. Finally, the relative proportion of Wolbachia genome sequence matches to the raw, unfiltered C. johnstoni assembly ("Spades" assembly in Table 2) was quantified using PROmer version 3.07 (Kurtz et al., 2004).

The analysis code used in this study is available from GitHub and is archived with Zenodo (Doyle & McKann, 2021).

### Data availability

#### Genomic resources

European Nucleotide Archive: Raw sequence data, genome and annotation are deposited in the ENA. Accession number PRJEB47283; https://identifiers.org/ena.embl:PRJEB47283.

The assembly will also be made available at WormBase ParaSite (https://parasite.wormbase.org/), the primary repository for helminth genomes and annotations.

---

### Analysis code

Analysis code is available from: https://github.com/stephenrdoyle/cercopithifilaria_johnstoni.

### License

BSD 3-Clause “New” or “Revised” License

### Acknowledgements

We thank Will Ritchie for assistance in establishing the project, Jacqueline Orián and Phuc Dang for help with the animal dissections to recover parasites, the Grant Lab for valuable discussions throughout the project, and the Pathogen Informatics team (Wellcome Sanger Institute) for informatic assistance. For the purpose of Open Access, the authors have applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

---

### References

Armos S, Doyle SR, Osei-Atweneboana MY, et al.: Significant heterogeneity in Wolbachia copy number within and between populations of Onchocerca volvulus. Parasit Vectors. 2017;10(1):188.

Bernt M, Donath A, Jühling F, et al.: MITOS: improved de novo metazoan mitochondrial genome annotation. Mol Phylogenet Evol. 2013;69(2):313–9.

Bolger AM, Lohse M, Usadel B: Trimmmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114–20.

Brüna T, Hoff KJ, Lomsadze A, et al.: BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. NAR Genom Bioinform. 2021;8(1):iqaa108.

Choi YJ, Yagi R, McNulty SN, et al.: Genomic diversity in Onchocerca volvulus and its Wolbachia endosymbiont. Nat Microbiol. 2016;2:16207.

Cotton JA, Bennuru S, Grote A, et al.: The genome of Onchocerca volvulus, agent of river blindness. Nat Microbiol. 2016;2:16216.

Darling AE, Mau B, Perna NT: progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. PLoS One. 2010;5(6):e11147.

Desjardins CA, Cerqueira GC, Goldberg JM, et al.: Genomics of Loa loa, a Wolbachia-free filarial parasite of humans. Nat Genet. 2013;45(5):495–500.

Doye S, McCann K: stephenrdoyle/cercopithifilaria_johnstoni: (v1.1). Zenodo. http://www.dois.org.org/10.5281/zenodo.5746893

Ferri E, Bain O, Barbuto M, et al.: New Insights into the Evolution of Wolbachia Infections in Filarial Nemathodes Inferred from a Large Range of Screened Species. PLoS One. 2011;6(6):e20843.

Foster JM, Grote A, Matichak J, et al.: Sex chromosome evolution in parasitic nematodes of humans. Nat Commun. 2020;11(1):1964.

Geib SM, Hall B, Derego T, et al.: Genome Annotation Generator: a simple tool for generating and correcting WGS annotation tables for NCBI submission. Gigascience. 2018;7(4):1-5.

German M, Gansauge MT, Weigert A, et al.: Phylogenomic analyses uncover origin and spread of the Wolbachia pandemic. Nat. Commun. 2014;5:5117.

Howe KL, Bolt BJ, Shafie M, et al.: WormBase ParaSite—a comprehensive resource for helminth genomics. Mol Biochem Parasitol. 2017;215:2-10.

Kurtz S, Phillippy A, Delcher AL, et al.: Versatile and open software for comparing large genomes. Genome Biol. 2004;5(2):R12.

Laetsch DR, Blaxter ML: BlobTools: Interrogation of genome assembles (version 1: peer review: 2 approved with reservations). F1000Res. 2017;6:1287.

Lefoulon E, Clark LW, Guerrera R, et al.: Diminutive, degraded but dissimilar: Wolbachia genomes from filarial nematodes do not conform to a single paradigm. Microb Genom. 2020;6(2):mgn000487.

Mackerras M: Two new species of Dipetalonema (Nematoda: Filarioidea) from Australian marsupials. (Proceedings of the Royal Society of Queensland). 1954.

Maddison W, Maddison D: MESQUITE: a modular system for evolutionary analysis. 2019.

McNulty SN, Mullin AS, Vaughan JA, et al.: Comparing the mitochondrial genomes of Wolbachia-dependent and independent filarial nematode species. BMC Genomics. 2012;13:145.

McNulty Abstract | Publisher Full Text | Free Full Text

Minh BQ, Schmidt HA, Chernomor O, et al.: IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Mol Biol Evol. 2020;37(5):1500-1534.

Nerling M, Jareborg N, Dainat J: EMBLmyGFF3: a converter facilitating genome annotation submission to European Nucleotide Archive. BMC Res...
Open Peer Review

Current Peer Review Status: ✔ ✔ ✔

Version 2

Reviewer Report 06 December 2021

https://doi.org/10.21956/wellcomeopenres.19331.r47422

© 2021 Wasmuth J. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

James Wasmuth ID

1 Department of Ecosystem and Public Health, Faculty of Veterinary Medicine, University of Calgary, Calgary, Canada
2 Host-Parasite Interactions (HPI) Research Training Network, University of Calgary, Calgary, Canada

I thank the authors for their careful consideration of my comments. The additional information they provide in the publication easily satisfies my earlier concerns.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Genome informatics, nematode evolution

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 06 December 2021

https://doi.org/10.21956/wellcomeopenres.19331.r47424

© 2021 Young N. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Neil Young ID

Department of Veterinary Biosciences, Melbourne Veterinary School, The University of Melbourne, Melbourne, Vic, Australia

Thanks for the comprehensive response to the comments. I believe the version 2 of the manuscript addresses the aims/objectives clearly. It also provides more detail on the position of
this nematode to guide the use of this resource for future comparative genomic and genetic studies.

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Parasite genomics and genetics.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.
Yes

Are the protocols appropriate and is the work technically sound?
Yes

Are sufficient details of methods and materials provided to allow replication by others?
Yes

Are the datasets clearly presented in a useable and accessible format?
Yes

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Parasite genomics and genetics.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 01 Dec 2021

Stephen Doyle,

Reviewer 3 – Neil Young

The article by McKann *et al.* reports an annotated draft genome of Cercopithifilaria johnstoni, a filarial nematode parasitising Australian mammals. Despite the fragmented nature of the assembly, the genome is of significant value to the research community because of its taxonomic position, its adaptive evolution to parasitise marsupials and its potential use as a laboratory model system for onchocerchiasis. Combined fundamental and applied applications make it a valuable nematode genomic resource.

I was surprised to see genome completeness scores over 94% with an assembly using only short-read sequence data. Could this relate to short introns with less repetitive elements?

*Response:* *We were also pleasantly surprised by the relatively high genome completeness statistics for an Illumina-only assembly. We attribute it in part to the low diversity of sequencing a freshly-collected individual parasite at high coverage with relatively long Illumina reads (3x300 bp). The genome at 76 Mb is on the smaller side compared to other nematodes, and while its overall repetitive content was not noticeably different from other filarial nematodes, it must compact its genome content into a smaller space suggesting it is “less complex” to some degree.*

*Regarding the intron lengths - we do find that C. johnstoni introns are shorter than O. volvulus introns (See table below on intron length stats). To what extent this is due to assembly contiguity (i.e., a technical effect - the O. volvulus genome is assembled into chromosomal-scale scaffolds and so provides a more robust framework for longer gene models), or true biological differences, is difficult to determine without having a more*
As this parasite is a useful model for onchocerciasis, it would be good to show the completeness of gene models specific to parasite-host interactions in Onchocerca and related species.

Response: We provide an estimate of the overall geneset completeness, as indicated by the BUSCO scores for the genome and proteome in Table 1. Arguably, these data show that the genome and proteome are highly representative based on conserved orthologs “expected” to be present, and relative to closely related species. We agree that to further establish C. johnstoni as a model for onchocerciasis, a better understanding of the genes involved in host-parasite interactions is needed; these data are in fact the focus of a separate follow up publication. As a Wellcome Open Research Data Note aims to focus specifically on the data themselves and “not... analyses or conclusions”, we initially (and now again, subsequently after peer review) decided against presenting these downstream analyses of the genome resources.

The mt genome was also assembled and annotated but there was no description of the mt genome herein. A mt phylogenomic tree might provide the reader with more context of the taxonomic position of this parasite.

Response: The reviewer is correct – we did not specifically describe the mitochondrial genome. However, we agree that a phylogeny using the mitochondrial genome would illustrate where C. johnstoni is placed relative to other filarial species.

To address this comment, we now include this phylogeny in Figure 3.

Did the lack of RNAseq data affect gene model predictions? If not, then the findings herein would be strong support for relying only on amino acid sequence homology for training ab initio gene predictors. It would simplify efforts to complete the genome annotations for some taxa.

Response: This is a difficult question to respond to specifically, given we didn't generate RNA-seq data for C. johnstoni. However, the BUSCO predictions were respectable given we did use amino acid homology from a broad range of metazoan species.

To explore this idea as a purely academic exercise, we reannotated the genomes of a closely related but less-well annotated species (Acanthocheilonema viteae) and a species with a high-quality genome / good annotation (Onchocerca volvulus). The predicted proteins inferred from these new annotations were assessed using BUSCO. We note that both A. viteae and O. volvulus are not (yet) included in the OrthoDB database from which the metazoan proteins used to generate the hints were derived; thus, the de novo
Annotations described below are not biased by pre-existing species-specific protein models.

**A. viteae**
- Genes:
  - original: 10,397
  - de novo: 12,056
- BUSCOS:
  - original: C: 88.3 [S:86.5, D:1.8], F:8.8, M:2.9
  - de novo: C: 92.4 [S:82.6, D:9.8], F:5.6, M:2.0
- Conclusion: New models are more complete, less fragmented, fewer missing, however, there is a higher duplication rate, which may relate to new alternative transcripts present.

**O. volvulus**
- Genes
  - original: 12,109
  - de novo: 12473
- BUSCOs:
  - original: C:99.2 [S:98.3, D:0.9], F:0.8, M:0.0
  - de novo: C:98.8 [S:88.9, D:9.9], F:1.0, M:0.2
- Conclusion: Minor differences overall - slightly fewer complete models, slightly more fragmented, and slightly more missing. Higher duplication rates. However, O. volvulus is outstanding already, and the de novo annotation is still very good overall.

These results suggest, based on a relatively simple metric of the proportion of conserved orthologous genes, that using a large collection of diverse metazoan proteins as hints for Braker2 can improve existing annotations and does a respectable job when compared with a well-curated genome annotation. Therefore, it is likely that this represents a valid approach for annotation of genomes from species where collecting additional species-species evidence, ie, RNA-seq, is difficult. This needs further testing, which is outside the scope of this work.

This validation exercise of the approach used to annotate the C. johnstoni genome as we describe provides further support for the high BUSCO scores we report and completeness of the C. johnstoni genome and annotation.

**Competing Interests:** No competing interests were disclosed.
Jane Hodgkinson
Veterinary Parasitology, Institute of Infection, Veterinary and Ecological Sciences, University of Liverpool, Liverpool, UK

The authors present a genome for the nematode *Cercopithifilaria johnstoni*, a parasite of considerable interest in its own right and as a comparator for other filarial nematodes.

In my opinion all the methodologies are appropriate and every attempt has been made to produce a genome assembly of the best quality with the available sequence data. Table 1 clearly identifies that the quality of the assembly of *Cercopithifilaria johnstoni* as presented, is comparable with the quality of published genomes for other filaria; indeed it is towards the top end (6/18) in terms of completeness and contiguity.

I have no reservations in recommended this manuscript for indexing in its current form.

**Is the rationale for creating the dataset(s) clearly described?**
Yes

**Are the protocols appropriate and is the work technically sound?**
Yes

**Are sufficient details of methods and materials provided to allow replication by others?**
Yes

**Are the datasets clearly presented in a useable and accessible format?**
Yes

*Competing Interests:* No competing interests were disclosed.

*Reviewer Expertise:* Molecular helminthology, anthelmintic resistance

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 01 Dec 2021

Stephen Doyle,

*Response: We are grateful for the positive appraisal of our work.*

*Competing Interests:* No competing interests were disclosed.

Reviewer Report 26 October 2021

https://doi.org/10.21956/wellcomeopenres.19074.r46394
James Wasmuth

1 Department of Ecosystem and Public Health, Faculty of Veterinary Medicine, University of Calgary, Calgary, Canada
2 Host-Parasite Interactions (HPI) Research Training Network, University of Calgary, Calgary, Canada

The authors have done an excellent job in describing the sequencing, assembly and annotation of the genome of a parasitic nematode, whose broad host range recommends it for understanding how host-parasite interactions evolve. While the sequencing is only short-read (300 bp), the depth of sequencing and careful assembly gives us confidence in the gene models. The level of detail in the methods should be considered the new standard of reporting. I enjoyed the data in table 2, which demonstrates the value of careful assembly. I have three requests in any future version:

1. It would be helpful to know the phylogenetic placement of *C. johnstoni* in the filarial nematodes from this paper. Perhaps Table 1 could include the phylogenetic relationships.

2a. In the search for *Wolbachia* in *C. johnstoni* the authors found that 0.02% of reads mapped to Rickettsiales. What is the % of reads from the *O. volvulus* sequencing project that maps to Rickettsiales? This comparison is necessary as this report will be cited as evidence of *Wolbachia* loss.

2b. For the other *Wolbachia* searches, it is unclear to me if the authors used the assembled contigs before or after blobtools decontamination. If after, it is not surprising that there is so little evidence of matches.

Is the rationale for creating the dataset(s) clearly described?
Yes

Are the protocols appropriate and is the work technically sound?
Yes

Are sufficient details of methods and materials provided to allow replication by others?
Yes

Are the datasets clearly presented in a useable and accessible format?
Yes

*Competing Interests*: No competing interests were disclosed.

*Reviewer Expertise*: Genome informatics, nematode evolution

I confirm that I have read this submission and believe that I have an appropriate level of
expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 01 Dec 2021

Stephen Doyle,

Reviewer 1 – James Wasmuth

The authors have done an excellent job in describing the sequencing, assembly and annotation of the genome of a parasitic nematode, whose broad host range recommends it for understanding how host-parasite interactions evolve. While the sequencing is only short-read (300 bp), the depth of sequencing and careful assembly gives us confidence in the gene models. The level of detail in the methods should be considered the new standard of reporting. I enjoyed the data in table 2, which demonstrates the value of careful assembly.

I have three requests in any future version:

1. It would be helpful to know the phylogenetic placement of C. johnstoni in the filarial nematodes from this paper. Perhaps Table 1 could include the phylogenetic relationships.

Response: We agree that this is would be useful to show.

We have now included a phylogeny of filarial species based on whole mitochondrial genome alignment as Figure 3.

2a. In the search for Wolbachia in C. johnstoni the authors found that 0.02% of reads mapped to Rickettsiales. What is the % of reads from the O. volvulus sequencing project that maps to Rickettsiales? This comparison is necessary as this report will be cited as evidence of Wolbachia loss.

Response: This is a really good question and one that we had not originally asked.

To address this comment, we determined the proportion of reads classified as Wolbachia from 32 O. volvulus whole-genome sequencing datasets described in Choi et al. 2016 (https://doi.org/10.1038/nmicrobiol.2016.207). We also improved the sensitivity of analysis using a new custom kraken database with all known filarial Wolbachia genomes added to it, including the Wolbachia genome from O. volvulus. This was important, as an initial analysis of a single O. volvulus read set performed poorly with the original kraken database used.

The new analysis revealed only a single C. johnstoni sequencing read classified as Wolbachia (rather than Rickettsiales as we reported originally), whereas, on average, 1.98% of O. volvulus reads classified as Wolbachia (range: 0.08-13.26%, average library size = 34 million reads). Considering the O. volvulus Wolbachia genome is ~1 Mb and the nuclear genome ~100 Mb, it suggests there are ~ 2 Wolbachia genomes for every nuclear genome, which is within the range we have observed previously estimated from mapped reads to
the nuclear and Wolbachia genomes (see https://doi.org/10.1186/s13071-017-2126-4).

These new results are now included in the manuscript.

2b. For the other Wolbachia searches, it is unclear to me if the authors used the assembled contigs before or after blobtools decontamination. If after, it is not surprising that there is so little evidence of matches.

Response: Well observed. We originally analysed matches between Wolbachia and the assembled genomes pre- and post-decontamination, but in the end, only reported the analysis of the final genome assembly.

To address this comment, we now report the analysis performed on the Spades assembly prior to blobtools processing; encouragingly, we find only 1.4% of Wolbachia matches in the Spades-only assembly, which is consistent with the 1.38% we originally reported using the decontaminated genome.

Collectively, these results strengthen our argument that C. johnstoni does not harbour Wolbachia.

Competing Interests: No competing interests were disclosed.