



DATA NOTE

The genome sequence of *Heligmosomum mixtum* Schulz, 1929 (Rhabditida: Heligmosomidae)

[version 1; peer review: 3 approved]

Abosede Olarewaju ¹, Agnieszka Kloch¹, Lewis Stevens², Manuela Kieninger², Michael Paulini²,

Wellcome Sanger Institute Tree of Life Management, Samples and Laboratory team,

Wellcome Sanger Institute Scientific Operations: Sequencing Operations,

Wellcome Sanger Institute Tree of Life Core Informatics team,

Tree of Life Core Informatics collective, Mark L. Blaxter ²

¹Faculty of Biology University of Warsaw, Warsaw, Poland

²Tree of Life Programme, Wellcome Sanger Institute, Hinxton, England, UK

V1 First published: 17 Apr 2026, 11:225
<https://doi.org/10.12688/wellcomeopenres.26318.1>

Latest published: 17 Apr 2026, 11:225
<https://doi.org/10.12688/wellcomeopenres.26318.1>

Abstract

We present a genome assembly of *Heligmosomum mixtum* (Nematoda; Chromadorea; Rhabditida; Heligmosomidae). The assembly consists of two haplotypes with total lengths of 741.91 megabases and 754.62 megabases. Most of haplotype 1 (98.33%) is scaffolded into 6 chromosomal pseudomolecules, including the X sex chromosome. Haplotype 2 was assembled to scaffold level. The mitochondrial genome has been assembled, with a length of 13.53 kilobases.

Keywords

Heligmosomum mixtum, genome sequence, chromosomal, Rhabditida



This article is included in the [Tree of Life gateway](#).

Open Peer Review

Approval Status

	1	2	3
version 1 17 Apr 2026	 view	 view	 view

1. **Erich M. Schwarz** , Cornell University, Ithaca, USA
2. **James Wasmuth** , University of Calgary, Calgary, Canada
3. **Jun Kim** , Chungnam National University, Daejeon, South Korea

Any reports and responses or comments on the article can be found at the end of the article.

Corresponding author: Lewis Stevens (lewis.stevens@sanger.ac.uk)

Author roles: **Olarewaju A:** Investigation, Methodology, Resources, Writing – Original Draft Preparation, Writing – Review & Editing; **Kloch A:** Conceptualization, Funding Acquisition, Project Administration, Supervision, Writing – Review & Editing; **Stevens L:** Conceptualization, Investigation, Methodology, Project Administration, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Kieninger M:** Investigation, Methodology, Writing – Review & Editing; **Paulini M:** Investigation, Methodology, Writing – Review & Editing; **Blaxter ML:** Supervision

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by Wellcome through core funding to the Wellcome Sanger Institute (220540). *The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*

Copyright: © 2026 Olarewaju A *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: Olarewaju A, Kloch A, Stevens L *et al.* **The genome sequence of *Heligmosomum mixtum* Schulz, 1929 (Rhabditida: Heligmosomidae) [version 1; peer review: 3 approved]** Wellcome Open Research 2026, 11:225 <https://doi.org/10.12688/wellcomeopenres.26318.1>

First published: 17 Apr 2026, 11:225 <https://doi.org/10.12688/wellcomeopenres.26318.1>

Species taxonomy

Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Protostomia; Ecdysozoa; Nematoda; Chromadorea; Rhabditida; Rhabditina; Rhabditomorpha; Strongyloidea; Heligmosomidae; *Heligmosomum*; *Heligmosomum mixtum* Schulz, 1929 (NCBI:txid375988).

Background

Heligmosomum mixtum is a common gastrointestinal parasite of bank voles (*Clethrionomys* = *Myodes glareolus* Schreber, 1780) and is widely used in parasitology and ecology (Bajer *et al.*, 2005; Behnke *et al.*, 2001, 2009; Kloch *et al.*, 2010). More recently, it has been a focus of eco-immunology and phylogeography research (McManus *et al.*, 2021; Sakka *et al.*, 2014). The genus *Heligmosomum* comprises 12 nominal species, which are nested within the clades of the paraphyletic *Heligmosomoides* genus (Alnaqeb *et al.*, 2022; Harris *et al.*, 2015; Zaleśny *et al.*, 2014). These gastrointestinal parasites reside in the small intestines, particularly the duodenum, of voles and mice distributed throughout the Palearctic and Nearctic regions, i.e., the Holarctic region (Durette-Desset, 1985; Harris *et al.*, 2015). In continental Europe, these parasites are typically the dominant species found, except in the eastern region where the species from the *Heligmosomoides* genus can be predominant (Haukisalmi & Henttonen, 2000). While *M. glareolus* is the main host of *H. mixtum* (Butmyrin *et al.*, 2005), they have been found in other rodent species such as *Clethrionomys rutilus*, *Microtus arvalis*, *Microtus agrestis*, *Microtus socialis*, and *Meriones persicus* (Grikienė, 2005; Haukisalmi *et al.*, 1996; Kia *et al.*, 2010), as well as *Apodemus flavicollis* (Mažeika *et al.*, 2003).

H. mixtum are dioecious parasites, exhibit a direct life cycle, and inhabit the intestinal lumen of the duodenum (Asakawa, 1987; Bryant, 1973; Sukhdeo *et al.*, 1984), primarily feeding on gut content (Haukisalmi & Henttonen, 1993). They reproduce sexually with three free-living larval stages and a 4th, parasitic larval stage that resides in the intestinal wall (Asakawa, 1987). Infection occurs when hosts ingest food or water contaminated with the infective larval stage (L3) of *H. mixtum*. The life cycle begins when the adult females produce eggs, which are expelled into the environment through the host's faeces. The eggs hatch into the first-stage larvae (L1). These larvae undergo two additional moults, developing into second-stage (L2) and third-stage larvae (L3), which are free-living and can survive in various environmental conditions until a suitable host ingests them. Once ingested by a rodent, the L3 larvae migrate through the mucosa to the intestines, shedding their sheath and growing to L4 and L5 before developing into adult worms (Asakawa, 1987; Ehrenford, 1954). Adult *H. mixtum* reside in the intestinal lumen, where they reproduce sexually, completing the cycle by producing new eggs that are excreted with the host's faeces. Larval migration through the intestinal mucosa during various life stages can cause mild inflammation, as observed with *Heligmosomum costellatum* in common voles (Janova *et al.*, 2010), a close relative of the bank vole. While specific details regarding the lifespan and mating behaviours of *H. mixtum* are limited (Kloch *et al.*, 2015), it is believed that many helminth parasites, including *H. mixtum*, exhibit relatively long lifespans comparable to their rodent hosts (Behnke *et al.*, 2009; Haukisalmi *et al.*, 1996), potentially employing immunosuppression strategies to persist within hosts (Behnke *et al.*, 2009; Hewitson *et al.*, 2011; Maizels *et al.*, 2012).

Sakka *et al.* (2014) showed that the genetic analysis of the cytochrome b gene (cytb) in mitochondrial DNA reveals a weak phylogeographical structure for *H. mixtum* compared to its primary host, *M. glareolus*, along with low genetic diversity relative to its counterpart *Heligmosomoides polygyrus*, suggesting a rapidly expanding population with a demographic history characterised by a unimodal distribution linked to past bottlenecks during the last glacial maximum.

Methods

Sample acquisition

We collected *H. mixtum* from *M. glareolus* captured in Okuniew, Warsaw, in central Poland (latitude 52.279, longitude 21.32) in Autumn 2023. Voles were caught live in locally constructed wooden traps, with a small metal internal platform triggered by a vole to release a metal door (Bajer *et al.*, 2005; Behnke *et al.*, 2001). Wooden traps were set out at 20 m intervals in parallel lines, 10 m on either side of tracks running through the site and 2 traps were placed within 2–3 m of one another at each point (Bajer *et al.*, 2005; Behnke *et al.*, 2001). The traps were baited with a mixture of wheat and peanut butter, as well as fruits like apples and carrots, to provide water. Captured voles were weighed (± 1 g), sexed, and euthanised following the European ethical standard (Directive 2010/63/EU) through anaesthetic overdose and cervical dislocation. The voles were sectioned, and their internal organs and intestinal tract were extracted and examined for intestinal parasites. Parasite species identification was carried out as described in Kloch *et al.* (2010). Briefly, in the laboratory, each intestine was divided into its anatomical sections (stomach, small intestine, caecum, and colon) and carefully examined under a microscope. All *H. mixtum* individuals were removed, counted, and snap-frozen, then stored at -80°C until being transported to the Wellcome Sanger Institute for genome sequencing.

Nucleic acid extraction

High molecular weight DNA was extracted from a frozen male adult of *Heligmosomum mixtum* (nxHelMixt5). DNA extraction was performed using a modified Monarch HMW DNA Extraction Kit (Kieninger, 2025). A total of 40 ng of sheared DNA was used for library preparation.

PacBio HiFi library preparation and sequencing

Library preparation and sequencing were performed at the WSI Scientific Operations core. Prior to library preparation, the DNA was fragmented to ~10 kb. Ultra-low-input (ULI) libraries were prepared using the PacBio SMRTbell® Express Template Prep Kit 2.0 and gDNA Sample Amplification Kit. Samples were normalised to 20 ng DNA. Single-strand overhang removal, DNA damage repair, and end-repair/A-tailing were performed according to the manufacturer's instructions, followed by adapter ligation. A 0.85× pre-PCR clean-up was carried out with Promega ProNex beads.

The DNA was evenly divided into two aliquots for dual PCR (reactions A and B), both following the manufacturer's protocol. A 0.85× post-PCR clean-up was performed with ProNex beads. DNA concentration was measured using a Qubit Fluorometer v4.0 (Thermo Fisher Scientific) with the Qubit HS Assay Kit, and fragment size was assessed on an Agilent Femto Pulse Automated Pulsed Field CE Instrument (Agilent Technologies) using the gDNA 55 kb BAC analysis kit. PCR reactions A and B were then pooled, ensuring a total mass of ≥500 ng in 47.4 µL.

The pooled sample underwent another round of DNA damage repair, end-repair/A-tailing, and hairpin adapter ligation. A 1× clean-up was performed with ProNex beads, followed by DNA quantification using the Qubit and fragment size analysis using the Agilent Femto Pulse. Size selection was performed on the Sage Sciences PippinHT system, with target fragment size determined by Femto Pulse analysis (typically 4–9 kb). Size-selected libraries were cleaned with 1.0× ProNex beads and normalised to 2 nM before sequencing.

The sample was sequenced on a Revio instrument (Pacific Biosciences). The prepared library was normalised to 2 nM, and 15 µL was used for making complexes. Primers were annealed and polymerases bound to generate circularised complexes, following the manufacturer's instructions. Complexes were purified using 1.2X SMRTbell beads, then

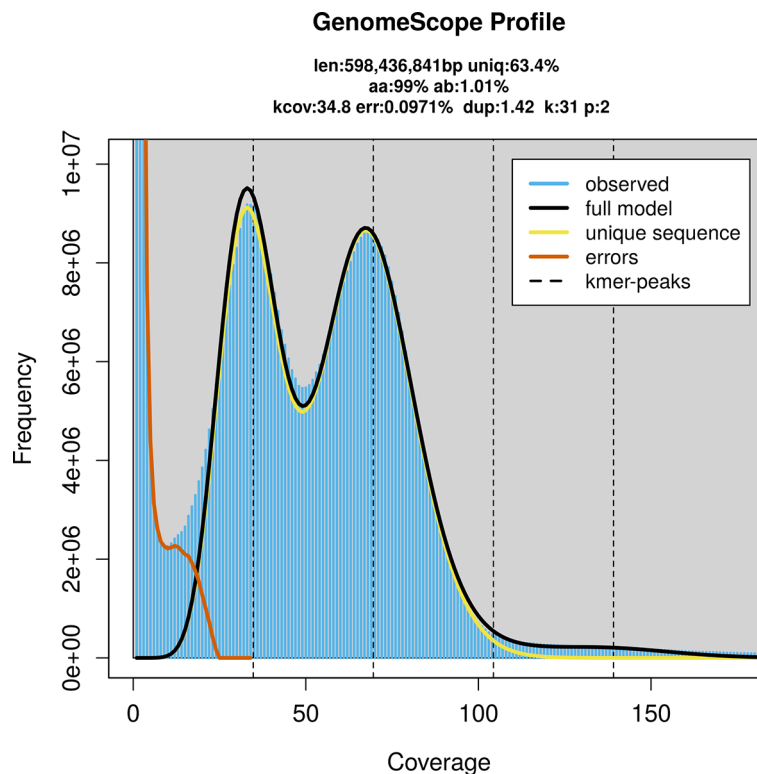


Figure 1. Frequency distribution of *k*-mers generated using GenomeScope2. The plot shows observed and modelled *k*-mer spectra, providing estimates of genome size, heterozygosity, and repeat content based on unassembled sequencing reads.

Table 1. Specimen and sequencing data for BioProject PRJEB83168.

Platform	PacBio HiFi	Hi-C
ToLID	nxHelMixt5	nxHelMixt1
Specimen ID	SAN20002475	SAN20002477
BioSample (source individual)	SAMEA115287570	SAMEA115287566
BioSample (tissue)	SAMEA115287576	SAMEA115287572
Tissue	whole organism	whole organism
Instrument	Revio	Illumina NovaSeq X
Run accessions	ERR14048106	ERR14056209
Read count total	3.75 million	385.33 million
Base count total	44.43 Gb	116.37 Gb

Table 2. Genome assembly statistics.

Genome assembly	Haplotype 1	Haplotype 2
Assembly name	nxHelMixt5.hap1.1	nxHelMixt5.hap2.1
Assembly accession	GCA_965643505.1	GCA_965643595.1
Assembly level	chromosome	scaffold
Span (Mb)	741.91	754.62
Number of chromosomes	6	-
Number of contigs	2 791	15 381
Contig N50	0.5 Mb	0.11 Mb
Number of scaffolds	197	10 392
Scaffold N50	127.18 Mb	95.46 Mb
Longest scaffold length (Mb)	130.85	-
Sex chromosomes	X	-
Organelles	Mitochondrion: 13.53 kb	-

diluted to the Revio loading concentration (200–300 pM) and spiked with a Revio sequencing internal control. The sample was sequenced on a Revio 25 M SMRT cell. The SMRT Link software (Pacific Biosciences), a web-based workflow manager, was used to configure and monitor the run and to carry out primary and secondary data analysis.

Hi-C

Sample preparation and crosslinking

Hi-C data were generated from the whole organism tissue of nxHelMixt1 using the Arima-HiC v2 kit (Arima Genomics). Tissue was finely ground using the Covaris cryoPREP Dry Pulverizer (Covaris), and then subjected to nuclei isolation. Nuclei were isolated using a modified protocol based on the Qiagen QProteome Cell Compartment Kit (Qiagen), in which only the Lysis and CE2 buffers were used, with QIAshredder spin columns. After isolation, nuclei were fixed using formaldehyde to a final concentration of 2% to crosslink the DNA. The crosslinked DNA was then digested and biotinylated according to the manufacturer's instructions. A clean-up step was performed with SPRIselect beads before library preparation. DNA concentration was quantified using the Qubit Fluorometer v4.0 (Thermo Fisher Scientific) and the Qubit HS Assay Kit, following the manufacturer's instructions.

Hi-C library preparation and sequencing

Biotinylated DNA constructs were fragmented using a Covaris E220 sonicator and size selected to 400–600 bp using SPRIselect beads. DNA was enriched with Arima-HiC v2 kit Enrichment beads. End repair, A-tailing, and adapter ligation were carried out with the NEBNext Ultra II DNA Library Prep Kit (New England Biolabs), following a modified

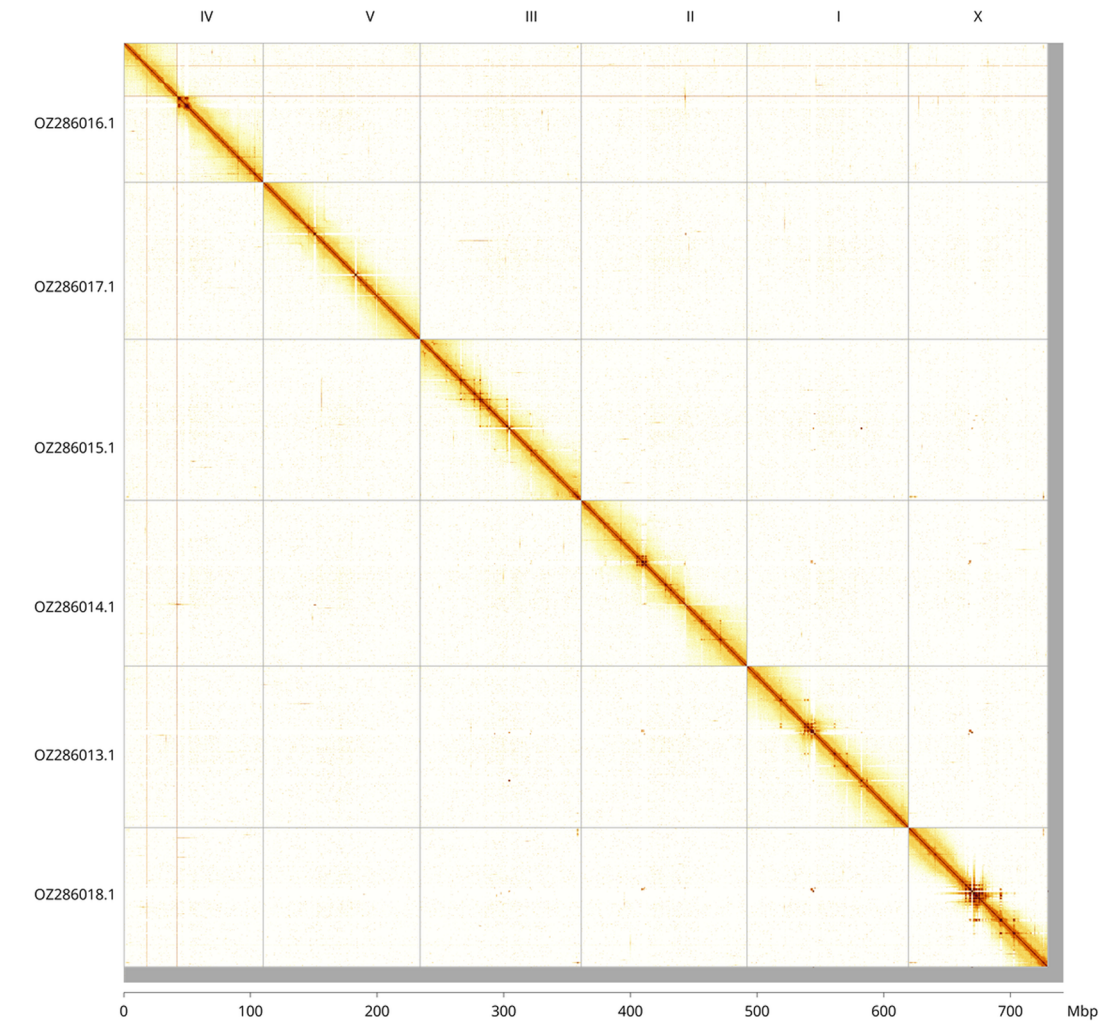


Figure 2. Hi-C contact map of the *Heligmosomum mixtum* genome assembly. Assembled chromosomes are shown in order of size and labelled along the axes, with a megabase scale shown below. The plot was generated using PretextSnapshot.

Table 3. Chromosomal pseudomolecules in the haplotype 1 genome assembly of *Heligmosomum mixtum* nxHelMixt5.

INSDC accession	Molecule	Length (Mb)	GC%
OZ286018.1	X	109.94	46.50
OZ286013.1	I	110.24	46.50
OZ286014.1	II	123.88	45.50
OZ286015.1	III	127.18	46.50
OZ286016.1	IV	130.85	46
OZ286017.1	V	127.43	46

protocol where library preparation occurs while DNA remains bound to the Enrichment beads. Library amplification was performed using KAPA HiFi HotStart mix and a custom Unique Dual Index (UDI) barcode set (Integrated DNA Technologies). Depending on sample concentration and biotinylation percentage determined at the crosslinking stage, libraries were amplified with 10–16 PCR cycles. Post-PCR clean-up was performed with SPRISelect beads. Libraries

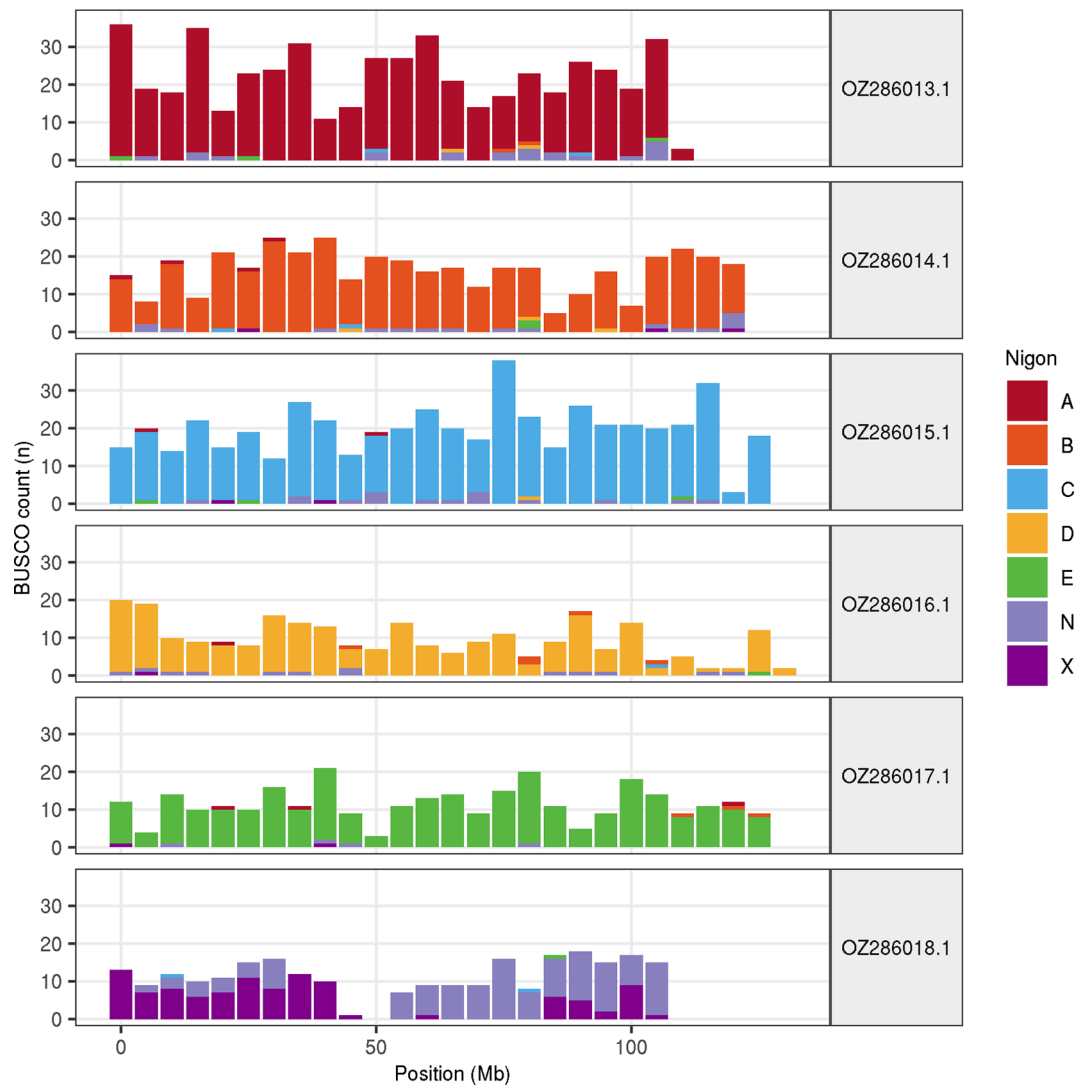


Figure 3. Nigon element painting of chromosomes in the nxHelMmixt5.hap1.1 assembly of *Heligmosomum mixtum*. Counts of Benchmarking using Single Copy Orthologues (BUSCO) loci in 500 kb windows are coloured by their allocation to the seven Nigon elements (A-E, N, X).

were quantified using the AccuClear Ultra High Sensitivity dsDNA Standards Assay Kit (Biotium) and a FLUOstar Omega plate reader (BMG Labtech).

Prior to sequencing, libraries were normalised to 10 ng/ μ L. Normalised libraries were quantified again to create equimolar and/or weighted 2.8 nM pools. Pool concentrations were checked using the Agilent 4200 TapeStation (Agilent) with High Sensitivity D500 reagents before sequencing. Sequencing was performed using paired-end 150 bp reads on the Illumina NovaSeq X.

Genome assembly

Prior to assembly of the PacBio HiFi reads, a database of k -mer counts ($k = 31$) was generated from the filtered reads using FastK. GenomeScope2 (Ranallo-Benavidez *et al.*, 2020) was used to analyse the k -mer frequency distributions, providing estimates of genome size, heterozygosity, and repeat content.

The HiFi reads were assembled using Hifiasm in Hi-C phasing mode (Cheng *et al.*, 2021, 2022), producing two haplotypes. Hi-C reads (Rao *et al.*, 2014) were mapped to the primary contigs using bwa-mem2 (Vasimuddin *et al.*, 2019). Contigs were further scaffolded with Hi-C data in YaHS (Zhou *et al.*, 2023), using the --break option for handling

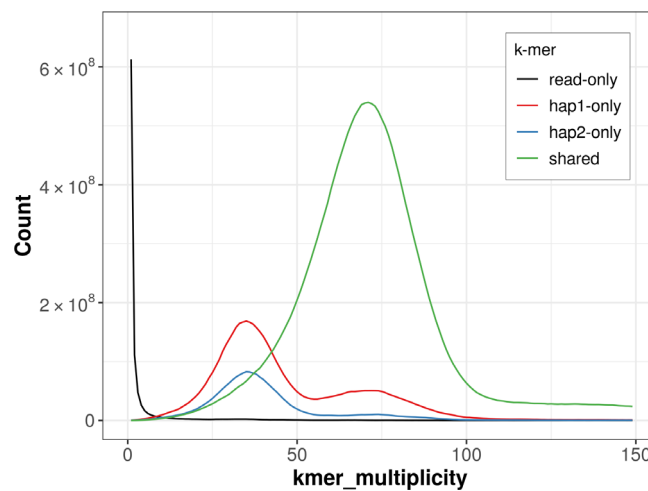


Figure 4. Evaluation of *k*-mer completeness using MerquryFK. This plot illustrates the recovery of *k*-mers from the original read data in the final assemblies. The horizontal axis represents *k*-mer multiplicity, and the vertical axis shows the number of *k*-mers. The black curve represents *k*-mers that appear in the reads but are not assembled. The green curve corresponds to *k*-mers shared by both haplotypes, and the red and blue curves show *k*-mers found only in one of the haplotypes.

potential misassemblies. The scaffolded assemblies were evaluated using Gfastats (Formenti *et al.*, 2022), BUSCO (Manni *et al.*, 2021) and MERQURY.FK (Rhie *et al.*, 2020). The mitochondrial genome was assembled using MitoHiFi (Uliano-Silva *et al.*, 2023).

Assembly curation

The assembly was decontaminated using the Assembly Screen for Cobionts and Contaminants (ASCC) pipeline. TreeVal was used to generate the flat files and maps for use in curation. Manual curation was conducted primarily in PretextView and HiGlass (Kerpedjiev *et al.*, 2018). Scaffolds were visually inspected and corrected as described by Howe *et al.* (2021). Manual corrections included 227 breaks and 337 joins. This reduced the scaffold count by 30.5%, increased the scaffold N50 by 4.8%, and reduced the total assembly length by 0.8%. During curation, we observed that there is a haplotypic inversion on chromosome III (3–11.3 Mbp). The curation process is documented at <https://gitlab.com/wtsi-grit/rapid-curation>. PretextViewSnapshot was used to generate a Hi-C contact map of the final assembly.

Assembly analysis and quality assessment

The distribution of Nigon elements, which represent the seven ancestral linkage groups in rhabditid nematodes (Gonzalez de la Rosa *et al.*, 2021), was visualised by providing the 'full_table.tsv' file from the BUSCO nematoda_odb10 output to the R script `vis_ALG.R`, using a window size of 2 Mb.

The Merqury.FK tool (Rhie *et al.*, 2020) was used to evaluate *k*-mer completeness and assembly quality for both haplotypes using the *k*-mer databases ($k = 31$) computed prior to genome assembly. The analysis outputs included assembly QV scores and completeness statistics.

The genome was analysed using the BlobToolKit pipeline, a Nextflow implementation of the earlier Snakemake version (Challis *et al.*, 2020). The pipeline aligns PacBio reads using minimap2 (Li, 2018) and SAMtools (Danecek *et al.*, 2021) to generate coverage tracks. It runs BUSCO (Manni *et al.*, 2021) using lineages identified from NCBI Taxonomy (Schoch *et al.*, 2020). For the three domain-level lineages, BUSCO genes are aligned to the UniProt Reference Proteomes database (Bateman *et al.*, 2023) using DIAMOND blastp (Buchfink *et al.*, 2021). The genome is divided into chunks based on the density of BUSCO genes from the closest taxonomic lineage, and each chunk is aligned to the UniProt Reference Proteomes database with DIAMOND blastx. Sequences without hits are chunked using seqtk and aligned to the NT database with blastn (Altschul *et al.*, 1990). The BlobToolKit suite consolidates all outputs into a blobdir for visualisation. The BlobToolKit pipeline was developed using nf-core tooling (Ewels *et al.*, 2020) and MultiQC (Ewels *et al.*, 2016), with package management via Conda and Bioconda (Grüning *et al.*, 2018), and containerisation through Docker (Merkel, 2014) and Singularity (Kurtzer *et al.*, 2017).

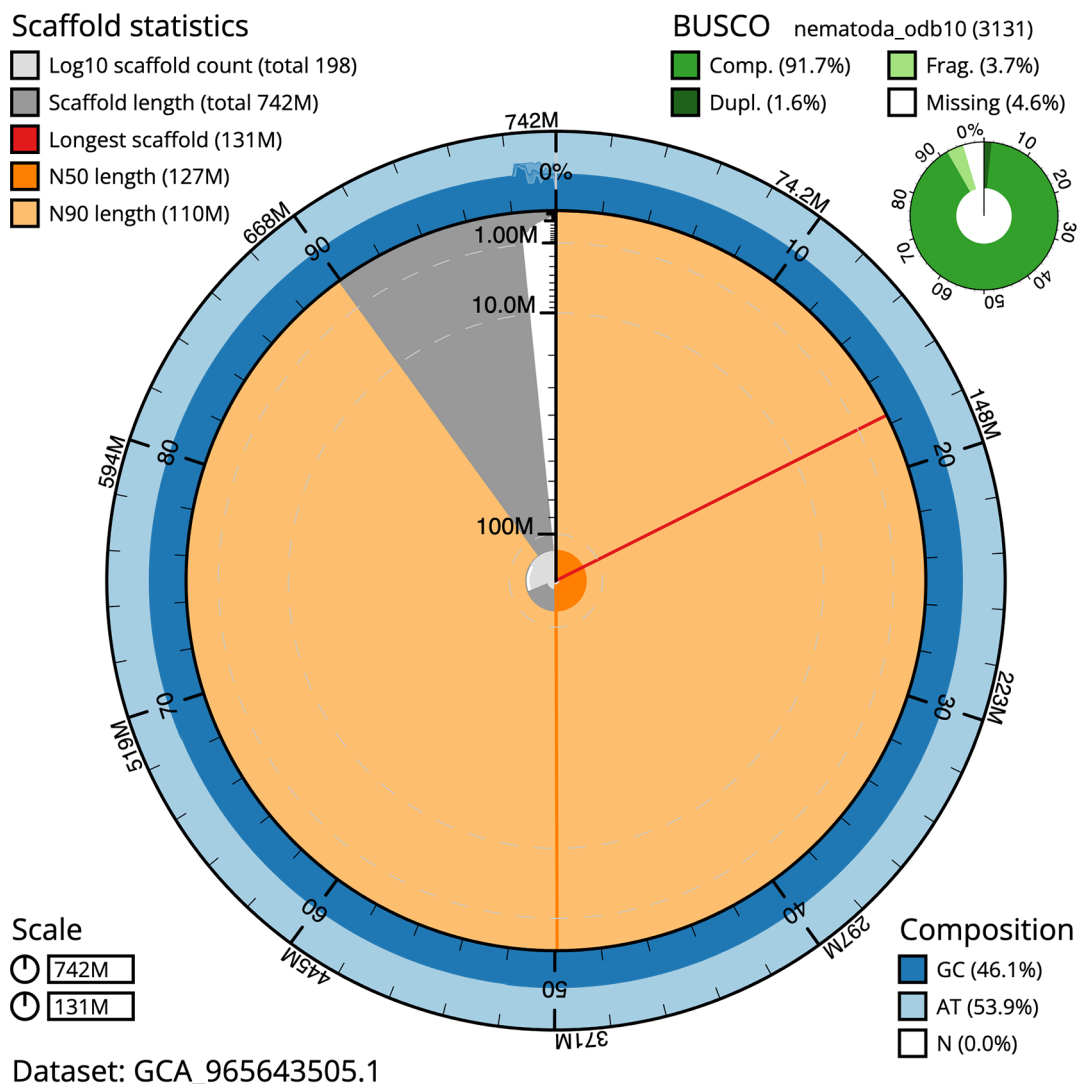


Figure 5. Assembly metrics for nxHelMixt5.hap1.1. The BlobToolKit snail plot provides an overview of assembly metrics and BUSCO gene completeness. The circumference represents the length of the whole genome sequence, and the main plot is divided into 1,000 bins around the circumference. The outermost blue tracks display the distribution of GC, AT, and N percentages across the bins. Scaffolds are arranged clockwise from longest to shortest and are depicted in dark grey. The longest scaffold is indicated by the red arc, and the deeper orange and pale orange arcs represent the N50 and N90 lengths. A light grey spiral at the centre shows the cumulative scaffold count on a logarithmic scale. A summary of complete, fragmented, duplicated, and missing BUSCO genes in the set is presented at the top right. An interactive version of this figure can be accessed on the [BlobToolKit viewer](#).

Genome sequence report

Sequence data

The genome of a specimen of *Heligmosomum mixtum* was sequenced using Pacific Biosciences single-molecule HiFi long reads, generating 44.43 Gb (gigabases) from 3.75 million reads, which were used to assemble the genome. GenomeScope2.0 analysis estimated the haploid genome size at 599.27 Mb, with a heterozygosity of 1.01% and repeat content of 36.60% (Figure 1). These estimates guided expectations for the assembly. Based on the estimated genome size, the sequencing data provided approximately 70× coverage. Hi-C sequencing produced 116.37 Gb from 385.33 million reads, which were used to scaffold the assembly. Table 1 summarises the specimen and sequencing details.

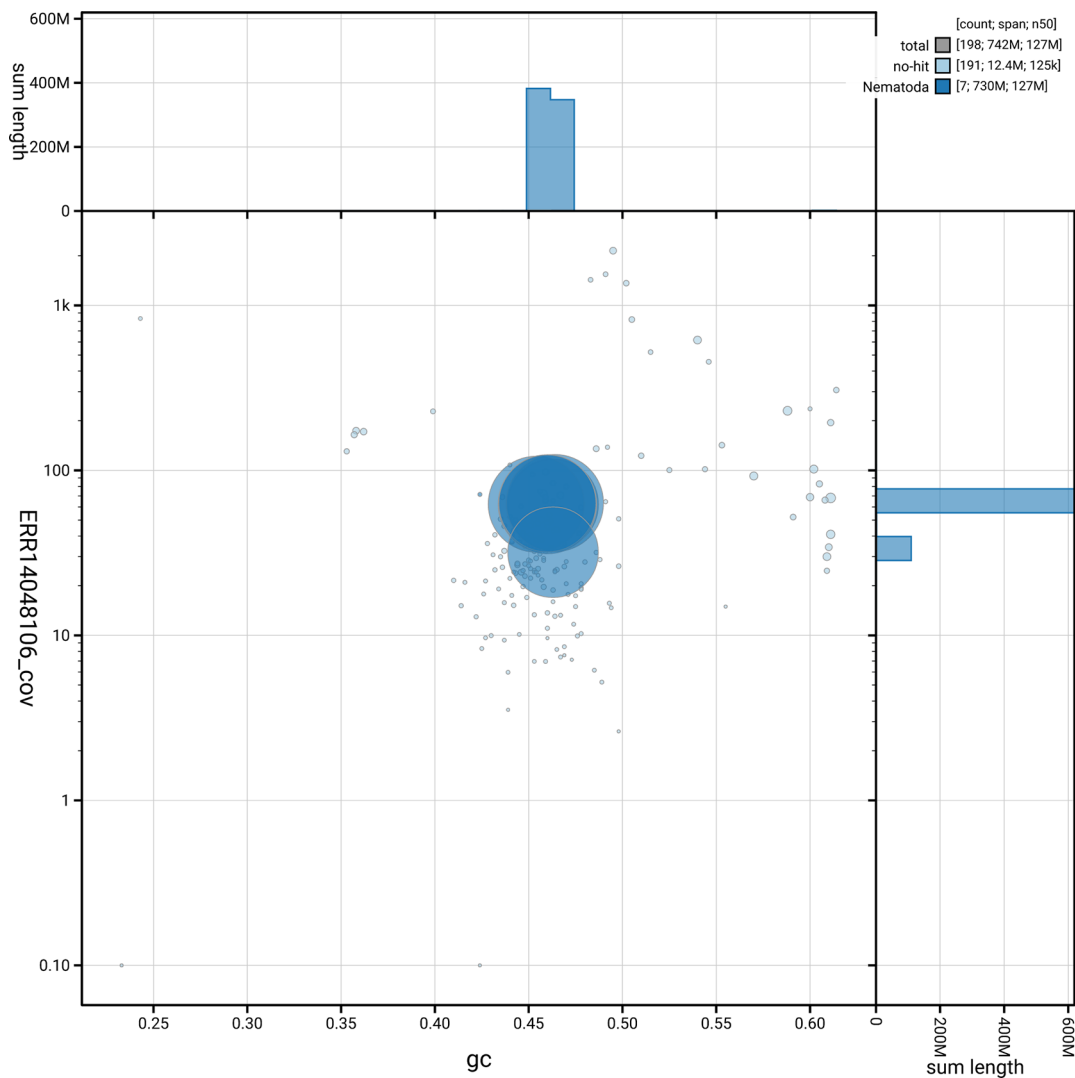


Figure 6. BlobToolKit GC-coverage plot for nxHelMixt5.hap1.1. Blob plot showing sequence coverage (vertical axis) and GC content (horizontal axis). The circles represent scaffolds, with the size proportional to scaffold length and the colour representing phylum membership. The histograms along the axes display the total length of sequences distributed across different levels of coverage and GC content. An interactive version of this figure is available on the [BlobToolKit viewer](#).

Assembly statistics

The genome was assembled into two haplotypes using Hi-C phasing. Haplotype 1 was curated to chromosome level, while haplotype 2 was assembled to scaffold level. The final assembly has a total length of 741.91 Mb in 197 scaffolds, with 2 594 gaps, and a scaffold N50 of 127.18 Mb (Table 2).

Most of the assembly sequence (98.33%) was assigned to 6 chromosomal-level scaffolds, representing 5 autosomes and the X sex chromosome. These chromosome-level scaffolds, confirmed by Hi-C data, are named according to synteny (Figure 2; Table 3).

The mitochondrial genome was also assembled. This sequence is included as a contig in the multifasta file of the genome submission and as a standalone record.

Chromosome painting with Nigon elements illustrates the distribution of orthologues along chromosomes and highlights patterns of chromosomal evolution relative to ancestral linkage groups (Figure 3).

Table 4. Earth Biogenome Project summary metrics for the *Heligmosomum mixtum* assembly.

Measure	Value	Benchmark
EBP summary (haplotype 1)	5.C.Q58	6.C.Q40
Contig N50 length	0.50 Mb	≥ 1 Mb
Scaffold N50 length	127.18 Mb	= chromosome N50
Consensus quality (QV)	Haplotype 1: 58.2; haplotype 2: 57.0; combined: 57.5	≥ 40
k-mer completeness	Haplotype 1: 88.23%; Haplotype 2: 73.52%; combined: 99.23%	≥ 95%
BUSCO	C:91.7% [S:90.1%, D:1.6%], F:3.7%, M:4.6%, n:3 131	S > 90%; D < 5%
Percentage of assembly assigned to chromosomes	98.33%	≥ 90%

Notes: The EBP summary uses log10(Contig N50); chromosome-level (C) or log10(Scaffold N50); Q (Merquy QV). BUSCO: C = complete; S = single-copy; D = duplicated; F = fragmented; M = missing; n = orthologues.

Assembly quality metrics

For haplotype 1, the estimated QV is 58.2, and for haplotype 2, 57.0. When the two haplotypes are combined, the assembly achieves an estimated QV of 57.5. The k-mer completeness is 88.23% for haplotype 1, 73.52% for haplotype 2, and 99.23% for the combined haplotypes (Figure 4).

BUSCO analysis using the nematoda_odb10 reference set (n = 3 131) identified 91.7% of the expected gene set (single = 90.1%, duplicated = 1.6%) for haplotype 1. The snail plot in Figure 5 summarises the scaffold length distribution and other assembly statistics for haplotype 1. The blob plot in Figure 6 shows the distribution of scaffolds by GC proportion and coverage for haplotype 1.

Table 4 lists the assembly metric benchmarks adapted from Rhie et al. (2021) and the Earth BioGenome Project Report on Assembly Standards September 2024. The EBP metric calculated for the haplotype 1 is 5.C.Q58.

Author information

- Members of the Wellcome Sanger Institute Tree of Life Management, Samples and Laboratory team
- Members of Wellcome Sanger Institute Scientific Operations – Sequencing Operations
- Members of the Wellcome Sanger Institute Tree of Life Core Informatics team
- Members of the Tree of Life Core Informatics collective

Wellcome Sanger Institute – Legal and Governance

The materials that have contributed to this genome note have been supplied by a Tree of Life collaborator. The Wellcome Sanger Institute employs a process whereby due diligence is carried out proportionate to the nature of the materials themselves, and the circumstances under which they have been/are to be collected and provided for use. The purpose of this is to address and mitigate any potential legal and/or ethical implications of receipt and use of the materials as part of the research project, and to ensure that in doing so, we align with best practice wherever possible. The overarching areas of consideration are:

- Ethical review of provenance and sourcing of the material.
- Legality of collection, transfer and use (national and international).

Each transfer of samples is undertaken according to a Research Collaboration Agreement or Material Transfer Agreement entered into by the Tree of Life collaborator, Genome Research Limited (operating as the Wellcome Sanger Institute), and in some circumstances, other Tree of Life collaborators.

Table 5. Software versions and sources.

Software	Version	Source
BLAST	2.14.0	ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/
BlobToolKit	4.4.6	https://github.com/blobtoolkit/blobtoolkit
BUSCO	5.8.3	https://gitlab.com/ezlab/busco
bwa-mem2	2.2.1	https://github.com/bwa-mem2/bwa-mem2
DIAMOND	2.1.8	https://github.com/bbuchfink/diamond
fasta_windows	0.2.4	https://github.com/tolkit/fasta_windows
FastK	1.1	https://github.com/thegenemyers/FASTK
GenomeScope2.0	2.0.1	https://github.com/tbenavi1/genomescope2.0
Gfastats	1.3.6	https://github.com/vgl-hub/gfastats
Hifiasm	0.19.8-r603	https://github.com/chhylp123/hifiasm
HiGlass	1.13.4	https://github.com/higlass/higlass
MercuryFK	1.1.2	https://github.com/thegenemyers/MERQUERY.FK
Minimap2	2.28-r1209	https://github.com/lh3/minimap2
MitoHiFi	3	https://github.com/marcelauliano/MitoHiFi
MultiQC	1.14; 1.17 and 1.18	https://github.com/MultiQC/MultiQC
Nextflow	24.10.4	https://github.com/nextflow-io/nextflow
PretextSnapshot	0.0.5	https://github.com/sanger-tol/PretextSnapshot
PretextView	1.0.3	https://github.com/sanger-tol/PretextView
samtools	1.21	https://github.com/samtools/samtools
sanger-tol/ascc	0.1.0	https://github.com/sanger-tol/ascc
sanger-tol/blobtoolkit	v0.8.0	https://github.com/sanger-tol/blobtoolkit
sanger-tol/curationpretext	1.4.2	https://github.com/sanger-tol/curationpretext
Seqtk	1.3	https://github.com/lh3/seqtk
Singularity	3.9.0	https://github.com/sylabs/singularity
TreeVal	1.4.0	https://github.com/sanger-tol/treeval
YaHS	1.2.2	https://github.com/c-zhou/yahs

Data availability

European Nucleotide Archive: *Heligmosomum mixtum*. Accession number [PRJEB83168](#). The genome sequence is released openly for reuse. The *Heligmosomum mixtum* genome sequencing initiative is part of the Sanger Institute Tree of Life Programme (PRJEB43745) and 959NG (PRJEB81973). All raw sequence data and the assembly have been deposited in INSDC databases. The genome will be annotated using available RNA-Seq data and presented through the [Ensembl](#) pipeline at the European Bioinformatics Institute. Raw data and assembly accession identifiers are reported in [Tables 1](#) and [2](#).

Pipelines used for genome assembly at the WSI Tree of Life are available at <https://pipelines.tol.sanger.ac.uk/pipelines>. [Table 5](#) lists software versions used in this study.

References

- Alnaqeb H, Greiman SE, Vandegrift KJ, *et al.*: **A molecular reconstruction of holarctic Heligmosomidae reveals a new species of *Heligmosomoides* (Nematoda: Heligmosomidae) in *Peromyscus maniculatus* (Neotominae) from Canada.** *Syst. Biodivers.* 2022; 20: 1–19. [Publisher Full Text](#)
- Altschul SF, Gish W, Miller W, *et al.*: **Basic Local Alignment Search Tool.** *J. Mol. Biol.* 1990; 215(3): 403–410. [PubMed Abstract](#) | [Publisher Full Text](#)
- Asakawa M: **Genus *Heligmosomoides* Hall, 1916 (Heligmosomidae: Nematoda) from the Japanese wood mice, *Apodemus* spp. III.** The

- life-cycle of *Heligmosomoides kurilensis kobayashii* (Nadtochii, 1966) in ICR mice and preliminary experimental infection to jirds.** *Journal of the College of Dairying*. 1987; **12**: 131–140.
- Bajer A, et al.: **Medium-term temporal stability of the helminth component community structure in bank voles (*Clethrionomys glareolus*) from the Mazury Lake District region of Poland.** *Parasitology*. 2005; **130**: 213–228.
[Publisher Full Text](#)
- Bateman A, Martin M-J, Orchard S, et al.: **UniProt: The Universal Protein Knowledgebase in 2023.** *Nucleic Acids Res.* 2023; **51**(D1): D523–D531.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Behnke J, et al.: **Variation in the helminth community structure in bank voles (*Clethrionomys glareolus*) from three comparable localities in the Mazury Lake District region of Poland.** *Parasitology*. 2001; **123**: 401–414.
[Publisher Full Text](#)
- Behnke J, et al.: **Helminth species richness in wild wood mice, *Apodemus sylvaticus*, is enhanced by the presence of the intestinal nematode *Heligmosomoides polygyrus*.** *Parasitology*. 2009; **136**: 793–804.
[Publisher Full Text](#)
- Bryant V: **The life cycle of *Nematospiroides dubius* Baylis, 1926 (Nematoda: Heligmosomidae).** *J. Helminthol.* 1973; **47**: 263–268.
[Publisher Full Text](#)
- Buchfink B, Reuter K, Drost H-G: **Sensitive protein alignments at tree-of-life scale using DIAMOND.** *Nat. Methods*. 2021; **18**(4): 366–368.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Butnaryn S, Ieshko E, Anikanova V, et al.: **Patterns of host-parasite interactions between the nematode *Heligmosomum mixtum* (Schulz, 1952) and the bank vole (*Clethrionomys glareolus* Schreber, 1780).** *Parazitologija*. 2005; **39**: 414–422.
- Challis R, Richards E, Rajan J, et al.: **BlobToolKit – interactive quality assessment of genome assemblies.** *G3 Genes | Genomes | Genetics*. 2020; **10**(4): 1361–1374.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Cheng H, Concepcion GT, Feng X, et al.: **Haplotype-resolved *de novo* assembly using phased assembly graphs with Hifiasm.** *Nat. Methods*. 2021; **18**(2): 170–175.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Cheng H, Jarvis ED, Fedrigo O, et al.: **Haplotype-resolved assembly of diploid genomes without parental data.** *Nat. Biotechnol.* 2022; **40**(9): 1332–1335.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Danecek P, Bonfield JK, Liddle J, et al.: **Twelve years of SAMtools and BCFTools.** *GigaScience*. 2021; **10**(2).
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Durette-Desset M: **Trichostrongyloid nematodes and their vertebrate hosts: Reconstruction of the phylogeny of a parasitic group.** *Adv. Parasitol.* 1985; **24**: 239–306.
[Publisher Full Text](#)
- Ehrenford F: **The life cycle of *Nematospiroides dubius* Baylis (Nematoda: Heligmosomidae).** *J. Parasitol.* 1954; **40**: 480.
[Publisher Full Text](#)
- Ewels P, Magnusson M, Lundin S, et al.: **MultiQC: Summarize analysis results for multiple tools and samples in a single report.** *Bioinformatics*. 2016; **32**(19): 3047–3048.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Ewels PA, Peltzer A, Fillinger S, et al.: **The nf-core framework for community-curated bioinformatics pipelines.** *Nat. Biotechnol.* 2020; **38**(3): 276–278.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Formenti G, Abueg L, Brajuka A, et al.: **Gfastats: Conversion, evaluation and manipulation of genome sequences using assembly graphs.** *Bioinformatics*. 2022; **38**(17): 4214–4216.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Gonzalez de la Rosa PM, Thomson M, Trivedi U, et al.: **A telomere-to-telomere assembly of *Osccheius tipulae* and the evolution of rhabditid nematode chromosomes.** *G3 Genes | Genomes | Genet.* 2021; **11**(1): jkaa020.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Grikeniū J: **Investigations into endoparasites of small mammals in the environs of Lake Drūkšiai.** *Acta Zool Lituan.* 2005; **15**: 109–114.
[Publisher Full Text](#)
- Grüning B, Dale R, Sjödin A, et al.: **Bioconda: Sustainable and comprehensive software distribution for the life sciences.** *Nat. Methods*. 2018; **15**(7): 475–476.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Harris P, et al.: **The status of *Heligmosomoides americanus*, representative of an American clade of vole-infecting nematodes.** *J. Parasitol.* 2015; **101**: 382–385.
[Publisher Full Text](#)
- Haukisalmi V, Henttonen H: **Coexistence in helminths of the bank vole *Clethrionomys glareolus*. I. Patterns of co-occurrence.** *J. Anim. Ecol.* 1993; **62**: 221.
[Publisher Full Text](#)
- Haukisalmi V, Henttonen H: **Variability of helminth assemblages and populations in the bank vole *Clethrionomys glareolus*.** *Pol. J. Ecol.* 2000; **48**: 219–231.
- Haukisalmi V, Henttonen H, Vikman P: **Variability of sex ratio, mating probability and egg production in an intestinal nematode in its fluctuating host population.** *Int. J. Parasitol.* 1996; **26**: 755–763.
[Publisher Full Text](#)
- Hewitson J, et al.: **Proteomic analysis of secretory products from the model gastrointestinal nematode *Heligmosomoides polygyrus* reveals dominance of venom allergen-like (VAL) proteins.** *J. Proteome*. 2011; **74**: 1573–1594.
[Publisher Full Text](#)
- Howe K, Chow W, Collins J, et al.: **Significantly improving the quality of genome assemblies through curation.** *GigaScience*. 2021; **10**(1).
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Janova E, Skoric M, Heroldova M, et al.: **Determinants of the prevalence of *Heligmosomum costellatum* (Heligmosomidae: Trichostrongyloidea) in a common vole population in southern Moravia, Czech Republic.** *J. Helminthol.* 2010; **84**: 410–414.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Kerpedjiev P, Abdennur N, Lekschas F, et al.: **HiGlass: Web-based visual exploration and analysis of genome interaction maps.** *Genome Biol.* 2018; **19**(1): 125.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Kia E, et al.: **Endoparasites of rodents and their zoonotic importance in Germi, Dashte-Mogan, Ardabil Province, Iran.** *Iran. J. Parasitol.* 2010; **5**: 15–20.
- Kieninger M: **HMW DNA extraction from frozen nematodes (bulk).** 2025.
[Reference Source](#)
- Kloch A, Babik W, Bajer A, et al.: **Effects of an MHC-DRB genotype and allele number on the load of gut parasites in the bank vole *Myodes glareolus*.** *Mol. Ecol.* 2010; **19**(Suppl 1): 255–265.
[Publisher Full Text](#)
- Kloch A, Michalski A, Bajer A, et al.: **Biased sex ratio among worms of the family Heligmosomidae — searching for a mechanism.** *Int. J. Parasitol.* 2015; **45**: 939–945.
[Publisher Full Text](#)
- Kurtzer GM, Sochat V, Bauer MW: **Singularity: Scientific containers for mobility of compute.** *PLOS ONE*. 2017; **12**(5): e0177459.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Li H: **Minimap2: Pairwise alignment for nucleotide sequences.** *Bioinformatics*. 2018; **34**(18): 3094–3100.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Maizels R, et al.: **Immune modulation and modulators in *Heligmosomoides polygyrus* infection.** *Exp. Parasitol.* 2012; **132**: 76–89.
[Publisher Full Text](#)
- Manni M, Berkeley MR, Seppey M, et al.: **BUSCO update: Novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes.** *Mol. Biol. Evol.* 2021; **38**(10): 4647–4654.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Mažeika V, Paulauskas A, Balčiauskas L: **New data on the helminth fauna of rodents of Lithuania.** *Acta Zoologica Lituanica*. 2003; **13**: 41–47.
[Publisher Full Text](#)
- McManus A, Holland C, Henttonen H, et al.: **The invasive bank vole (*Myodes glareolus*): A model system for studying parasites and ecoimmunology during a biological invasion.** *Animals*. 2021; **11**: 2529.
[Publisher Full Text](#)
- Merkel D: **Docker: Lightweight Linux containers for consistent development and deployment.** *Linux J.* 2014; **2014**(239).
[Publisher Full Text](#)
- Ranallo-Benavidez TR, Jaron KS, Schatz MC: **GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes.** *Nat. Commun.* 2020; **11**(1): 1432.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Rao SSP, Huntley MH, Durand NC, et al.: **A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping.** *Cell*. 2014; **159**(7): 1665–1680.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Rhie A, McCarthy SA, Fedrigo O, et al.: **Towards complete and error-free genome assemblies of all vertebrate species.** *Nature*. 2021; **592**(7856): 737–746.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Rhie A, Walenz BP, Koren S, et al.: **Merqury: Reference-free quality, completeness, and phasing assessment for genome assemblies.** *Genome Biol.* 2020; **21**(1).
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Sakka H, Henttonen H, Baraket G, *et al.*: **Phylogeography analysis and molecular evolution patterns of the nematode parasite *Heligmosomum mixtum* based on mitochondrial DNA sequences.** *Acta Parasitol.* 2014; **60**: 85–98.
[Publisher Full Text](#)

Schoch CL, Ciufo S, Domrachev M, *et al.*: **NCBI taxonomy: A comprehensive update on curation, resources and tools.** *Database.* 2020; **2020**: baaa062.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Sukhdeo M, O'Grady R, Hsu S: **The site selected by the larvae of *Heligmosomoides polygyrus*.** *J. Helminthol.* 1984; **58**: 19–23.
[Publisher Full Text](#)

Uliano-Silva M, Ferreira JGRN, Krashenninnikova K, *et al.*: **MitoHiFi: A Python pipeline for mitochondrial genome assembly from PacBio**

high fidelity reads. *BMC Bioinformatics.* 2023; **24**(1): 288.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Vasimuddin M, Misra S, Li H, *et al.*: **Efficient architecture-aware acceleration of BWA-MEM for multicore systems.** *2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS).* IEEE; 2019; 314–324.
[Publisher Full Text](#)

Zaleśny G, Hildebrand J, Paziewska-Harris A, *et al.*: ***Heligmosomoides neopolygyrus* Asakawa & Ohbayashi, 1986, a cryptic Asian nematode infecting the striped field mouse *Apodemus agrarius* in Central Europe.** *Parasit. Vectors.* 2014; **7**: 457.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Zhou C, McCarthy SA, Durbin R: **YaHS: Yet another Hi-C scaffolding tool.** *Bioinformatics.* 2023; **39**(1).
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Open Peer Review

Current Peer Review Status:   

Version 1

Reviewer Report 18 May 2026

<https://doi.org/10.21956/wellcomeopenres.28983.r153703>

© 2026 Kim 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.



Jun Kim 

Chungnam National University, Daejeon, Daejeon, South Korea

This report presents a high-quality genome assembly of *Heligmosomum mixtum*. As the Tree of Life team has already established a highly reliable workflow for data generation and analysis, I do not have any major comments regarding the analytical procedures themselves. I would, however, like to ask the authors to consider the following two points.

"Most of the assembly sequence (98.33%) was assigned to 6 chromosomal-level scaffolds, representing 5 autosomes and the X sex chromosome."

--> I would appreciate it if the authors could examine the telomeric repeats at the ends of the chromosomes. In addition, the read depth of the X sex chromosome should also be evaluated. My understanding is that this species probably has an XX/XO sex determination system, and it would be helpful if this point could be described in the manuscript. If the XX/XO system is indeed correct, the read depth of the X chromosome would be expected to be approximately half that of the autosomes, so confirmation of this would strengthen the study. Although the use of whole-genome amplification likely results in uneven read depth, it would still be informative to assess whether the X chromosome shows roughly half the coverage of the autosomes and whether there are any regions with unusually elevated read depth.

"For haplotype 1, the estimated QV is 58.2, and for haplotype 2, 57.0. When the two haplotypes are combined, the assembly achieves an estimated QV of 57.5. The k-mer completeness is 88.23% for haplotype 1, 73.52% for haplotype 2, and 99.23% for the combined haplotypes (Figure 4)."

--> It appears that only HiFi data were generated. If that is the case, it may be worth noting that the HiFi reads used as input for QV estimation with MerquryFK were likely the same HiFi reads used for the genome assembly itself, and that this may have led to an overestimation of the QV.

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 assembly

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 13 May 2026

<https://doi.org/10.21956/wellcomeopenres.28983.r153704>

© 2026 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 

University of Calgary, Calgary, Canada

This is a solid and useful genome resource for *Heligmosomum mixtum*, a species with clear relevance for parasitology, eco-immunology, and comparative nematode genomics. The manuscript is clearly written, and the rationale for generating the dataset is well articulated. Methods are described with sufficient detail to support reproducibility, and the combination of PacBio HiFi data, Hi-C scaffolding, and manual curation has resulted in a high-quality chromosome-level assembly.

The assembly metrics are strong. The haplotype-resolved assembly is also a strength, particularly for a wild-derived parasite where heterozygosity is expected to be biologically meaningful. One point to note is that BUSCO completeness (~92%) suggests there is still a fraction of conserved gene content that is either missing or fragmented. This is not unusual, but it does indicate some remaining scope for improvement, whether through further polishing or downstream annotation.

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: Nematode genomics

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 05 May 2026

<https://doi.org/10.21956/wellcomeopenres.28983.r153709>

© 2026 Schwarz E. 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.



Erich M. Schwarz 

Cornell University, Ithaca, New York, USA

This genome assembly is done to the usual high Tree of Life standard and leaves little to comment upon: technically, and in the clarity of its writing, and in its selection of a useful resource for comparative genomics of model gastrointestinal nematode parasites, it is pleasing to read. The remaining analysis that will need to be done by other people using this is protein-coding gene prediction, but improvements in software make this an increasingly easy deficit to remedy.

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: Nematode genomics, in both non-parasitic and parasitic nematode species

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.
