DATA NOTE

The genome sequence of the click beetle, *Hemicrepidius hirtus* (Herbst, 1784) (Coleoptera: Elateridae)

[version 1; peer review: 3 approved]

Liam M. Crowley[1], James McCulloch[1,2],
University of Oxford and Wytham Woods 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]University of Oxford, Oxford, England, UK
[2]Wellcome Sanger Institute, Hinxton, England, UK

v1 **First published:** 03 Oct 2025, **10**:536
https://doi.org/10.12688/wellcomeopenres.24924.1
Latest published: 03 Oct 2025, **10**:536
https://doi.org/10.12688/wellcomeopenres.24924.1

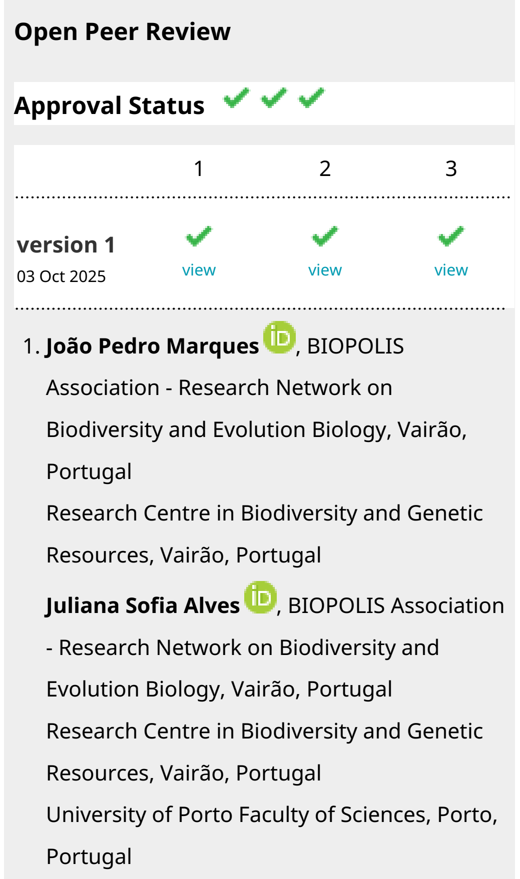
Abstract

We present a genome assembly from an individual male *Hemicrepidius hirtus* (click beetle; Arthropoda; Insecta; Coleoptera; Elateridae). The genome sequence has a total length of 377.55 megabases. Most of the assembly (96.79%) is scaffolded into 10 chromosomal pseudomolecules, including the X sex chromosome. The mitochondrial genome has also been assembled, with a length of 15.9 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

Hemicrepidius hirtus; click beetle; genome sequence; chromosomal; Coleoptera

This article is included in the Tree of Life gateway.



Open Peer Review

Approval Status ✔ ✔ ✔

	1	2	3
version 1 03 Oct 2025	✔ view	✔ view	✔ view

1. **João Pedro Marques**, BIOPOLIS Association - Research Network on Biodiversity and Evolution Biology, Vairão, Portugal
   Research Centre in Biodiversity and Genetic Resources, Vairão, Portugal
   Juliana Sofia Alves, BIOPOLIS Association - Research Network on Biodiversity and Evolution Biology, Vairão, Portugal
   Research Centre in Biodiversity and Genetic Resources, Vairão, Portugal
   University of Porto Faculty of Sciences, Porto, Portugal

2. **Zachary Cohen**, Florida Atlantic University College of Engineering and Computer Science (Ringgold ID: 306684), Boca Raton, USA

3. **Chenyang Cai**, Nanjing Institute of Geology and Palaeontology Chinese Academy of Sciences, Nanjing, China

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: Crowley LM: Investigation, Resources; **McCulloch J**: Investigation, Resources, Writing – Original Draft Preparation;

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.

Copyright: © 2025 Crowley LM *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: Crowley LM, McCulloch J, University of Oxford and Wytham Woods Genome Acquisition Lab *et al.* **The genome sequence of the click beetle, *Hemicrepidius hirtus* (Herbst, 1784) (Coleoptera: Elateridae) [version 1; peer review: 3 approved]** Wellcome Open Research 2025, **10**:536 https://doi.org/10.12688/wellcomeopenres.24924.1

First published: 03 Oct 2025, **10**:536 https://doi.org/10.12688/wellcomeopenres.24924.1

Species taxonomy

Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Protostomia; Ecdysozoa; Panarthropoda; Arthropoda; Mandibulata; Pancrustacea; Hexapoda; Insecta; Dicondylia; Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga; Elateriformia; Elateroidea; Elateridae; Denticollinae; *Hemicrepidius*; *Hemicrepidius hirtus* (Herbst, 1784) (NCBI:txid1132043)

Background

Hempicrepidius hirtus is a species of click beetle in the family Elateridae. This is the only species in its genus in the UK. As an adult, it can be distinguished from other British species in the superfamily Elateroidea by the combination of the following characteristics: antennae not clubbed, strongly serrate, with the 1st antennal segment shorter than the space between the antennal insertions, the 2nd segment shorter than the 4th segment, but the 2nd and 3rd antennal segments together as long as the 4th segment; eyes not strikingly convex; pronotum black and broadest nearer to the hind angles, which project backwards past the pronotal hind margin; elytra striate, not confusingly punctured, and not patterned by pubescence; and the 4th tarsal segment on the mid and hind tarsi weakly bilobed and about a third the size of the 3rd segment (Telfer & Joy, 2010).

According to records on the NBN Atlas (NBN Trust, 2024), *H. hirtus* is a widespread species in the UK, with records from Northern Ireland and as far north as Orkney (Scotland). Outside of the UK, this species has been recorded across Europe and in western Asia, and as a non-native species in North America (Douglas, 2011). There was initially concern that this species may act as an agricultural pest in its introduced range, but stable isotope analysis from within its native distribution suggests that it is more likely to be predatory in the larval stage (Traugott *et al.*, 2008). The habitat preferences of this species appear to be diverse, and it can be found in both forested and open habitats (Ruchin *et al.*, 2018).

Here we present a chromosome-level genome sequence for *Hemicrepidius hirtus*, based on a male specimen from Wytham Woods, Oxfordshire, England. This assembly is the first high-quality genome for the genus *Hemicrepidius* as of August 2025 (data obtained via NCBI datasets, O'Leary *et al.*, 2024). This assembly was generated as part of the Darwin Tree of Life Project, which aims to generate high-quality reference genomes for all named eukaryotic species in Britain and Ireland to support research, conservation, and the sustainable use of biodiversity (Blaxter *et al.*, 2022).

Methods

Sample acquisition and DNA barcoding

The specimen used for genome sequencing was an adult male *Hemicrepidius hirtus* (specimen ID Ox002292, ToLID icHemNige1; Figure 1), collected from Wytham Woods, Oxfordshire, UK (latitude 51.772, longitude –1.338) on 2022-07-05. The specimen was collected by Liam Crowley and James McCulloch and identified by Liam Crowley. The same specimen was used for RNA sequencing. For the Darwin Tree of Life sampling and metadata approach, refer to Lawniczak *et al*. (2022).

Figure 1. Photograph of the *Hemicrepidius hirtus* (icHemNige1) specimen used for genome sequencing.

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.

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 icHemNige1 sample was weighed and triaged to determine the appropriate extraction protocol. Tissue from the head and 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. DNA was sheared into an average fragment size of 12–20 kb following the Megaruptor®3 for LI PacBio protocol. 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. For this sample, the final post-shearing DNA had a Qubit concentration of 9.58 ng/μL and a yield of 431.10 ng, with a fragment size of 12.4 kb. The 260/280 spectrophotometric ratio was 1.78, and the 260/230 ratio was 1.15.

RNA was extracted from abdomen tissue of icHemNige1 in the Tree of Life Laboratory at the WSI using the RNA Extraction: Automated MagMax™ *mir*Vana protocol. The RNA concentration

was assessed using a Nanodrop spectrophotometer and a Qubit Fluorometer using the Qubit RNA Broad-Range Assay kit. Analysis of the integrity of the RNA was done using the Agilent RNA 6000 Pico Kit and Eukaryotic Total RNA assay.

PacBio HiFi library preparation and sequencing

Library preparation and sequencing were performed at the WSI Scientific Operations core. Libraries were prepared using the SMRTbell Prep Kit 3.0 (Pacific Biosciences, California, USA), following the manufacturer's instructions. The kit includes reagents for end repair/A-tailing, adapter ligation, post-ligation SMRTbell bead clean-up, and nuclease treatment. Size selection and clean-up were performed using diluted AMPure PB beads (Pacific Biosciences). DNA concentration was quantified using a Qubit Fluorometer v4.0 (ThermoFisher Scientific) and the Qubit 1X dsDNA HS assay kit. Final library fragment size was assessed with the Agilent Femto Pulse Automated Pulsed Field CE Instrument (Agilent Technologies) using the gDNA 55 kb BAC analysis kit.

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 icHemNige1 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.

RNA library preparation and sequencing

Libraries were prepared using the NEBNext® Ultra™ II Directional RNA Library Prep Kit for Illumina (New England Biolabs), following the manufacturer's instructions. Poly(A) mRNA in the total RNA solution was isolated using oligo(dT) beads, converted to cDNA, and uniquely indexed; 14 PCR cycles were performed. Libraries were size-selected to produce fragments between 100–300 bp. Libraries were quantified, normalised, pooled to a final concentration of 2.8 nM, and diluted to 150 pM for loading. Sequencing was carried out on the Illumina NovaSeq 6000 to generate 150-bp paired-end reads.

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 --primary option. Haplotypic duplications were identified and removed using purge_dups (Guan *et al.*, 2020). 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.

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 1 break, 13 joins, and removal of 1 haplotypic duplication. 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 for the primary and alternate 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 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 *Hemicrepidius hirtus* specimen generated 25.74 Gb (gigabases) from 2.34 million reads, which were used to assemble the genome. GenomeScope2.0 analysis estimated the haploid genome size at 357.58 Mb, with a heterozygosity of 0.96% and repeat content of 39.63% (Figure 2). These estimates guided expectations for the assembly. Based on the estimated genome size, the sequencing data provided approximately 69× coverage. Hi-C sequencing produced 135.46 Gb from 897.12 million reads, which were used to scaffold the

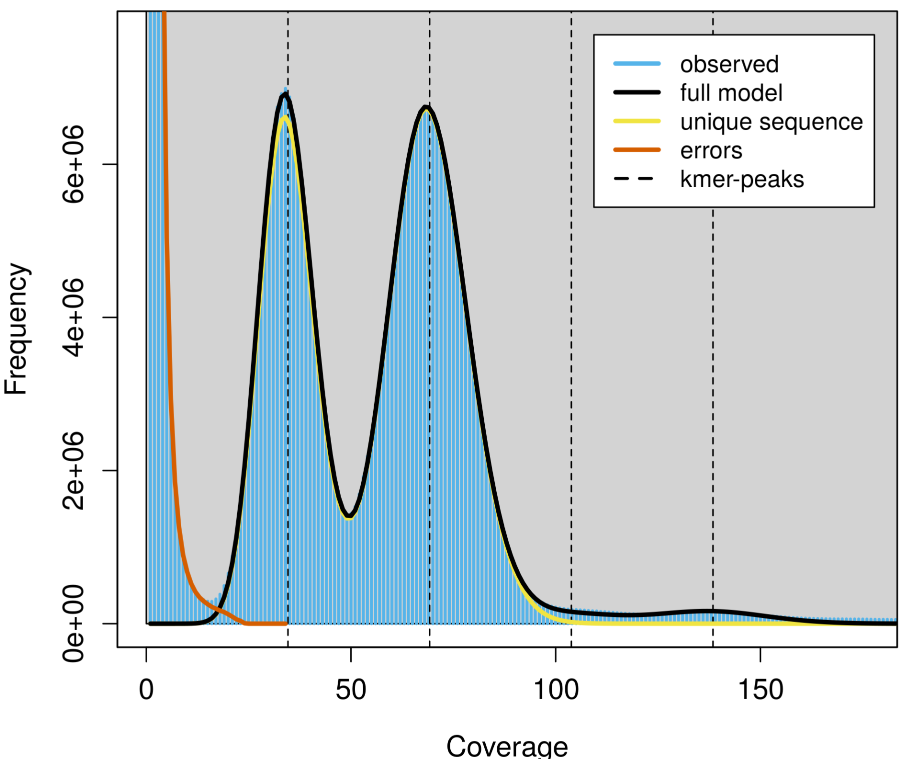


Figure 2. Frequency distribution of *k*-mers generated using GenomeScope2. The plot shows observed and modelled *k*-mer spectra, providing estimates of genome size, heterozygosity, and repeat content based on unassembled sequencing reads.

assembly. RNA sequencing data were also generated and are available in public sequence repositories. Table 1 summarises the specimen and sequencing details.

Assembly statistics

The primary haplotype was assembled, and contigs corresponding to an alternate haplotype were also deposited in INSDC databases. The final assembly has a total length of 377.55 Mb in 196 scaffolds, with 68 gaps, and a scaffold N50 of 35.74 Mb (Table 2).

Most of the assembly sequence (96.79%) was assigned to 10 chromosomal-level scaffolds, representing 9 autosomes and the X sex chromosome. These chromosome-level scaffolds, confirmed by Hi-C data, are named according to size (Figure 3; Table 3). Chromosome X was assigned by read coverage statistics. No Y sequence could be identified.

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 combined primary and alternate assemblies achieve an estimated QV of 64.1. The *k*-mer completeness is 82.40% for the primary assembly, 75.90% for the alternate haplotype, and 99.12% for the combined assemblies (Figure 4).

BUSCO (v5.4.3) analysis using the endopterygota_odb10 reference set (n = 2,124) identified a completeness score of 99.2% (single = 97.7%, duplicated = 1.5%). 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 1. Specimen and sequencing data for BioProject PRJEB61499.

Platform	PacBio HiFi	Hi-C	RNA-seq
ToLID	icHemNige1	icHemNige1	icHemNige1
Specimen ID	Ox002292	Ox002292	Ox002292
BioSample (source individual)	SAMEA112232526	SAMEA112232526	SAMEA112232526
BioSample (tissue)	SAMEA112232970	SAMEA112232970	SAMEA112232971
Tissue	head and thorax	head and thorax	abdomen
Instrument	Sequel IIe	Illumina NovaSeq 6000	Illumina NovaSeq 6000
Run accessions	ERR11263500	ERR11271519	ERR12245565
Read count total	2.34 million	897.12 million	75.58 million
Base count total	25.74 Gb	135.46 Gb	11.41 Gb

Table 2. Genome assembly statistics.

Assembly name	icHemNige1.1
Assembly accession	GCA_963082805.1
Alternate haplotype accession	GCA_963082985.1
Assembly level	chromosome
Span (Mb)	377.55
Number of chromosomes	10
Number of contigs	264
Contig N50	7.64 Mb
Number of scaffolds	196
Scaffold N50	35.74 Mb
Sex chromosomes	X
Organelles	Mitochondrion: 15.9 kb

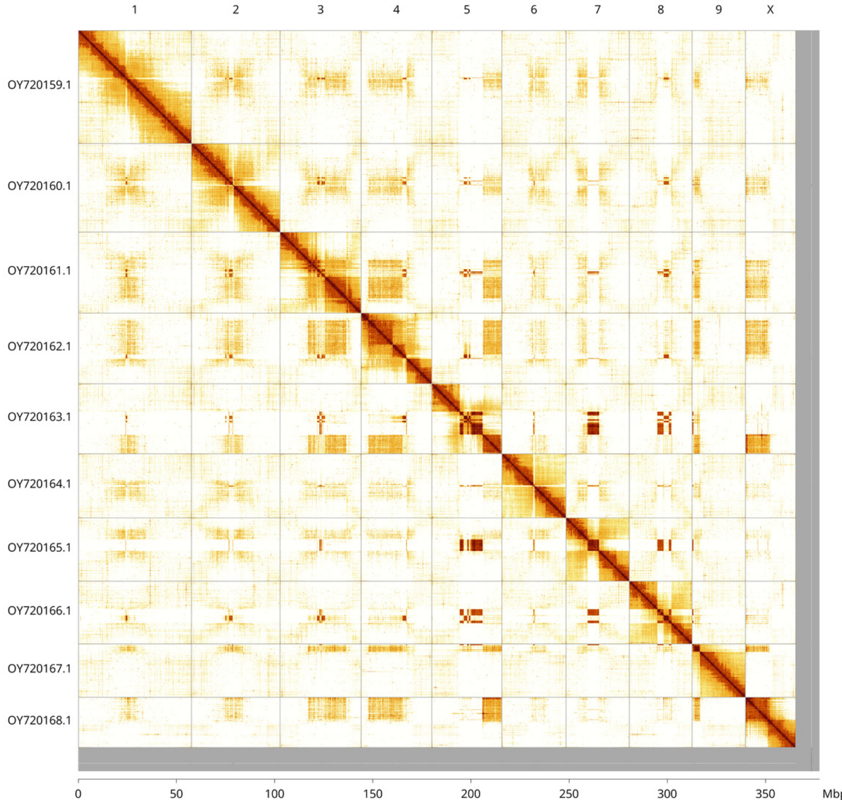


Figure 3. Hi-C contact map of the *Hemicrepidius hirtus* genome assembly. Assembled chromosomes are shown in order of size and labelled along the axes, with a megabase scale shown below. The plot was generated using PretextSnapshot.

Table 3. Chromosomal pseudomolecules in the primary genome assembly of *Hemicrepidius hirtus* icHemNige1.

INSDC accession	Molecule	Length (Mb)	GC%
OY720159.1	1	57.74	33.50
OY720160.1	2	45.08	34
OY720161.1	3	41.26	35
OY720162.1	4	36.02	35.50
OY720163.1	5	35.74	36
OY720164.1	6	32.61	34
OY720165.1	7	32.31	34
OY720166.1	8	31.92	35
OY720167.1	9	27.15	34
OY720168.1	X	25.60	34

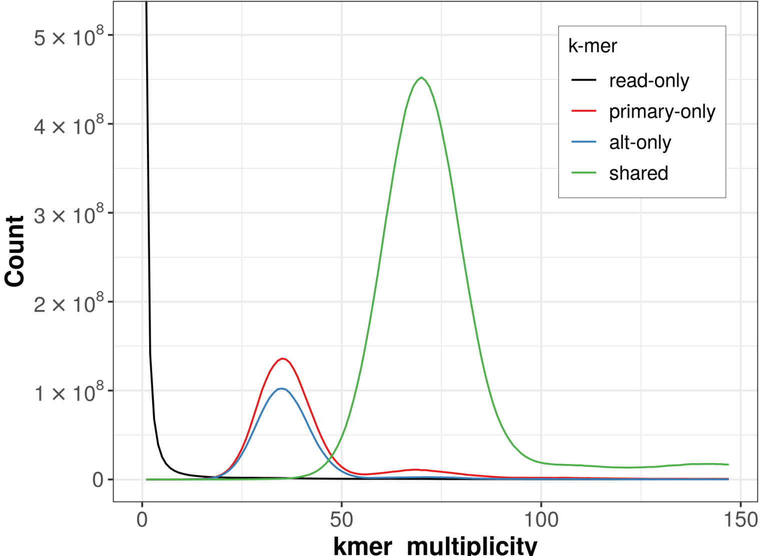


Figure 4. Evaluation of *k*-mer completeness using MerquryFK. This plot illustrates the recovery of *k*-mers from the original read data in the final assemblies. The horizontal axis represents *k*-mer multiplicity, and the vertical axis shows the number of *k*-mers. The black curve represents *k*-mers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

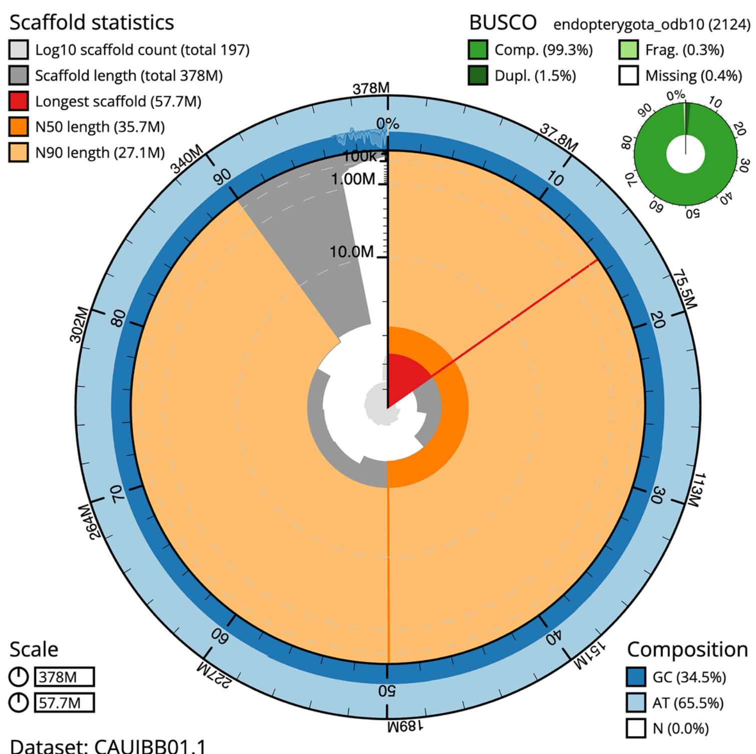


Figure 5. Assembly metrics for icHemNige1.1. The BlobToolKit snail plot provides an overview of assembly metrics and BUSCO gene completeness. The circumference represents the length of the whole genome sequence, and the main plot is divided into 1 000 bins around the circumference. The outermost blue 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.

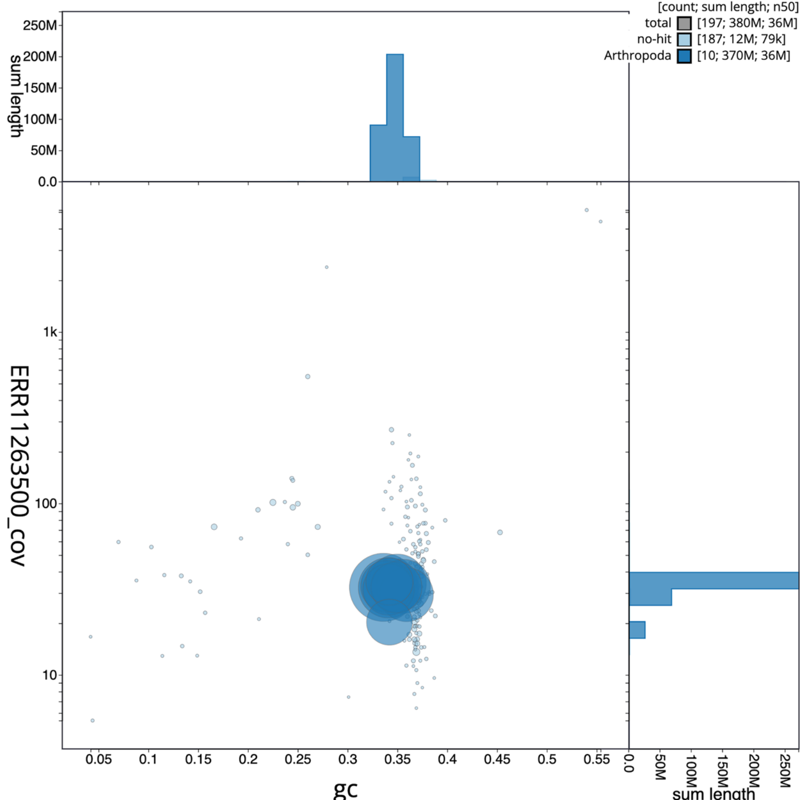


Figure 6. BlobToolKit GC-coverage plot for icHemNige1.1. Blob plot showing sequence coverage (vertical axis) and GC content (horizontal axis). The circles represent scaffolds, with the size proportional to scaffold length and the colour representing phylum membership. The histograms along the axes display 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 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 the primary assembly, is **6.C.Q64**, meeting the recommended reference standard.

Table 4. Earth Biogenome Project summary metrics for the *Hemicrepidius hirtus* assembly.

Measure	Value	Benchmark
EBP summary (primary)	6.C.Q64	6.C.Q40
Contig N50 length	7.64 Mb	≥ 1 Mb
Scaffold N50 length	35.74 Mb	= chromosome N50
Consensus quality (QV)	Primary: 64.2; alternate: 63.9; combined: 64.1	≥ 40
k-mer completeness	Primary: 82.40%; alternate:75.90%; combined: 99.12%	≥ 95%
BUSCO	C=99.3% [S=97.8%, D=1.5%]; F=0.3%; M=0.4%.	S > 90%; D < 5%
Percentage of assembly assigned to chromosomes	96.79%	≥ 90%

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 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.

Data availability

European Nucleotide Archive: Hemicrepidius hirtus. Accession number PRJEB61499. The genome sequence is released openly for reuse. The *Hemicrepidius hirtus* 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 University of Oxford and Wytham Woods 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+/
bwa-mem2	2.2.1	https://github.com/bwa-mem2/bwa-mem2
Cooler	0.8.11	https://github.com/open2c/cooler
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-r375	https://github.com/chhylp123/hifiasm
HiGlass	1.13.4	https://github.com/higlass/higlass
MerquryFK	1.1.2	https://github.com/thegenemyers/MERQURY.FK
MitoHiFi	3	https://github.com/marcelauliano/MitoHiFi
MultiQC	1.14; 1.17 and 1.18	https://github.com/MultiQC/MultiQC
PretextSnapshot	N/A	https://github.com/sanger-tol/PretextSnapshot
PretextView	0.2.5	https://github.com/sanger-tol/PretextView
purge_dups	1.2.5	https://github.com/dfguan/purge_dups

Software	Version	Source
sanger-tol/ascc	0.1.0	https://github.com/sanger-tol/ascc
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	1.2a.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**

Blaxter M, Mieszkowska N, Di Palma F, *et al.*: **Sequence locally, think globally: the Darwin Tree of Life project.** *Proc Natl Acad Sci U S A.* 2022; **119**(4): e2115642118.
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**

Douglas H: **New records of European wireworm pests and other click beetles (Coleoptera: Elateridae) in Canada and USA.** *J Ent Soc Ont.* 2011; **142**: 11–17.
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**

Guan D, McCarthy SA, Wood J, *et al.*: **Identifying and removing haplotypic duplication in primary genome assemblies.** *Bioinformatics.* 2020; **36**(9): 2896–2898.
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**

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**

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

NBN Trust: ***Hemicrepidius hirtus* map on the NBN Atlas**. 2024.
Reference Source

O'Leary NA, Cox E, Holmes JB, *et al.*: **Exploring and retrieving sequence and metadata for species across the Tree of Life with NCBI Datasets.** *Sci Data.* 2024; **11**(1): 732.
PubMed Abstract | **Publisher Full Text** | **Free 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): 245.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Ruchin AB, Egorocv LV, Semishin GB: **Fauna of click beetles (Coleoptera: Elateridae) in the interfluve of Rivers Moksha and Sura, Republic of

Mordovia, Russia. *Biodiversitas.* 2018; **19**(4): 1352–1365.
Publisher 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**

Telfer M, Joy NM: **Joy's keys to Elateridae, Eucnemidae and Throscidae (original from 1932, edited by Mark Telfer)**. 2010.

Traugott M, Schallhart N, Kaufmann R, *et al.*: **The feeding ecology of elaterid larvae in Central European arable land: new perspectives based on naturally occurring stable isotopes.** *Soil Biol Biochem.* 2008; **40**(2): 342–349.
Publisher 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, 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 05 January 2026

https://doi.org/10.21956/wellcomeopenres.27454.r141507

© **2026 Cai C.** 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.

Chenyang Cai
Nanjing Institute of Geology and Palaeontology Chinese Academy of Sciences, Nanjing, Jiangsu, China

This manuscript presents a high-quality, chromosome-level genome assembly for *Hemicrepidius hirtus*, the only UK species of its genus and an ecologically noteworthy click beetle. The study follows the established Darwin Tree of Life (DToL) workflows. Please clarify whether the mitochondrial genome was validated using coverage depth or annotation completeness.

For antennal characters: *1st antennal segment 2nd segment shorter than the 4th segment, but the 2nd and 3rd antennal segments together as long as the 4th segment;*
'segment' should be replaced by 'antennomere'

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: phylogenomics

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 02 January 2026

https://doi.org/10.21956/wellcomeopenres.27454.r141505

© **2026 Cohen Z.** 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.

Zachary Cohen
Florida Atlantic University College of Engineering and Computer Science (Ringgold ID: 306684), Boca Raton, Florida, USA

Another impressive report on the industry standard for genome assemblies. By providing the genome for the only extant species in the Hemicrepidius genus, these data will be invaluable for further understanding on species diversity & the consequence of isolation. This work expertly leverages HiFi and Hi-C reads to generate a chromosome level genome assembly while isolating the mitochondria and X chromosome. Additionally, they provide the community with RNA seq data for subsequent analysis.

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: genomics, entomology, adapation

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 30 December 2025

https://doi.org/10.21956/wellcomeopenres.27454.r141510

© **2025 Marques J et al.** 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.

João Pedro Marques
[1] BIOPOLIS Association - Research Network on Biodiversity and Evolution Biology, Vairão, Porto, Portugal
[2] Research Centre in Biodiversity and Genetic Resources, Vairão, Porto, Portugal

Juliana Sofia Alves
[1] BIOPOLIS Association - Research Network on Biodiversity and Evolution Biology, Vairão, Porto, Portugal
[2] Research Centre in Biodiversity and Genetic Resources, Vairão, Porto, Portugal
[3] University of Porto Faculty of Sciences, Porto, Porto District, Portugal

The manuscript presents a high-quality chromosome-level genome assembly for *Hemicrepidius hirtus*. The genome was generated using current state-of-the-art methods within the Darwin Tree of Life project and is accompanied by a clear and thorough technical description.

Overall, the manuscript is well written, technically sound, and fully appropriate for a genome note. The assembly quality is very high and meets recommended community standards. This genome will be a valuable resource for future studies on beetle genomics, evolution, and ecology. Our comments below are mainly minor and aim to improve clarity and biological context.

Minor comments and clarifications

1. **Background section:** Some morphological details described in the Background (e.g. pronotum characteristics and segment proportions) could not be fully found in the cited reference. We suggest adding one or two additional references to support these descriptions.
2. **Genome assembly description:** In the genome assembly section, it would be useful to clarify whether duplicate purging within *hifiasm* was applied, and at which level. In addition, explicitly stating that an unphased assembly pipeline was used would improve clarity, as referring only to the --primary option may not be clear to all readers.
3. **Nucleic acid extraction protocol:** The authors used the Automated MagAttract v2 protocol for HMW DNA extraction. A short explanation of why this protocol was chosen instead of the MagAttract protocol for small arthropods would be helpful.
4. **Typographical error:** On page 3, *Hempicrepidius hirtus* should be corrected to *Hemicrepidius hirtus*.

Biological context: Although we recognise the concise nature of genome notes, adding a small amount of biological interpretation would help readers better understand the results:

- The relatively high heterozygosity inferred from the k-mer analysis and GenomeScope plot (~0.96%) could be briefly mentioned.
- The assembly consists of 10 chromosomes (9 autosomes and the X chromosome). A short comment on whether this chromosome number matches expectations for this species, genus, or family would be useful.
- The GC content of the genome is relatively low (~34–36%). Indicating whether this is typical for *Elateridae* or related beetle groups would add context.

In conclusion, this is a strong and well-executed genome note. Addressing the minor points above would further improve clarity and usefulness.

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: Evolutionary Biology, Population Genomics and Genome Assembly

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