



DATA NOTE

The genome sequence of the Barred Hook-tip, *Watsonalla cultraria* (Fabricius, 1775) (Lepidoptera: Drepanidae)

[version 1; peer review: 2 approved, 1 approved with reservations]

Liam M. Crowley¹, Finley Hutchinson², Douglas Boyes³⁺,
University of Oxford and Wytham Woods Genome Acquisition Lab,
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

¹University of Oxford, Oxford, England, UK

²University of Exeter, Penryn, Cornwall, England, UK

³UK Centre for Ecology & Hydrology, Wallingford, England, UK

⁺ Deceased author

V1 First published: 05 Nov 2025, 10:618
<https://doi.org/10.12688/wellcomeopenres.25058.1>
Latest published: 05 Nov 2025, 10:618
<https://doi.org/10.12688/wellcomeopenres.25058.1>

Abstract

We present a genome assembly from a male specimen of *Watsonalla cultraria* (Barred Hook-tip; Arthropoda; Insecta; Lepidoptera; Drepanidae). The genome sequence has a total length of 319.38 megabases. Most of the assembly (99.94%) is scaffolded into 31 chromosomal pseudomolecules, including the Z sex chromosome. Gene annotation of this assembly on Ensembl identified 16 011 protein-coding genes. The mitochondrial genome has also been assembled, with a length of 15.21 kilobases.

Keywords

Watsonalla cultraria; Barred Hook-tip; genome sequence; chromosomal; Lepidoptera



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

Open Peer Review

Approval Status ?

	1	2	3
version 1			?
05 Nov 2025	view	view	view
1. Ozren Polašek	University of Split, Split, Croatia		
2. Olli-Pekka Smolander	id University of Helsinki, Helsinki, Finland		
3. Charles Plessy	id Okinawa Institute of Science and Technology, Okinawa, Japan		
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; Hutchinson F: Investigation, Resources; Boyes D: Investigation, Resources;

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.*

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, Hutchinson F, Boyes D *et al.* **The genome sequence of the Barred Hook-tip, *Watsonalla cultraria* (Fabricius, 1775) (Lepidoptera: Drepanidae) [version 1; peer review: 2 approved, 1 approved with reservations]** Wellcome Open Research 2025, **10:618** <https://doi.org/10.12688/wellcomeopenres.25058.1>

First published: 05 Nov 2025, **10:618** <https://doi.org/10.12688/wellcomeopenres.25058.1>

Species taxonomy

Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Protostomia; Ecdysozoa; Panarthropoda; Arthropoda; Mandibulata; Pancrustacea; Hexapoda; Insecta; Dicondylia; Pterygota; Neoptera; Endopterygota; Amphiesmenoptera; Lepidoptera; Glossata; Neolepidoptera; Heteroneura; Ditrysia; Obtectomera; Drepanoidea; Drepanidae; Drepaninae; *Watsonalla*; *Watsonalla cultraria* (Fabricius, 1775) (NCBI:txid721166)

Background

Watsonalla cultraria (Barred Hook-tip) is a local woodland moth associated with mature beech, and is most frequent where extensive beech stands occur. (Waring *et al.*, 2017)

Adults have two flight periods (May to June and mid-July to early September). It overwinters as a pupa in a white cocoon within a curled beech leaf; larvae occur mainly June to September. Males may fly by day near the foodplant; both sexes fly at night and come to light. (Waring *et al.*, 2017)

In Britain it is well distributed in the south, more local in the Midlands, Wales and northern England, with recent spread noted in Yorkshire and Lancashire. It was first recorded on the Isle of Man in 2008; in Ireland it appears to be a recent colonist (first Co. Down record 1999; Irish Republic from 2007, very local), and it was first noted on Guernsey in 2013. Records on GBIF show a mainly western and central European distribution with many coastal and lowland records across the United Kingdom, Ireland and adjacent seas. (GBIF Secretariat, 2025; Waring *et al.*, 2017)

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 – we present a chromosomally complete genome sequence for *Watsonalla cultraria*, the Barred Hook-tip. This genome was assembled using the Tree of Life pipeline from a specimen collected in Wytham Woods, Oxfordshire, United Kingdom (Figure 1).

Fewer than 20 genomes have been published for Drepanidae as of August 2025, including two genomes for the genus *Watsonalla* (data obtained via NCBI datasets, O'Leary *et al.*, 2024). This assembly adds chromosome-scale data for the lineage. It is currently the only genome assembly available for *Watsonalla cultraria* and is the RefSeq reference genome for the species (O'Leary *et al.*, 2016).

Methods

Sample acquisition and DNA barcoding

The specimen used for genome sequencing was an adult male *Watsonalla cultraria* (specimen ID Ox003068, ToLID ilWatCult2; Figure 1), collected from Wytham Woods, Oxfordshire, UK (latitude 51.772, longitude -1.338) on 2022-07-22. The specimen was collected by Finley Hutchinson and Liam Crowley and formally identified by Finley Hutchinson. A second specimen used for Hi-C sequencing (specimen ID



Figure 1. Photograph of the *Watsonalla cultraria* (ilWatCult2) specimen from which samples were taken for genome sequencing.

Ox000407, ToLID ilWatCult1) was collected from the same location on 2020-05-22. This specimen was collected and identified by Douglas Boyes. 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.

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 ilWatCult2 sample was weighed and triaged to determine the appropriate extraction protocol. Tissue from the whole organism was homogenised by powermashing using a PowerMasher II tissue disruptor. HMW DNA was extracted 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 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 13.8 ng/µL and a yield of 1 794.00 ng.

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 on a Revio instrument (Pacific Biosciences). The prepared library was normalised to 2 nM, and 15 µL was used for making complexes. Primers were annealed and polymerases bound to generate circularised complexes, following the manufacturer's instructions. Complexes were purified using 1.2X SMRTbell beads, then diluted to the Revio loading concentration (200–300 pM) and spiked with a Revio sequencing internal control. The sample was sequenced on a Revio 25M SMRT cell. The SMRT Link software (Pacific Biosciences), a web-based workflow manager, was used to configure and monitor the run and to carry out primary and secondary data analysis.

Hi-C

Sample preparation and crosslinking

The Hi-C sample was prepared from 20–50 mg of frozen tissue of the iWatCult1 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 were created. 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 HiSeq X Ten.

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 (Rhee *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 15 breaks, 14 joins, and removal of two haplotypic duplications. 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 (Rhee *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 *Watsonalla cultraria* specimen generated 22.33 Gb (gigabases) from 2.11 million reads, which were used to assemble the genome. GenomeScope2.0 analysis estimated the haploid genome size at 326.32 Mb, with a heterozygosity of 0.06% and repeat content of 20.56% ([Figure 2](#)). These estimates guided expectations for the assembly. Based on the estimated genome size, the sequencing data provided approximately 66 \times coverage. Hi-C sequencing produced 80.44 Gb from 532.70 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 319.38 Mb in 36 scaffolds, with 56 gaps, and a scaffold N50 of 11.4 Mb ([Table 2](#)).

Most of the assembly sequence (99.94%) was assigned to 31 chromosomal-level scaffolds, representing 30 autosomes and

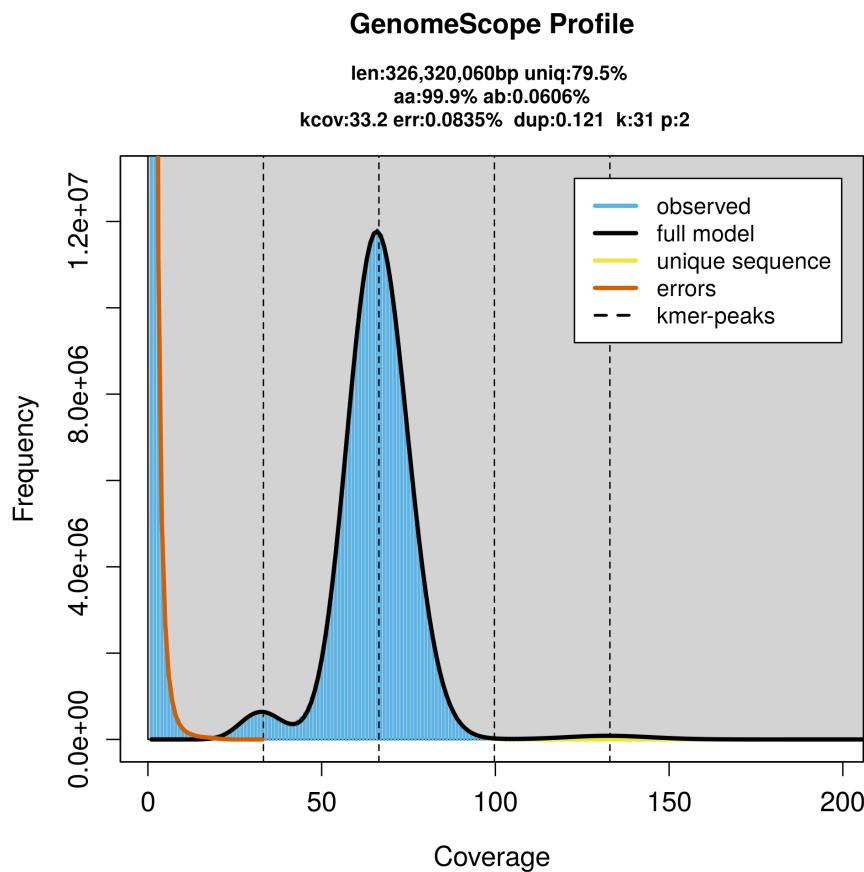


Figure 2. Frequency distribution of k -mers generated using GenomeScope2. The plot shows observed and modelled k -mer spectra, providing estimates of genome size, heterozygosity, and repeat content based on unassembled sequencing reads.

Table 1. Specimen and sequencing data for BioProject PRJEB73407.

Platform	PacBio HiFi	Hi-C
ToLID	ilWatCult2	ilWatCult1
Specimen ID	Ox003068	Ox000407
BioSample (source individual)	SAMEA112775017	SAMEA7520529
BioSample (tissue)	SAMEA112775080	SAMEA7520624
Tissue	whole organism	whole organism
Instrument	Revio	HiSeq X Ten
Run accessions	ERR12721059	ERR12723477; ERR12723476; ERR12723478
Read count total	2.11 million	532.70 million
Base count total	22.33 Gb	80.44 Gb

Table 2. Genome assembly statistics.

Assembly name	ilWatCult2.1
Assembly accession	GCA_964035505.1
Alternate haplotype accession	GCA_964035515.1
Assembly level	chromosome
Span (Mb)	319.38
Number of chromosomes	31
Number of contigs	92
Contig N50	6.83 Mb
Number of scaffolds	36
Scaffold N50	11.4 Mb
Sex chromosomes	Z
Organelles	Mitochondrion: 15.21 kb

the Z sex chromosome. These chromosome-level scaffolds, confirmed by Hi-C data, are named according to size (Figure 3; Table 3).

The mitochondrial genome was also assembled (length 15.21 kb, OZ037742.1). 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 61.8. The *k*-mer completeness is 97.74% for the primary assembly, 88.30% for the alternate haplotype, and 99.32% for the combined assemblies (Figure 4).

BUSCO v.5.5.0 analysis using the lepidoptera_odb10 reference set ($n = 5\,286$) identified 98.7% of the expected gene set (single = 98.5%, 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 Rie *et al.* (2021) and the Earth BioGenome Project Report on Assembly Standards September 2024. The EBP metric, calculated for the primary assembly, is **6.C.Q63**, meeting the recommended reference standard.

Genome annotation report

The *Watsonalla cultraria* genome assembly (GCA_964035505.1) was annotated by Ensembl at the European Bioinformatics Institute (EBI). This annotation includes 16 209 transcribed mRNAs from 16 011 protein-coding genes. The average transcript length is 6 238.38 bp, with an average of 6.14 exons

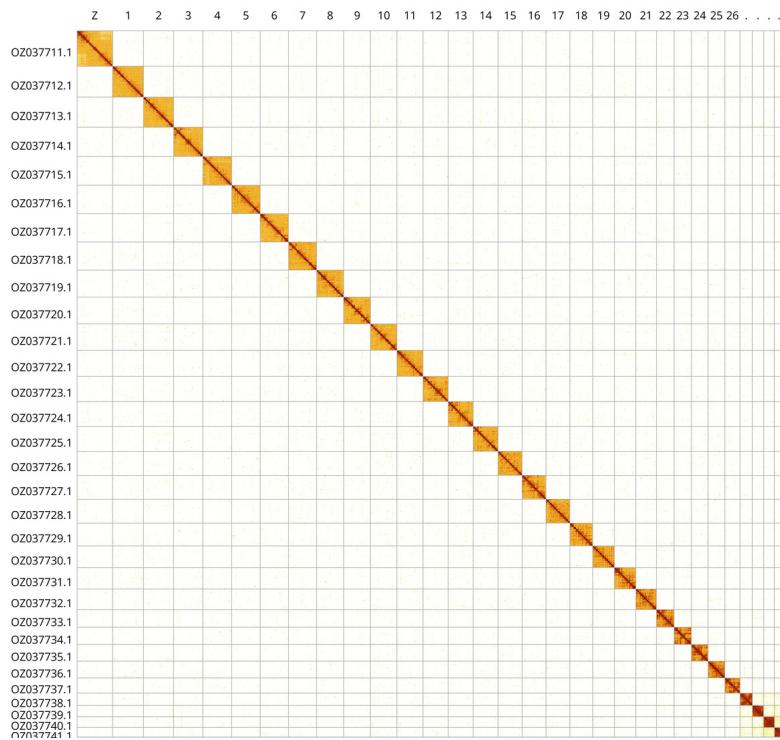


Figure 3. Hi-C contact map of the *Watsonalla cultraria* genome assembly. Assembled chromosomes are shown in order of size and labelled along the axes, with a megabase scale shown below. The plot was generated using PretextSnapshot.

Table 3. Chromosomal pseudomolecules in the primary genome assembly of *Watsonalla cultraria* iWatCult2.

INSDC accession	Molecule	Length (Mb)	GC%
OZ037712.1	1	13.96	37
OZ037713.1	2	13.62	37.50
OZ037714.1	3	13.13	37.50
OZ037715.1	4	13.02	36.50
OZ037716.1	5	12.86	37
OZ037717.1	6	12.80	37.50
OZ037718.1	7	12.65	37
OZ037719.1	8	12.26	36.50
OZ037720.1	9	12.11	36.50
OZ037721.1	10	11.96	37
OZ037722.1	11	11.70	37.50
OZ037723.1	12	11.40	37
OZ037724.1	13	11.31	37
OZ037725.1	14	11.14	37

INSDC accession	Molecule	Length (Mb)	GC%
OZ037726.1	15	10.93	37
OZ037727.1	16	10.76	37
OZ037728.1	17	10.75	37.50
OZ037729.1	18	10.31	38
OZ037730.1	19	9.78	37
OZ037731.1	20	9.73	37.50
OZ037732.1	21	9.23	37
OZ037733.1	22	7.97	37.50
OZ037734.1	23	7.86	38
OZ037735.1	24	7.55	38
OZ037736.1	25	7.54	37.50
OZ037737.1	26	6.83	37.50
OZ037738.1	27	5.60	38.50
OZ037739.1	28	5.10	38.50
OZ037740.1	29	4.81	39.50
OZ037741.1	30	4.37	39
OZ037711.1	Z	16.14	37

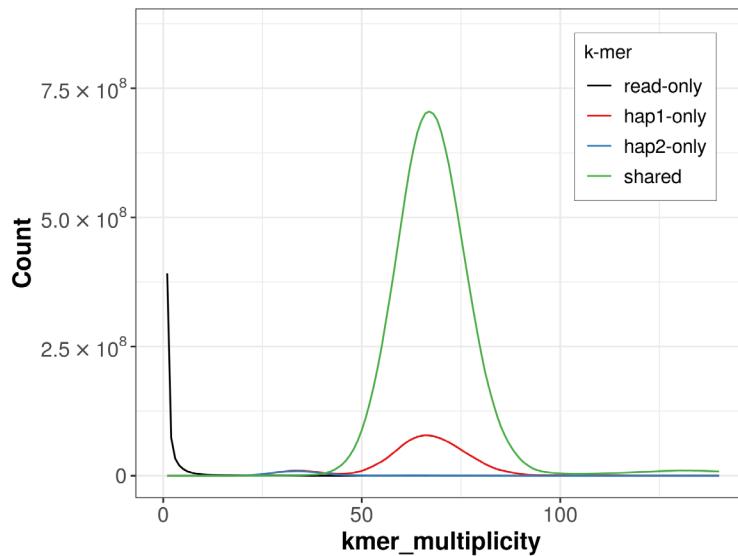


Figure 4. Evaluation of k-mer completeness using MerquryFK. This plot illustrates the recovery of k-mers from the original read data in the final assemblies. The horizontal axis represents k-mer multiplicity, and the vertical axis shows the number of k-mers. The black curve represents k-mers that appear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

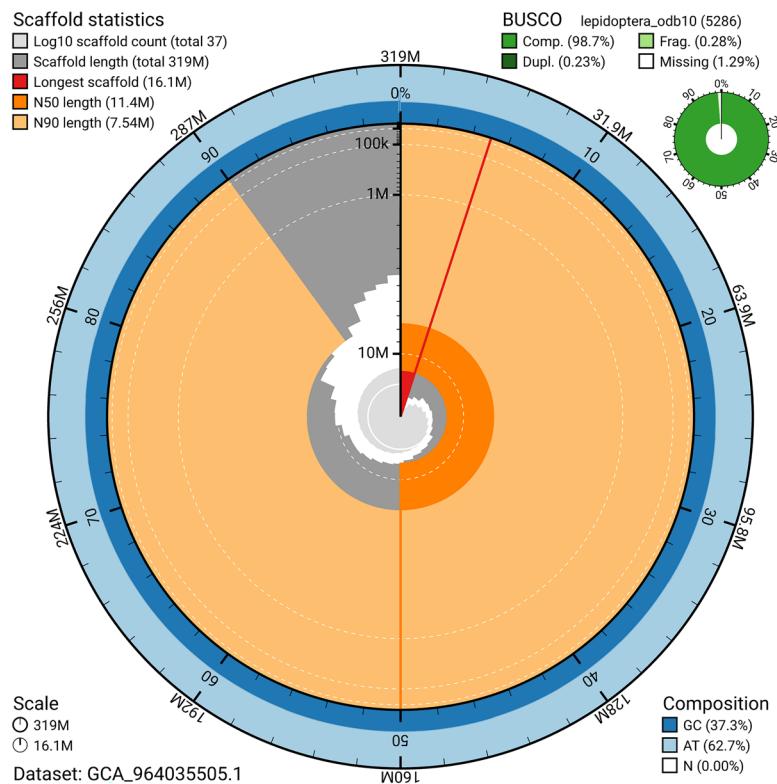


Figure 5. Assembly metrics for iWatCult2.1. The BlobToolKit snail plot provides an overview of assembly metrics and BUSCO gene completeness. The circumference represents the length of the whole genome sequence, and the main plot is divided into 1 000 bins around the circumference. The outermost blue 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 lepidoptera_odb10 set is presented at the top right. An interactive version of this figure can be accessed on the [BlobToolKit viewer](#).

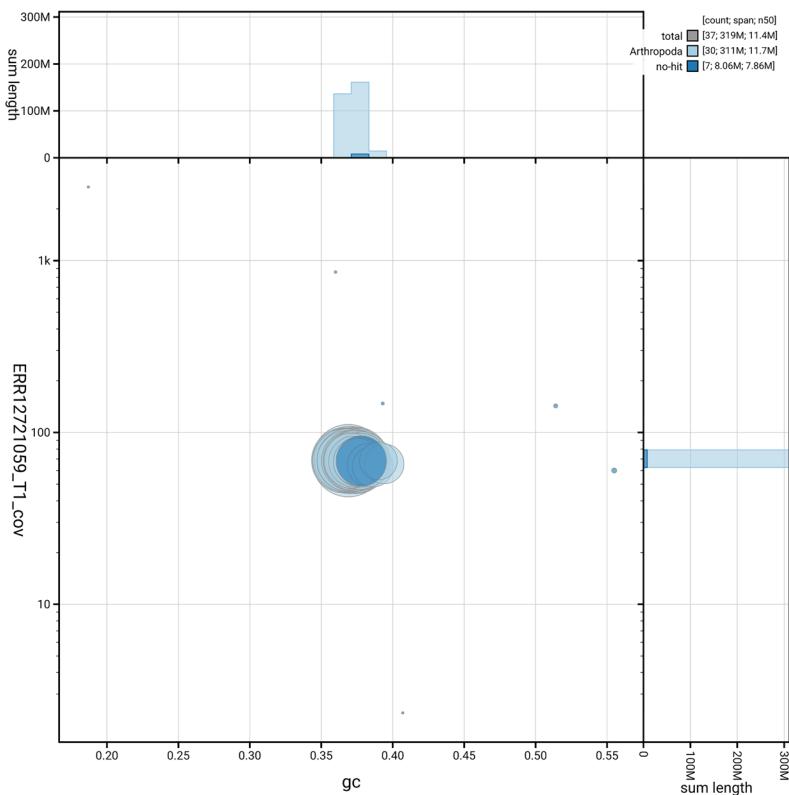


Figure 6. BlobToolKit GC-coverage plot for iWatCult2.1. Blob plot showing sequence coverage (vertical axis) and GC content (horizontal axis). The circles represent scaffolds, with the size proportional to scaffold length and the colour representing phylum membership. The histograms along the axes display 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 *Watsonalla cultraria* assembly.

Measure	Value	Benchmark
EBP summary (primary)	6.C.Q63	6.C.Q40
Contig N50 length	6.83 Mb	≥ 1 Mb
Scaffold N50 length	11.40 Mb	= chromosome N50
Consensus quality (QV)	Primary: 63.2; alternate: 61.3; combined: 62.1	≥ 40
k-mer completeness	Primary: 97.74%; alternate: 88.30%; combined: 99.32%	≥ 95%
BUSCO	C:98.7% [S:98.5%; D:0.2%]; F:0.3%; M:1.0%; n:5 286	S > 90%; D < 5%
Percentage of assembly assigned to chromosomes	99.94%	≥ 90%

per transcript. For further information about the annotation, please refer to the [annotation page](#) on Ensembl.

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: Watsonalla cultraria (barred hooktip). Accession number [PRJEB73407](#). The genome sequence is released openly for reuse. The *Watsonalla cultraria* genome sequencing initiative is part of the Darwin Tree of Life

Project (PRJEB40665), the Sanger Institute Tree of Life Programme (PRJEB43745) and Project Psyche (PRJEB71705). All raw sequence data and the assembly have been deposited in INSDC databases. 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+/?
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/thegeomyers/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	0.19.8-r603	https://github.com/chhylp123/hifiasm
HiGlass	1.13.4	https://github.com/higlass/higlass
MerquryFK	1.1.2	https://github.com/thegeomyers/MERQURY.FK
Minimap2	2.24-r1122	https://github.com/lh3/minimap2

Software	Version	Source
MitoHiFi	3	https://github.com/marcelauliano/MitoHiFi
MultiQC	1.14; 1.17 and 1.18	https://github.com/MultiQC/MultiQC
Nextflow	23.10.0	https://github.com/nextflow-io/nextflow
PretextSnapshot	-	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
samtools	1.19.2	https://github.com/samtools/samtools
sanger-tol/ascc	0.1.0	https://github.com/sanger-tol/ascc
sanger-tol/blob toolkit	0.6.0	https://github.com/sanger-tol/blob toolkit
Seqtk	1.3	https://github.com/lh3/seqtk
Singularity	3.9.0	https://github.com/sylabs/singularity
TreeVal	1.2.0	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](#)

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](#)

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](#)

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](#)

O'Leary NA, Wright MW, Brister JR, *et al.*: Reference Sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* 2016; **44**(D1): D733–45.
[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](#)

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](#)

Waring P, Townsend M, Lewington R: **Field guide to the Moths of Great Britain and Ireland.** London, UK: Bloomsbury, 2017.
[Reference Source](#)

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 07 January 2026

<https://doi.org/10.21956/wellcomeopenres.27624.r143619>

© 2026 Plessy 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.



Charles Plessy 

Okinawa Institute of Science and Technology, Okinawa, Japan

The high quality of the data is supported by all the standard quality controls done in this report, which follows the standard Darwin Tree of Life template for genome data notes. The BlobToolKit plot is one of the cleanest ones I have seen, the assembly metrics are excellent, and the single-copy orthologue genes (BUSCO) are found at a high accuracy. I have compared the genome presented here to the one of *W. binaria* and the resulting pairwise plot displays a very sharp diagonal, which rules out major scaffolding errors.

This said, I was disappointed that the sequenced specimen was a male, which means that the W chromosome was not sequenced for this species. I think that the article should 1) give a very brief introduction on the use of the ZW sex determination system in this clade, and 2) explain why a female was not sequenced.

The heterozygosity is very low (0.06%) compared with what I am used to with other wild organisms. Can the authors comment on whether this is unexpected for this species as well?

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, however I have significant reservations, as outlined above.

Reviewer Report 05 January 2026

<https://doi.org/10.21956/wellcomeopenres.27624.r138833>

© 2026 Smolander O. 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.

✓ **Olli-Pekka Smolander** 

University of Helsinki, Helsinki, Finland

The manuscript (data note) presents a high-quality, chromosomal-level genome assembly for *Watsonalla cultraria* (the Barred Hook-tip moth). The reported genome size is approximately 319 Mb with a scaffold N50 of 11.4 Mb. The genome has been assembled using a combination of PacBio HiFi long-read sequencing and Hi-C chromatin conformation capture. The assembly is highly contiguous, with 99.94% of the sequence assigned to 31 chromosomal pseudomolecules, including the Z sex chromosome. The mitochondrial genome has also been assembled and included in the assembly and as a separate database entry. Gene annotation has been done using Ensembl pipeline, with 16,011 protein-coding genes predicted. The mitochondrial assembly length (15.21 kb) aligns well with other recent Drepanidae assemblies, such as *Tethea ocularis* (15.28 kb) and *Tethea albicostata* (15.31 kb).

The reported methodology is scientifically sound and consistent with the current best practices for reference genome assembly. The sampling strategy is clearly described, with adequate data provided for the collected specimens. The study used two different specimens for the PacBio and Hi-C sequencing steps which is understandable due to the size of the organism. Furthermore, it supported by the results that this is not a significant issue, looking at the the low observed heterozygosity (0.06%) and the high quality of the final scaffolding (suggesting the practical absence of chimeric sequences). The manuscript provides sufficient detail regarding the library preparation, sequencing platforms (PacBio Revio and HiSeq X Ten), and the bioinformatics pipeline (*Hifiasm*, *purge_dups*, *YaHS*, etc.), ensuring reproducibility. The software versions and specific parameters are also appropriately documented.

The data quality metrics presented are excellent and meet the standards required for reference-quality genomes. The assembly achieves a consensus Quality Value (QV) of 61.8, which significantly exceeds the Earth BioGenome Project benchmark of QV40, indicating a very high base-level accuracy. The completeness of the assembly is further supported by a BUSCO score of 98.7% (single = 98.5%, duplicated = 0.2%) against the lepidoptera_odb10 set, indicating that the vast majority of expected orthologs are present as single copies and intact. The Hi-C contact map validates the chromosomal scaffolding and the BlobToolKit analysis confirms the absence of

significant contamination from non-target organisms. The assignment of 31 autosomes and the Z chromosome is consistent with expected karyotypes in Lepidoptera, and the identification of the Z chromosome in the male (ZZ) specimen is biologically appropriate. The assembly metrics (QV ~62, N50 11.4 Mb) compare favourably with other recent Lepidoptera assemblies produced by the consortium

Overall, the manuscript provides a clear and concise contribution to the genomic resources available for the family Drepanidae. All raw sequencing data and assembly accessions are clearly listed for the ENA. There are no major flaws in the study design, execution, or presentation of the data. Consequently, I have no criticisms. The article is scientifically sound in its current form.

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: Multiomics, genomics, microbiome, bioinformatics, computational biology, ML/AI

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 29 December 2025

<https://doi.org/10.21956/wellcomeopenres.27624.r138828>

© 2025 Polašek O. 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.



Ozren Polašek

University of Split, Split, Croatia

The manuscript provides an account of sequencing of a moth species, contributing to wider understanding of biodiversity. The species determination is the single problem, as the angle of the photograph may be slightly misleading, as it does not provide a decent account of the tip. The applied sequencing methodology is appropriate and well-defined, thus leaving only a possible

problem in species assignment.

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: Hymenoptera taxonomy

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.
