DATA NOTE

The genome sequence of the Southern White Admiral, *Limenitis reducta* (Staudinger, 1901) (Lepidoptera: Nymphalidae)

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

Tinatin Chkhartishvili[1], Daniel Linke[2], Kay Lucek[3], Charlotte J. Wright[4], Joana I. Meier[4], Mark L. Blaxter[4],
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, Project Psyche Community

[1]Institute of Zoology, Ilia State University, Tbilisi, Tbilisi, Georgia
[2]Biology Centre of the Czech Academy of Sciences, Institute of Entomology, České Budějovice, Czech Republic
[3]University of Neuchâtel, Neuchâtel, Switzerland
[4]Tree of Life Programme, Wellcome Sanger Institute, Hinxton, England, UK

v1 **First published:** 09 Feb 2026, **11**:97
https://doi.org/10.12688/wellcomeopenres.25845.1
Latest published: 09 Feb 2026, **11**:97
https://doi.org/10.12688/wellcomeopenres.25845.1

Abstract

We present a genome assembly from a female specimen of *Limenitis reducta* (Southern White Admiral; Arthropoda; Insecta; Lepidoptera; Nymphalidae). The assembly contains two haplotypes with total lengths of 397.72 megabases and 361.87 megabases. Most of haplotype 1 (98.92%) is scaffolded into 31 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 15.18 kilobases. Gene annotation of this assembly on Ensembl identified 12 058 protein-coding genes. This work is part of Project Psyche, a collaborative programme generating genomes for European butterflies and moths.

Keywords

Limenitis reducta; Southern White Admiral; genome sequence; chromosomal; Lepidoptera

Open Peer Review

Approval Status ✓ ✓ ? ?

	1	2	3	4
version 1 09 Feb 2026	✓ view	✓ view	? view	? view

1. **Stuart J.E. Baird**, Academy of Sciences of the Czech Republic, Národní, Czech Republic
2. **Ariel Chipman**, Hebrew University of Jerusalem Department of Ecology Evolution and Behavior (Ringgold ID: 549289), Jerusalem, Israel
3. **Anwesha Das**, Sanaka Educational Trust's Group of Institutions (SETGOI), West Bengal, India
4. **Yukun Sun**, The Field Museum, Chicago,

This article is included in the Tree of Life gateway.

USA

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

Corresponding author: Charlotte J. Wright (cw22@sanger.ac.uk)

Author roles: Chkhartishvili T: Investigation, Resources; **Linke D**: Writing – Original Draft Preparation; **Lucek K**: Supervision, Writing – Review & Editing; **Wright CJ**: Funding Acquisition, Project Administration, Supervision; **Meier JI**: Funding Acquisition, Project Administration, Supervision; **Blaxter ML**: Funding Acquisition, Project Administration, Supervision;

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by Wellcome through core funding to the Wellcome Sanger Institute (220540). DL was funded by Czech Science Foundation (GACR) grant 22-35084J and Technology Agency of the Czech Republic (TACR) grant SS07010197. KL received funding from the Swiss National Science Foundation (SNSF) Grant ID 202869.
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: Chkhartishvili T, Linke D, Lucek K *et al.* **The genome sequence of the Southern White Admiral, *Limenitis reducta* (Staudinger, 1901) (Lepidoptera: Nymphalidae) [version 1; peer review: 2 approved, 2 approved with reservations]** Wellcome Open Research 2026, **11**:97 https://doi.org/10.12688/wellcomeopenres.25845.1

First published: 09 Feb 2026, **11**:97 https://doi.org/10.12688/wellcomeopenres.25845.1

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; Papilionoidea; Nymphalidae; Limenitidinae; Limenitidini; *Limenitis*; *Limenitis reducta* (Staudinger, 1901) (NCBI:txid191421)

Background

Limenitis reducta, known as the Southern White Admiral, is a butterfly species within the family Nymphalidae. Its range extends from Northern Portugal, central and eastern Spain to southern France, Switzerland, South Germany, Slovakia and Moldova, including the entire Balkan and Italian Peninsula (Tolman & Lewington, 2009). An isolated population is present in Brittany. *L. reducta* is absent from the Balearic Islands and Crete but occurs on most other Mediterranean islands (Tolman & Lewington, 2009). In the east it extends into Turkey, the Levant, the Caucasus and northern Iran. The species typically inhabits light woodlands, woodland glades, floodplains and forest edges (Crișan *et al.*, 2011; Tolman & Lewington, 2009), at elevations up to 1 400 m.

The wingspan of *L. reducta* is 46–54 mm. The upperside of the wings is dark brown to black with a metallic blue sheen and a prominent white transverse band. The underside is reddish with a silvery basal area and a row of white markings.

Larvae feed on various *Lonicera* species (honeysuckle) (Hinneberg *et al.*, 2022). Adults are univoltine in the northern part of its range (flying from mid-June to early August) and bivoltine or even trivoltine in the southern part (mid-May to mid-August, depending on elevation) (Tolman & Lewington, 2009). The larvae hibernates within a hibernaculum formed from a leaf attached to the host-plant stem using silk (Hinneberg *et al.*, 2022; Settele *et al.*, 2015).

Currently, *L. reducta* is assessed as Least Concern by the IUCN, with a stable population trend (IUCN, 2025). However, the species is highly threatened in the north of its range, in Germany it is considered critically endangered, mainly due to habitat loss (Hinneberg *et al.*, 2022; Reinhardt & Bolz, 2011).

We present a chromosome-level genome sequence of *Limenitis reducta*, sequenced as part of Project Psyche. The sequence data were derived from a female specimen (Figure 1) collected from Zekari Pass, Imereti, Georgia.

Methods

Sample acquisition

The specimen used for genome sequencing was an adult female *Limenitis reducta* (specimen ID SAN28000042, ToLID ilLimRedu1; Figure 1), collected from Zekari Pass, Imereti, Georgia (latitude 41.7937, longitude 42.8455) on 2023-08-23. The specimen was collected and identified by Tinatin Chkhartishvili (Ilia State University).

Figure 1. Voucher photograph of the *Limenitis reducta* (ilLimRedu1) specimen used for genome sequencing.

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 ilLimRedu1 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 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 44.22 ng/μL and a yield of 2 078.34 ng, with a fragment size of 15.7 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), according to 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.

Specimen details, sequencing platforms, and data yields are summarised in Table 1.

Hi-C

Sample preparation and crosslinking

The Hi-C sample was prepared from 20–50 mg of frozen tissue from the head of the ilLimRedu1 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 X.

Specimen details, sequencing platforms, and data yields are summarised in Table 1.

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).

Table 1. Specimen and sequencing data for BioProject PRJEB79421.

Platform	PacBio HiFi	Hi-C
ToLID	ilLimRedu1	ilLimRedu1
Specimen ID	SAN28000042	SAN28000042
BioSample (source individual)	SAMEA115109510	SAMEA115109510
BioSample (tissue)	SAMEA115434404	SAMEA115434403
Tissue	thorax	head
Instrument	Revio	Illumina NovaSeq X
Run accessions	ERR13615491	ERR13621420
Read count total	1.71 million	864.50 million
Base count total	14.98 Gb	130.54 Gb

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 five breaks and 65 joins. This reduced the scaffold count by 21.1%, increased the scaffold N50 by 2.9%, and increased the total assembly length by 1.4%. The curation process is described at https://gitlab.com/wtsi-grit/rapid-curation. PretextSnapshot was used to generate a Hi-C contact map of the final assembly.

Assembly quality assessment

The Merqury.FK tool (Rhie *et al.*, 2020), run in a Singularity container (Kurtzer *et al.*, 2017), was used 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 (Di Tommaso *et al.*, 2017) 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).

We used lep_busco_painter to paint Merian elements along chromosomes (Wright *et al.*, 2024). Merian elements represent the 32 ancestral linkage groups in Lepidoptera. The painting process utilised BUSCO gene locations from the lepidoptera_odb10 set (Kriventseva *et al.*, 2019) and chromosome lengths from NCBI Datasets. Each complete BUSCO gene (both single-copy and duplicated) was assigned to a Merian element based on a reference database, then plotted along chromosomes drawn to scale.

Genome sequence report

Sequence data

PacBio sequencing of the *Limenitis reducta* specimen generated 14.98 Gb (gigabases) from 1.71 million reads, which were used to assemble the genome. GenomeScope2.0 analysis estimated the haploid genome size at 377.24 Mb, with a heterozygosity of 0.55% and repeat content of 22.43% (Figure 2). These estimates guided expectations for the assembly. Based on the estimated genome size, the sequencing data provided approximately 39× coverage. Hi-C sequencing produced 130.54 Gb from 864.50 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 397.72 Mb in 164 scaffolds, with 154 gaps, and a scaffold N50 of 13.25 Mb (Table 2).

Most of the assembly sequence (98.92%) was assigned to 31 chromosomal-level scaffolds, representing 29 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). Chromosome painting with Merian elements illustrates the distribution of orthologues along chromosomes and highlights patterns of chromosomal evolution relative to Lepidopteran ancestral linkage groups (Figure 4). Contigs along chromosome 22 between 3.5 Mb and 9.2 Mb have uncertain order and orientation.

The mitochondrial genome was also assembled (length 15.18 kb, OZ186299.1). This sequence is included as a contig in the multifasta file of the genome submission and as a standalone record.

Assembly quality metrics

For haplotype 1, the estimated QV is 63.1, and for haplotype 2, 63.0. When the two haplotypes are combined, the assembly achieves an estimated QV of 63.1. The *k*-mer completeness is 88.26% for haplotype 1, 82.98% for haplotype 2, and 99.33% for the combined haplotypes (Figure 5). BUSCO analysis using the lepidoptera_odb10 reference set (n = 5 286) identified 98.6% of the expected gene set (single = 98.3%, duplicated = 0.3%) in haplotype 1. For haplotype 2, BUSCO analysis identified 94.5% of the expected gene set (single = 94.3%, duplicated = 0.1%). The snail plot in Figure 6 summarises the scaffold length distribution and other assembly statistics for haplotype 1. The blob plot in Figure 7 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) the Earth BioGenome Project Report on

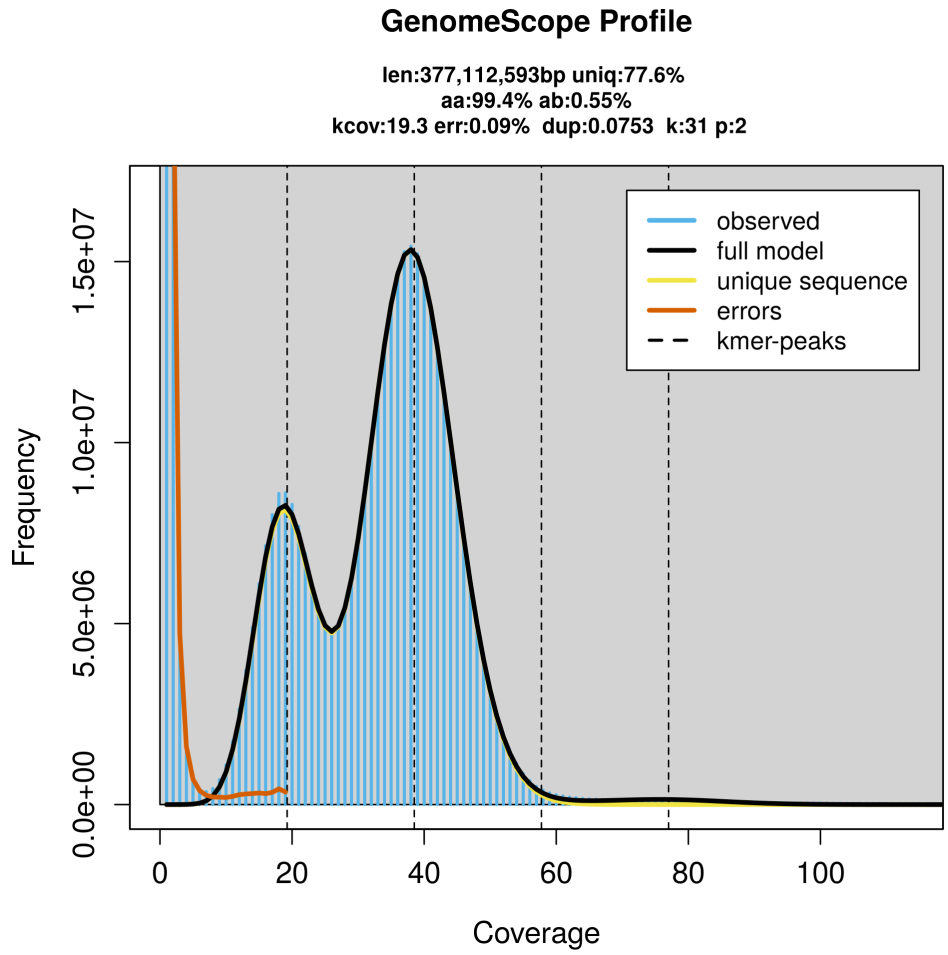

Figure 2. Frequency distribution of *k*-mers generated using GenomeScope2. The plot shows observed and modelled *k*-mer spectra, providing estimates of genome size, heterozygosity, and repeat content based on unassembled sequencing reads.

Table 2. Genome assembly statistics.

Assembly name	ilLimRedu1.hap1.1	ilLimRedu1.hap2.1
Assembly accession	GCA_964271765.1	GCA_964271735.1
Assembly level	chromosome	scaffold
Span (Mb)	397.72	361.87
Number of chromosomes	31	scaffold-level
Number of contigs	318	254
Contig N50	4.8 Mb	3.88 Mb
Number of scaffolds	164	157
Scaffold N50	13.25 Mb	12.89 Mb
Longest scaffold length (Mb)	21.7	-
Sex chromosomes	W and Z	-
Organelles	Mitochondrion: 15.18 kb	-

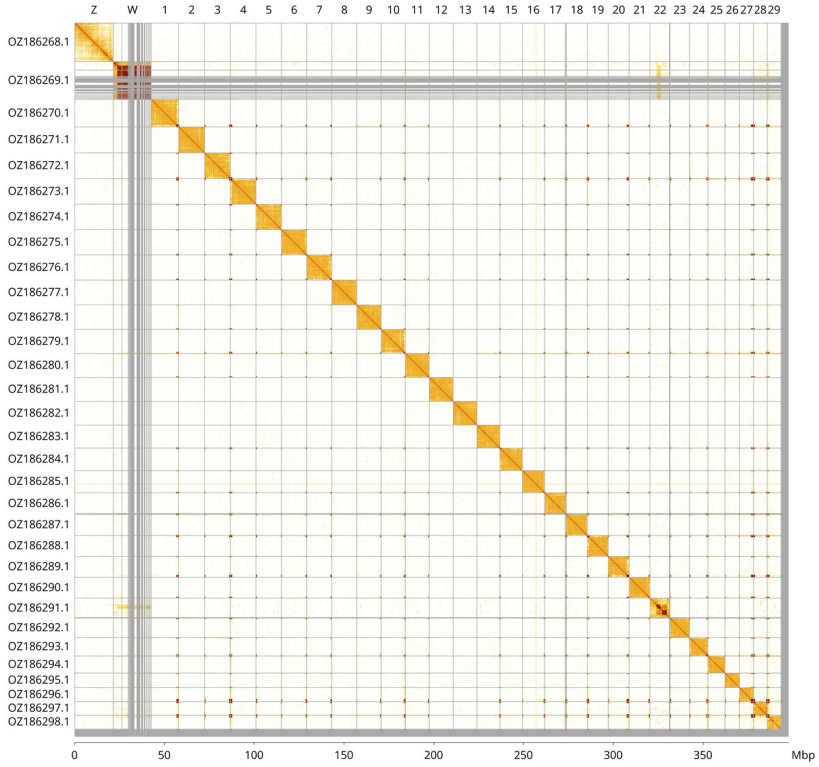

Figure 3. Hi-C contact map of the *Limenitis reducta* genome assembly. Assembled chromosomes are shown in order of size and labelled along the axes, with a megabase scale shown below. The plot was generated using PretextSnapshot.

Table 3. Chromosomal pseudomolecules in the haplotype 1 genome assembly of *Limenitis reducta* ilLimRedu1.

INSDC accession	Molecule	Length (Mb)	GC%	Assigned Merian elements
OZ186270.1	1	15.25	32	M1
OZ186271.1	2	14.42	32	M2
OZ186272.1	3	14.41	31.50	M16
OZ186273.1	4	14.29	31.50	M9
OZ186274.1	5	14.11	32	M3
OZ186275.1	6	14.08	32	M5
OZ186276.1	7	14.04	32	M12
OZ186277.1	8	13.84	32.50	M8
OZ186278.1	9	13.61	31.50	M18
OZ186279.1	10	13.43	32	M4
OZ186280.1	11	13.33	32	M7
OZ186281.1	12	13.31	32.50	M17;M20
OZ186282.1	13	13.25	32	M19;M28
OZ186283.1	14	12.82	32	M21
OZ186284.1	15	12.42	32	M6
OZ186285.1	16	12.42	32	M22
OZ186286.1	17	12.13	32.50	M15
OZ186287.1	18	11.70	32.50	M11
OZ186288.1	19	11.68	32	M13
OZ186289.1	20	11.61	31.50	M23
OZ186290.1	21	11.60	32	M10
OZ186291.1	22	11.17	38.50	M30
OZ186292.1	23	10.77	32.50	M14
OZ186293.1	24	10.25	32	M24
OZ186294.1	25	9.55	33	M26
OZ186295.1	26	8.14	33	M27
OZ186296.1	27	7.70	32	M25
OZ186297.1	28	7.68	33	M31
OZ186298.1	29	7.64	33	M29
OZ186269.1	W	21.07	34	-
OZ186268.1	Z	21.70	32.50	MZ

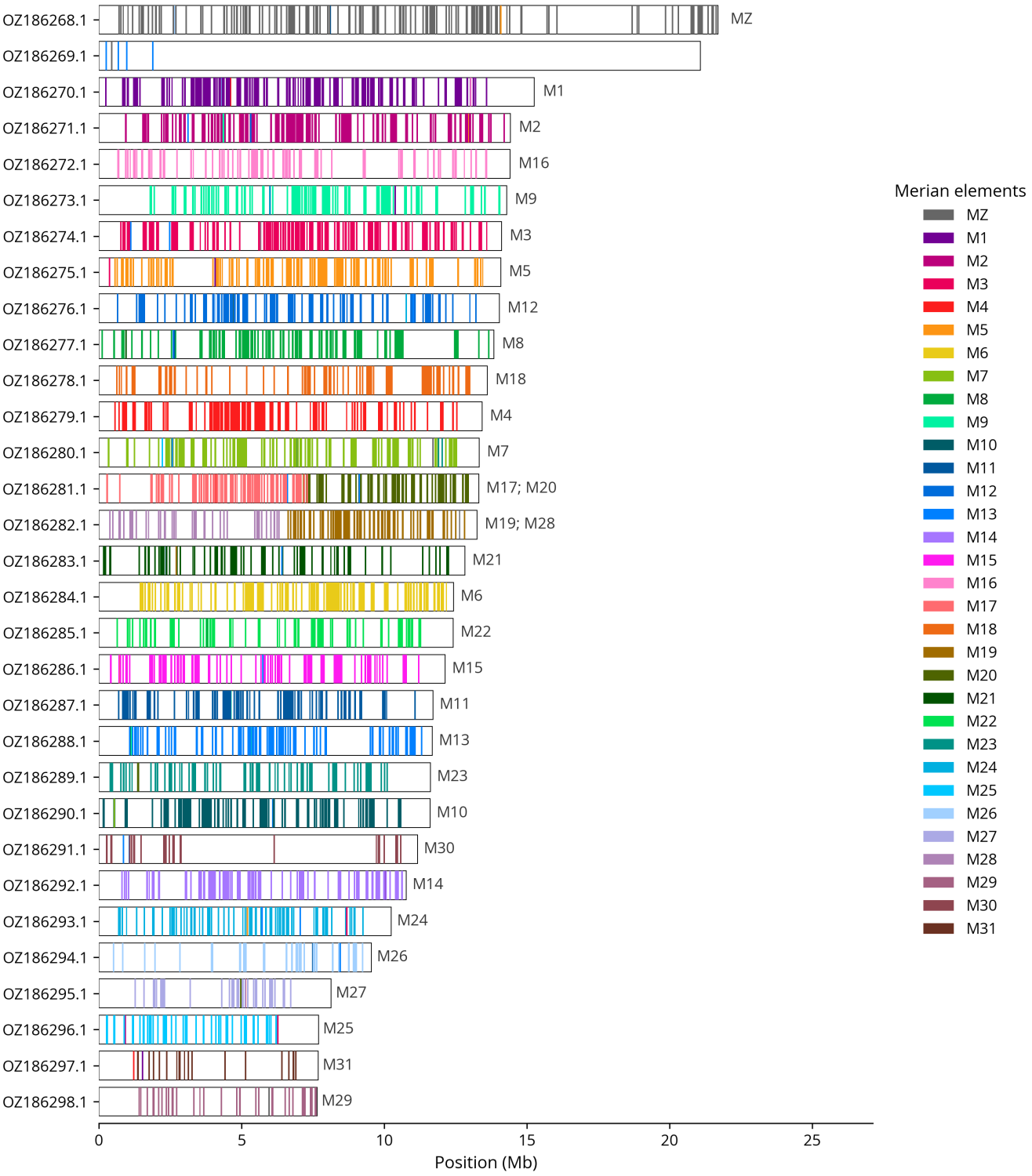

Figure 4. Merian elements painted across chromosomes in the ilLimRedu1.hap1.1 assembly of *Limenitis reducta*. Chromosomes are drawn to scale, with the positions of orthologues shown as coloured bars. Each orthologue is coloured by the Merian element that it belongs to. All orthologues which could be assigned to Merian elements are shown.

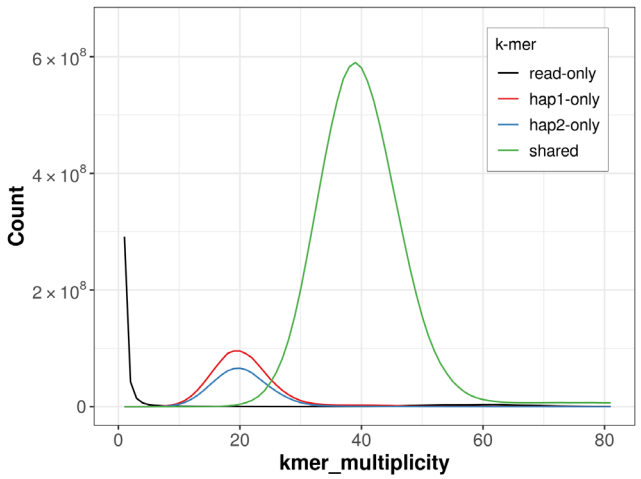

Figure 5. Evaluation of *k*-mer completeness using MerquryFK. This plot illustrates the recovery of *k*-mers from the original read data in the final assemblies. The horizontal axis represents *k*-mer multiplicity, and the vertical axis shows the number of *k*-mers. The black curve represents *k*-mers that appear in the reads but are not assembled. The green curve (the homozygous peak) corresponds to *k*-mers shared by both haplotypes and the red and blue curves (the heterozygous peaks) show *k*-mers found only in one of the haplotypes.

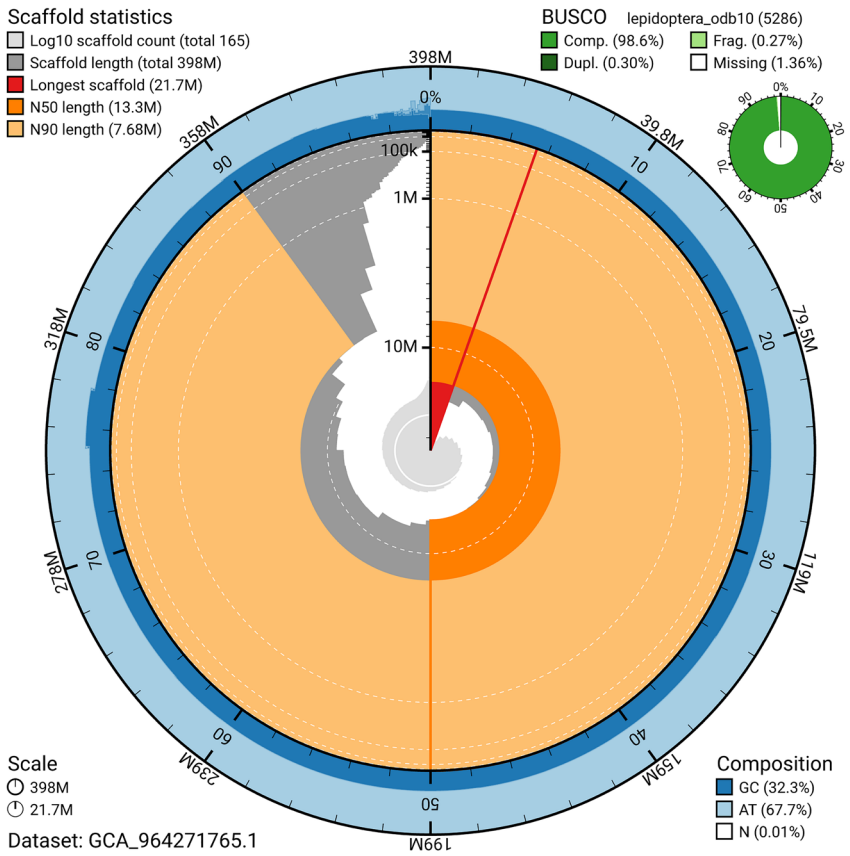

Figure 6. Assembly metrics for ilLimRedu1.hap1.1. The BlobToolKit snail plot provides an overview of assembly metrics and BUSCO gene completeness. The circumference represents the length of the whole genome sequence, and the main plot is divided into 1,000 bins around the circumference. The outermost 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.

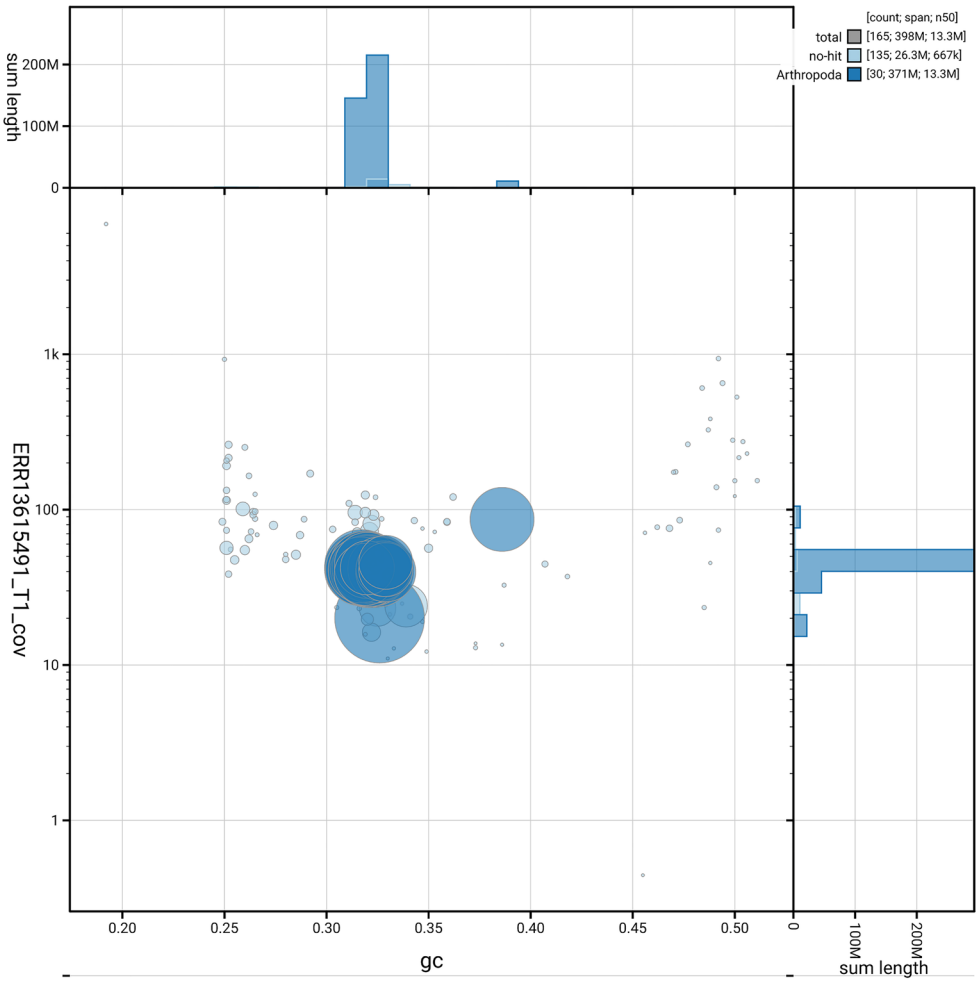

Figure 7. BlobToolKit GC-coverage plot for ilLimRedu1.hap1.1. Blob plot showing sequence coverage (vertical axis) and GC content (horizontal axis). The circles represent scaffolds, with the size proportional to scaffold length and the colour representing phylum membership. The histograms along the axes 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 *Limenitis reducta* assembly.

Measure	Value	Benchmark
EBP summary (haplotype 1)	6.C.Q63	6.C.Q40
Contig N50 length	4.80 Mb	≥ 1 Mb
Scaffold N50 length	13.25 Mb	= chromosome N50
Consensus quality (QV)	Haplotype 1: 63.1; haplotype 2: 63.0; combined: 63.1	≥ 40
k-mer completeness	Haplotype 1: 88.26%; Haplotype 2: 82.98%; combined: 99.33%	≥ 95%
BUSCO	C:98.6% [S:98.3%; D:0.3%]; F:0.3%; M:1.1%; n:5 286	S > 90%; D < 5%
Percentage of assembly assigned to chromosomes	98.92%	≥ 90%

Notes: 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

Assembly Standards September 2024. The EBP metric, calculated for the haplotype 1, is **6.C.Q63**, meeting the recommended reference standard.

Genome annotation report

The *Limenitis reducta* genome assembly (GCA_964271765.1) was annotated by Ensembl at the European Bioinformatics Institute (EBI). This annotation includes 22 915 transcribed mRNAs from 12 058 protein-coding and 2 402 non-coding genes. The average transcript length is 13 023.43 bp, with an average of 1.58 coding transcripts per gene and 7.86 exons per transcript. For further information, please refer to the Ensembl annotation page.

Wellcome Sanger Institute – Legal and Governance

The materials that have contributed to this genome note have been supplied by a Tree of Life collaborator. 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 undertaken according to a Research Collaboration Agreement or Material Transfer Agreement entered into by the Tree of Life collaborator, Genome Research Limited (operating as the Wellcome Sanger Institute), and in some circumstances, other Tree of Life collaborators.

Data availability

European Nucleotide Archive: Limenitis reducta (southern white admiral). Accession number PRJEB79421. The genome sequence is released openly for reuse. The *Limenitis reducta* genome sequencing initiative is part of 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.

Pipelines used for genome assembly at the WSI Tree of Life are available at https://pipelines.tol.sanger.ac.uk/pipelines. Table 5 lists software versions used in this study.

Author information

Contributors are listed at the following links:

- 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 Project Psyche Community

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
Hifiasm	0.19.8-r603	https://github.com/chhylp123/hifiasm
HiGlass	1.13.4	https://github.com/higlass/higlass

Software	Version	Source
lep_busco_painter	1.0.0	https://github.com/charlottewright/lep_busco_painter
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

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

Crişan A, Craioveanu C, Rákosy L: **Effects of alterations of Romania's forestry stock area and structure on diurnal Lepidoptera – a GIS approach.** In: *Proceedings of the International Conference Integrated Management of Environmental Resources – Suceava, 4–6 November 2011.* Editura Universităţii Ştefan cel Mare Suceava, 2011; 73–84.
Reference Source

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

Di Tommaso P, Chatzou M, Floden EW, *et al.*: **Nextflow enables reproducible computational workflows.** *Nat Biotechnol.* 2017; **35**(4): 316–319.
PubMed Abstract | Publisher 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

Hinneberg H, Döring J, Hermann G, *et al.*: **Multi-surveyor capture-mark-recapture as a powerful tool for butterfly population monitoring in the pre-imaginal stage.** *Ecol Evol.* 2022; **12**(8): e9140.
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

IUCN: **The IUCN Red List of Threatened Species**. Version 2025-1, 2025.
Reference Source

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

Kriventseva EV, Kuznetsov D, Tegenfeldt F, *et al.*: **OrthoDB v10: sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs.** *Nucleic Acids Res.*

2019; **47**(D1): D807–D811.
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): 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**

Reinhardt R, Bolz R: **Rote Liste Und Gesamtartenliste Der Tagfalter (Rhopalocera) Deutschlands**. Bundesamt für Naturschutz, Bonn, 2011; **70**(3).

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**

Settele J, Steiner R, Reinhardt R, *et al.*: **Schmetterlinge: Die Tagfalter Deutschlands.** Stuttgart, Germany: Verlag Eugen Ulmer, 2015.
Reference Source

Tolman T, Lewington R: **Collins butterfly guide: the most complete guide to the Butterflies of Britain and Europe**. HarperCollins Publishers, 2009.
Reference Source

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 06 May 2026

https://doi.org/10.21956/wellcomeopenres.28446.r154174

?

Yukun Sun
The Field Museum, Chicago, Illinois, USA

This Data Note presents a chromosome-level, haplotype-resolved genome assembly for a female *Limenitis reducta*specimen. The work was produced through Project Psyche and the Wellcome Sanger Institute Tree of Life Programme. The authors used PacBio HiFi reads, Arima-HiC data, Hifiasm, YaHS, manual curation, MitoHiFi, BlobToolKit, Merqury.FK, BUSCO, and Ensembl annotation. The data are deposited under PRJEB79421. The article fits the Wellcome Open Research model, where review should assess academic merit and validity rather than novelty. Overall, the assembly is strong and well documented. Haplotype 1 spans 397.72 Mb across 164 scaffolds. Most sequence, 98.92%, is assigned to 31 chromosomal pseudomolecules. The reported QV values are high. BUSCO completeness is also high. The Hi-C map, Merian-element painting, MerquryFK plot, and BlobToolKit plots support the assembly quality. The software version table also improves reproducibility. In my view, the article is scientifically sound. I do not see issues that require new analysis before the article can be considered valid. However, several clarifications would make the paper more useful for readers and downstream users.

Major points

1. Explain the unresolved region on chromosome 22. The text says contigs from 3.5 Mb to 9.2 Mb have an unclear order. That is about 5.7 Mb on an 11.17 Mb chromosome. So roughly half of the chromosome is uncertain. Users doing synteny, recombination, or structural-variant work need to know why. Likely causes are weak Hi-C signal, repeats, heterochromatin near the centromere, or haplotype switching. The authors should state the most likely cause. If the cause is unclear, they should say so. One or two sentences in the Assembly curation section would be enough. A short note on whether ONT reads or optical mapping could fix the region would also help.

2. Make the reason for sequencing this species clearer**.** The Background has good ecological detail. But the reason for choosing this species stays implicit. One short sentence in the Abstract or Background would help. For example, *L. reducta* is Least Concern globally but critically endangered in Germany. That contrast makes it useful for regional conservation genomics. The genome also fits comparative chromosome work in Nymphalidae.

3. Clarify the k-mer completeness benchmark. Table 4 lists ≥95% as the target. The reported values are 88.26% for haplotype 1 and 82.98% for haplotype 2. Only the combined value (99.33%) meets the target. This split is normal for phased diploid assemblies. But the table can read as if the benchmark is missed. A short footnote, or one extra clause in the text, would fix this.

Minor points

4. Note the limits of automated annotation. Ensembl reports 12,058 protein-coding genes and 2,402 non-coding genes. There is no species-specific RNA-seq input. So UTRs, alternative isoforms, and lineage-specific genes may be incomplete. One short sentence pointing readers to this would help downstream users.

5. Consider releasing curation artifacts. The raw Hi-C reads are public, which is good. But Pretext maps or curation BAMs would help others re-curate chromosome 22. A version tag on the rapid-curation repo would also help reproducibility.

6. Style and formatting. Use "haplotype 1" and "haplotype 2" everywhere. Avoid switching to "primary" or "secondary" outside the accession names. Numbers such as "12 058", "22 915", and "1 400 m" use thin spaces and render unevenly across viewers. Pick one format and use it throughout.

7. Small wording. "The larvae hibernates" should read "The larva hibernates" or "Larvae hibernate".

I did not re-run the assembly, scaffolding, contamination screen, BUSCO, Merqury.FK, BlobToolKit, MitoHiFi, or Ensembl pipelines. I did not inspect the raw reads. I did not test single gene models. My review is based on the text, figures, tables, methods, software list, and public accessions.

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, comparative genomics, phylogenomics, bioinformatics, long-read sequencing, chromosome-level assembly assessment, genome annotation, and evolutionary 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 30 April 2026

https://doi.org/10.21956/wellcomeopenres.28446.r154169

?

Anwesha Das
Sanaka Educational Trust's Group of Institutions (SETGOI), West Bengal, India

The manuscript entitled as "The genome sequence of the Southern White Admiral, *Limenitis reducta* (Staudinger, 1901) (Lepidoptera: Nymphalidae)" presents a chromosome-level, haplotype-resolved genome assembly for the Southern White Admiral butterfly (*Limenitis reducta*), generated as part of Project Psyche and the Wellcome Sanger Institute Tree of Life programme. Using PacBio HiFi long-read sequencing and Hi-C scaffolding, the authors assembled two haplotypes, with haplotype 1 curated to 31 chromosomal pseudomolecules (29 autosomes plus W and Z sex chromosomes). The final assembly totals 397.72 Mb for haplotype 1, with very high consensus quality (QV ≈ 63), strong BUSCO completeness (98.6%), and extensive chromosomal assignment (98.92%). The authors additionally report assembly metrics in line with Earth BioGenome Project reference standards, include a fully assembled mitochondrial genome (15.18 kb), and provide Ensembl-based genome annotation identifying 12,058 protein-coding genes. The manuscript largely follows established Data Note conventions, providing extensive methodological detail, clear quality assessment, and open access to all underlying sequencing data and assemblies. Overall, the work constitutes a high-quality genomic resource for Lepidoptera research, conservation genomics, and comparative chromosome evolution studies. This is a technically strong and clearly presented genome resource that meets or exceeds current standards for reference-quality assemblies in insects. The assembly quality metrics, methodological transparency, and data availability are excellent. The manuscript aligns well with expectations for a Data Note and will be valuable to a wide range of users. The study is scientifically sound; however, a small number of clarifications and contextual improvements would significantly strengthen the manuscript's usefulness and transparency, particularly for non-specialist readers or downstream users.

Major Points

1. Explicit explanation of the unresolved region on chromosome 22
The manuscript notes that: "Contigs along chromosome 22 between 3.5 Mb and 9.2 Mb have uncertain order and orientation." This is a notable exception in an otherwise chromosome-complete assembly, yet no explanation is provided.
Unresolved regions may affect downstream analyses (e.g., recombination, structural variation, gene order, or comparative synteny studies). Users need to understand whether the uncertainty is

due to biological features (e.g., repeats, heterochromatin), technical limitations, or insufficient Hi-C signal.

Briefly explain why this region remains unresolved. For example: Low Hi-C contact density, Conflicting Hi-C signals, High repeat content or structural complexity, Potential misassembly between haplotypes. If the reason is unclear, state this explicitly. Indicate whether future resequencing or alternative scaffolding approaches could resolve the region. This can be addressed in one or two sentences in the Genome sequence report or Assembly curation sections.

2. Clearer rationale for generating this genome in the abstract or background
The rationale for sequencing *L. reducta* is implicit (Project Psyche, biodiversity genomics) but only briefly stated. Data Notes serve a broad readership. Explicit rationale helps users understand the biological, ecological, or evolutionary relevance of the species and dataset. Add one concise sentence to the abstract or background clarifying the motivation, e.g.: Its biogeographic distribution, Conservation relevance (e.g. regional endangerment), Importance for Lepidoptera chromosome evolution. This does not require new analyses, only contextual framing.

3. Brief contextualisation of chromosome number and sex chromosomes
The assembly reports 29 autosomes plus W and Z sex chromosomes, but this is not contextualised relative to other *Limenitis* or Nymphalidae species. Given the inclusion of Merian element painting and the relevance of chromosome evolution in Lepidoptera, readers would benefit from minimal comparative context. Add one short sentence noting whether this karyotype is typical or notable within: *Limenitis*, Nymphalidae, Lepidoptera more broadly. No comparative analysis is required-just a statement of expectation or consistency.

Minor Points and Suggestions

1. Terminology consistency
Ensure consistent use of: “haplotype 1 / haplotype 2” vs. “primary / secondary assembly”. This avoids confusion for non-specialist readers.

2. Gene annotation interpretation
The Ensembl annotation statistics are clearly reported, but a short sentence stating that annotation follows standard automated Ensembl pipelines (with their known strengths/limitations) would be helpful for users interpreting gene models.

3. Minor stylistic edits
A careful proofread could eliminate a small number of typographical issues (e.g. spacing around units such as “1 400 m” or “22 915”).

I did not independently re-run assembly or annotation pipelines. I did not directly inspect raw sequencing reads or validate contamination filtering beyond the reported BlobToolKit outputs. Functional accuracy of individual gene models was not assessed (appropriate for a Data Note). These limitations are inherent to manuscript-based peer review and do not detract from the technical quality presented.

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: Computational chemistry, medicinal chemistry, pharmacology, assays.

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 29 April 2026

https://doi.org/10.21956/wellcomeopenres.28446.r154173

Ariel Chipman
Hebrew University of Jerusalem Department of Ecology Evolution and Behavior (Ringgold ID: 549289), Jerusalem, Jerusalem District, Israel

The authors report a chromosome-level genome sequence of the nymphal butterfly *Limenitis reducta* (White Admiral), from a female specimen collected in the mountains of Georgia. The sequencing and assembly follow the standard Tree of Life protocols and guidelines. The analysis of the sequenced genome shows a high-quality assembly, with very high completeness scores and BUSCO scores.
The written report is clear and concise as expected from this type of article. I have no comments or suggestions for change.

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: Arthropod 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 25 March 2026

https://doi.org/10.21956/wellcomeopenres.28446.r147845

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

The report largely conforms to the excellent DToL template. The rationale for creating the dataset is only fleetingly mentioned in the abstract. Unusually for DToL, there is a clearly un-assembled region: "Contigs along chromosome 22 between 3.5 Mb and 9.2 Mb have uncertain order and orientation." The reader should be told why, or if the reason is unclear, the reader should be told why it is unclear.

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.