

Genome assembly and annotation of the mermithid nematode *Mermis nigrescens*

Upendra R. Bhattarai¹, Robert Poulin², Neil J. Gemmell^{1*}, Eddy Dowle^{1*}

¹ Department of Anatomy, University of Otago, Dunedin 9016, New Zealand

² Department of Zoology, University of Otago, Dunedin 9016, New Zealand

* Corresponding authors

Keywords: *Mermis nigrescens*, Nematode, Hybrid genome assembly, Repeatome, Genome annotation

Abstract

Genetic studies of nematodes have been dominated by *Caenorhabditis elegans* as a model species. Lack of genomic resources has been a limiting factor for expansion of genetic research to other groups of nematodes. Here, we report a draft genome assembly of a mermithid nematode, *Mermis nigrescens*. Mermithidae are insect parasitic nematodes with hosts including a wide range of terrestrial arthropods. We sequenced, assembled, and annotated the whole genome of *M. nigrescens* using nanopore long-reads and 10X chromium link-reads. The assembly is 524 Mb in size consisting of 867 scaffolds. The N50 value is 2.42 Mb, and half of the assembly is in the 30 longest scaffolds. The assembly BUSCO score from the eukaryotic database (eukaryota_odb10) indicates that the genome is 86.7% complete and 5.1% partial. The genome has a high level of heterozygosity (6.6%) with a repeat content of 78.7%. mRNA-seq reads from different sized nematodes (≤ 2 cm, 3.5-7 cm, and >7 cm body length) representing different developmental stages were also generated and used for the genome annotation. Using ab initio and evidence-based gene model predictions, 12,313 protein-coding genes and 24,186 mRNAs were annotated. These genomic resources will help researchers investigate the various aspects of the biology and host-parasite interactions of mermithid nematodes.

Introduction

Nematodes are a highly diverse group with over one million estimated species (Scheffers et al. 2012; Smythe et al. 2019). They are ubiquitous, occurring in habitats ranging from deep oceans

to mountain peaks (Schratzberger et al. 2019). Most animals and plants host at least one species of parasitic nematode (Blaxter and Koutsovoulos 2015). Despite representing an important component of all natural ecosystems (Ferris 2010; Cardoso et al. 2016), most nematode species remain undescribed (Dobson et al. 2008).

Nematodes fall into two classes based on both molecular and morphological systematics: Chromadorea and Enoplea (Blaxter et al. 1998; Liu et al. 2013). They have a conserved body plan with similar structural organisation and cellular morphologies (Basyoni and Rizk 2016). However, comparative analyses have highlighted the molecular and physiological diversity within the phylum (Mitrevu et al. 2005; Coghlan et al. 2019). *Caenorhabditis elegans* is a Chromadorean that has been at the forefront of genetic research as a model species. In contrast, there exists very little genomic information from the members of class Enoplea. Most studies on Enoplea focus on *Trichinella spiralis* (Enoplea: Trichocephalida) because of its importance as a mammalian parasite (Mitrevu et al. 2011). There are also several insect parasitic taxa within the Enoplea, but only the genome of *Romanomermis culicivorax* (Dorylaimia: Mermithidae) is publicly available (Schiffer et al. 2013). The genomes of *T. spiralis* and *R. culicivorax* differ significantly in size (64Mb vs 320Mb) and other attributes like the proportion of repetitive sequences and transposable elements, highlighting the genetic diversity within the Enoplea (Schiffer et al. 2013).

Mermithids are obligate endoparasitic nematodes. They represent a large portion of the understudied nematode class Enoplea. There are over 100 described mermithid species; however, their biology and distribution are poorly studied (Poinar 1975; Presswell et al. 2015). Past studies have paid special attention to mermithid species that have potential as biological control agents against insect pests (Petersen 1973; Abagli et al. 2019). They have adapted to a wide range of arthropod hosts, from terrestrial to aquatic taxa (Yeates and Buckley 2009; Zamani 2014; Kubo et al. 2016; Tong et al. 2021). Of the few well-studied species, many have broad geographical distributions. For example, *Isomermis* species infect several Simuliid hosts throughout Africa, Europe, and North America (Gradinarov 2014). Similarly, *Mermis nigrescens* has been reported from different parts of Europe, Asia, North and South America, Australia, and New Zealand, infecting various hosts across its broad range (Presswell et al. 2015).

Furthermore, several mermithids manipulate their host's behaviour for their own benefit (Herbison et al. 2019). Adult mermithids are free-living; after completion of their development within an arthropod host, they must emerge from the host to mate and lay eggs. As they are

prone to desiccation, they must emerge in water or water-saturated substrate. Since many mermithids parasitize terrestrial hosts, the need to emerge in water has selected for mermithids capable of inducing water-seeking behaviour in their host. For example, the mermithid nematode *Strelkovimermis spiculatus* modifies the behaviour of its adult female mosquito host to make it seek water instead of a blood meal, providing a dispersal means for the nematodes and ensuring their return to a suitable habitat for reproduction (Allahverdipour et al. 2019). Similarly, *M. nigrescens* induces water-seeking behaviour in its terrestrial earwig host (*Forficula auricularia*), enabling the nematodes to emerge into a favourable environment to continue their life cycle (Presswell et al. 2015; Herbison et al. 2019).

Here, we present a high-quality genome assembly of *Mermis nigrescens*. The life cycle, morphological, and molecular characteristics of *M. nigrescens* have been well studied (Baker and Capinera 1997; Presswell et al. 2015). However, the lack of whole-genome information greatly limits the genetic investigations of its biology and adaptations, such as its ability to alter host behaviour. This genome will also fill the knowledge and resource gap for Enoplean genomes. It will further facilitate the application of a range of molecular tools and approaches to study the genetic underpinnings of the wide geographical distribution and successful host exploitation of mermithid nematodes.

Methods

Sampling and Nucleotide extractions

Nematodes (*Mermis nigrescens*) were dissected out of the European Earwigs (*Forficula auricularia*) collected from the Dunedin Botanic Garden (Latitude: -45° 51' 27.59" S; Longitude: 170° 31' 15.56" E) and reared in a temperature-controlled room (Temperature: cycling from 15 to 12 °C, day/night; Photoperiod of L:D 16:8) in the Department of Zoology, University of Otago, Dunedin. Nematodes thus obtained were snap-frozen in liquid nitrogen and stored at -80°C until further use. Individual nematodes were used for each DNA extraction. Individuals were not sexed prior to sequencing due to the difficulty in achieving this in a manner timely enough to maintain high-quality RNA and DNA samples. DNA extracted using DNeasy® Blood & Tissue Kit (Qiagen, USA) was used for nanopore sequencing. DNA extracted using Nanobind Tissue Big DNA kit (Circulomics, USA) was used for 10x sequencing library preparation. RNase treatment to remove RNA was performed using 4µl of RNase A (10mg/ml) per 200µl of template following DNA extraction. DNA samples were

quantified and quality assessed using a Qubit 2.0 Fluorometer (Life Technologies, USA) and Nanodrop. High-quality DNA samples were stored at -20°C and were used within a week of extraction.

Total RNA from different sized nematodes (≤ 2 cm, 3.5-7cm, and > 7 cm body length) representing different developmental stages was extracted using Direct-zol™ RNA MicroPrep kit (Zymo Research) following the manufacturer's instruction and stored at -80°C until further use. RNA from five individuals from each size category (small, medium, and large) were individually extracted and later pooled at equimolarity to make a single sequencing library for each category. Samples were quantified and quality checked on a Qubit 2.0 Fluorometer (Life Technologies, USA) and Nanodrop. RNA integrity for the pooled samples was assessed using a Fragment Analyzer (Advanced Analytical Technologies Inc., USA) at the Otago Genomics Facility (OGF), University of Otago, Dunedin, New Zealand. The RNA quality number (RQN) values ranged from 8.3 to 9.5, indicating high-quality samples.

Library preparation and sequencing

A long read sequencing library for Oxford minion was prepared with 416.5 ng of genomic DNA using a ligation sequencing kit (SQK-LSK109) (Oxford Nanopore Technologies, Oxford, UK) following the manufacturer's instructions. Lambda phage (DNA CS) was used as a positive control. The prepared library was sequenced with R9 chemistry MinION flow cell (FLO-MIN106) (Oxford Nanopore Technologies, UK).

A linked read library was prepared at the Genetic Analysis Services (GAS), University of Otago (Dunedin, New Zealand). DNA was size-selected for fragments over 40 Kbp using a Blue Pippin (Sage Science, USA). A chromium 10x linked read (10x Genomics, USA) library was prepared following the manufacturer's instructions. The library was sequenced on the Illumina Nova-seq platform to generate paired-end reads (2x151bp) at the Garvan Institute, Australia.

TruSeq stranded mRNA libraries were prepared and sequenced in OGF as 2x100 bp paired-end reads across two Rapid V2 flowcell lanes of HiSeq 2500.

Genome size estimation

Genome size was estimated with flow cytometry and a genomic k-mer based approach. Flow cytometry was carried on two whole worms at Flowjoanna (Palmerston North, NZ, USA). The sample preparation protocol for the flowcytometry is described in our earlier study (Bhattarai et al. 2022). Rooster red blood cells (RRBC) from the domestic chicken (*Gallus gallus*) was

used a reference. Samples were then processed on a FACSCalibur (BD Biosciences, USA) and the data were analyzed using Flowjo (BD Biosciences, USA).

The short-read data from the 10x Chromium library were used for the k-mer based approach. We used scaff_reads script from Scaff10x (v.5.0) to remove the 10x link adapters and low-quality reads, while any remaining Illumina adapters were removed using Trimmomatic (v.0.39) with options: LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:35. A k-mer size of 21 was used with KMC (v.3.1.1) (Kokot et al. 2017) to produce a histogram that was visualised in the Genomescope (v.2.0) web browser.

Bioinformatic pipeline

We used various software and packages to assemble and refine the genome (Figure 1). The pipeline and the scripts used in this project are available on GitHub (https://github.com/upendrabhattarai/Nematode_Genome_Project). The bioinformatics software and packages were run in New Zealand eScience Infrastructure (NeSI).

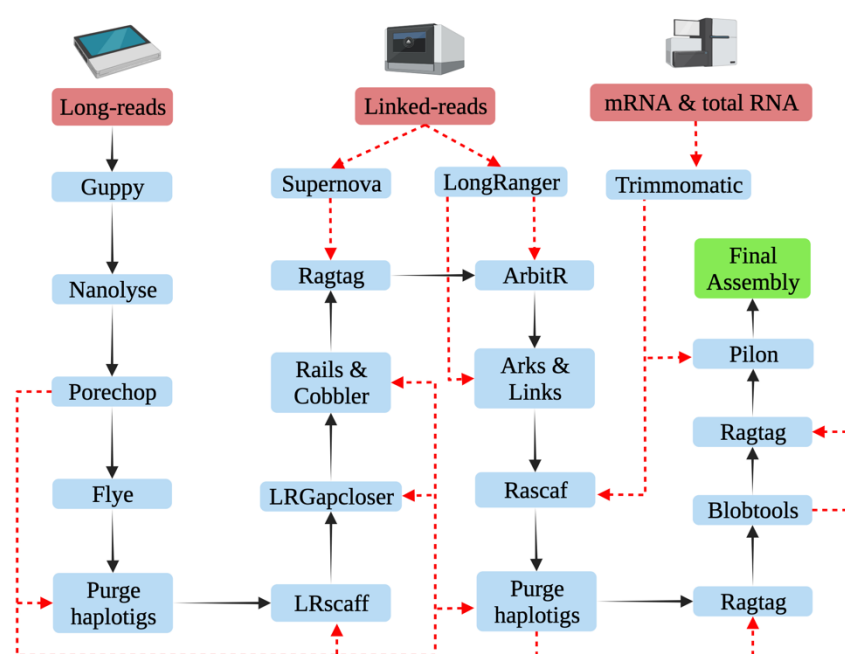


Figure 1: Schematic representation of the assembly pipeline for the *Mermis nigrescens* genome. The black arrows represent the workflow and the red dotted lines represent the input data in the pipeline (Created with BioRender.com).

Genome and transcriptome assembly

All the Nanopore reads were base called using Guppy (v.5.0.7) and assembled using Flye (v.2.9) (Kolmogorov et al. 2019) with default parameters. The paired end reads from the

chromium library were assembled using supernova (Weisenfeld et al. 2017). BUSCO (v.4.1.4) (Simão et al. 2015) and Quast (v.5.0.2) (Gurevich et al. 2013) were used to evaluate the assembly statistics. The assembly from long reads outperformed the linked reads for the BUSCO completeness and assembly contiguity. Therefore, long read assembly was used as a primary assembly, and the linked read assembly was used for scaffolding later in the pipeline (Figure 1). Multiple scaffolding and gap-closing steps were performed using Lrscaf (v.1.1.11) (Qin et al. 2019), Rails (v.1.5.1) and Cobbler (v.0.6.1) (Warren 2016), Ragtag (v.2.1.0) (Alonge et al. 2019), ArbitR (v.0.2) (Hiltunen et al. 2021), Arks (v.1.0.4) (Coombe et al. 2018), Links (v.1.8.7) (Warren et al. 2015) and Lrgapcloser (Xu et al. 2019). mRNA-seq reads obtained for genome annotation purposes, and total RNA-seq reads sequenced for another project (manuscript under preparation) were also used for scaffolding the assembly with Rascaf (Song et al. 2016). Purgehaplotigs (Roach et al. 2018) removed haplotigs from the assembly. Removed haplotig sequences were used to further scaffold the assembly with Ragtag (v.2.1.0). Blobtools2 (Laetsch and Blaxter 2017) was used to remove and discard bacterial reads and remove low coverage (<5x coverage) and short (<1000bp) scaffolds. Low coverage and short scaffolds were then used to scaffold the filtered assembly with Ragtag (v.2.1.0) (Alonge et al. 2019). Finally, Pilon (v.1.24) was used to polish the assembly using mRNA-seq data.

The adapters and low quality reads from mRNA-seq data were filtered using Trimmomatic (v.0.38) (Bolger et al. 2014) with options HEADCROP:15 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:20 MINLEN:35. Clean data was de novo assembled with Trinity (v.2.13.2) (Grabherr et al. 2011) using all the default parameters.

Repetitive content masking

A custom repeat library was generated using de novo and homology-based identifiers, including LTRharvest (Ellinghaus et al. 2008), LTRdigest (Steinbiss et al. 2009), RepeatModeler (Flynn et al. 2020), TransposonPSI (Brian Haas 2010), SINEBase (Vassetzky and Kramerov 2013), and MITE-Tracker (Crescente et al. 2018). Individual libraries from each software were concatenated, and sequences with more than 80% similarity were merged to remove redundancy using usearch (v.11.0.667) (Edgar 2010). The library was then classified using RepeatClassifier. Sequences with unknown categories in the library were mapped against the reviewed nematode database from UniProtKB/Swiss-Prot database (e-value <1e-01); if not annotated as repeat sequences, they were removed from the library. The final repeat library was used in RepeatMasker (v.4.1.0) (Chen 2004) to mask the repeats and generate a report. The repeat library was used as an input to the MAKER2 pipeline (Holt and Yandell 2011).

Genome annotation

The assembly was annotated using evidence-based and ab initio gene model predictions through MAKER2 (v.2.31.9) (Holt and Yandell 2011) and Braker (v.2.16) (Hoff et al. 2019) pipelines. The first round of MAKER2 was run with 180,630 mRNA transcripts *denovo* assembled with the Trinity pipeline (Grabherr et al. 2011) along with 511,117 mRNA and 1,148,233 protein sequences from all the *Dorylaimia* species available in NCBI and WormBase (<https://wormbase.org>) databases. mRNA-seq reads and output from GenMark-ES were used to train Augustus with the Braker pipeline. SNAP was trained after each round of MAKER2. Trained SNAP and Augustus were used for two more iterations of the MAKER2 pipeline.

We used InterProScan (v.5.51-85.0) (Jones et al. 2014) for the functional annotation on the predicted protein sequences from MAKER and retrieved InterPro ID, PFAM domains, and Gene Ontology (GO) terms. Furthermore, the Uniprot database with BLAST was used to assign gene descriptors to each transcript based on the best hit.

Results and Discussion

Genome size estimates

The flowcytometry estimate of the nematode genome size is 814.185±24.2 Mb (mean ±SD), however, the nematode samples yielded very low nuclei counts, i.e. as low as <200 total nuclei for one of the samples. Given that this is a whole animal digest and the very low nuclei counts in the samples, the calculated pg/nucleus may not be accurate. To get better genome size estimates by flowcytometry in this nematode, we recommend increasing the sample size and replicated measurements from the nuclei suspension in future experiments.

The k-mer based analysis estimated the genome size to be 336.07 Mb with 6.56% of heterozygosity (Figure 2). A high level of heterozygosity can result in genome assemblers generating multiple copies of the region, leading to larger genome size and reduced contiguity (Guan et al. 2020; Asalone et al. 2020). This may explain why the assembled genome size (524.2 Mb) is substantially larger than the k-mer estimates. Furthermore, as *M. nigrescens* can also reproduce parthenogenetically (Christie J. R. 1937), we assume that the high level of heterozygosity in their genome is associated with their parthenogenetic mode of reproduction. Similarly, a highly heterozygous genome (5.7%) was also reported for another parthenogenetic nematode, *Diploscapter coronatus* (Hiraki et al. 2017).

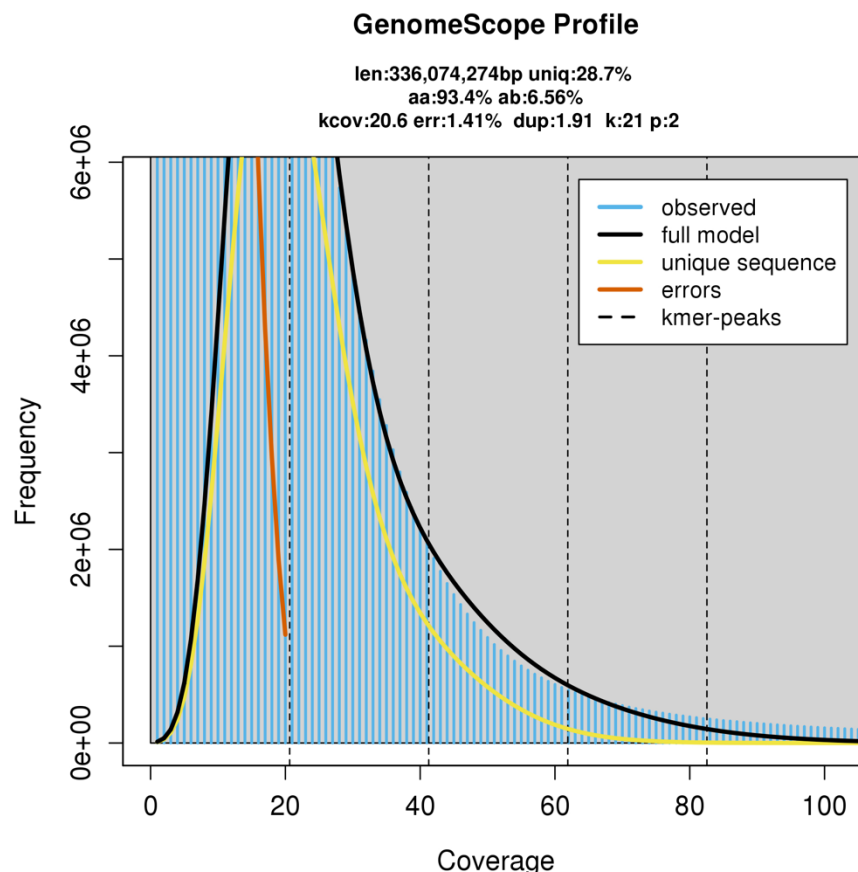


Figure 2: The k-mer analysis of *Mermis nigrescens* genome using short reads. A k-mer size of 21 was used in KMC (v.3.1.1) to produce a histogram and was visualised in the Genomescope (v.2.0) web browser. It shows a genome size of 336.07 Mb with heterozygosity of 6.56%.

Genome and transcriptome assembly

A total of 6.9 Gb of sequencing data was generated from a Nanopore minion flowcell run consisting of 530,888 reads with an N50 length of 17,849 bp (Table 1). Flye (v.2.7.1) produced the primary long read-assembly with 16,414 contigs with N50 of 94,974 bp and a total length of ~714 Mb. Quast reported 88.45% complete BUSCO from the Eukaryote database for this assembly.

Table 1. Sequencing output from Oxford nanopore minion.

Reads	Bases	Median Read Length	N50 length	Median Read Quality
530,888	6,935,586,657	11,833	17,849	13.63

The 10x chromium library yielded 450.3 million paired-end reads. These linked reads with the Supernova assembler produced an assembly of 582.5 Mb with 147,205 scaffolds and N50 of

9,664 bp. Quast reported only 20.13% of complete BUSCO genes for this assembly (Table 2). Comparing the contiguity and completeness of these two assemblies, we used Flye assembly as a primary assembly and the supernova assembly was used in scaffolding steps later in the pipeline (Figure 1).

Table 2. Assembly statistics for the genome assembly of *Mermis nigrescens*.

	Assembly length	No. of scaffolds	N50	L50	Ns per 100 kbp	Busco % from Quast (eukaryota odb10)	
						Complete	Partial
Flye	714,265,163	16,414	94,974	2,079	0.45	88.45	4.29
Supernova	582,450,576	147,205	9,664	12,318	732.59	20.13	21.45
Final Assembly	524,220,005	867	2,429,002	30	240.96	86.70	5.10

The assembly statistics after each round of processing are presented in Table 3. The final assembly has a size of 524.2 Mb with 867 scaffolds. The assembly has the N50 of 2,429,002 bp, and 50% of the assembly is in the 30 longest scaffolds. The k-mer based analysis estimated the genome size to be 336 Mb with a heterozygosity of 6.56%.. Gaps in the assembly account for 240.97 bp per 100 Kbp length. The final assembly has a complete BUSCO score of 86.7% and a partial BUSCO score of 5.1% to the eukaryotic database and (Figure 3A).

Table 3. Assembly statistics after each round of processing.

Assembly steps	Assembly length	No. of scaffolds	N50	L50	Ns per 100 kbp	Busco % from Quast (eukaryota odb10)	
						Complete	Partial
Flye	714,265,163	16,414	94,974	2,079	0.45	88.45	4.29
Purgehaplotigs	547,497,121	8,101	125,056	1,325	0.57	85.15	7.59
Lrscaff	739,840,003	5,500	263,548	890	2122.06	88.45	3.96
LRGapcloser	739,845,718	5,500	263,608	890	283.38	88.45	3.96
Rails & Cobbler	740,070,390	5,454	266,049	885	256.75	89.11	3.30
Ragtag	740,076,890	5,389	266,578	883	257.63	89.11	3.30
ArbitR	738,644,384	4,958	308,460	738	262.84	89.11	3.30
ARKS & LINKS	738,665,484	4,747	328,769	687	265.69	88.45	3.96
Rascaf	738,576,542	4,343	388,132	575	253.68	88.45	3.96
Purgehaplotigs	587,968,296	3,113	443,562	412	244.20	86.14	5.61
Ragtag	588,012,496	2,671	533,508	328	251.70	86.14	5.61
Blobtools	524,163,746	1541	553,974	285	228.13	80.86	5.94
Ragtag	524,231,146	867	2,428,984	30	240.95	80.86	6.27
Pilon	524,220,005	867	2,429,002	30	240.96	86.70	5.10

mRNA-seq libraries generated 288.9 million paired-end reads. They were *de novo* assembled using Trinity (v.2.13.2) (Grabherr et al. 2011) producing 180,630 transcripts with N50 of 1,362 bp including 103,144 trinity genes. The Busco score from the nematoda_odb10 database was 63% for the transcriptome assembly and 55.2% complete for the genome assembly (Figure 3B). This suggests that the nematode Busco database (nematoda_odb10) does not provide a good representation of the *M. nigrescens* genome, further highlighting the high genetic diversity among nematodes.

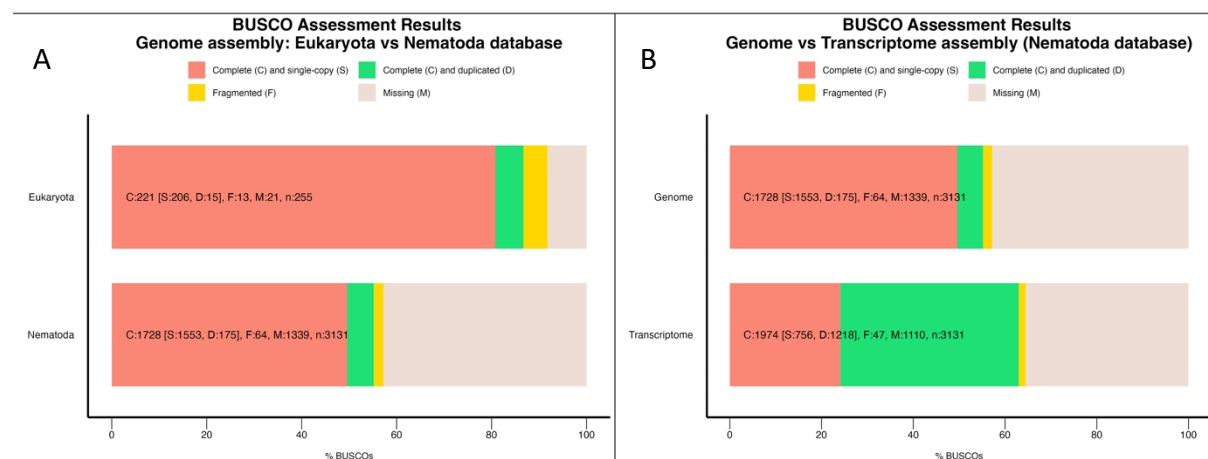


Figure 3: The BUSCO (v.4.1.4) analysis report for the genome and transcriptome assembly of *Mermis nigrescens*. A) Comparison of the BUSCO scores from the Eukaryote (eukaryota_odb10) and the Nematode (nematoda_odb10) database for the genome. B) Comparison of BUSCO score for the genome and transcriptome assemblies with the nematoda_odb10 database.

Romanomermis culicivorax is the only mermithid nematode with a publicly available whole-genome assembly (Schiffer et al. 2013) to date. Its genome assembly has a 322 Mb size with N50 of 17,632 bp. It has 62,537 scaffolds and 37.3% complete Busco genes (nematoda_odb10, n=3131). Between these two assemblies, the *M. nigrescens*' genome assembly is more contiguous and complete and can be used as a better representative of mermithid nematodes.

Genes and repeats annotation

A total of 12,313 protein coding genes and 24,186 mRNAs were identified in the genome. The mean gene length is 12,224 bp, and the total gene length is 150.5 Mb. The longest gene annotated is 206,129 bp (Table 4). Functional annotation resulted in 7,496 genes annotated to the InterPro and Pfam databases and 4,623 annotated to Gene Ontology terms (Table 5). Quality of annotation was measured using Annotation Edit Distance (AED) score, and 96% of

the annotated genes have an AED score of 0.5 or less, ensuring highly confident gene prediction (Figure 4).

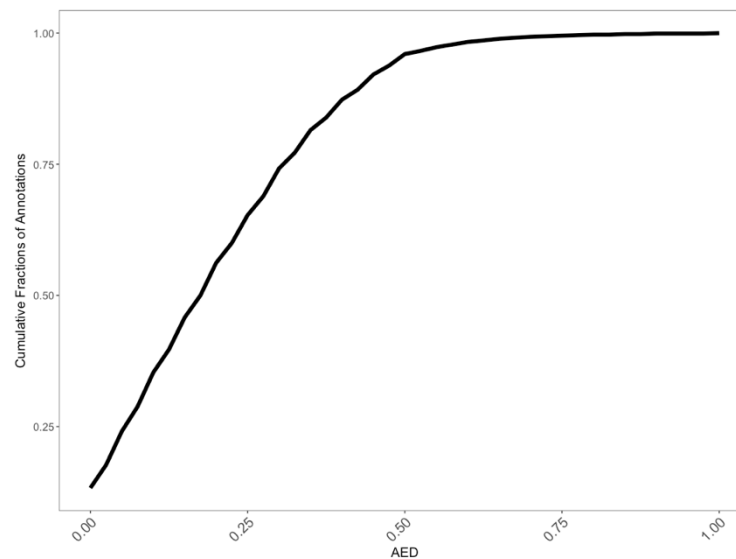


Figure 4: Annotation quality with AED scores. The y-axis shows the cumulative fractions of annotations and the x-axis their AED scores.

Table 4. Genome annotation summary for *Mermis nigrescens*.

Number of genes	12,313
Number of mRNAs	24,186
Number of exons	152,566
Number of introns	1,073,659
Number of CDS	24,185
Total gene length	150,517,544
Total mRNA length	298,339,285
Total exon length	30,891,521
Total intron length	2,450,417,037
Total CDS length	23,760,135
Longest gene	206,129
Longest mRNA	206,129
Longest exon	6,519
Longest intron	539,449
Longest CDS	19,164
mean gene length	12,224
mean mRNA length	12,335
mean exon length	202
mean intron length	2,282
mean CDS length	982

Table 5. Functional annotation statistics from different databases.

Database	No. of terms linked to mRNA	No. of mRNA with term	No. of gene with term
InterPro	21,702	16,007	7,496
Gene Ontology	20,530	10,180	4,623
Pfam	21,954	16,007	7,496

The repeat contents in the genome amount to 412.78 Mb comprising 78.74% of the whole assembly. It includes 17.86% of retroelements, 21.97% of DNA transposons, 1.09% rolling circles and 37.18% unclassified repeatomes (Table 6). Compared to other nematode species, this is a very high proportion of repeats in the genome. For example, the repeat contents in *R. culcivorax* genome is 47% (Schiffer et al. 2013), in *Trichinella spiralis* genome it is 19.8% (Mitrevva et al. 2011) and in *Caenorhabditis elegans* genome it is 17% (*C. elegans* Sequencing Consortium 1998). At present, we do not know the significance of such a highly repetitive genome in *M. nigrescens*. However, further investigations might shed light on its biological and evolutionary significance.

Table 6. Repeat content analysis in *Mermis nigrescens* genome assembly.

No. of sequences:	867			
Total length (bp):	524,220,005			
GC level:	36.63%			
Bases masked:	412,783,310 bp (78.74%)			
		Numbers*	Length (bp)	Percentage
Retroelements		305,036	93,643,203	17.86%
	SINEs:	2,942	369,353	0.07%
	Penelope	25,117	7,072,788	1.35%
	LINEs:	147,865	40,170,331	7.66%
	CRE/SLACS	-	-	0.00%
	L2/CR1/Rex	85,577	23,760,597	4.53%
	R1/LOA/Jockey	80	54,133	0.01%
	R2/R4/NeSL	166	29,796	0.01%
	RTE/Bov-B	13,943	3,447,682	0.66%
	L1/CIN4	13,163	2,091,210	0.40%
	LTR elements:	154,229	53,103,519	10.13%
	BEL/Pao	8,192	2,959,770	0.56%
	Ty1/Copia	731	84,377	0.02%
	Gypsy/DIRS1	98,159	41,038,229	7.83%
	Retroviral	17,348	3,301,532	0.63%
DNA transposons		503,758	115,186,810	21.97%

	hobo-Activator	105,906	20,614,222	3.93%
	Tc1-IS630-Pogo	36,016	7,078,658	1.35%
	En-Spm	-	-	0.00%
	MuDR-IS905	-	-	0.00%
	PiggyBac	1,737	704,735	0.13%
	Tourist/Harbinger	25,116	5,166,803	0.99%
	Other (Mirage, P-element, Transib)	-	-	0.00%
Rolling-circles		28,792	5,739,966	1.09%
Unclassified:		777,756	194,902,752	37.18%
Total interspersed repeats:			403,732,765	77.02%
Small RNA:		15,557	2,652,405	0.51%
Satellites:		2,895	658,174	0.13%
Simple repeats:		-	-	0.00%
Low complexity:		-	-	0.00%

* most repeats fragmented by insertions or deletions have been counted as one element.

Conclusions

This study presented a high quality genome assembly, annotation, and analysis of *M. nigrescens*. The genome was assembled using the long and linked-read sequencing data. Transcriptomic data from different developmental stages of the nematode was also generated. The *M. nigrescens* genome showed very high level of repeat content compared to other nematode genomes. These new resources and information will contribute to the better understanding of the genomic architecture and biology of mermithid nematodes and their adaptation to the broad range of insect the infect.

Reference:

- Abagli AZ, Alavo TBC, Perez-Pacheco R, Platzer EG (2019) Efficacy of the mermithid nematode, *Romanomermis iyengari*, for the biocontrol of *Anopheles gambiae*, the major malaria vector in sub-Saharan Africa. Parasit Vectors 12:253. <https://doi.org/10.1186/s13071-019-3508-6>
- Allahverdipour HH, Talaei-Hassanloui R, Karimi J, et al (2019) Behavior manipulation of mosquitoes by a mermithid nematode. J Invertebr Pathol 168:107273. <https://doi.org/10.1016/j.jip.2019.107273>
- Alonge M, Soyk S, Ramakrishnan S, et al (2019) RaGOO: fast and accurate reference-guided scaffolding of draft genomes. Genome Biol 20:224. <https://doi.org/10.1186/s13059-019-1829-6>
- Asalone KC, Ryan KM, Yamadi M, et al (2020) Regional sequence expansion or collapse in heterozygous genome assemblies. PLoS Comput Biol 16:e1008104
- Baker GL, Capinera JL (1997) Nematodes and Nematomorphs as control agents of Grasshoppers and Locusts. Memoirs of the Entomological Society of Canada 129:157–211. <https://doi.org/10.4039/entm129171157-1>
- Basyoni MMA, Rizk EMA (2016) Nematodes ultrastructure: complex systems and processes. Journal of Parasitic Diseases 40:1130–1140. <https://doi.org/10.1007/s12639-015-0707-8>
- Bhattarai UR, Katuwal M, Poulin R, et al (2022) Genome assembly and annotation of the European earwig *Forficula auricularia* (subspecies B). G3 Genes|Genomes|Genetics 12:. <https://doi.org/10.1093/g3journal/jkac199>
- Blaxter M, Koutsovoulos G (2015) The evolution of parasitism in Nematoda. Parasitology 142:S26–S39. <https://doi.org/10.1017/S0031182014000791>
- Blaxter ML, De Ley P, Garey JR, et al (1998) A molecular evolutionary framework for the phylum Nematoda. Nature 392:71–75. <https://doi.org/10.1038/32160>
- Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
- Brian Haas (2010) TransposonPSI: An Application of PSI-Blast to Mine (Retro-)Transposon ORF Homologies. <https://transposonpsi.sourceforge.net/>

- C elegans Sequencing Consortium (1998) Genome Sequence of the Nematode *C. elegans* : A Platform for Investigating Biology. Science (1979) 282:2012–2018. <https://doi.org/10.1126/science.282.5396.2012>
- Chen N (2004) Using RepeatMasker to Identify Repetitive Elements in Genomic Sequences. Curr Protoc Bioinformatics 5:4.10.1–4.10.14. <https://doi.org/10.1002/0471250953.bi0410s05>
- Christie J. R. (1937) *Mermis subnigrescens*, A nematode parasite. J Agric Res 55:353–364
- Coghlan A, Tyagi R, Cotton JA, et al (2019) Comparative genomics of the major parasitic worms. Nat Genet 51:163–174. <https://doi.org/10.1038/s41588-018-0262-1>
- Coombe L, Zhang J, Vandervalk BP, et al (2018) ARKS: chromosome-scale scaffolding of human genome drafts with linked read kmers. BMC Bioinformatics 19:234. <https://doi.org/10.1186/s12859-018-2243-x>
- Crescente JM, Zavallo D, Helguera M, Vanzetti LS (2018) MITE Tracker: an accurate approach to identify miniature inverted-repeat transposable elements in large genomes. BMC Bioinformatics 19:348. <https://doi.org/10.1186/s12859-018-2376-y>
- Dobson A, Lafferty KD, Kuris AM, et al (2008) Homage to Linnaeus: How many parasites? How many hosts? Proceedings of the National Academy of Sciences 105:11482 LP – 11489. <https://doi.org/10.1073/pnas.0803232105>
- Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26:2460–2461. <https://doi.org/10.1093/bioinformatics/btq461>
- Ellinghaus D, Kurtz S, Willhoeft U (2008) LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. BMC Bioinformatics 9:18. <https://doi.org/10.1186/1471-2105-9-18>
- Ferris H (2010) Contribution of nematodes to the structure and function of the soil food web. J Nematol 42:63–7
- Flynn JM, Hubley R, Goubert C, et al (2020) RepeatModeler2 for automated genomic discovery of transposable element families. Proceedings of the National Academy of Sciences 117:9451–9457. <https://doi.org/10.1073/pnas.1921046117>

- Grabherr MG, Haas BJ, Yassour M, et al (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* 29:644–652. <https://doi.org/10.1038/nbt.1883>
- Gradinarov D (2014) The mermithid species *Isomermis lairdi* (Nematoda, Mermithidae), previously only known in Africa, found in Europe. *Zookeys* 454:1–11. <https://doi.org/10.3897/zookeys.454.7577>
- Guan D, McCarthy SA, Wood J, et al (2020) Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics* 36:2896–2898. <https://doi.org/10.1093/bioinformatics/btaa025>
- Gurevich A, Saveliev V, Vyahhi N, Tesler G (2013) QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29:1072–1075. <https://doi.org/10.1093/bioinformatics/btt086>
- Herbison REH, Evans S, Doherty J-F, Poulin R (2019) Let's go swimming: mermithid-infected earwigs exhibit positive hydrotaxis. *Parasitology* 146:1631–1635. <https://doi.org/10.1017/S0031182019001045>
- Hiltunen M, Ryberg M, Johannesson H (2021) ARBitR: an overlap-aware genome assembly scaffolder for linked reads. *Bioinformatics* 37:2203–2205. <https://doi.org/10.1093/bioinformatics/btaa975>
- Hiraki H, Kagoshima H, Kraus C, et al (2017) Genome analysis of *Diploscapter coronatus*: insights into molecular peculiarities of a nematode with parthenogenetic reproduction. *BMC Genomics* 18:478. <https://doi.org/10.1186/s12864-017-3860-x>
- Hoff KJ, Lomsadze A, Borodovsky M, Stanke M (2019) Whole-Genome Annotation with BRAKER. In: *Methods in molecular biology* (Clifton, N.J.). pp 65–95
- Holt C, Yandell M (2011) MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* 12:491. <https://doi.org/10.1186/1471-2105-12-491>
- Jones P, Binns D, Chang H-Y, et al (2014) InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30:1236–1240. <https://doi.org/10.1093/bioinformatics/btu031>
- Kokot M, Długosz M, Deorowicz S (2017) KMC 3: counting and manipulating k-mer statistics. *Bioinformatics* 33:2759–2761. <https://doi.org/10.1093/bioinformatics/btx304>

- Kolmogorov M, Yuan J, Lin Y, Pevzner PA (2019) Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol* 37:540–546. <https://doi.org/10.1038/s41587-019-0072-8>
- Kubo R, Ugajin A, Ono M (2016) Molecular phylogenetic analysis of mermithid nematodes (Mermithida: Mermithidae) discovered from Japanese bumblebee (Hymenoptera: Bombinae) and behavioral observation of an infected bumblebee. *Appl Entomol Zool* 51:549–554. <https://doi.org/10.1007/s13355-016-0430-7>
- Laetsch DR, Blaxter ML (2017) BlobTools: Interrogation of genome assemblies. *F1000Res* 6:1287. <https://doi.org/10.12688/f1000research.12232.1>
- Liu G-H, Shao R, Li J-Y, et al (2013) The complete mitochondrial genomes of three parasitic nematodes of birds: a unique gene order and insights into nematode phylogeny. *BMC Genomics* 14:414. <https://doi.org/10.1186/1471-2164-14-414>
- Mitreva M, Blaxter ML, Bird DM, McCarter JP (2005) Comparative genomics of nematodes. *Trends in Genetics* 21:573–581. <https://doi.org/10.1016/j.tig.2005.08.003>
- Mitreva M, Jasmer DP, Zarlenga DS, et al (2011) The draft genome of the parasitic nematode *Trichinella spiralis*. *Nat Genet* 43:228–235. <https://doi.org/10.1038/ng.769>
- Petersen JJ (1973) Role of mermithid nematodes in biological control of mosquitoes. *Exp Parasitol* 33:239–247. [https://doi.org/10.1016/0014-4894\(73\)90030-1](https://doi.org/10.1016/0014-4894(73)90030-1)
- Poinar GO (1975) Entomogenous nematodes: a manual and host list of insect-nematode associations. Colombia, Brill.
- Presswell B, Evans S, Poulin R, Jorge F (2015) Morphological and molecular characterization of *Mermis nigrescens* Dujardin, (Nematoda: Mermithidae) parasitizing the introduced European earwig (Dermaptera: Forficulidae) in New Zealand. *J Helminthol* 89:267–276. <https://doi.org/10.1017/S0022149X14000017>
- Qin M, Wu S, Li A, et al (2019) LRScaf: improving draft genomes using long noisy reads. *BMC Genomics* 20:955. <https://doi.org/10.1186/s12864-019-6337-2>
- Roach MJ, Schmidt SA, Borneman AR (2018) Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. *BMC Bioinformatics* 19:460. <https://doi.org/10.1186/s12859-018-2485-7>

- S. O. Cardoso M, M. R. Pedrosa E, Ferris H, et al (2016) Nematode Fauna of Tropical Rainforest in Brazil: A Descriptive and Seasonal Approach. J Nematol 48:116–125. <https://doi.org/10.21307/jofnem-2017-017>
- Scheffers BR, Joppa LN, Pimm SL, Laurance WF (2012) What we know and don't know about Earth's missing biodiversity. Trends Ecol Evol 27:501–510. <https://doi.org/10.1016/j.tree.2012.05.008>
- Schiffer PH, Kroihner M, Kraus C, et al (2013) The genome of *Romanomermis culicivorax*: revealing fundamental changes in the core developmental genetic toolkit in Nematoda. BMC Genomics 14:923. <https://doi.org/10.1186/1471-2164-14-923>
- Schratzberger M, Holterman M, van Oevelen D, Helder J (2019) A Worm's World: Ecological Flexibility Pays Off for Free-Living Nematodes in Sediments and Soils. Bioscience 69:867–876. <https://doi.org/10.1093/biosci/biz086>
- Simão FA, Waterhouse RM, Ioannidis P, et al (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31:3210–3212. <https://doi.org/10.1093/bioinformatics/btv351>
- Smythe AB, Holovachov O, Kocot KM (2019) Improved phylogenomic sampling of free-living nematodes enhances resolution of higher-level nematode phylogeny. BMC Evol Biol 19:121. <https://doi.org/10.1186/s12862-019-1444-x>
- Song L, Shankar DS, Florea L (2016) Rascaf: Improving Genome Assembly with RNA Sequencing Data. Plant Genome 9(3). <https://doi.org/10.3835/plantgenome2016.03.0027>
- Steinbiss S, Willhoeft U, Gremme G, Kurtz S (2009) Fine-grained annotation and classification of de novo predicted LTR retrotransposons. Nucleic Acids Res 37:7002–7013. <https://doi.org/10.1093/nar/gkp759>
- Tong X, Kanzaki N, Akimoto S (2021) First record of a mermithid nematode (Nematoda: Mermithidae) parasitizing winged females of gall-forming aphids (emiptera: phididae: riosomatinae). Entomol Sci 24:354–360. <https://doi.org/10.1111/ens.12486>
- Vassetzky NS, Kramerov DA (2013) SINEBase: a database and tool for SINE analysis. Nucleic Acids Res 41:D83–D89. <https://doi.org/10.1093/nar/gks1263>
- Warren RL (2016) RAILS and Cobbler: Scaffolding and automated finishing of draft genomes using long DNA sequences. The Journal of Open Source Software 1:116. <https://doi.org/10.21105/joss.00116>

- Warren RL, Yang C, Vandervalk BP, et al (2015) LINKS: Scalable, alignment-free scaffolding of draft genomes with long reads. *Gigascience* 4:35. <https://doi.org/10.1186/s13742-015-0076-3>
- Weisenfeld NI, Kumar V, Shah P, et al (2017) Direct determination of diploid genome sequences. *Genome Res* 27:757–767. <https://doi.org/10.1101/gr.214874.116>
- Xu G-C, Xu T-J, Zhu R, et al (2019) LR_Gapcloser: a tiling path-based gap closer that uses long reads to complete genome assembly. *Gigascience* 8 (1), giy157 <https://doi.org/10.1093/gigascience/giy157>
- Yeates GW, Buckley TR (2009) First records of mermithid nematodes (Nematoda: Mermithidae) parasitising stick insects (Insecta: Phasmatodea). *N Z J Zool* 36:35–39. <https://doi.org/10.1080/03014220909510137>
- Zamani A (2014) First report of a mermithid nematode (Enoplea: Mermithida) parasitizing the crab spider *Heriaeus spinipalpus* (Araneae: Thomisidae). *Acta Arachnol* 63:63–64. <https://doi.org/10.2476/asjaa.63.63>