DATA NOTE

The genome sequence of the flat clown beetle, *Hololepta plana* (Sulzer, 1776) (Coleoptera: Histeridae)

[version 1; peer review: 2 approved]

Michael F. Geiser[1], Jordan J. Rainey[2], Dmitry Telnov[1,3,4], Maxwell V. L. Barclay[1], 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]Natural History Museum, London, England, UK
[2]Biology Centre, Czech Academy of Sciences, České Budějovice, Czech Republic
[3]Daugavpils University, Daugavpils, Latvia
[4]Institute of Biology, University of Latvia, Rīga, Latvia

v1 **First published:** 18 May 2026, **11**:289
https://doi.org/10.12688/wellcomeopenres.26588.1
Latest published: 18 May 2026, **11**:289
https://doi.org/10.12688/wellcomeopenres.26588.1

Abstract

We present a genome assembly from an individual female *Hololepta plana* (flat clown beetle; Arthropoda; Insecta; Coleoptera; Histeridae). The genome sequence has a total length of 139.51 megabases. Most of the assembly (97.93%) is scaffolded into 12 chromosomal pseudomolecules, including the X sex chromosome. The mitochondrial genome has also been assembled, with a length of 17.73 kilobases. This assembly was generated as part of the Darwin Tree of Life project, which produces genomes for eukaryotic species found in Britain and Ireland.

Keywords

Hololepta plana, flat clown beetle, genome sequence, chromosomal, Coleoptera

This article is included in the Tree of Life gateway.

Open Peer Review

Approval Status ✓ ✓

	1	2
version 1 18 May 2026	✓ view	✓ view

1. **Rajib Deb**, ICAR-National Research Centre, Guwahat, India

2. **Henrik Lantz**, SciLifeLab, Uppsala University, Uppsala, Sweden

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: Geiser MF: Investigation, Resources; **Rainey JJ**: Writing – Original Draft Preparation; **Telnov D**: Writing – Original Draft Preparation; **Barclay MVL**: Resources, 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, https://doi.org/10.35802/218328].
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: Geiser MF, Rainey JJ, Telnov D *et al.* **The genome sequence of the flat clown beetle, *Hololepta plana* (Sulzer, 1776) (Coleoptera: Histeridae) [version 1; peer review: 2 approved]** Wellcome Open Research 2026, **11**:289 https://doi.org/10.12688/wellcomeopenres.26588.1

First published: 18 May 2026, **11**:289 https://doi.org/10.12688/wellcomeopenres.26588.1

Species taxonomy

Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Protostomia; Ecdysozoa; Panarthropoda; Arthropoda; Mandibulata; Pancrustacea; Altocrustacea; Allotriocarida; Hexapoda; Insecta; Dicondylia; Pterygota; Neoptera; Eumetabola; Endopterygota; Aparaglossata; Neuropteroidea; Coleoptera; Polyphaga; Staphyliniformia; Histeroidea; Histeridae; Histerinae; Hololeptini; *Hololepta*; *Hololepta plana* (Sulzer, 1776) (NCBI:txid290657).

Background

Hololepta plana (Sulzer, 1776) is a member of the family Histeridae, commonly known as the clown beetles. This species is the only representative of the tribe Hololeptini Hope, 1840 in the UK. The combined characters of large adult size (approximately 8.5 mm), a flattened body, and a reduced prosternal projection leaving the ventral surface of the head exposed make *H. plana* one of the most recognisable of the 54 histerid species in Britain (Lane, 2017).

Hololepta plana is widely distributed across Europe and extends into Asia (Lackner *et al.*, 2015). Sánchez-Ruiz *et al.* (2002) documented its range extending as far south as southern Spain. The northernmost known record is from Mäntsälä, Finland, reported by iNaturalist user "janijarvi" in 2022, with supporting images. The species is widespread in the boreal forest zone of Europe in the Baltic countries (D. Telnov, personal observations).

The species was first recorded in Britain from Norfolk in 2009 (Allen & Hance, 2009). Additional records have been added since then, mostly from south-western England. Lane (2021) summarises the British records and Barclay (2018) expanded on this. It has been recorded in vice-counties including West Norfolk (Allen & Hance, 2009; Collier & Lane, 2015; Lane *et al.*, 2021), Middlesex (Barclay, 2018; Lane *et al.*, 2021; Telfer, 2014), West Suffolk (Lane, 2017; Lane *et al.*, 2021), Cambridgeshire (Lane *et al.*, 2021), Buckinghamshire and Surrey (Lane *et al.*, 2021). Lane's species status review (Lane, 2017) gives the IUCN status of *H. plana* as Data Deficient and the GB Rarity Status as Nationally Rare. It seems likely that it is a recent immigrant in Britain, particularly given the complete absence of old records for such a conspicuous species.

Adults and larvae of *H. plana* prey on subcortical dipteran larvae (Barclay, 2018; Trach, 2013). Their flattened bodies allow them to live beneath bark (Kryzhanovsky & Reikhardt, 1976). In Britain, *H. plana* is most often found beneath the sappy, laminating bark of dead poplar trees (*Populus* spp.) (Lane, 2017). Barclay (2018) recorded the species from a dead tree of heaven (*Ailanthus altissima*), which also supported a large population of subcortical dipteran larvae. Other recorded hosts include oak and pine (*Quercus* sp. and *Pinus sylvestris*) and poplars including *Populus tremula* and *P. alba* (Nagaraju *et al.*, 2022). In Latvia, the species has also been recorded beneath the bark of Norway maple (*Acer platanoides*), although it appears to prefer European aspen (*Populus tremula*) and other poplars (D. Telnov, unpublished data). Live specimens of *H. plana* were recently found from logs imported into India from Belgium (Nagaraju *et al.*, 2022), suggesting that the species may have arrived in Britain via logs imported from mainland Europe.

A chromosome-level genome sequence for *H. plana* is presented here. The assembly was produced using the Tree of Life pipeline from a specimen collected in Middlesex, England (Figure 1). The genome sequence will support investigations into the origin of the British population of *H. plana* and uncover likely connections with subpopulations of the species in continental Europe.

Methods

Sample acquisition and DNA barcoding

The specimen used for genome sequencing was an adult female *Hololepta plana* (specimen ID NHMUK014436789, ToLID icHolPlan1; Figure 1), collected from Wormwood Scrubs Park, Hammersmith And Fulham, Middlesex, England, United Kingdom (latitude 51.52, longitude −0.25) on 2021-07-17. The specimen was collected and identified by Michael Geiser (Natural History Museum).

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.

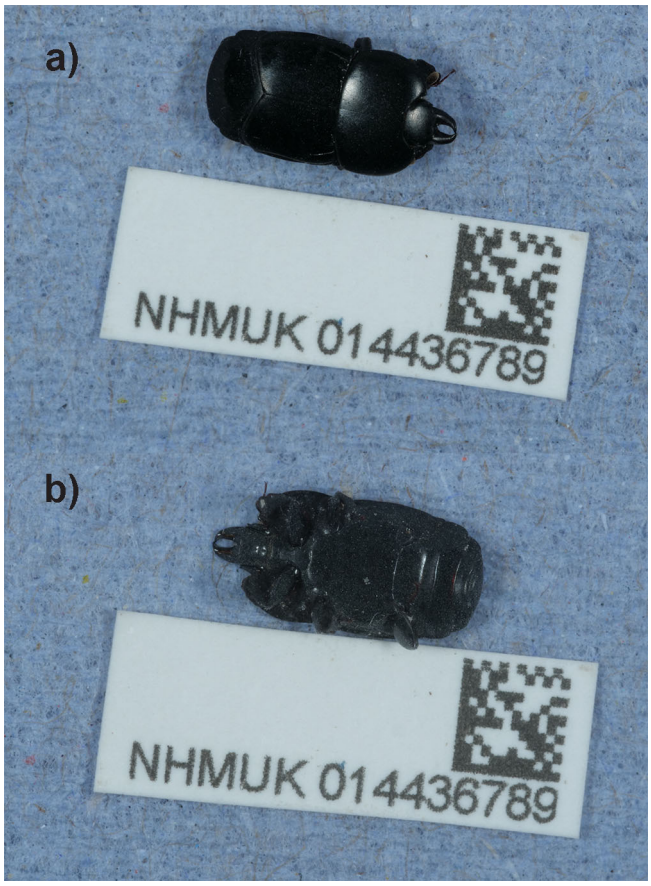

Figure 1. Photograph of the *Hololepta plana* (icHolPlan1) specimen used for genome sequencing.

Nucleic acid extraction

Detailed protocols for nucleic acid extraction developed at the Wellcome Sanger Institute (WSI) Tree of Life Core Laboratory are available on protocols.io (Howard *et al.*, 2025). The icHolPlan1 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. High molecular weight (HMW) DNA was extracted using the Automated MagAttract v1 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 automated 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 0.856 ng/μL and a yield of 333.84 ng.

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 abdomen tissue of the icHolPlan1 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

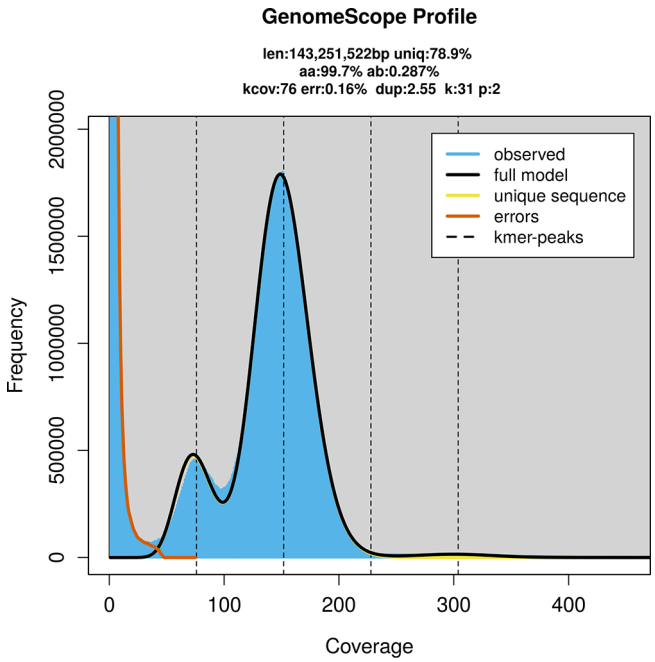

Figure 2. Frequency distribution of *k*-mers generated using GenomeScope2. The plot shows observed and modelled *k*-mer spectra, providing estimates of genome size, heterozygosity, and repeat content based on unassembled sequencing reads.

Table 1. **Specimen and sequencing data for *hololepta plana* (BioProject PRJEB64074).**

Platform	PacBio HiFi	Hi-C
ToLID	icHolPlan1	icHolPlan1
Specimen ID	NHMUK014436789	NHMUK014436789
BioSample (source individual)	SAMEA111458819	SAMEA111458819
BioSample (tissue)	SAMEA111458840	SAMEA111458851
Tissue	head and thorax	abdomen
Instrument	Sequel IIe	Illumina NovaSeq 6000
Run accessions	ERR11673234	ERR11679387
Read count total	2.46 million	325.17 million
Base count total	24.50 Gb	98.20 Gb

Table 2. **Genome assembly data for *hololepta plana.***

Genome assembly	Primary assembly
Assembly name	icHolPlan1.2
Assembly accession	GCA_963695495.2
Alternate haplotype accession	GCA_963700325.2
Assembly level	chromosome
Span (Mb)	139.51
Number of chromosomes	12
Number of contigs	145
Contig N50	2.2 Mb
Number of scaffolds	45
Scaffold N50	14.29 Mb
Sex chromosomes	X
Organelles	Mitochondrion: 17.73 kb

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 to create equimolar and/or weighted 2.8 nM pools. Pool concentrations were checked using the Agilent 4200 TapeStation (Agilent) with High Sensitivity D500 reagents before sequencing. Sequencing was performed using paired-end 150 bp reads on the Illumina NovaSeq 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 --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 MerquryFK (Rhie *et al.*, 2020).

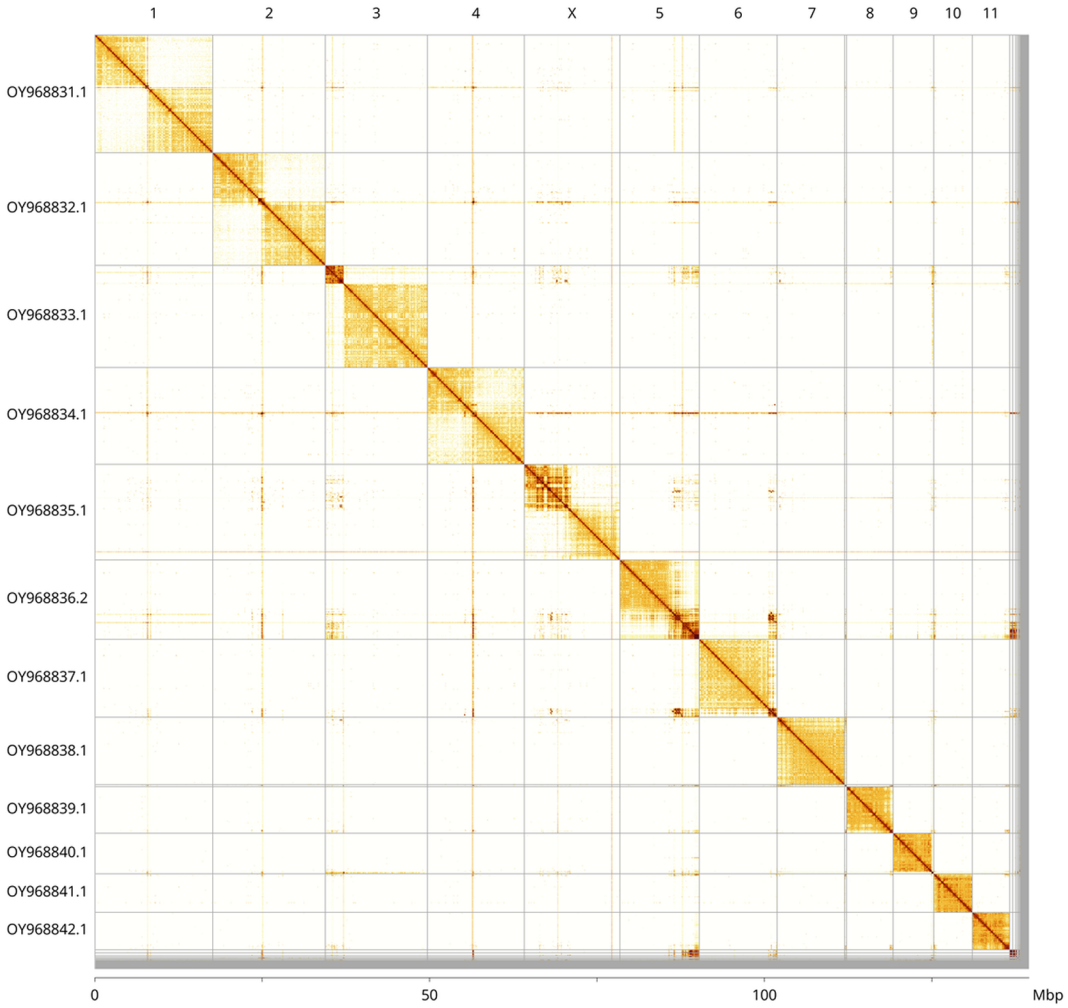

Figure 3. Hi-C contact map of the *Hololepta plana* genome assembly. Assembled chromosomes are shown in order of size and labelled along the axes, with a megabase scale shown below. The plot was generated using PretextSnapshot.

Table 3. Chromosomal pseudomolecules in the primary genome assembly of *Hololepta plana* icHolPlan1.

INSDC accession	Molecule	Length (Mb)	GC%
OY968831.1	1	17.62	34.50
OY968832.1	2	16.83	35.50
OY968833.1	3	15.23	36
OY968834.1	4	14.46	35.50
OY968836.2	5	11.85	38
OY968837.1	6	11.63	36
OY968838.1	7	10.37	38
OY968839.1	8	6.98	39.50
OY968840.1	9	6.04	41
OY968841.1	10	5.74	41.50
OY968842.1	11	5.58	41.50
OY968835.1	X	14.29	35.50

The mitochondrial genome was assembled using MitoHiFi (Uliano-Silva *et al.*, 2023).

Assembly curation

The assembly was decontaminated using the Assembly Screen for Cobionts and Contaminants (ASCC) pipeline. TreeVal was used to generate the flat files and maps for use in curation. Manual curation was conducted primarily in PretextView and HiGlass (Kerpedjiev *et al.*, 2018). Scaffolds were visually inspected and corrected as described by Howe *et al.* (2021). Manual corrections included ten breaks, 35 joins, and removal of three haplotypic duplications. This reduced the scaffold count by 94.0%, increased the scaffold N50 by 41.5%, and reduced the total assembly length by 9.6%. The curation process is described 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 MerquryFK 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 database ($k = 31$) 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 *Hololepta plana* specimen generated 24.50 Gb (gigabases) from 2.46 million reads, which were used to assemble the genome. GenomeScope2.0 analysis estimated the haploid genome size at 143.25 Mb, with a heterozygosity of 0.29% and repeat content of 21.12% (Figure 2). These estimates guided expectations for the assembly. Based on the estimated genome size, the sequencing data provided approximately 152× coverage. Hi-C sequencing

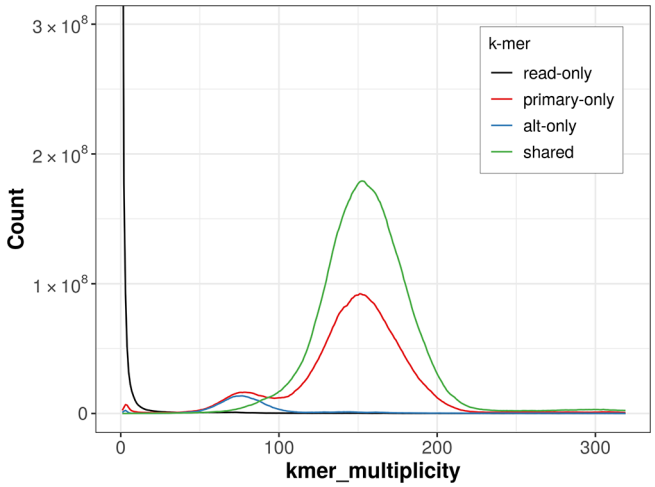

Figure 4. Evaluation of *k*-mer completeness using MerquryFK. This plot illustrates the recovery of *k*-mers from the original read data in the final assemblies. The horizontal axis represents *k*-mer multiplicity, and the vertical axis shows the number of *k*-mers. The black curve represents *k*-mers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

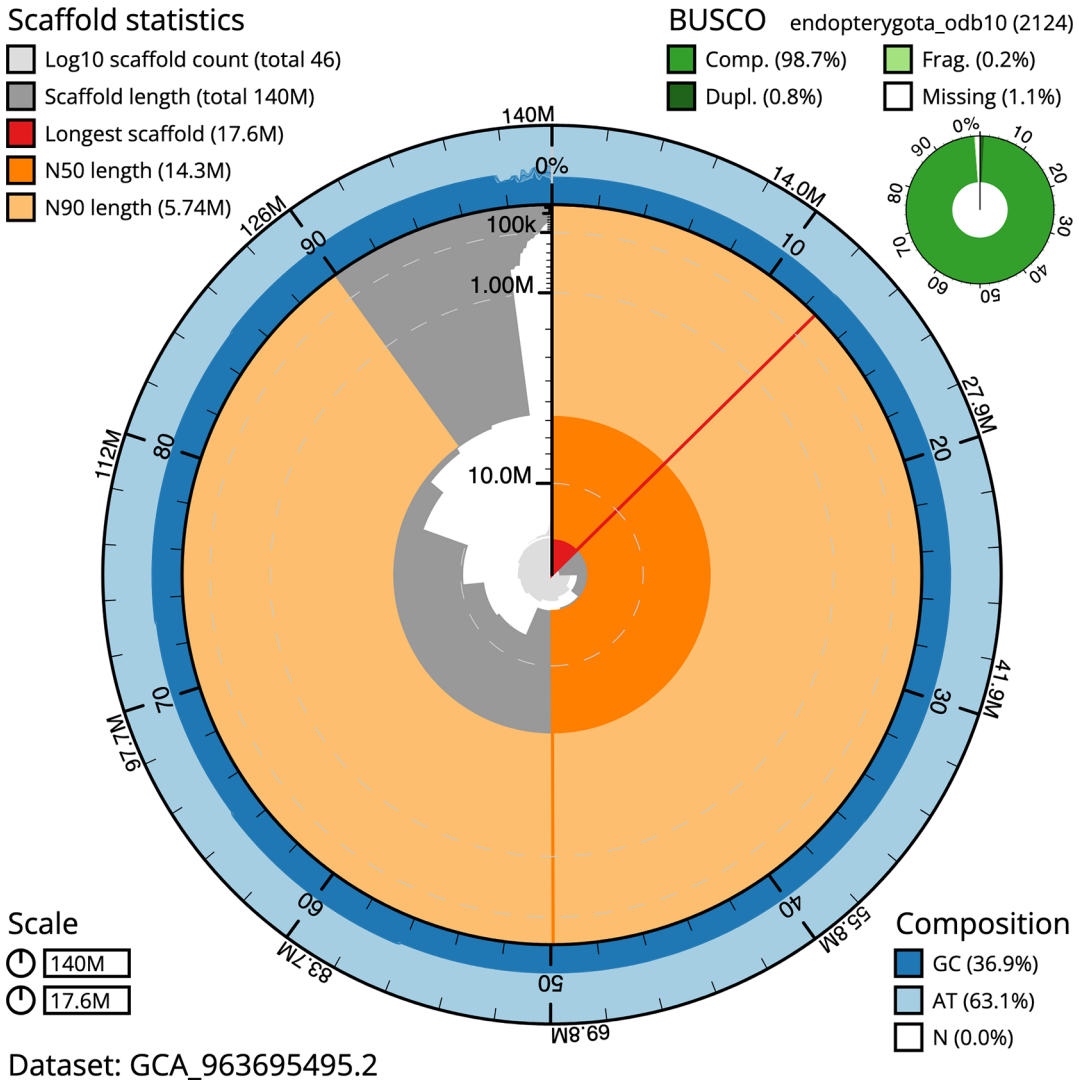

Figure 5. Assembly metrics for icHolPlan1.2. The BlobToolKit snail plot provides an overview of assembly metrics and BUSCO gene completeness. The circumference represents the length of the whole genome sequence, and the main plot is divided into 1 000 bins around the circumference. The outermost blue 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 endopterygota_odb10 set is presented at the top right. An interactive version of this figure can be accessed on the BlobToolKit viewer.

produced 98.20 Gb from 325.17 million reads, which were used to scaffold the assembly. 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 139.51 Mb in 45 scaffolds, with 100 gaps, and a scaffold N50 of 14.29 Mb (Table 2).

Most of the assembly sequence (97.93%) was assigned to 12 chromosomal-level scaffolds, representing 11 autosomes and the X sex chromosome. These chromosome-level scaffolds, confirmed by Hi-C data, are named according to size (Figure 3; Table 3). X chromosome identified based on synteny with the genome of *Leptodirus hochenwartii* (GCA_947310635.1) (Delić *et al.*, 2025).

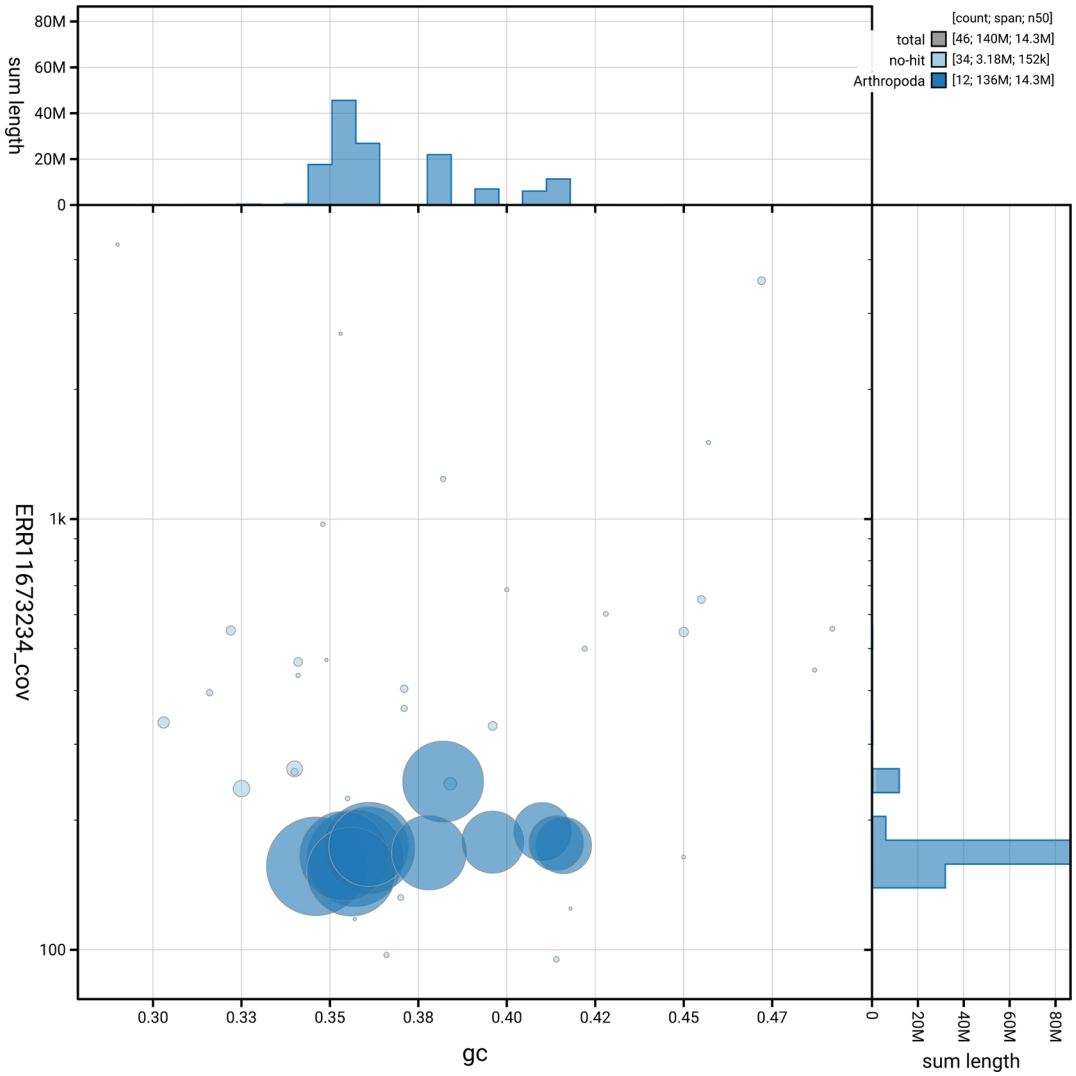

Figure 6. BlobToolKit blob plot for icHolPlan1.2. The plot shows base coverage (vertical axis) and GC content (horizontal axis). The circles represent scaffolds, with the size proportional to scaffold length and the colour representing phylum membership. The histograms along the axes display the total length of sequences distributed across different levels of coverage and GC content. An interactive version of this figure is available on the BlobToolKit viewer.

The mitochondrial genome was also assembled (length 17.73 kb, OY968843.1). This sequence is included as a contig in the multifasta file of the genome submission and as a standalone record.

Assembly quality metrics

The combined primary and alternate assemblies achieve an estimated QV of 55.1. The *k*-mer completeness is 93.62% for the primary assembly, 64.96% for the alternate haplotype, and 99.13% for the combined assemblies (Figure 4).

BUSCO v.6.0.0 analysis using the endopterygota_odb10 reference set (n = 2 124) identified 98.7% of the expected gene set (single = 97.9%, duplicated = 0.8%). 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) and the Earth BioGenome Project Report on Assembly Standards January 2026. The EBP metric, calculated for the primary assembly, is **6.C.Q54**, meeting the recommended reference standard.

Table 4. **Earth Biogenome Project summary metrics for the *Hololepta plana* assembly.**

Measure	Value	Benchmark
EBP summary (primary)	6.C.Q54	6.C.Q40
Contig N50 length	2.20 Mb	≥ 1 Mb
Scaffold N50 length	14.29 Mb	= chromosome N50
Consensus quality (QV)	Primary: 54.1; alternate: 55.9; combined: 55.1	≥ 40
k-mer completeness	Primary: 93.62%; alternate: 64.96%; combined: 99.13%	≥ 95%
BUSCO	C:98.7% [S:97.9%, D:0.8%], F:0.2%, M:1.1%, n:2 124	S > 90%; D < 5%
Percentage of assembly assigned to chromosomes	97.93%	≥ 90%

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

Table 5. **Software versions and sources used for *Hololepta plana*.**

Software	Version	Source
BLAST	2.14.0	ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/
BlobToolKit	4.4.6	https://github.com/blobtoolkit/blobtoolkit
BUSCO	6.0.0	https://gitlab.com/ezlab/busco
bwa-mem2	2.2.1	https://github.com/bwa-mem2/bwa-mem2
DIAMOND	2.1.8	https://github.com/bbuchfink/diamond
fasta_windows	0.2.4	https://github.com/tolkit/fasta_windows
FastK	1.1	https://github.com/thegenemyers/FASTK
GenomeScope2.0	2.0.1	https://github.com/tbenavi1/genomescope2.0
Gfastats	1.3.6	https://github.com/vgl-hub/gfastats
Hifiasm	0.19.5-r587	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	2.28-r1209	https://github.com/lh3/minimap2
MitoHiFi	3	https://github.com/marcelauliano/MitoHiFi
MultiQC	1.14; 1.17 and 1.18	https://github.com/MultiQC/MultiQC
Nextflow	24.10.4	https://github.com/nextflow-io/nextflow
PretextSnapshot	0.0.5	https://github.com/sanger-tol/PretextSnapshot
PretextView	1.0.3	https://github.com/sanger-tol/PretextView
purge_dups	1.2.5	https://github.com/dfguan/purge_dups
samtools	1.21	https://github.com/samtools/samtools
sanger-tol/ascc	0.1.0	https://github.com/sanger-tol/ascc
sanger-tol/blobtoolkit	v0.9.0	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	1.2a.2	https://github.com/c-zhou/yahs

Data availability

European Nucleotide Archive: Hololepta plana. Accession number PRJEB64074. The genome sequence is released openly for reuse. The *Hololepta plana* 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 Tables 1 and 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

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.

References

Allen AJ, Hance D: ***Hololepta plana* (Sulzer, 1776) (Histeridae) in Norfolk – new to Britain.** *Coleopterist.* 2009; **18**: 153–154.

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

Barclay MVL: ***Hololepta plana* (Sulzer) (Histeridae) from Margravine Cemetery, West London.** *Coleopterist.* 2018; **27**: 27–28.

Bateman A, Martin M-J, Orchard S, *et al.*: **UniProt: The Universal Protein Knowledgebase in 2023.** *Nucleic Acids Res.* 2023; **51**(D1): D523–D531.
PubMed Abstract | Publisher Full Text | Free Full Text

Buchfink B, Reuter K, Drost H-G: **Sensitive protein alignments at tree-of-life scale using DIAMOND.** *Nat. Methods.* 2021; **18**(4): 366–368.
PubMed Abstract | Publisher Full Text | Free Full Text

Challis R, Richards E, Rajan J, *et al.*: **BlobToolKit – interactive quality assessment of genome assemblies.** *G3: Genes, Genomes, Genetics.* 2020; **10**(4): 1361–1374.
PubMed Abstract | Publisher Full Text | Free Full Text

Cheng H, Concepcion GT, Feng X, *et al.*: **Haplotype-resolved *de novo* assembly using phased assembly graphs with Hifiasm.** *Nat. Methods.* 2021; **18**(2): 170–175.
PubMed Abstract | Publisher Full Text | Free Full Text

Collier MJ, Lane S: **Recent Norfolk beetle records, including twelve additions to the county list.** *Coleopterist.* 2015; **24**: 46–52.

Crowley L, Allen H, Barnes I, *et al.*: **A sampling strategy for genome sequencing the British terrestrial Arthropod fauna.** *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

Delić TWellcome Sanger Institute Tree of Life Management, Samples and Laboratory teamWellcome Sanger Institute Scientific Operations: Sequencing Operations, *et al.*: **The genome sequence of a cave beetle, *Leptodirus hochenwartii* F.J.Schmidt, 1832.** *Wellcome Open Res.* 2025; **10**: 159.
PubMed Abstract | Publisher Full Text | Free Full Text

Ewels P, Magnusson M, Lundin S, *et al.*: **MultiQC: Summarize analysis results for multiple tools and samples in a single report.** *Bioinformatics.* 2016; **32**(19): 3047–3048.
PubMed Abstract | Publisher Full Text | Free Full Text

Ewels PA, Peltzer A, Fillinger S, *et al.*: **The nf-core framework for community-curated bioinformatics pipelines.** *Nat. Biotechnol.* 2020; **38**(3): 276–278.
PubMed Abstract | Publisher Full Text

Formenti G, Abueg L, Brajuka A, *et al.*: **Gfastats: Conversion, evaluation and manipulation of genome sequences using assembly graphs.** *Bioinformatics.* 2022; **38**(17): 4214–4216.
PubMed Abstract | Publisher Full Text | Free Full Text

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

Kryzhanovsky OL, Reikhardt AN: *Beetles of the Superfamily Histeroidea (Families Sphaeritidae, Histeridae, Synteliidae).* Leningrad: Nauka Press; 1976.

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

Lackner T, Mazur S, Newton AF: **Family Histeridae Gyllenhal, 1808.** Löbl I, Löbl D, editors. *Catalogue of Palaearctic Coleoptera 2. Hydrophiloidea Staphylinoidea (Revised and Updated).* Stenstrup: Apollo Books; 2015; 76–130.

Lane SA: ***A Review of the Status of the Beetles of Great Britain – the Clown Beetles and False Clown Beetles – Histeridae and Sphaeritidae.*** 2017.

Lane SA, Lucas CBH, Whiffin AL: *The Histeridae, Sphaeritidae and Silphidae of Britain and Ireland.* Field Studies Council; 2021.

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.
Publisher Full Text

Nagaraju DK, Kasturi N, Singh M, *et al.*: **Interception of *Hololepta plana* (Sulzer) (Coleoptera, Histeridae) a rare beetle on poplar logs imported from Belgium.** *Insect Environ.* 2022; **25**: 18–20.

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

Sánchez-Ruiz A, Utrilla JA, Sánchez-Ruiz M, *et al.*: **Ampliación de la distribución conocida de *Hololepta* (s. Str.) *plana* (Sulzer, 1776) en la península Ibérica (Coleoptera: Histeridae).** *Boletín de La SEA.* 2002; **31**: 202.

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

Telfer M: **Tales of the unexpected.** 2014.
Reference Source

Trach VA: **On the morphology, biology, and distribution of *Lobogynioides andreinii* (Acari, Mesostigmata, Diplogyniidae).** *Entomol. Rev.* 2013; **93**: 105–112.
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.** *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.** *2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS).* IEEE, 2019; 314–24.
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 04 June 2026

https://doi.org/10.21956/wellcomeopenres.29281.r156252

✔ **Henrik Lantz**
SciLifeLab, Uppsala University, Uppsala, Sweden

This manuscript presents the genome sequence of the flat clown beetle, Hololepta plana, assembled as a part of the Darwin Tree of Life project. The species is rare in Britain, and was only first collected there in 2009, making experts think the species most likely is a recent immigrant. The species is large compared to closely related species and has specific morphological features which makes it a distinct species in the family Histeridae.

The assembled genome is of high quality and can be used to determine the origin of the British population and how it is related to other populations in continental Europe. 97.93% of the assembly (a high number) is scaffolded into 12 chromosomal pseudomolecules, including the sex determination X chromosome. BUSCO score is also very high (98.7%) and the assembly easily meets all of the Earth Biogenome Project criteria except possibly the scaffold N50 value as not 100% of the scaffolds are assigned to chromosomes, but that difference is negligible.

Methods follow a high international standard and are in general well-described, except for some few cases commented below.

1. Figure 1, and text: The specimen used has been documented by a picture which is good, and the text explains how tissues were dissected and used for DNA extraction and barcoding, but there is no information if any material has been saved in a museum and/or biobank. The morphology of the species seems distinctive enough that it should be possible to determine the identity of the specimen from the photo, but I would prefer that a high quality photo is preserved in a Natural History Museum and also any remaining tissue, if available.

2. Figure 3. The HiC contact map shows some remaining contacts between different scaffolds, e.g., between OY968836.2 and OY968837.1. This is most likely caused by repeats and not really something that needs to be corrected, but it would be good to have the results of the curation discussed more in detail in the text.

3. The BlobToolKit plot in Figure 6 shows some small blobs outside of the main cluster with no hits

in the database, hinting at remaining contamination. While this is unlikely to cause any big problems in downstream analyses, I would prefer if the results of the de-contamination would be discussed in the text.

4. Simply stating that the mitochondrion was assembled using MitoHiFi is not enough to ensure reproducibility. Settings should be mentioned, or even better, scripts should be supplied in a GitHub repo.

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?
Partly

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

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Genome assembly, genome annotation, phylogenetics, taxonomy, systematics.

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 03 June 2026

https://doi.org/10.21956/wellcomeopenres.29281.r156249

Rajib Deb
ICAR-National Research Centre, Guwahat, Assam, India

The rationale for generating the genome dataset is clearly explained and biologically justified.
The sequencing, assembly, and curation protocols are appropriate and technically sound, using current standard methodologies and software tools.
Methods and materials are described in sufficient detail to enable reproducibility, including software versions, sequencing platforms, and accession numbers
The genome assembly is of high quality, with strong BUSCO completeness and chromosome-level scaffolding statistics. The datasets are clearly presented and openly accessible through public

repositories, making them usable for future research.
However, following issues may be resolved:

- The manuscript could briefly clarify whether any residual unplaced scaffolds contain repetitive or low-complexity regions.
- A short statement on expected future genome annotation outcomes or comparative genomic applications would further enhance the utility of the note.
- Figure legends are generally clear, but adding a brief interpretation of the Hi-C contact map quality in the text would improve accessibility for non-specialist readers.
- Minor typographical and formatting consistency checks (e.g., spacing around percentages and symbols) may improve readability.

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

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.