DATA NOTE

The genome sequence of the Moorland Grey, *Eudonia murana* (Curtis, 1827) (Lepidoptera: Crambidae)

[version 1; peer review: 2 approved]

Vladimir Blagoderov[1], Gavin R. Broad[2],
Natural History Museum Genome Acquisition Lab,
Darwin Tree of Life Barcoding Collective,
Wellcome Sanger Institute Tree of Life Management, Samples and Laboratory team,
Wellcome Sanger Institute Scientific Operations: Sequencing Operations,
Wellcome Sanger Institute Tree of Life Core Informatics team,
Tree of Life Core Informatics collective, Darwin Tree of Life Consortium

[1]National Museums Scotland, Edinburgh, Scotland, UK
[2]Natural History Museum, London, England, UK

v1 **First published:** 22 Dec 2025, **10**:686
https://doi.org/10.12688/wellcomeopenres.25207.1
Latest published: 22 Dec 2025, **10**:686
https://doi.org/10.12688/wellcomeopenres.25207.1

Abstract

We present a genome assembly from an individual female *Eudonia murana* (Moorland Grey; Arthropoda; Insecta; Lepidoptera; Crambidae). The assembly contains two haplotypes with total lengths of 729.58 megabases and 550.40 megabases. Most of haplotype 1 (98.55%) is scaffolded into 31 chromosomal pseudomolecules, including the W, Z_1 and Z_2 sex chromosomes. Most of haplotype 2 (95.39%) is scaffolded into 28 chromosomal pseudomolecules. The mitochondrial genome has also been assembled, with a length of 15.34 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

Eudonia murana; Moorland Grey; genome sequence; chromosomal; Lepidoptera

This article is included in the Tree of Life gateway.

Open Peer Review

Approval Status ✓ ✓

	1	2
version 1 22 Dec 2025	✓ view	✓ view

1. **Marko Mutanen**, University of Oulu, Oulu, Finland
2. **Danilo Trabudo do Amaral**, Universidade Federal do ABC Centro de Ciencias Naturais e Humanas (Ringgold ID: 425753), Santo André, Brazil

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: Blagoderov V: Investigation, Resources; **Broad GR**: 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.

How to cite this article: Blagoderov V, Broad GR, Natural History Museum Genome Acquisition Lab *et al.* **The genome sequence of the Moorland Grey, *Eudonia murana* (Curtis, 1827) (Lepidoptera: Crambidae) [version 1; peer review: 2 approved]** Wellcome Open Research 2025, **10**:686 https://doi.org/10.12688/wellcomeopenres.25207.1

First published: 22 Dec 2025, **10**:686 https://doi.org/10.12688/wellcomeopenres.25207.1

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; Pyraloidea; Crambidae; Scopariinae; *Eudonia*; *Eudonia murana* (Curtis, 1827) (NCBI:txid1594301)

Background

Eudonia murana (Curtis, 1827), the Moorland Grey, is a crambid micromoth of upland, rocky moorland. In Britain it is local, with doubtful records from the south-west and lowland north-west; confusion with *E. truncicolella* is frequent (Kimber, 2025; Sterling *et al.*, 2023). Across Europe, GBIF records indicate a mainly Fennoscandian distribution (GBIF Secretariat, 2025).

Forewings of the Moorland Grey are elongate and pointed, white, irregularly speckled or peppered blackish, with whitish cross-lines at about one-third and two-thirds that curve around an obscure dark "8"/X-shape; hindwings pale whitish (Sterling *et al.*, 2023). Adults fly from late May to mid-September. Larvae feed in a silken tube on mosses including *Hypnum cupressiforme*, *Dicranum scoparium*, *Bryum capillare* and *Grimmia pulvinata* (Sterling *et al.*, 2023). Separation from *E. truncicolella* and *Scoparia ambigualis* can be difficult, even with dissection, and should be treated with care (Kimber, 2025; Sterling *et al.*, 2023).

We present a chromosome-level genome sequence for *E. murana*. The assembly was generated with the Tree of Life pipeline from a specimen collected from Beinn Eighe (Figure 1) as part of the Darwin Tree of Life programme to produce reference genomes for all named UK and Irish eukaryotes (Darwin Tree of Life Project Consortium, 2022).

Figure 1. Photograph of the *Eudonia murana* (ilEudMura1) specimen used for genome sequencing.

Methods

Sample acquisition and DNA barcoding

The specimen used for genome sequencing was an adult female *Eudonia murana* (specimen ID NHMUK015060364, ToLID ilEudMura1; Figure 1), collected from Beinn Eighe, Nature Scot Visitor Centre, Kinlochewe, Wester Ross, Scotland, UK (latitude 57.61, longitude –5.31) on 2022-08-23. The specimen was collected by Vladimir Blagoderov (Natural History Museum Scotland) and identified by Gavin Broad (Natural History Museum). For the Darwin Tree of Life sampling and metadata approach, refer to Lawniczak *et al.* (2022).

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

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 ilEudMura1 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 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 11.33 ng/μL and a yield of 532.51 ng, with a fragment size of 15.2 kb.

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 from the head and thorax of the ilEudMura1 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 Illumina NovaSeq X.

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 in Hi-C phasing mode (Cheng *et al.*, 2021; Cheng *et al.*, 2022), producing two haplotypes. Hi-C reads (Rao *et al.*, 2014) were mapped to the primary contigs using bwa-mem2 (Vasimuddin *et al.*, 2019). Contigs were further scaffolded with Hi-C data in YaHS (Zhou *et al.*, 2023), using 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 35 breaks and 481 joins. This reduced the scaffold count by 55.8%, increased the scaffold N50 by 5.9%, and increased the total assembly length by 4.7%. 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 both 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 *Eudonia murana* specimen generated 54.66 Gb (gigabases) from 5.31 million reads, which were used to assemble the genome. GenomeScope2.0 analysis estimated the haploid genome size at 640.00 Mb, with a heterozygosity of 0.96% and repeat content of 38.45% (Figure 2). These estimates guided expectations for the assembly. Based on the estimated genome size, the sequencing data provided approximately 83× coverage. Hi-C sequencing produced 116.93 Gb from 774.34 million reads, which were used to scaffold the assembly. Table 1 summarises the specimen and sequencing details.

Assembly statistics

The genome was assembled into two haplotypes using Hi-C phasing. Haplotype 1 was curated to chromosome level, while haplotype 2 was assembled to scaffold level. The final assembly has a total length of 729.58 Mb in 210 scaffolds, with 611 gaps, and a scaffold N50 of 22.1 Mb (Table 2).

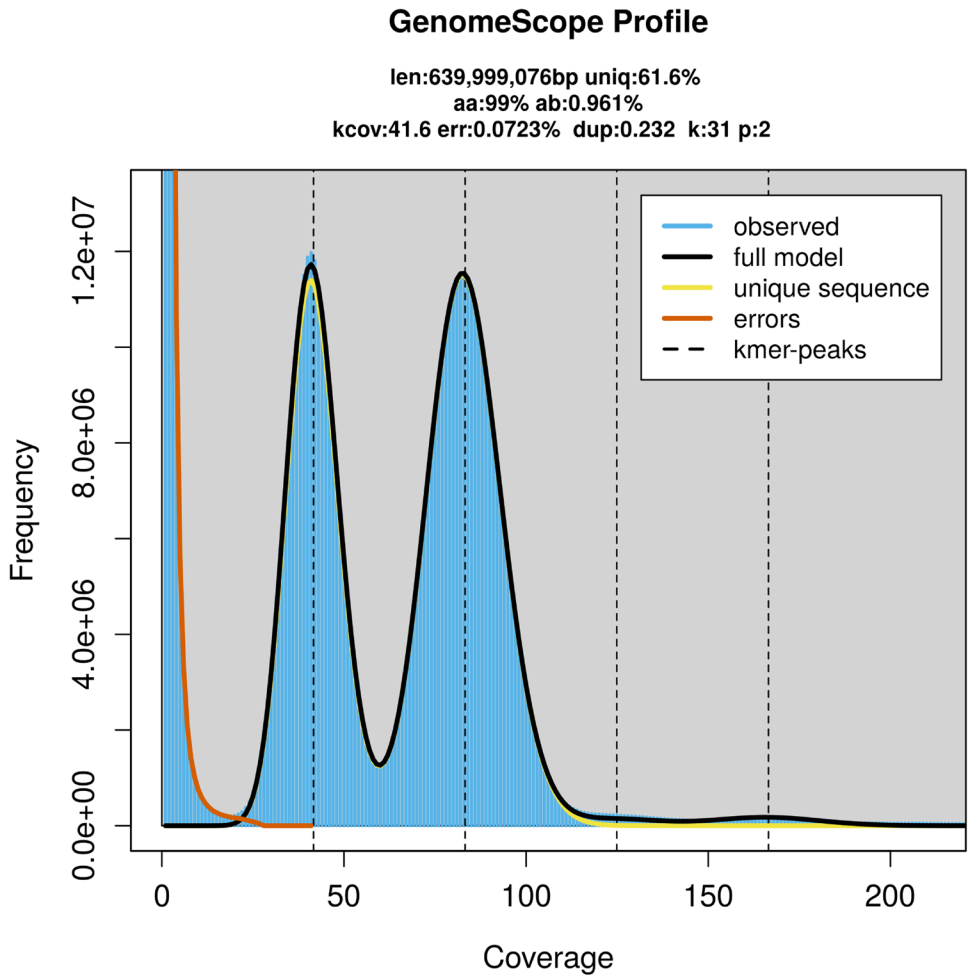

Figure 2. Frequency distribution of *k*-mers generated using GenomeScope2. The plot shows observed and modelled *k*-mer spectra, providing estimates of genome size, heterozygosity, and repeat content based on unassembled sequencing reads.

Table 1. Specimen and sequencing data for BioProject PRJEB82145.

Platform	PacBio HiFi	Hi-C
ToLID	ilEudMura1	ilEudMura1
Specimen ID	NHMUK015060364	NHMUK015060364
BioSample (source individual)	SAMEA115574703	SAMEA115574703
BioSample (tissue)	SAMEA115599800	SAMEA115599800
Tissue	head and thorax	head and thorax
Instrument	Revio	Illumina NovaSeq X
Run accessions	ERR13946450	ERR13947469
Read count total	5.31 million	774.34 million
Base count total	54.66 Gb	116.93 Gb

Table 2. Genome assembly statistics.

Assembly name	ilEudMura1.hap1.1	ilEudMura1.hap2.1
Assembly accession	GCA_964468325.1	GCA_964468265.1
Assembly level	chromosome	chromosome
Span (Mb)	729.58	550.40
Number of chromosomes	31	28
Number of contigs	821	661
Contig N50	3.92 Mb	3.94 Mb
Number of scaffolds	210	460
Scaffold N50	22.1 Mb	20.74 Mb
Longest scaffold length (Mb)	87.64	26.29
Sex chromosomes	W; Z1; and Z2	-
Organelles	Mitochondrion: 15.34 kb	-

Most of the assembly sequence (98.55%) was assigned to 31 chromosomal-level scaffolds, representing 28 autosomes and the W, Z_1, and Z_2 sex chromosomes. These chromosome-level scaffolds, confirmed by Hi-C data, are named according to size (Figure 3; Table 3). Chromosomes Z_1, Z_2 and W were identified by copy number in the diploid genome and Hi-C signal. The order and orientation of the scaffolds making up Chromosome W is uncertain. Chromosome Z_1 was identified as the canonical Z by BUSCO gene painting with ancestral Merian elements (Wright *et al.*, 2024). The W chromosome was identified based on read coverage analysis and its single-copy status within a merged diploid map. The exact order and orientation of the contigs on chromosome W are unknown.

The mitochondrial genome was also assembled (length 15.34 kb, OZ207822.1). This sequence is included as a contig in the multifasta file of the genome submission and as a standalone record.

For haplotype 1, the estimated QV is 62.3, and for haplotype 2, 62.7. When the two haplotypes are combined, the assembly achieves an estimated QV of 62.5. The *k*-mer completeness is 85.81% for haplotype 1, 73.20% for haplotype 2, and 99.45% for the combined haplotypes (Figure 4).

BUSCO analysis using the lepidoptera_odb10 reference set (n = 5 286) identified 98.6% of the expected gene set (single = 97.8%, duplicated = 0.8%) for haplotype 1. The snail plot in Figure 5 summarises the scaffold length distribution and other assembly statistics for haplotype 1. The blob plot in Figure 6 shows the distribution of scaffolds by GC proportion and coverage for haplotype 1.

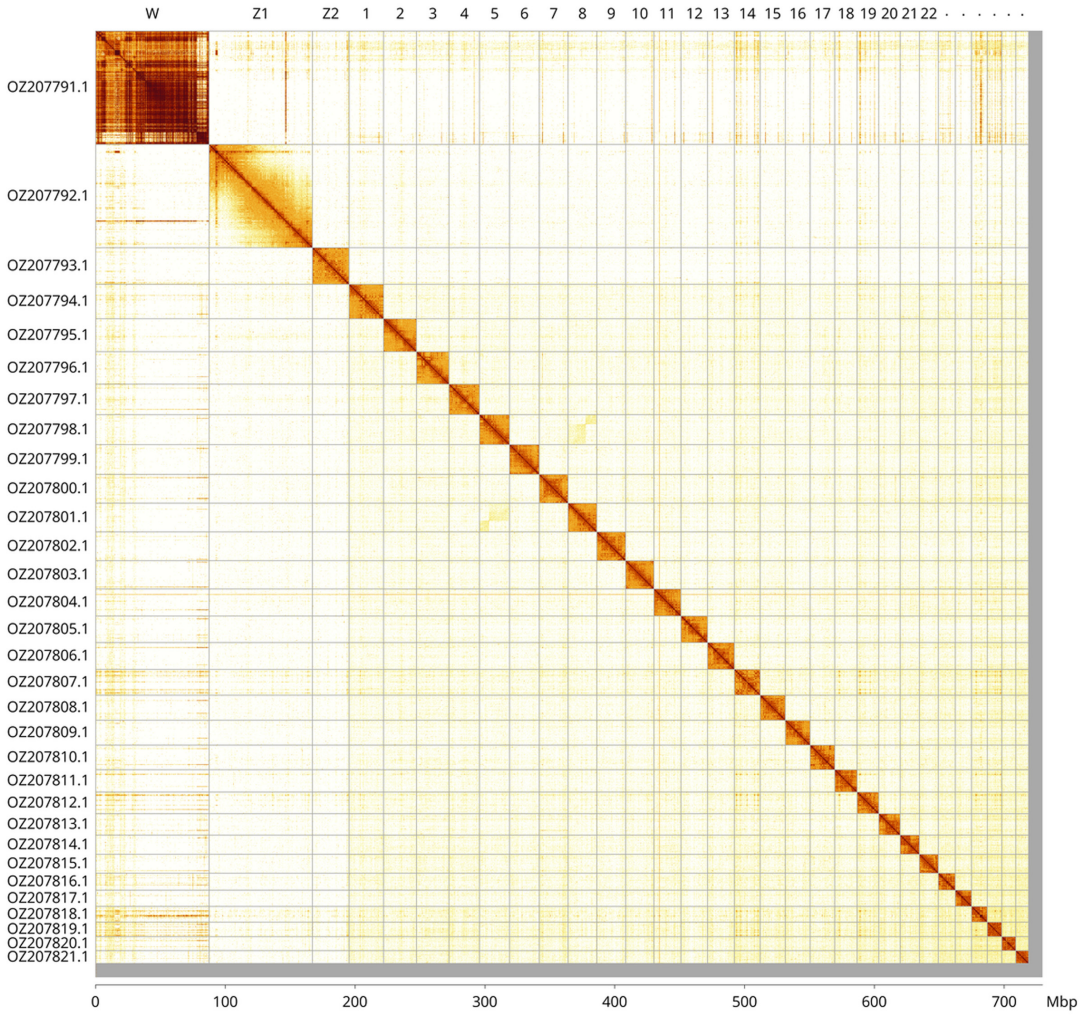

Figure 3. Hi-C contact map of the *Eudonia murana* genome assembly. Assembled chromosomes are shown in order of size and labelled along the axes, with a megabase scale shown below. The plot was generated using PretextSnapshot.

Table 4 lists the assembly metric benchmarks adapted from Rhie *et al.* (2021) and the Earth BioGenome Project Report on Assembly Standards September 2024. The EBP metric, calculated for the haplotype 1, is **6.C.Q62**, meeting the recommended reference standard.

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.

Table 3. Chromosomal pseudomolecules in both haplotypes of the genome assembly of *Eudonia murana*, ilEudMura1.

Haplotype 1				Haplotype 2			
INSDC accession	**Name**	**Length (Mb)**	**GC%**	**INSDC accession**	**Name**	**Length (Mb)**	**GC%**
OZ207794.1	1	26.45	37.50	OZ207763.1	1	26.29	37.50
OZ207795.1	2	25.48	37.50	OZ207764.1	2	25.64	37.50
OZ207796.1	3	25.07	38	OZ207765.1	3	25.31	37.50
OZ207797.1	4	23.42	37.50	OZ207766.1	4	23.25	37.50
OZ207798.1	5	23.16	37.50	OZ207767.1	5	22.82	37
OZ207799.1	6	22.85	37.50	OZ207768.1	6	22.75	37.50
OZ207800.1	7	22.28	37.50	OZ207769.1	7	22.24	37.50
OZ207801.1	8	22.10	37.50	OZ207770.1	8	22.09	37.50
OZ207802.1	9	22.09	37.50	OZ207771.1	9	21.97	37.50
OZ207803.1	10	21.83	37.50	OZ207772.1	10	21.91	37.50
OZ207804.1	11	20.87	38	OZ207773.1	11	20.96	37.50
OZ207805.1	12	20.55	37.50	OZ207774.1	12	20.74	37.50
OZ207806.1	13	20.51	37.50	OZ207775.1	13	20.61	37.50
OZ207807.1	14	19.90	38	OZ207776.1	14	20.04	38
OZ207808.1	15	19.40	38	OZ207777.1	15	19.43	37.50
OZ207809.1	16	19.16	38	OZ207778.1	16	19.19	38
OZ207810.1	17	18.95	38	OZ207779.1	17	18.72	38
OZ207811.1	18	17.28	38	OZ207780.1	18	17.40	37.50
OZ207812.1	19	16.59	38	OZ207781.1	19	16.85	38.50
OZ207813.1	20	16.57	38	OZ207782.1	20	16.43	38.50
OZ207814.1	21	14.70	38	OZ207783.1	21	14.77	38
OZ207815.1	22	14.61	38	OZ207784.1	22	14.68	38
OZ207816.1	23	13.01	38	OZ207785.1	23	13.83	38.50
OZ207817.1	24	12.64	38.50	OZ207786.1	24	13.11	38.50
OZ207818.1	25	12.14	40	OZ207787.1	25	12.23	40
OZ207819.1	26	11.06	38.50	OZ207788.1	26	10.98	39
OZ207820.1	27	10.82	38.50	OZ207789.1	27	10.76	38
OZ207821.1	28	9.98	38.50	OZ207790.1	28	10.02	38.50
OZ207791.1	W	87.64	35.50				
OZ207792.1	Z1	79.71	37.50				
OZ207793.1	Z2	28.16	38				

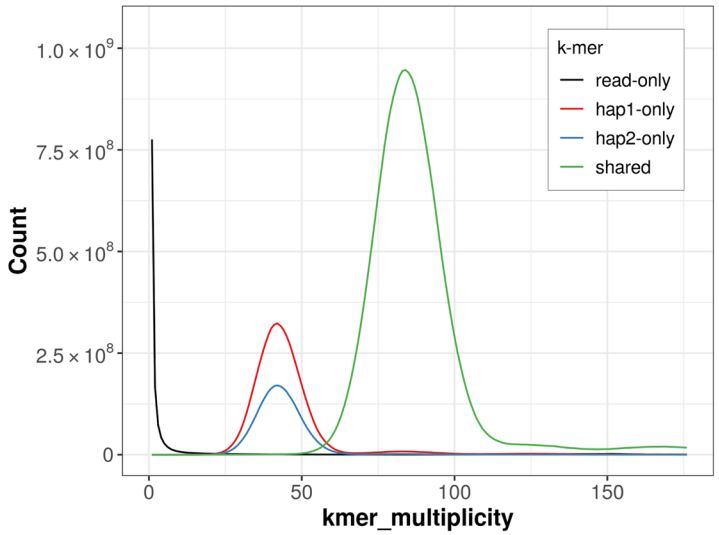

Figure 4. Evaluation of *k*-mer completeness using MerquryFK. This plot illustrates the recovery of *k*-mers from the original read data in the final assemblies. The horizontal axis represents *k*-mer multiplicity, and the vertical axis shows the number of *k*-mers. The black curve represents *k*-mers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

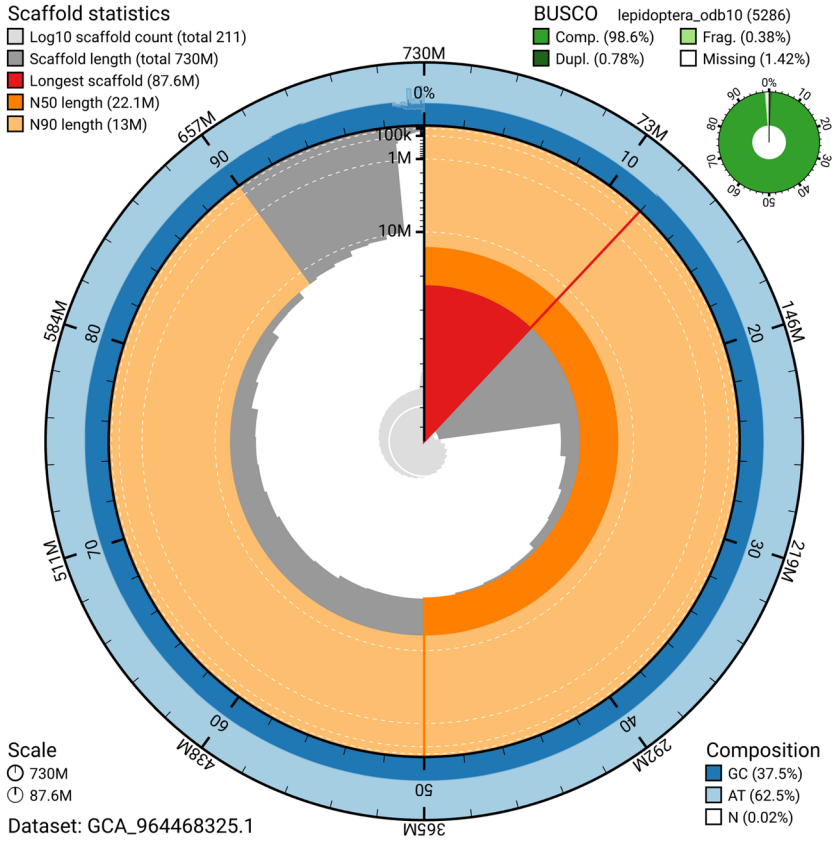

Figure 5. Assembly metrics for ilEudMura1.hap1.1. The BlobToolKit snail plot provides an overview of assembly metrics and BUSCO gene completeness. The circumference represents the length of the whole genome sequence, and the main plot is divided into 1 000 bins around the circumference. The outermost 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.

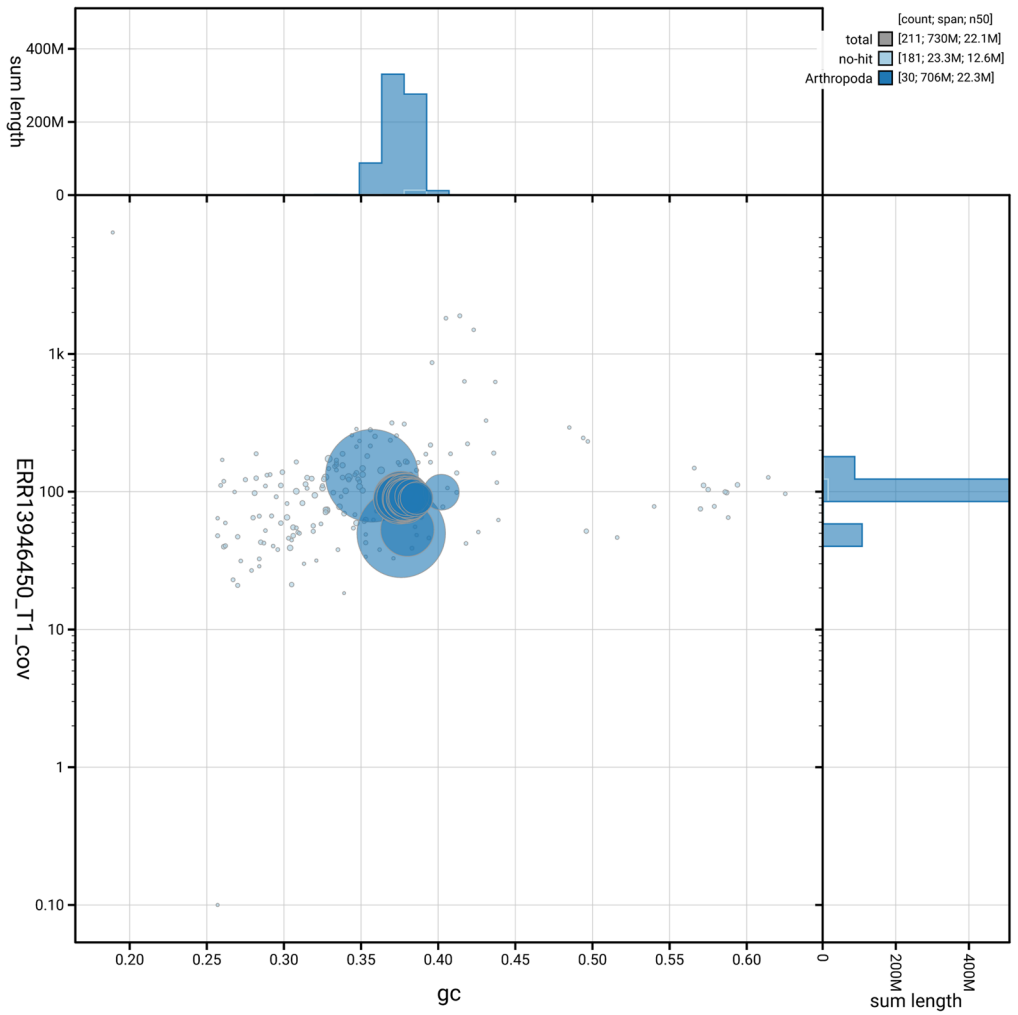

Figure 6. BlobToolKit GC-coverage plot for ilEudMura1.hap1.1. Blob plot showing sequence coverage (vertical axis) and GC content (horizontal axis). The circles represent scaffolds, with the size proportional to scaffold length and the colour representing phylum membership. The histograms along the axes 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 *Eudonia murana* assembly.

Measure	Value	Benchmark
EBP summary (haplotype 1)	6.C.Q62	6.C.Q40
Contig N50 length	3.92 Mb	≥ 1 Mb
Scaffold N50 length	22.10 Mb	= chromosome N50
Consensus quality (QV)	Haplotype 1: 62.3; haplotype 2: 62.7; combined: 62.5	≥ 40
k-mer completeness	Haplotype 1: 85.81%; Haplotype 2: 73.20%; combined: 99.45%	≥ 95%
BUSCO	C:98.6% [S:97.8%; D:0.8%]; F:0.4%; M:1.0%; n:5 286	S > 90%; D < 5%
Percentage of assembly assigned to chromosomes	98.55%	≥ 90%

Data availability

European Nucleotide Archive: Eudonia murana. Accession number PRJEB82145. The genome sequence is released openly for reuse. The *Eudonia murana* 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. 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.

Table 5. Software versions and sources.

Software	Version	Source
BEDTools	2.30.0	https://github.com/arq5x/bedtools2
BLAST	2.14.0	ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/
BlobToolKit	4.3.9	https://github.com/blobtoolkit/blobtoolkit
BUSCO	5.5.0	https://gitlab.com/ezlab/busco
bwa-mem2	2.2.1	https://github.com/bwa-mem2/bwa-mem2
Cooler	0.8.11	https://github.com/open2c/cooler
DIAMOND	2.1.8	https://github.com/bbuchfink/diamond
fasta_windows	0.2.4	https://github.com/tolkit/fasta_windows
FastK	1.1	https://github.com/thegenemyers/FASTK
GenomeScope2.0	2.0.1	https://github.com/tbenavi1/genomescope2.0
Gfastats	1.3.6	https://github.com/vgl-hub/gfastats
GoaT CLI	0.2.5	https://github.com/genomehubs/goat-cli
Hifiasm	0.19.8-r603	https://github.com/chhylp123/hifiasm
HiGlass	1.13.4	https://github.com/higlass/higlass
MerquryFK	1.1.2	https://github.com/thegenemyers/MERQURY.FK
Minimap2	2.24-r1122	https://github.com/lh3/minimap2
MitoHiFi	3	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	0.0.5	https://github.com/sanger-tol/PretextSnapshot
PretextView	0.2.5	https://github.com/sanger-tol/PretextView
samtools	1.19.2	https://github.com/samtools/samtools
sanger-tol/ascc	0.1.0	https://github.com/sanger-tol/ascc
sanger-tol/blobtoolkit	0.6.0	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

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

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

Cheng H, Jarvis ED, Fedrigo O, *et al.*: **Haplotype-resolved assembly of diploid genomes without parental data.** *Nat Biotechnol.* 2022; **40**(9): 1332–1335.
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

Darwin Tree of Life Project Consortium: **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

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

GBIF Secretariat: ***Eudonia murana* (Curtis, 1827).** In: GBIF Backbone Taxonomy. 2025.
Reference Source

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

Kimber I: ***Eudonia murana* (Moorland Grey).** 2025.
Reference Source

Kurtzer GM, Sochat V, Bauer MW: **Singularity: scientific containers for mobility of compute.** *PLoS One.* 2017; **12**(5): e0177459.
PubMed Abstract | Publisher Full Text | Free Full Text

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

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

Manni M, Berkeley MR, Seppey M, *et al.*: **BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes.** *Mol Biol Evol.* 2021; **38**(10): 4647–4654.
PubMed Abstract | Publisher Full Text | Free Full Text

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

Ranallo-Benavidez TR, Jaron KS, Schatz MC: **GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes.** *Nat Commun.* 2020; **11**(1): 1432.
PubMed Abstract | Publisher Full Text | Free Full Text

Rao SSP, Huntley MH, Durand NC, *et al.*: **A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping.** *Cell.* 2014; **159**(7): 1665–1680.
PubMed Abstract | Publisher Full Text | Free Full Text

Rhie A, McCarthy SA, Fedrigo O, *et al.*: **Towards complete and error-free genome assemblies of all vertebrate species.** *Nature.* 2021; **592**(7856): 737–746.
PubMed Abstract | Publisher Full Text | Free Full Text

Rhie A, Walenz BP, Koren S, *et al.*: **Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies.** *Genome Biol.* 2020; **21**(1): 245.
PubMed Abstract | Publisher Full Text | Free Full Text

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

Sterling P, Parsons M, Lewington R: **Field guide to the Micro-Moths of Great Britain and Ireland.** London: Bloomsbury Publishing, 2023.
Reference Source

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

Wright CJ, Stevens L, Mackintosh A, *et al.*: **Comparative genomics reveals the dynamics of chromosome evolution in Lepidoptera.** *Nat Ecol Evol.* 2024; **8**(4): 777–790.
PubMed Abstract | Publisher Full Text | Free Full Text

Zhou C, McCarthy SA, Durbin R: **YaHS: Yet another Hi-C Scaffolding tool.** *Bioinformatics.* 2023; **39**(1): btac808.
PubMed Abstract | Publisher Full Text | Free Full Text

Open Peer Review

Current Peer Review Status: ✔ ✔

Version 1

Reviewer Report 26 February 2026

https://doi.org/10.21956/wellcomeopenres.27783.r148701

✔ **Danilo Trabudo do Amaral**
Universidade Federal do ABC Centro de Ciencias Naturais e Humanas (Ringgold ID: 425753), Santo André, State of São Paulo, Brazil

This study presents a high-quality, chromosome-level genome assembly of *Eudonia murana*, generated using state-of-the-art long-read and Hi-C technologies. The work represents an important contribution to the Darwin Tree of Life initiative by providing a reliable genomic resource for future ecological, evolutionary, and comparative studies. The methodology is robust, and the data are well documented and openly accessible.
As minor suggestions, the authors could strengthen the manuscript by including a brief comparison with closely related species, adding a short discussion on the evolutionary implications of the Z1/Z2/W sex chromosome system, and clarifying the differences in completeness between the two haplotypes.

Is the rationale for creating the dataset(s) clearly described?
Yes

Are the protocols appropriate and is the work technically sound?
Yes

Are sufficient details of methods and materials provided to allow replication by others?
Yes

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

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Bioinformatics

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 26 February 2026

https://doi.org/10.21956/wellcomeopenres.27783.r146697

Marko Mutanen
University of Oulu, Oulu, Finland

This study presents the genome sequence for a crambid moth Eudonia murana. This species is a member of a taxonomically difficult group of species, but as far as is known, DNA barcode provides a safe way to distinguish this species from all of its close relatives. It is therefore good that the identification was confirmed by barcoding. Interestingly, the samples specimens undoubtedly represent a second-generation specimen; a rare thing in this group of species but known for E. murana. Unlike stated in the introduction, this species is readily separate from Scoparia spp. in genitalia (dissection). It is very difficult to tell apart from E. truncicolella by genital characteristics.
The genome sequence is clearly of high quality and will serve as an excellent reference genome for the species. The snail plot figure, for example, demonstrates this high quality. For example, the BUSCO recovery rate is high. I have not reservations about the quality of the genome sequence.

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: Lepidoptera diversity, molecular taxonomy, sawfly diversity, DNA barcoding, phylogenetics

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.