DATA NOTE

The genome sequence of an ichneumonid wasp, *Netelia melanura* (Thomson, 1888) (Hymenoptera: Ichneumonidae)

[version 1; peer review: 3 approved]

Lucy Broad [1], Gavin R. Broad [2],
Natural History Museum Genome Acquisition Lab,
Darwin Tree of Life Barcoding Collective,
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, Darwin Tree of Life Consortium

[1]Independent researcher, Tonbridge, Kent, England, UK
[2]Natural History Museum, London, England, UK

v1 **First published:** 06 Oct 2025, **10**:556
https://doi.org/10.12688/wellcomeopenres.24927.1
Latest published: 06 Oct 2025, **10**:556
https://doi.org/10.12688/wellcomeopenres.24927.1

Abstract

We present a genome assembly from an individual male *Netelia melanura* (ichneumonid wasp; Arthropoda; Insecta; Hymenoptera; Ichneumonidae). The genome sequence has a total length of 253.87 megabases. Most of the assembly (94.81%) is scaffolded into 7 chromosomal pseudomolecules. The mitochondrial genome has also been assembled, with a length of 28.04 kilobases. This assembly was generated as part of the Darwin Tree of Life project, which produces reference genomes for eukaryotic species found in Britain and Ireland.

Keywords

Netelia melanura; ichneumonid wasp; genome sequence; chromosomal; Hymenoptera

This article is included in the Tree of Life gateway.

Open Peer Review

Approval Status ✓ ✓ ✓

	1	2	3
version 1 06 Oct 2025	✓ view	✓ view	✓ view

1. **Alexandra Cerqueira De Araujo**, University College London, London, UK
2. **Alejandro Zaldívar Riverón**, Universidad Nacional Autónoma de México, Ciudad de Mexico, Mexico
3. **Bhagya C. Thimmappa**, USDA-ARS Beltsville Agricultural Research Center, Beltsville, USA

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

Corresponding author: Darwin Tree of Life Consortium (mark.blaxter@sanger.ac.uk)

Author roles: Broad L: Investigation, Resources; **Broad GR**: Investigation, Resources, Writing – Original Draft Preparation, Writing – Review & Editing;

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by Wellcome through core funding to the Wellcome Sanger Institute (220540) and the Darwin Tree of Life Discretionary Award (218328).
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

How to cite this article: Broad L, Broad GR, Natural History Museum Genome Acquisition Lab *et al.* **The genome sequence of an ichneumonid wasp, *Netelia melanura* (Thomson, 1888) (Hymenoptera: Ichneumonidae) [version 1; peer review: 3 approved]** Wellcome Open Research 2025, **10**:556 https://doi.org/10.12688/wellcomeopenres.24927.1

First published: 06 Oct 2025, **10**:556 https://doi.org/10.12688/wellcomeopenres.24927.1

Species taxonomy

Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Protostomia; Ecdysozoa; Panarthropoda; Arthropoda; Mandibulata; Pancrustacea; Hexapoda; Insecta; Dicondylia; Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Ichneumonoidea; Ichneumonidae; Tryphoninae; Phytodietini; *Netelia*; *Netelia*; *Netelia melanura* (Thomson, 1888) (NCBI:txid2922079)

Background

Netelia melanura is a nocturnal ichneumonid wasp (Darwin wasp), one of several hundred species of *Netelia*, probably the largest genus of the subfamily Tryphoninae. As with other *Netelia* species, *N. melanura* is a koinobiont ectoparasitoid of Lepidoptera larvae. The host caterpillar is temporarily paralysed and an egg laid behind its head, the *Netelia* larva then feeds externally, completing its development once the host has formed a pupation retreat, with the *Netelia* spinning a dense, dark brown, silken cocoon in which to pupate. The taxonomy and identification of European *Netelia* are complicated, but *N. melanura* is relatively easy to identify, with the end of the metasoma sharply black, the inter-ocellar area (stemmaticum) black and the temples sharply narrowed behind the eyes. *Netelia opacula* (Thomson) males can look similar but the temples are wider and the genitalia are distinctly different: in *N. melanura* there is a rounded, ovoid pad whereas in *N. opacula* this pad forms two narrow 'arms'.

The taxonomy of European *Netelia* was tackled by Delrio (1975); despite this, some authors (e.g. Kasparyan & Tolkanitz, 1999) have used the name '*Netelia testacea*' for this species. Broad *et al.* (in prep.) are completing a review of European *Netelia* which should clarify the correct names for all the species, as well as summarising their ecology. *Netelia melanura* is a plurivoltine species, on the wing in Britain from late May to October. Although collected fairly frequently throughout Britain, mainly by light trapping, and found across the Palaearctic as far East as Japan (Konishi, 2005), the only reliable host record is of an unidentified species of Noctuidae feeding on low vegetation (Broad *et al.*, in prep.).

The genome of *Netelia melanura*, along with the increasing numbers of genomes for other *Netelia* species, will be of particular use in investigating population change in these easily sampled nocturnal parasitoids.

Methods

Sample acquisition and DNA barcoding

The specimen used for genome sequencing was an adult male *Netelia melanura* (specimen ID NHMUK014425920, ToLID iyNetMela1; Figure 1), collected from Bigbury, England, United Kingdom (latitude 50.3, longitude –3.87) on 2021-08-20. The specimen was collected by Lucy Broad and identified by Gavin Broad (Natural History Museum). For the Darwin Tree of Life sampling and metadata approach, refer to Lawniczak *et al.* (2022).

The initial identification was verified by an additional DNA barcoding process according to the framework developed by Twyford *et al.* (2024). A small sample was dissected from the specimen and stored in ethanol, while the remaining parts were shipped on dry ice to the Wellcome Sanger Institute (WSI) (see the protocol). The tissue was lysed, the COI marker region was amplified by PCR, and amplicons were sequenced and compared to the BOLD database, confirming the species identification (Crowley *et al.*, 2023). Following whole genome sequence generation, the relevant DNA barcode region was also used alongside the initial barcoding data for sample tracking at the WSI (Twyford *et al.*, 2024). The standard operating procedures for Darwin Tree of Life barcoding are available on protocols.io.

Figure 1. Photograph of the *Netelia melanura* (iyNetMela1) specimen used for genome sequencing.

Nucleic acid extraction

Protocols for high molecular weight (HMW) DNA extraction developed at the Wellcome Sanger Institute (WSI) Tree of Life Core Laboratory are available on protocols.io (Howard *et al.*, 2025). The iyNetMela1 sample was weighed and triaged to determine the appropriate extraction protocol. Tissue from the head | thorax was homogenised by powermashing using a PowerMasher II tissue disruptor.

HMW DNA was extracted in the WSI Scientific Operations core using the Automated MagAttract v2 protocol. We used centrifuge-mediated fragmentation to produce DNA fragments in the 8–10 kb range, following the Covaris g-TUBE protocol for ultra-low input (ULI). Sheared DNA was purified by manual SPRI (solid-phase reversible immobilisation). The concentration of the sheared and purified DNA was assessed using a Nanodrop spectrophotometer and Qubit Fluorometer using the Qubit dsDNA High Sensitivity Assay kit. Fragment size distribution was evaluated by running the sample on the FemtoPulse system.

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 using the Sequel IIe system (Pacific Biosciences, California, USA). The concentration of the library loaded onto the Sequel IIe was in the range 40–135 pM. The SMRT link software, a PacBio web-based end-to-end workflow manager, was used to set-up and monitor the run, and to perform primary and secondary analysis of the data upon completion.

Hi-C

Sample preparation and crosslinking

The Hi-C sample was prepared from 20–50 mg of frozen tissue from the head and thorax of the iyNetMela1 sample using the Arima-HiC v2 kit (Arima Genomics). Following the manufacturer's instructions, tissue was fixed and DNA crosslinked using TC buffer to a final formaldehyde concentration of 2%. The tissue was homogenised using the Diagnocine Power Masher-II. Crosslinked DNA was digested with a restriction enzyme master mix, biotinylated, and ligated. Clean-up was performed with SPRISelect beads before library preparation. DNA concentration was measured with the Qubit Fluorometer (Thermo Fisher Scientific) and Qubit HS Assay Kit. The biotinylation percentage was estimated using the Arima-HiC v2 QC beads.

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 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 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/μL. Normalised libraries were quantified again and 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 6000.

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 (Cheng *et al.*, 2021) with the flags --primary and -l0 (the latter is to switch off internal hifiasm purging). The Hi-C reads (Rao *et al.*, 2014) were mapped to the primary contigs using bwa-mem2 (Vasimuddin *et al.*, 2019), and the contigs were scaffolded in YaHS (Zhou *et al.*, 2023) with the --break option for handling 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), which runs MitoFinder (Allio *et al.*, 2020) and uses these annotations to select the final mitochondrial contig and to ensure the general quality of the sequence. The mitochondrial genome was assembled using OATK (Zhou *et al.*, 2025).

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 115 breaks and 262 joins. The curation process is documented at https://gitlab.com/wtsi-grit/rapid-curation. PretextSnapshot was used to generate a Hi-C contact map of the final assembly.

Assembly quality assessment

The Merqury.FK tool (Rhie *et al.*, 2020) was run in a Singularity container (Kurtzer *et al.*, 2017) to evaluate *k*-mer completeness and assembly quality 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 the 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 containerisation through Docker (Merkel, 2014) and Singularity (Kurtzer *et al.*, 2017).

Genome sequence report

Sequence data

PacBio sequencing of the *Netelia melanura* specimen generated 20.40 Gb (gigabases) from 2.25 million reads, which were used to assemble the genome. GenomeScope2.0 analysis estimated the haploid genome size at 218.52 Mb, with a heterozygosity of 0.24% and repeat content of 16.13% (Figure 2). These estimates guided expectations for the assembly. Based on the estimated genome size, the sequencing data provided approximately 83× coverage. Hi-C sequencing produced 101.16 Gb from 669.96 million reads, which were used to scaffold the assembly. Table 1 summarises the specimen and sequencing details.

Assembly statistics

This is a haploid assembly of a male specimen. The final assembly has a total length of 253.87 Mb in 631 scaffolds, with 1 516 gaps, and a scaffold N50 of 39.45 Mb (Table 2).

Most of the assembly sequence (94.81%) was assigned to 7 chromosomal-level scaffolds. These chromosome-level scaffolds, confirmed by Hi-C data, are named according to size (Figure 3; Table 3). A close relative has a karyotype of 6. This assembly may also have 6 chromosomes, but it has been recorded as 7 as there was not sufficient evidence to join Chromosome 6 and 7.

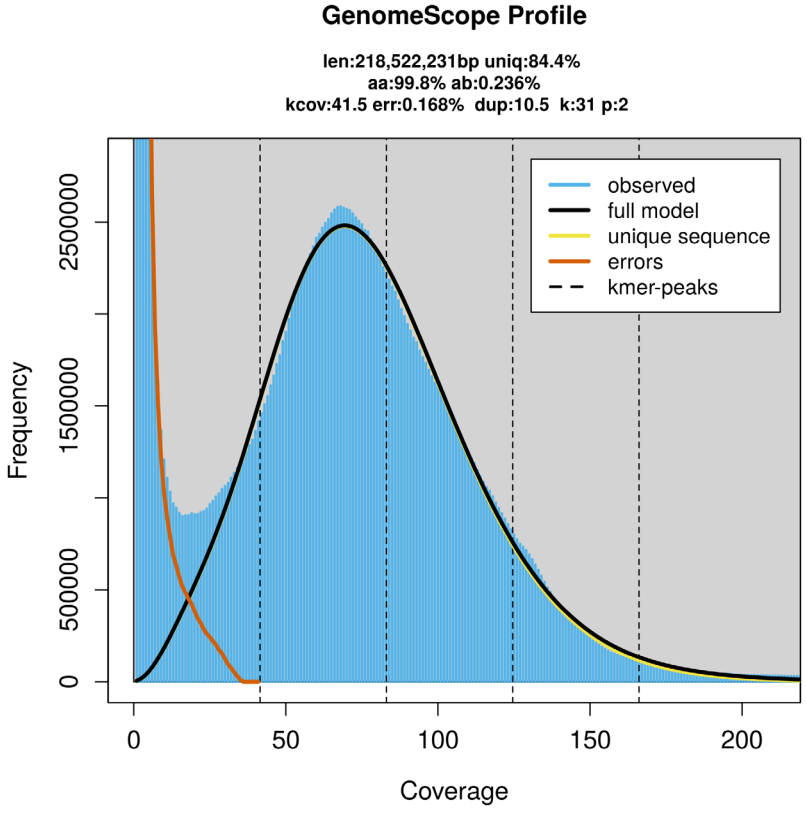

Figure 2. Frequency distribution of *k*-mers generated using GenomeScope2. The plot shows observed and modelled *k*-mer spectra, providing estimates of genome size, heterozygosity, and repeat content based on unassembled sequencing reads.

Table 1. Specimen and sequencing data for BioProject PRJEB66406.

Platform	PacBio HiFi	Hi-C
ToLID	iyNetMela1	iyNetMela1
Specimen ID	NHMUK014425920	NHMUK014425920
BioSample (source individual)	SAMEA14448407	SAMEA14448407
BioSample (tissue)	SAMEA14448782	SAMEA14448782
Tissue	head and thorax	head and thorax
Instrument	Sequel IIe	Illumina NovaSeq 6000
Run accessions	ERR12085115	ERR12102401
Read count total	2.25 million	669.96 million
Base count total	20.40 Gb	101.16 Gb

Table 2. Genome assembly statistics.

Assembly name	iyNetMela1.1
Assembly accession	GCA_964007305.1
Assembly level	chromosome
Span (Mb)	253.87
Number of chromosomes	7
Number of contigs	2 147
Contig N50	0.24 Mb
Number of scaffolds	631
Scaffold N50	39.45 Mb
Organelles	Mitochondrion: 28.04 kb

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.

The assembly has a QV of 58.4 and a *k*-mer completeness of 99.29%, calculated using Merqury.FK (Figure 4). BUSCO v.5.5.0 analysis using the hymenoptera_odb10 reference set (n = 5 991) identified 94.9% of the expected gene set (single = 94.8%, duplicated = 0.2%). The snail plot in Figure 5 summarises the scaffold length distribution and other assembly statistics for the primary assembly. The blob plot in Figure 6 shows the distribution of scaffolds by GC proportion and coverage.

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

Wellcome Sanger Institute – Legal and Governance

The materials that have contributed to this genome note have been supplied by a Darwin Tree of Life Partner. The submission of materials by a Darwin Tree of Life Partner is subject to the **'Darwin Tree of Life Project Sampling Code of Practice'**, which can be found in full on the Darwin Tree of Life website. By agreeing with and signing up to the Sampling Code of Practice, the Darwin Tree of Life Partner agrees they will meet the legal and ethical requirements and standards set out within this document in respect of all samples acquired for, and supplied to, the Darwin Tree of Life Project. Further, 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

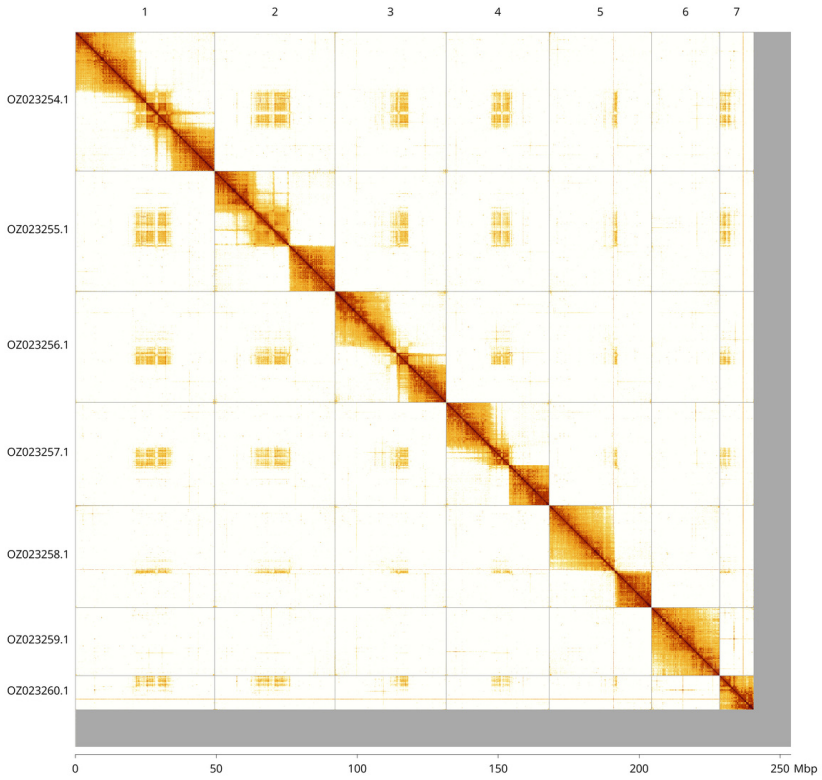

Figure 3. Hi-C contact map of the *Netelia melanura* genome assembly. Assembled chromosomes are shown in order of size and labelled along the axes, with a megabase scale shown below. The plot was generated using PretextSnapshot.

Table 3. Chromosomal pseudomolecules in the primary genome assembly of *Netelia melanura* iyNetMela1.

INSDC accession	Molecule	Length (Mb)	GC%
OZ023254.1	1	49.45	40.50
OZ023255.1	2	42.69	40.50
OZ023256.1	3	39.45	40
OZ023257.1	4	36.52	40
OZ023258.1	5	36.25	39
OZ023259.1	6	24.24	39
OZ023260.1	7	12.08	40.50

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

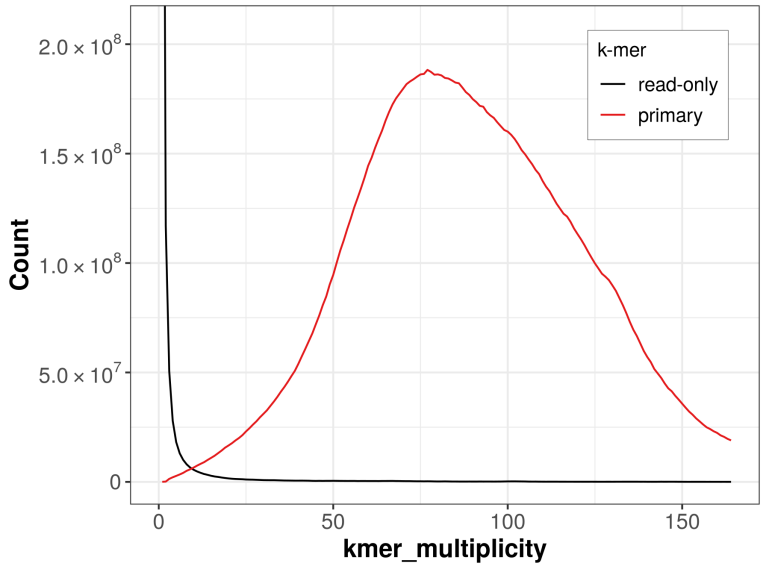

Figure 4. Evaluation of *k*-mer completeness using MerquryFK. This plot illustrates the recovery of *k*-mers from the original read data in the final assemblies. The horizontal axis represents *k*-mer multiplicity, and the vertical axis shows the number of *k*-mers. The black curve represents *k*-mers 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.

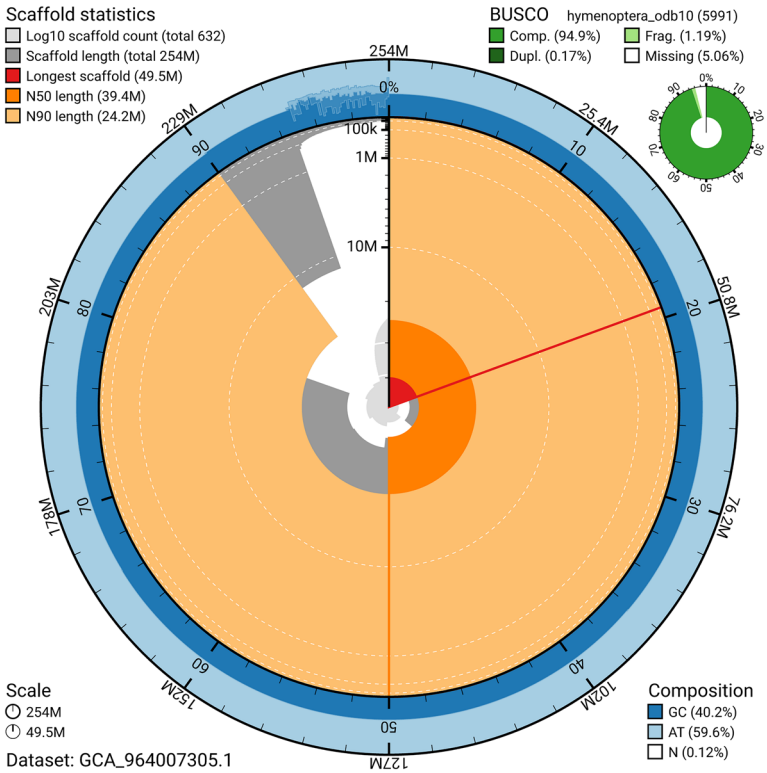

Figure 5. Assembly metrics for iyNetMela1.1. The BlobToolKit snail plot provides an overview of assembly metrics and BUSCO gene completeness. The circumference represents the length of the whole genome sequence, and the main plot is divided into 1 000 bins around the circumference. The outermost blue 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 hymenoptera_odb10 set is presented at the top right. An interactive version of this figure can be accessed on the BlobToolKit viewer.

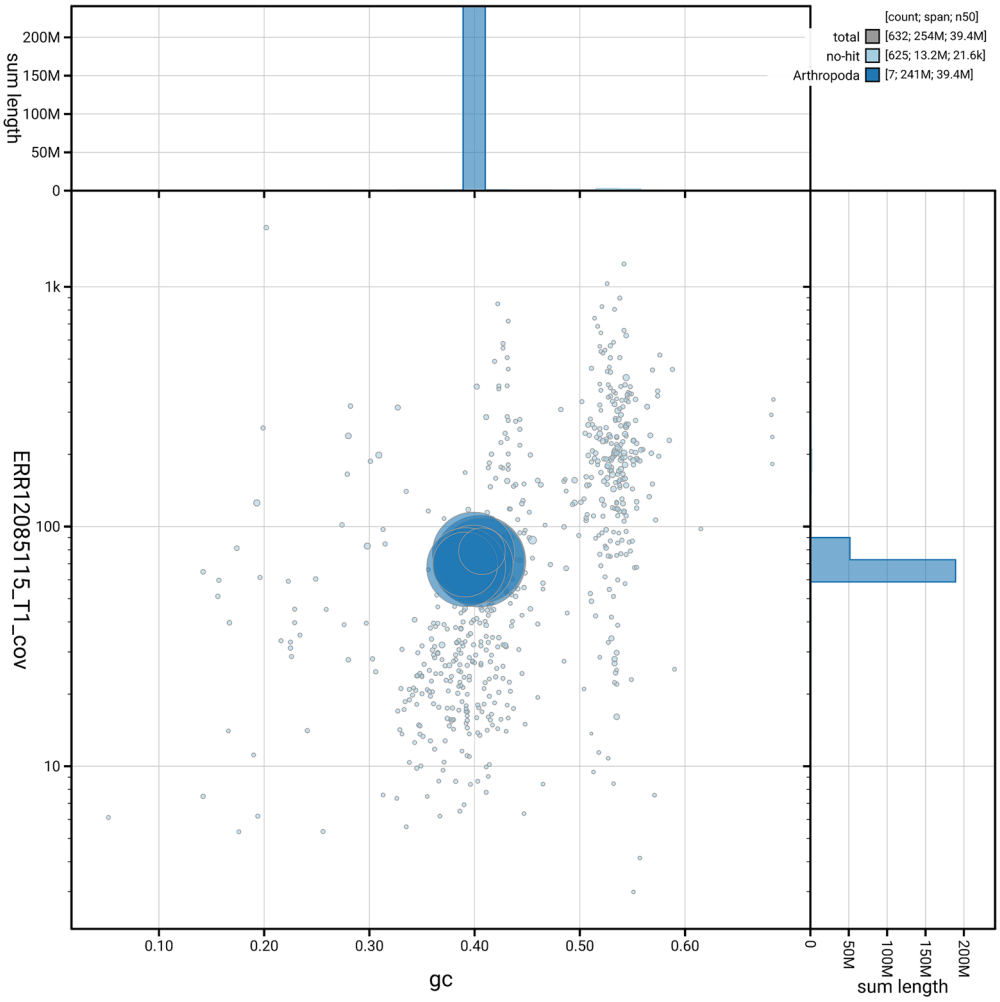

Figure 6. BlobToolKit GC-coverage plot for iyNetMela1.1. Blob plot showing sequence coverage (vertical axis) and GC content (horizontal axis). The circles represent scaffolds, with the size proportional to scaffold length and the colour representing phylum membership. The histograms along the axes display 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.

Table 4. Earth Biogenome Project summary metrics for the *Netelia melanura* assembly.

Measure	Value	Benchmark
EBP summary (primary)	5.C.Q58	6.C.Q40
Contig N50 length	0.24 Mb	≥ 1 Mb
Scaffold N50 length	39.45 Mb	= chromosome N50
Consensus quality (QV)	Primary: 58.4	≥ 40
k-mer completeness	Primary: 99.29%	≥ 95%
BUSCO	C:94.9% [S:94.8%; D:0.2%]; F:1.2%; M:3.9%; n:5 991	S > 90%; D < 5%
Percentage of assembly assigned to chromosomes	94.81%	≥ 90%

Data availability

European Nucleotide Archive: Netelia melanura. Accession number PRJEB66406. The genome sequence is released openly for reuse. The *Netelia melanura* genome sequencing initiative is part of the Darwin Tree of Life Project (PRJEB40665) and the Sanger Institute Tree of Life Programme (PRJEB43745). 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 Table 1 and Table 2.

Production code used in genome assembly at the WSI Tree of Life is available at https://github.com/sanger-tol. Table 5 lists software versions used in this study.

Author information

Contributors are listed at the following links:

- Members of the Natural History Museum Genome Acquisition Lab
- Members of the Darwin Tree of Life Barcoding collective
- 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
- Members of the Darwin Tree of Life Consortium

Table 5. Software versions and sources.

Software	Version	Source
BEDTools	2.30.0	https://github.com/arq5x/bedtools2
BLAST	2.14.0	ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/
BlobToolKit	4.3.9	https://github.com/blobtoolkit/blobtoolkit
BUSCO	5.5.0	https://gitlab.com/ezlab/busco
bwa-mem2	2.2.1	https://github.com/bwa-mem2/bwa-mem2
Cooler	0.8.11	https://github.com/open2c/cooler
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
GoaT CLI	0.2.5	https://github.com/genomehubs/goat-cli
Hifiasm	0.16.1	https://github.com/chhylp123/hifiasm
HiGlass	1.13.4	https://github.com/higlass/higlass
MerquryFK	1.1.2	https://github.com/thegenemyers/MERQURY.FK
Minimap2	2.24-r1122	https://github.com/lh3/minimap2
MitoHiFi	3.2	https://github.com/marcelauliano/MitoHiFi
Oatk	1	https://github.com/c-zhou/oatk
MultiQC	1.14; 1.17 and 1.18	https://github.com/MultiQC/MultiQC
Nextflow	23.04.1	https://github.com/nextflow-io/nextflow
PretextSnapshot	N/A	https://github.com/sanger-tol/PretextSnapshot
PretextView	0.2.5	https://github.com/sanger-tol/PretextView
samtools	1.19.2	https://github.com/samtools/samtools
sanger-tol/ascc	0.1.0	https://github.com/sanger-tol/ascc

Software	Version	Source
sanger-tol/blobtoolkit	0.4.0	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	1.1a.2	https://github.com/c-zhou/yahs

References

Allio R, Schomaker-Bastos A, Romiguier J, *et al.*: **MitoFinder: efficient automated large-scale extraction of mitogenomic data in target enrichment phylogenomics.** *Mol Ecol Resour.* 2020; **20**(4): 892–905.
PubMed Abstract | Publisher Full Text | Free 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

Bateman A, Martin MJ, 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

Buchfink B, Reuter K, Drost HG: **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

Challis R, Richards E, Rajan J, *et al.*: **BlobToolKit – interactive quality assessment of genome assemblies.** *G3 (Bethesda).* 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

Crowley L, Allen H, Barnes I, *et al.*: **A sampling strategy for genome sequencing the British terrestrial arthropod fauna [version 1; peer review: 2 approved].** *Wellcome Open Res.* 2023; **8**: 123.
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): giab008.
PubMed Abstract | Publisher Full Text | Free Full Text

Delrio G: **Révision Des Espèces Ouest-Paléarctiques Du Genre Netelia Gray (Hym., Ichneumonidae).** Studi Sassaresi Sez. III. Annali della Facolta di Agraria dell'Università di Sassari, 1975; **23**.
Reference Source

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

Howard C, Denton A, Jackson B, *et al.*: **On the path to reference genomes for all biodiversity: lessons learned and laboratory protocols created in the Sanger Tree of Life core laboratory over the first 2000 species.** *bioRxiv.* 2025.
Publisher Full Text

Howe K, Chow W, Collins J, *et al.*: **Significantly improving the quality of genome assemblies through curation.** *GigaScience.* 2021; **10**(1): giaa153.
PubMed Abstract | Publisher Full Text | Free Full Text

Kasparyan DR, Tolkanitz VI: **Ichneumonidae subfamily Tryphoninae: tribes Sphinctini, Phytodietini, Oedemopsini, Tryphonini (Addendum), Idiogrammatini. Subfamilies Eucerotinae, Adelognathinae (Addendum), Townesioninae (in Russian) [Fauna of Russia and Neighbouring Countries. Insecta Hymenoptera].** Saint Petersburg: Nauka, 1999; **III**.

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

Konishi K: **A preliminary revision of the subgenus *Netelia* of the genus *Netelia* from Japan (Hymenoptera, Ichneumonidae, Tryphoninae).** *Insecta Matsumurana n.s.* 2005; **62**: 45–121.
Reference Source

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

Lawniczak MKN, Davey RP, Rajan J, *et al.*: **Specimen and sample metadata standards for biodiversity genomics: a proposal from the Darwin Tree of Life project [version 1; peer review: 2 approved with reservations].** *Wellcome Open Res.* 2022; **7**: 187.
Publisher Full Text

Li H: **Minimap2: pairwise alignment for nucleotide sequences.** *Bioinformatics.* 2018; **34**(18): 3094–3100.
PubMed Abstract | Publisher Full Text | Free 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

Merkel D: **Docker: lightweight Linux containers for consistent development and deployment.** *Linux J.* 2014; **2014**(239): 2.
Reference Source

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): 245.
PubMed Abstract | Publisher Full Text | Free Full Text

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

Twyford AD, Beasley J, Barnes I, *et al.*: **A DNA barcoding framework for taxonomic verification in the Darwin Tree of Life project [version 1; peer review: 2 approved].** *Wellcome Open Res.* 2024; **9**: 339.
PubMed Abstract | Publisher Full Text | Free Full Text

Uliano-Silva M, Ferreira JGRN, Krasheninnikova 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.** In: *2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS).* IEEE, 2019; 314–324.
Publisher Full Text

Zhou C, Brown M, Blaxter M, *et al.*: **Oatk: a *de novo* assembly tool for complex plant organelle genomes.** *Genome Biol.* 2025; **26**(1): 235.
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): btac808.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Open Peer Review

Current Peer Review Status: ✔ ✔ ✔

Version 1

Reviewer Report 20 November 2025

https://doi.org/10.21956/wellcomeopenres.27457.r135385

✔ **Bhagya C. Thimmappa**
USDA-ARS Beltsville Agricultural Research Center, Beltsville, Maryland, USA

In this article, the authors perform whole-genome sequencing of the male ichneumonid wasp, *Netelia melanura,* and report the first nuclear and mitochondrial genome assemblies. The reported nuclear genome assembly at the chromosome level is 253.87 megabases in size, and the mitochondrial genome assembly is 28.04 kilobases in length. High-quality assembly was achieved using PacBio HiFi and Hi-C reads, followed by manual curation, resulting in a 94.9% BUSCO value. The data generated in this study have been made publicly available.

Minor comments:

1. In Table 2, please provide a hyperlink to the assembly accession number for easy access.
2. Correct the Figure 4 caption.
3. I see that transcriptome data are generated and made publicly available. However, providing nuclear and mitochondrial genome annotation information would further facilitate molecular and evolutionary studies.

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: Bioinformatics, genomics, transcriptomics, multi-omics

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 28 October 2025

https://doi.org/10.21956/wellcomeopenres.27457.r135788

Alejandro Zaldívar Riverón
Universidad Nacional Autónoma de México, Ciudad de Mexico, Mexico

In this article the authors report a complete nuclear genome sequence of the ichneumonid wasp species *Netelia melanura*. This genome assembly was conducted as in previous published studies carried out by the first author. The genome sequence report is comprehensive and the mitochondrial genome was also assembled. I recommend this manuscript to be accepted for its indexing in Welcome Open Research

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: Systematics, molecular 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 21 October 2025

https://doi.org/10.21956/wellcomeopenres.27457.r135382

Alexandra Cerqueira De Araujo
University College London, London, UK

The authors present here a chromosome-level assembly of the genome of *Netelia melanura* (Ichneumonidae, Hymenoptera), obtained from a single haploid male. This is the first assembly available for this species (three other chromosome-level assemblies are available on NCBI for this genus). This assembly is composed of 7 chromosome-level pseudomolecules (comprising 94.81% of the total assembly) and one mitochondrial chromosome. The assembly has a total length of 253.87 Mb and a scaffold N50 of 39.45 Mb. This assembly achieves an EBP reference standard of 5.C.Q58 and contains most of the orthologous genes found in hymenopterans (i.e. 94.9% of the complete BUSCO sequences). The code used to produce the data is available here https://github.com/sanger-tol.

Comments:

1. The authors mentioned that a close relative of *N. melanura* has a karyotype of 6 chromosomes. On NCBI, 3 other *Netelia* species have a genome assembly with 6 chromosomal pseudomolecules (*N. virgata*, *N. fuscicronis* and *N. thomsonii*, all resulting from the DToL project). In Gokhman V.E., 2011 (DOI: 10.15298/rusentj.20.3.09), Gokhman mentioned that no karyotype of 6 chromosomes was available at the time in parasitoid wasps. Then, Gokhman mentioned that the karyotype of *Netelia latungula* shows 7 chromosomes. Is there any reference for this close relative of *N. melanura* showing its karyotype of 6 chromosomes?
2. In the “Genome sequence report”: “Sequence data” section, it says that the repeat content of the haploid genome is 16.13% (followed by a link to Figure 2). I don’t see where this result can be obtained on figure 2. Should it not be around 15.6%? For demonstration, 84.4% of the assembly is composed of unique content, 100-84.4 = 15.6.
3. In the “Genome sequence report”: “Sequence data” section, the authors say that “the sequencing data provided approximately 83x coverage”. Should it not be 93x (20.40e+9/218.52e+6)? Unless some reads have been filtered before the analysis or another way of calculating the coverage has been used. If the total span of the assembly has been used for the calculation, the sentence in the first paragraph of “Sequence data” shouldn’t say “Based on the estimated genome size, the sequencing data provided approximately 83× coverage.”, which comes just after the k-mer based estimation.
4. In legend of figure 4, it mentions green and blue curves that are not visible on the figure (as it should be since hymenopteran males are haploid). This should be removed from the legend.
5. In the “Genome sequence report” section, the main text (I’m excluding the tables and figures here) contains software names with sometimes the version (e.g. BUSCO) and sometimes without it (e.g. Merqury.FK). This should be standardised throughout the text of this section.
6. Observation about the Hi-C map: if, within a chromosome, the contactless portions represent centromeres, it seems that the chromosomes arms around the centromere of the same chromosome don’t spatially interact with each other. There are some levels of contact

in trans. It also looks like some portions could have been mis-joined and could be split or inverted.

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: Comparative 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.