DATA NOTE

The genome sequence of the Red-necked Footman, *Atolmis rubricollis* (Linnaeus, 1758) (Lepidoptera: Erebidae)

[version 1; peer review: 3 approved, 1 approved with reservations]

Gavin R. Broad[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

v1 **First published:** 14 Apr 2026, **11**:219
https://doi.org/10.12688/wellcomeopenres.26265.1
Latest published: 14 Apr 2026, **11**:219
https://doi.org/10.12688/wellcomeopenres.26265.1

Abstract

We present a genome assembly from an individual female *Atolmis rubricollis* (Red-necked Footman; Arthropoda; Insecta; Lepidoptera; Erebidae). The assembly contains two haplotypes with total lengths of 647.88 megabases and 585.85 megabases. Most of haplotype 1 (99.59%) is scaffolded into 32 chromosomal pseudomolecules, including the W and Z sex chromosomes. Haplotype 2 was assembled to scaffold level. The mitochondrial genome has also been assembled, with a length of 15.43 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

Atolmis rubricollis, Red-necked Footman, genome sequence, chromosomal, Lepidoptera

This article is included in the Tree of Life gateway.

Open Peer Review

Approval Status ✓ ✓ ? ✓

	1	2	3	4
version 1 14 Apr 2026	✓ view	✓ view	? view	✓ view

1. **Oliver Rupp**, Justus-Liebig-University, Giessengiessen, Germany
2. **Stuart J.E. Baird**, Academy of Sciences of the Czech Republic, Národní, Czech Republic
3. **Annabel Whibley**, Bragato Research Institute, Lincoln, New Zealand
The University of Auckland School of Biological Sciences (Ringgold ID: 99026), Auckland, New Zealand
4. **Ricardo Rodriguez-de-la-vega**, Université Paris-Saclay, Gif-sur-Yvette, France

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

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

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

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by Wellcome through core funding to the Wellcome Sanger Institute (220540) and the Darwin Tree of Life Discretionary Award [218328, 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: Broad GR, Natural History Museum Genome Acquisition Lab, Darwin Tree of Life Barcoding Collective *et al.* **The genome sequence of the Red-necked Footman, *Atolmis rubricollis* (Linnaeus, 1758) (Lepidoptera: Erebidae) [version 1; peer review: 3 approved, 1 approved with reservations]** Wellcome Open Research 2026, **11**:219
https://doi.org/10.12688/wellcomeopenres.26265.1

First published: 14 Apr 2026, **11**:219 https://doi.org/10.12688/wellcomeopenres.26265.1

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; Noctuoidea; Erebidae; Arctiinae; Lithosiini; *Atolmis*; *Atolmis rubricollis* (Linnaeus, 1758) (NCBI:txid987889).

Background

Atolmis rubricollis, the Red-necked Footman, is a distinctive moth: the wings are a dark, sooty grey, usually tightly wrapped around the body, there is a red 'collar' behind the head, and the abdomen is bright yellow on the underside and the tip of the upperside. The overall body shape, of long, wrapped, rather narrow wings, resembling the traditional coats of footmen, is characteristic of many Lithosiini, a tribe of lichen-feeding moths in the family Erebidae, subfamily Arctiinae.

Larvae of *A. rubricollis* feed on lichens and algae on tree trunks and branches, from August to October in Britain (Henwood *et al.*, 2020). The larva is black, overlaid with many wavy greenish-yellow lines, furnished with hairy protrusions (verrucae). After over-wintering as a pupa, adults emerge in June and July. The adults are usually light-trapped in small numbers, but their habit of flying around the tops of trees in the daytime was noted long ago by Harris (1766).

Previously known in Britain as a southern species (South, 1907), since 1990 the distribution of *A. rubricollis* has increased by 66% and it is now widespread, except in northern Scotland (Randle *et al.*, 2019). In much of Ireland and Scotland, *A. rubricollis* breeds in conifer plantations (Waring *et al.*, 2017). Numbers are boosted by immigration from continental Europe; the sequenced individual was light-trapped on a night with various migrant moths and was also presumed to be an immigrant. *Atolmis rubricollis* is found across the Palaearctic (GBIF Secretariat, 2026).

We present a chromosome-level genome sequence for *Atolmis rubricollis*, generated using the Tree of Life pipeline from a specimen collected in Tonbridge, Kent, UK (Figure 1). As this is the only species of the genus *Atolmis*, this genome adds to a growing genomic resource for understanding the evolution of footman and tiger moths.

Methods

Sample acquisition and DNA barcoding

The specimen used for genome sequencing was an adult female *Atolmis rubricollis* (specimen ID NHMUK010634997, ToLID ilAtoRubr1; Figure 1), collected from Tonbridge, Kent, UK (latitude 51.1863, longitude 0.2865) on 2020-06-24. The specimen was collected and identified by Gavin Broad.

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 ilAtoRubr1 sample was weighed and triaged to determine the appropriate extraction protocol. Tissue from the thorax 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.

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

Figure 1. Photographs of the *Atolmis rubricollis* (ilAtoRubr1) specimen used for genome sequencing.

assay kit. Final library fragment size was assessed with the Agilent Femto Pulse Automated Pulsed Field CE Instrument (Agilent Technologies) using the gDNA 55 kb BAC analysis kit.

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

Hi-C

Sample preparation and crosslinking

The Hi-C sample was prepared from 20–50 mg of frozen tissue from the head of the ilAtoRubr1 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 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 in Hi-C phasing mode (Cheng *et al.*, 2021, 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

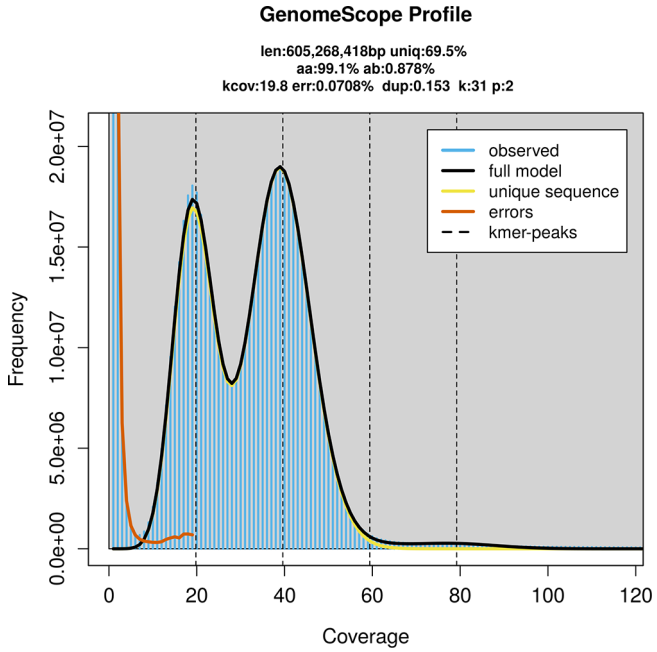

Figure 2. Frequency distribution of *k*-mers generated using GenomeScope2. The plot shows observed and modelled *k*-mer spectra, providing estimates of genome size, heterozygosity, and repeat content based on unassembled sequencing reads.

Table 1. Specimen and sequencing data for BioProject PRJEB83465.

Platform	PacBio HiFi	Hi-C
ToLID	ilAtoRubr1	ilAtoRubr1
Specimen ID	NHMUK010634997	NHMUK010634997
BioSample (source individual)	SAMEA7521515	SAMEA7521515
BioSample (tissue)	SAMEA7521584	SAMEA7521582
Tissue	abdomen	head
Instrument	Revio	Illumina NovaSeq 6000
Run accessions	ERR14104842; ERR14104841	ERR14075518
Read count total	2.79 million	381.88 million
Base count total	24.59 Gb	115.33 Gb

Table 2. **Genome assembly statistics.**

Genome assembly	Haplotype 1	Haplotype 2
Assembly name	ilAtoRubr1.hap1.1	ilAtoRubr1.hap2.1
Assembly accession	GCA_965117225.1	GCA_965117165.1
Assembly level	chromosome	scaffold
Span (Mb)	647.88	585.85
Number of chromosomes	32	-
Number of contigs	229	169
Contig N50	11.61 Mb	12.36 Mb
Number of scaffolds	82	108
Scaffold N50	21.15 Mb	21.04 Mb
Longest scaffold length (Mb)	39.12	-
Sex chromosomes	W and Z	-
Organelles	Mitochondrion: 15.43 kb	-

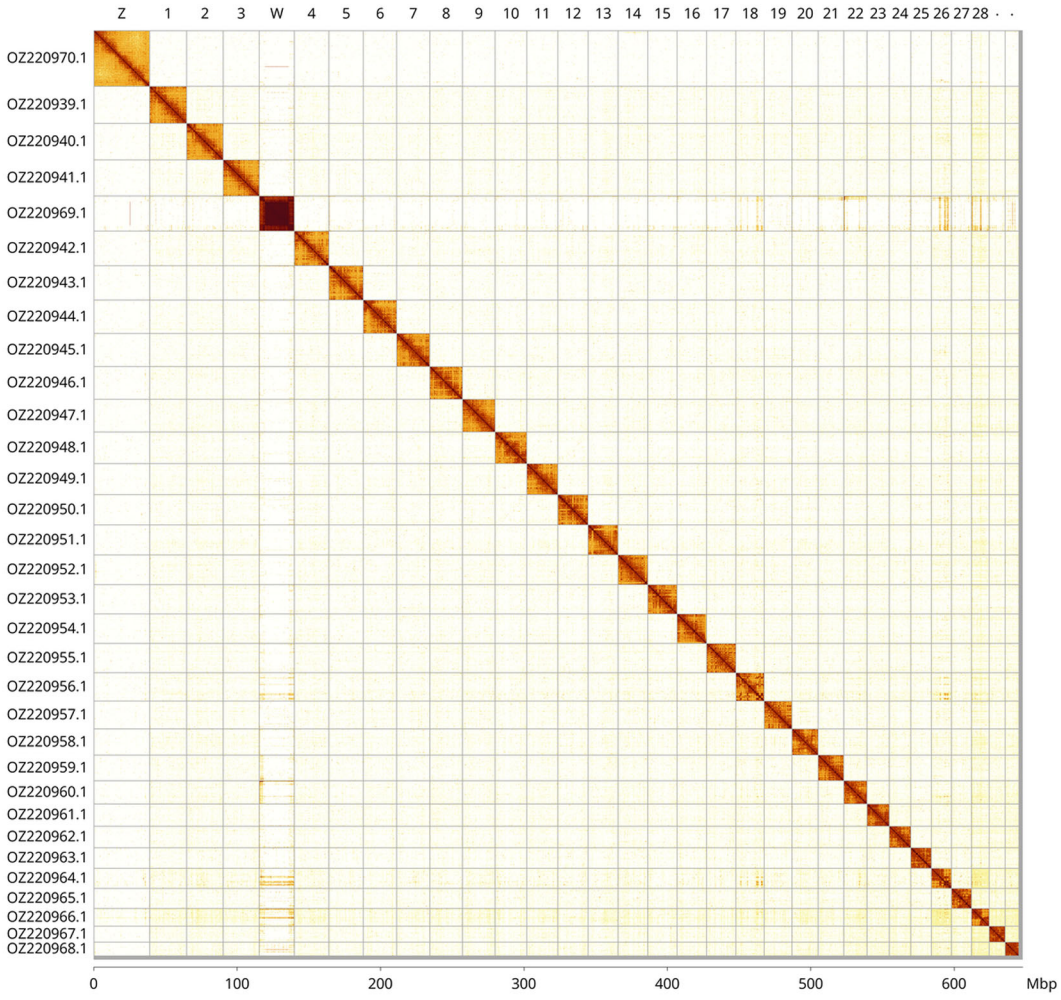

Figure 3. **Hi-C contact map of the *Atolmis rubricollis* genome assembly.** Assembled chromosomes are shown in order of size and labelled along the axes, with a megabase scale shown below. The plot was generated using PretextSnapshot.

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

Table 3. Chromosomal pseudomolecules in the haplotype 1 genome assembly of *Atolmis rubricollis* ilAtoRubr1.

INSDC accession	Molecule	Length (Mb)	GC%
OZ220939.1	1	25.85	38
OZ220940.1	2	25.32	38
OZ220941.1	3	25.26	38
OZ220942.1	4	24.22	38
OZ220943.1	5	23.89	38
OZ220944.1	6	23.39	38
OZ220945.1	7	23.04	38
OZ220946.1	8	22.81	38
OZ220947.1	9	22.74	38
OZ220948.1	10	22.04	37.50
OZ220949.1	11	21.60	38
OZ220950.1	12	21.15	38
OZ220951.1	13	20.96	38
OZ220952.1	14	20.65	38.50
OZ220953.1	15	20.53	38
OZ220954.1	16	20.46	38.50
OZ220955.1	17	20.46	38
OZ220956.1	18	19.72	38.50
OZ220957.1	19	19.46	38.50
OZ220958.1	20	18.16	39
OZ220959.1	21	18	38.50
OZ220960.1	22	16.13	39
OZ220961.1	23	15.46	39
OZ220962.1	24	15.03	39
OZ220963.1	25	14.37	39
OZ220964.1	26	14	40.50
OZ220965.1	27	13.96	39
OZ220966.1	28	12.42	40.50
OZ220967.1	29	11.13	40
OZ220968.1	30	9.53	40.50
OZ220969.1	W	24.37	41
OZ220970.1	Z	39.12	38

and HiGlass (Kerpedjiev *et al.*, 2018). Scaffolds were visually inspected and corrected as described by Howe *et al.* (2021). Manual corrections included 9 breaks and 131 joins. This reduced the scaffold count by 41.1% and increased the total assembly length by 1.6%. 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 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 *Atolmis rubricollis* specimen generated 24.59 Gb (gigabases) from 2.79 million reads, which were used to assemble the genome. GenomeScope2.0 analysis estimated the haploid genome size at 605.27 Mb, with a heterozygosity of 0.88% and repeat content of 30.58% (Figure 2). These estimates guided expectations for the assembly. Based on the estimated genome size, the sequencing data provided approximately 40× coverage. Hi-C sequencing produced 115.28 Gb from 763.47 million reads, which were used to scaffold the assembly. Table 1 summarises the specimen and sequencing details.

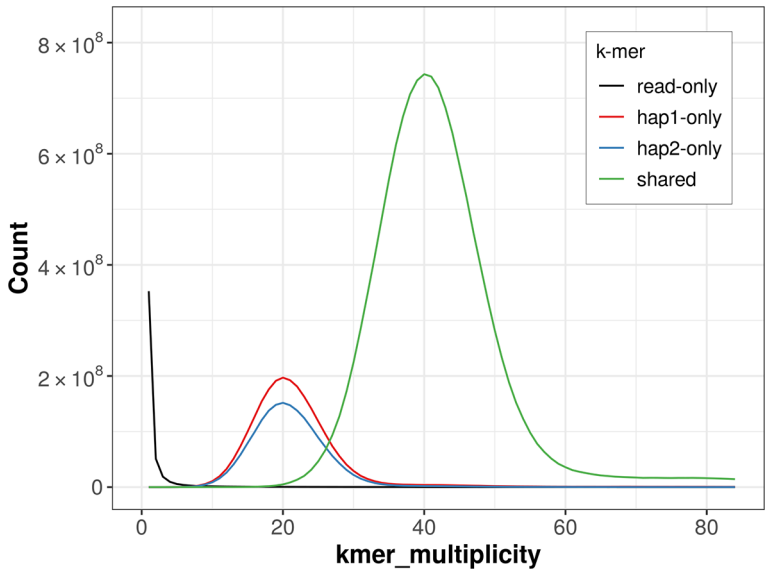

Figure 4. Evaluation of *k*-mer completeness using MerquryFK. This plot illustrates the recovery of *k*-mers from the original read data in the final assemblies. The horizontal axis represents *k*-mer multiplicity, and the vertical axis shows the number of *k*-mers. The black curve represents *k*-mers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

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 647.88 Mb in 82 scaffolds, with 147 gaps, and a scaffold N50 of 21.15 Mb (Table 2).

Most of the assembly sequence (99.59%) was assigned to 32 chromosomal-level scaffolds, representing 30 autosomes and the W and Z sex chromosomes. These chromosome-level scaffolds, confirmed by Hi-C data, are named according to size (Figure 3; Table 3). The exact order and orientation of the contigs on chromosome W are unknown. The Z and W chromosomes were identified based on read coverage and Hi-C signal.

The mitochondrial genome was also assembled (length 15.43 kb, OZ220971.1). This sequence is included as a contig in the multifasta file of the genome submission and as a standalone record.

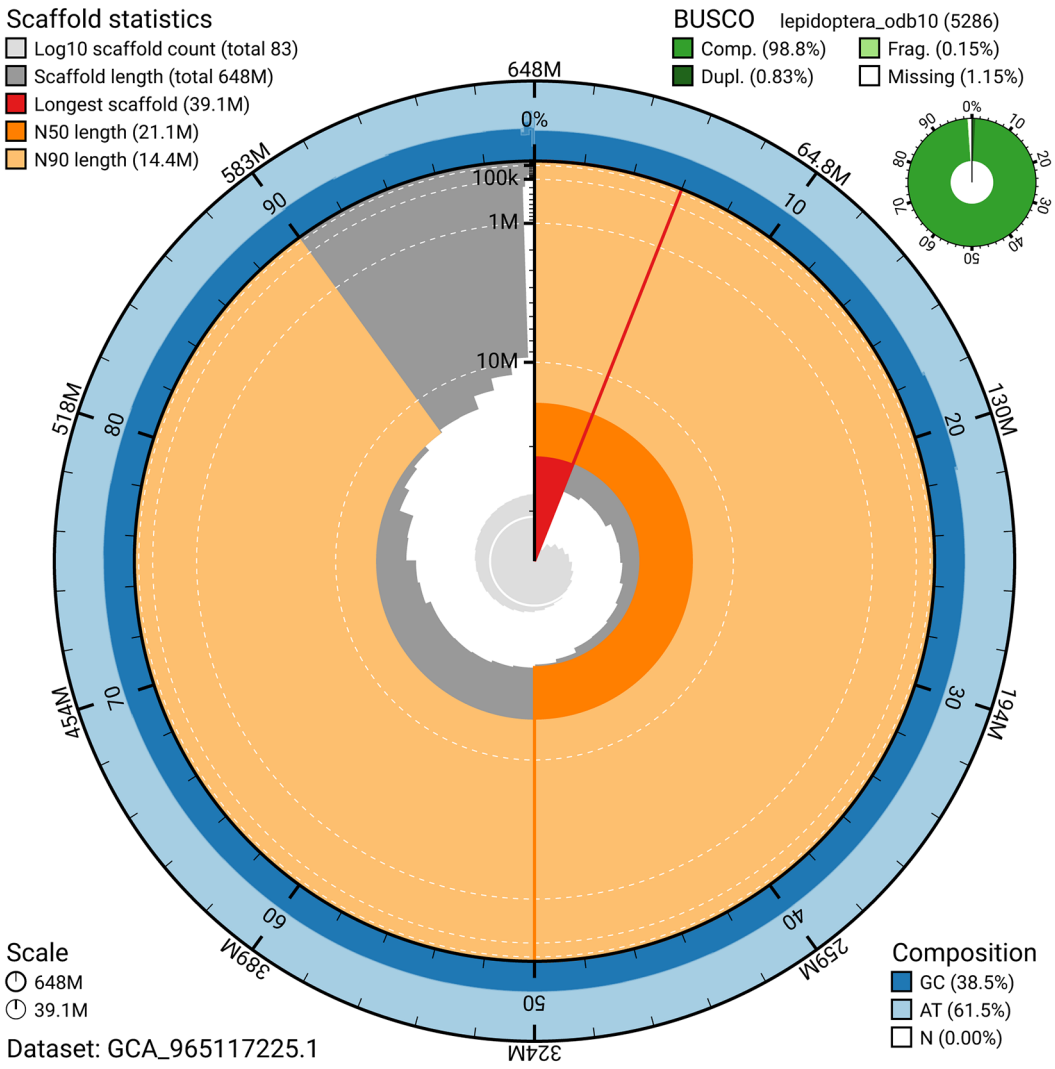

Figure 5. Assembly metrics for ilAtoRubr1.hap1.1. The BlobToolKit snail plot provides an overview of assembly metrics and BUSCO gene completeness. The circumference represents the length of the whole genome sequence, and the main plot is divided into 1 000 bins around the circumference. The outermost 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.

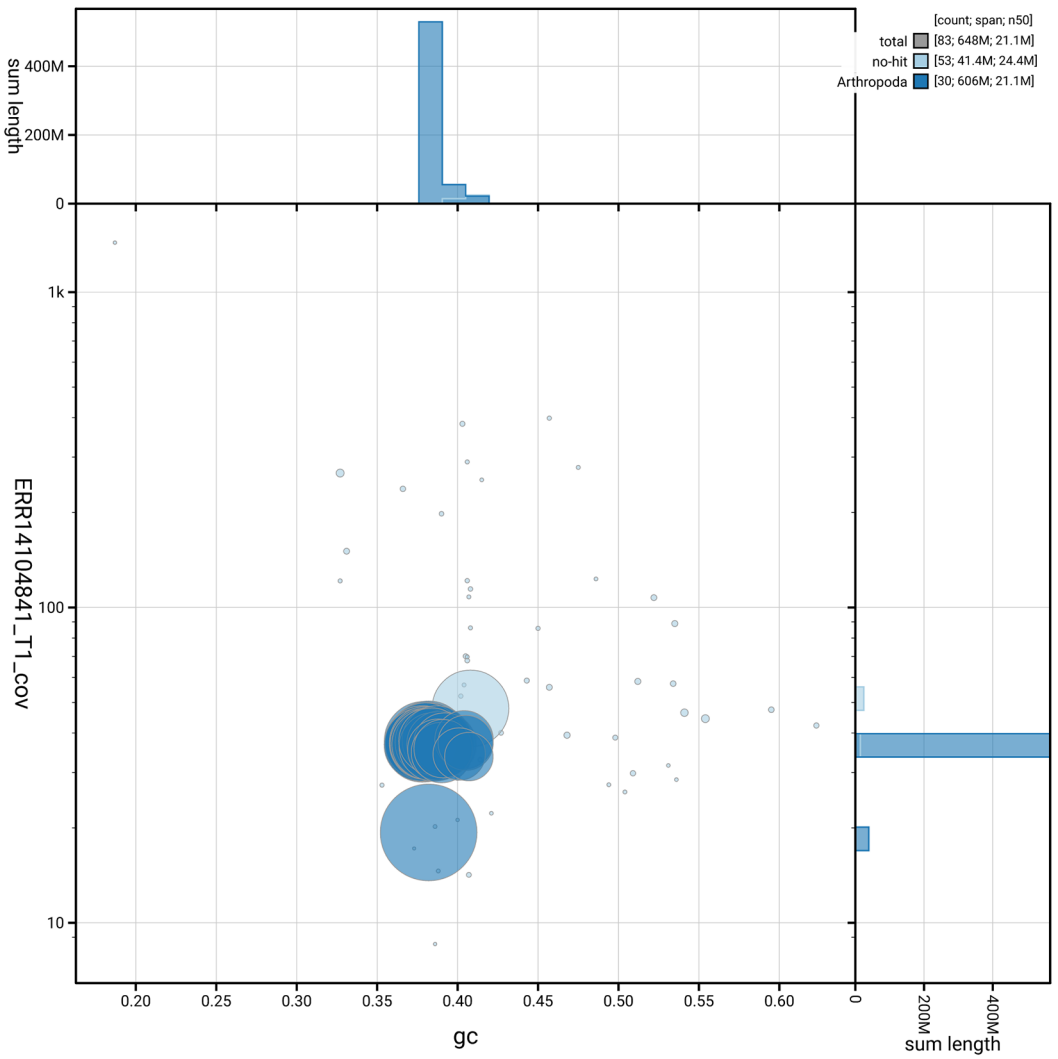

Figure 6. BlobToolKit GC-coverage plot for ilAtoRubr1.hap1.1. Blob plot showing sequence coverage (vertical axis) and GC content (horizontal axis). The circles represent scaffolds, with the size proportional to scaffold length and the colour representing phylum membership. The histograms along the axes 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.

For haplotype 1, the estimated QV is 66.1, and for haplotype 2, 65.9. When the two haplotypes are combined, the assembly achieves an estimated QV of 66.0. The *k*-mer completeness is 83.24% for haplotype 1, 78.26% for haplotype 2, and 99.74% for the combined haplotypes (Figure 4).

BUSCO analysis using the lepidoptera_odb10 reference set (n = 5 286) identified 98.8% of the expected gene set (single = 98.0%, 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.

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 **7.C.Q66**, meeting the recommended reference standard.

Table 4. **Earth Biogenome Project summary metrics for the *Atolmis rubricollis* assembly.**

Measure	Value	Benchmark
EBP summary (haplotype 1)	7.C.Q66	6.C.Q40
Contig N50 length	11.61 Mb	≥ 1 Mb
Scaffold N50 length	21.15 Mb	= chromosome N50
Consensus quality (QV)	Haplotype 1: 66.1; haplotype 2: 65.9; combined: 66.0	≥ 40
k-mer completeness	Haplotype 1: 83.24%; Haplotype 2: 78.26%; combined: 99.74%	≥ 95%
BUSCO	C:98.8% [S:98.0%, D:0.8%], F:0.2%, M:1.0%, n:5 286	S > 90%; D < 5%
Percentage of assembly assigned to chromosomes	99.59%	≥ 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.**

Software	Version	Source
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
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	0.19.8-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	1.0.3	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

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: Atolmis rubricollis (red-necked footman). Accession number PRJEB83465. The genome sequence is released openly for reuse. The *Atolmis rubricollis* 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 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.

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 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

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.** *Wellcome Open Research.* 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).
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: ***Atolmis rubricollis* (Linnaeus, 1758) in GBIF Backbone Taxonomy.** 2026.
Reference Source

Harris M: *The Aurelian, a Natural History of British Moths and Butterflies.* London: Moses Harris; 1766.

Henwood B, Sterling P, Lewington R: *Field Guide to the Caterpillars of Great Britain and Ireland.* London: Bloomsbury Wildlife; 2020.

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).
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

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).
Publisher 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

Randle Z, Evans-Hill LJ, Parsons MS, *et al.*: *Atlas of Britain & Ireland's Larger Moths.* Newbury: Pisces Publications; 2019.

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).
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.* 2020; **2020**: baaa062.
PubMed Abstract | Publisher Full Text | Free Full Text

South R: *The Moths of the British Isles.* London: Frederick Warne & Co; 1907.
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 Research.* 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

Waring P, Townsend M, Lewington R: *Field Guide to the Moths of Great Britain and Ireland.* London: Bloomsbury Publishing; 2017.

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

Open Peer Review

Current Peer Review Status: ✔ ✔ ? ✔

Version 1

Reviewer Report 22 May 2026

https://doi.org/10.21956/wellcomeopenres.28927.r153888

✔ **Ricardo Rodriguez-de-la-vega**
Université Paris-Saclay, Gif-sur-Yvette, France

Review Report on Broad et al., "The genome sequence of the Red-necked Footman..."

In this Darwin Tree of Life (DToL) project genome note, the authors describe the chromosome-level (EBP metric 7.C.Q66*) genome assembly of an *Atolmis rubricollis* specimen. This is the first *Atolmis* genome released by DToL, adding to the now extensive collection of Erebidae chromosome-level genome assemblies produced by DToL or the Psyche project (over 40 as of May 2025).
The reported assemblies are among the highest-scoring DToL genome notes, with near-perfect chromosome anchoring of scaffolds, as well as high BUSCO and k-mer completeness. The assembly combined PacBio and HiC reads obtained from the same specimen and followed the standard Wellcome Sanger Institute protocols for DNA extraction, library preparation, primary assembly, scaffolding, decontamination, and quality assessment—all of which are properly referenced and linked in the genome note.

While the presentation adheres to the highly standardized format of DToL genome notes, it is particularly well-written, and the linking and referencing of experimental and assembly protocols are especially commendable.

I would only highlight that the "no(BUSCO)hit" scaffold corresponds to the W chromosome, whose contiguity was not resolved. This is a common feature of Lepidoptera genomes released by the DToL.

I am including links to the Genomes on a Tree and Tree of Life QC pages, which I believe should be part of all DToL notes:

https://tolqc.cog.sanger.ac.uk/darwin/insects/Atolmis_rubricollis
https://goat.genomehubs.org/record?recordId=987889&result=taxon&taxonomy=ncbi

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, Bioinformatics, Evolution

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

Reviewer Report 14 May 2026

https://doi.org/10.21956/wellcomeopenres.28927.r153385

?

Annabel Whibley
[1] Grapevine Improvement, Bragato Research Institute, Lincoln, New Zealand
[2] The University of Auckland School of Biological Sciences (Ringgold ID: 99026), Auckland, Auckland, New Zealand

In this Data Note, Broad and colleagues present a reference genome assembly of the moth the Red-necked Footman, *Atolmis rubricollis*. The appearance, natural history and phylogenetic significance of the species (as the only representative of the genus) is well presented. Sample identification was verified by COI barcode analysis prior to sequencing. HiFi data was generated (but there is a conflict between exactly which platform was used) and Arima HiC data included for scaffolding and assembly curation, as is standard in DToL workflows.

The assembly quality stats are superb, the chromosomes appear confidently resolved in the HiC contact map and accuracy and completeness statistics are excellent for both haplotypes. The reporting follows DToL templating standards, is clear and comprehensive with public links active.

I will also include a comment I made in an earlier review which relates to the decision to present the reference as a “haplotype assembly” but to include both sex chromosomes with haplotype 1. While this is completely parallel to the primary/alt assembly mode and is also the most useful type of reference to generate for population work, the terminology doesn’t sit entirely comfortably with

me. I think there might be value in a different term or qualifier for a haplotype assembly that is augmented by the inclusion of both sex chromosomes in a single haplotype.

To address:
Please clarify whether Sequel IIe (as per Methods text) or Revio (as per Table 1) was used for the HiFi sequencing.

I think this is a templated decision, but I would prefer to see some text report of the haplotype 2 summary stats as well, in addition to the Table presentation.

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, 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, however I have significant reservations, as outlined above.

Reviewer Report 13 May 2026

https://doi.org/10.21956/wellcomeopenres.28927.r153887

Stuart J.E. Baird
Academy of Sciences of the Czech Republic, Národní, Staré Město, Czech Republic

The genome report follows the excellent DToL template. I find no typos, no grammar errors. Figure 6 shows us the blobtoolkit plot, with two outlier chromosomes. It would be nice for these to be identified for the reader...W,Z? It would be nice to have some comment/background for why the contigs of the W cannot be ordered. Apart from this I can find nothing to improve.

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: Admixture 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.

Reviewer Report 13 May 2026

https://doi.org/10.21956/wellcomeopenres.28927.r153882

Oliver Rupp
Justus-Liebig-University, Giessengiessen, Germany

This article presents the sequencing and haplotype phased assembly of the Red-necked Footman, Atolmis rubricollis genome.
The data creation and data analyses follow the state-of-the art methods for eukaryotic genome assemblies, including PacBio Hifi and Hi-C sequencing. The assembly with hifiasm in Hi-C mode was followed by Hi-C scaffolding and manual curation.
The assembly quality assessment using k-mer and orthology based methods like Merqury.FK or BUSCO showed a high quality sequence for both haplotypes.
The methods are clearly presented and the quality of the data is shown.
Overall no major issues are to report.

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: genome assembly

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.