



DATA NOTE

The genome sequence of *Erebia montana* Berg-Mohrenfalter (Lepidoptera: Nymphalidae)

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

Kay Lucek ¹, Irena Klečková ², Charlotte J. Wright ³, Joana I. Meier³, Mark L. Blaxter ³,
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

¹University of Neuchâtel, Neuchâtel, Switzerland
²Biology Centre CAS, Institute of Entomology, České Budějovice, Czech Republic
³Tree of Life, Wellcome Sanger Institute, Hinxton, England, UK

V1 First published: 18 Aug 2025, 10:449
<https://doi.org/10.12688/wellcomeopenres.24675.1>
Latest published: 18 Aug 2025, 10:449
<https://doi.org/10.12688/wellcomeopenres.24675.1>

Abstract

We present a genome assembly from a female specimen of *Erebia montana* (Arthropoda; Insecta; Lepidoptera; Nymphalidae). The assembly contains two haplotypes with total lengths of 597.21 megabases and 499.88 megabases. Most of haplotype 1 (98.25%) is scaffolded into 24 chromosomal pseudomolecules, including the W, Z₁, and Z₂ sex chromosomes. Haplotype 2 was assembled to scaffold level. The mitochondrial genome has also been assembled, with a length of 15.2 kilobases.

Keywords

Erebia montana, genome sequence, chromosomal, Lepidoptera



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

Open Peer Review

Approval Status

	1	2	3
version 1			
18 Aug 2025	view	view	view
1. Yang Mei , Zhejiang University, Hangzhou, China Jilin Agricultural University (Ringgold ID: 85112), Changchun, China			
2. Jasmine D Alqassar , The George Washington UniversityWashington, District of Columbia, USA			
3. Li-Wei Wu , Tunghai University, Taichung, Taiwan			
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: Lucek K: Investigation, Resources, Supervision; Klečková I: Writing – Original Draft Preparation; 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). KL was supported by Swiss National Science Foundation (SNSF) Grant ID 202869. IK was supported by Czech Science Foundation (GACR) grant 22-35084J and Technology Agency of the Czech Republic (TACR) grant SS07010197.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2025 Lucek K *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Lucek K, Klečková I, Wright CJ *et al.* **The genome sequence of *Erebia montana* Berg-Mohrenfalter (Lepidoptera: Nymphalidae) [version 1; peer review: 2 approved, 1 approved with reservations]** Wellcome Open Research 2025, 10:449 <https://doi.org/10.12688/wellcomeopenres.24675.1>

First published: 18 Aug 2025, 10:449 <https://doi.org/10.12688/wellcomeopenres.24675.1>

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; Satyrinae; Erebiini; *Erebia*; *Erebia montana* Berg-Mohrenfalter (NCBI:txid111905)

Background

Erebia montana is an endemic European species, found in the Alps and the Central Apennines. It belongs to the family Nymphalidae. The upperside of the wings is brownish with an orange band and black ocelli. The underside of the hindwing is brownish-grey and mottled, while the underside of the forewing is predominantly orange (Tolman & Lewington, 2008).

Erebia montana is univoltine and overwinters in the first larval stage. Like other *Erebia* species, it has four larval instars. Adults are on the wing from mid-July to early September, depending on elevation (Tolman & Lewington, 2008).

The species inhabits montane to alpine zones, typically in close proximity to rocky areas (Sonderegger, 2005). It prefers sunny, dry, and steep slopes. On hot days, males often rest on wet soil. The species is capable of efficient heat absorption from rocks (Kleckova *et al.*, 2014). In southern Ticino and the Münstertal, *E. montana* co-occurs with the closely related *E. styx*; however, the two species are not found in true sympatry. *Erebia styx* inhabits calcareous bedrock, whereas *E. montana* is restricted to crystalline substrates (Sonderegger, 2005).

We present a chromosome-level, haplotype-resolved genome sequence of *Erebia montana*, sequenced as part of Project Psyche. The sequence data were derived from a female specimen (Figure 1) collected from Stoos, Schwyz, Switzerland.



Figure 1. Voucher photograph of the *Erebia montana* (ilEreMont1) specimen used for genome sequencing.

Methods

Sample acquisition

The specimen used for genome sequencing was an adult female *Erebia montana* (specimen ID SAN28000023, ToLID ilEreMont1; Figure 1), collected from Stoos, Schwyz, Switzerland (latitude 46.9691, longitude 8.664; elevation 1463 m) on 08/08/2022. The specimen was collected and identified by Kay Lucek (University of Neuchâtel).

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 ilEreMont1 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 41.16 ng/μL and a yield of 1 934.52 ng, with a fragment size of 17.2 kb. The 260/280 spectrophotometric ratio was 1.91, and the 260/230 ratio was 2.88.

PacBio HiFi library preparation and sequencing

Library preparation and sequencing were performed at the WSI Scientific Operations core. Libraries were prepared using the SMRTbell Prep Kit 3.0 (Pacific Biosciences, California, USA), following the manufacturer's instructions. The kit includes reagents for end repair/A-tailing, adapter ligation, post-ligation SMRTbell bead clean-up, and nuclease treatment. Size selection and clean-up were performed using diluted AMPure PB beads (Pacific Biosciences). DNA concentration was quantified using a Qubit Fluorometer v4.0 (ThermoFisher Scientific) and the Qubit 1X dsDNA HS assay kit. Final library fragment size was assessed with the Agilent FemtoPulse 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 head tissue of the ilEreMont1 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 Diagenode Power Masher-II. Crosslinked DNA was digested with a restriction enzyme master mix, biotinylated, and ligated. Clean-up was performed with SPRIselect beads before library preparation. DNA concentration was measured with the Qubit Fluorometer (Thermo Fisher Scientific) and Qubit HS Assay Kit. The biotinylation percentage was estimated using the Arima-HiC v2 QC beads.

Hi-C library preparation and sequencing

Biotinylated DNA constructs were fragmented using a Covaris E220 sonicator and size selected to 400–600 bp using SPRIselect beads. DNA was enriched with Arima-HiC v2 kit Enrichment beads. End repair, A-tailing, and adapter ligation were carried out with the NEBNext Ultra II DNA Library Prep Kit (New England Biolabs), following a modified protocol where library preparation occurs while DNA remains bound to the Enrichment beads. Library amplification was performed using KAPA HiFi HotStart mix and a custom Unique Dual Index (UDI) barcode set (Integrated DNA Technologies). Depending on sample concentration and biotinylation percentage determined at the crosslinking stage, libraries were amplified with 10–16 PCR cycles. Post-PCR clean-up was performed with SPRIselect beads. Libraries were quantified using the AccuClear Ultra High Sensitivity dsDNA Standards Assay

Kit (Biotium) and a FLUOstar Omega plate reader (BMG Labtech).

Prior to sequencing, libraries were normalised to 10 ng/μL. Normalised libraries were quantified again and equimolar and/or weighted 2.8 nM pools. 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).

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 12 breaks, 43 joins, and removal of 2 haplotypic duplications. The curation process is documented at <https://gitlab.com/wtsi-grit/rapid-curation>. PretextViewSnapshot was used to generate a Hi-C contact map of the final assembly.

Assembly quality assessment

Chromosomal painting was performed using lep_buscoPainter using Merian elements, which represent the 32 ancestral linkage groups in Lepidoptera (Wright *et al.*, 2024). Painting was based on gene locations from the lepidoptera_odb10 BUSCO analysis and chromosome lengths from the genome index produced using SAMtools faidx (Danecek *et al.*, 2021).

Table 1. Specimen and sequencing data for BioProject.

Platform	PacBio HiFi	Hi-C
ToLID	ilEreMont1	ilEreMont1
Specimen ID	SAN28000023	SAN28000023
BioSample (source individual)	SAMEA115109513	SAMEA115109513
BioSample (tissue)	SAMEA115109647	SAMEA115109634
Tissue	thorax	head
Sequencing platform and model	Revio	Illumina NovaSeq X
Run accessions	ERR13485712	ERR13474177
Read count total	1.53 million	827.67 million
Base count total	16.56 Gb	124.98 Gb

Each complete BUSCO (including both single-copy and duplicated BUSCOs) was assigned to a Merian element using a reference database, and coloured positions were plotted along chromosomes drawn to scale.

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 databases ($k = 31$) computed prior to genome assembly. The analysis outputs included assembly QV scores and completeness statistics.

The genome was analysed using the BlobToolKit pipeline, a Nextflow implementation of the earlier Snakemake BlobToolKit pipeline (Challis *et al.*, 2020). The pipeline aligns PacBio reads using minimap2 (Li, 2018) and SAMtools (Daneczek *et al.*, 2021) to generate coverage tracks. Simultaneously, it queries the GoAT database (Challis *et al.*, 2023) to identify relevant BUSCO lineages and runs BUSCO (Manni *et al.*, 2021). For the three domain-level BUSCO 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 package management via Conda and Bioconda (Grüning *et al.*, 2018), and containerisation through Docker (Merkel, 2014) and Singularity (Kurtzer *et al.*, 2017).

Genome sequence report

Sequence data

The genome of a specimen of *Erebia montana* was sequenced using Pacific Biosciences single-molecule HiFi long reads, generating 16.56 Gb (gigabases) from 1.53 million reads, which were used to assemble the genome. GenomeScope2.0 analysis estimated the haploid genome size at 531.42 Mb, with a heterozygosity of 0.96% and repeat content of 39.49%. These estimates guided expectations for the assembly. Based on the estimated genome size, the sequencing data provided approximately 30× coverage. Hi-C sequencing produced 124.98 Gb from 827.67 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 597.21 Mb in 92 scaffolds, with 131 gaps, and a scaffold N50 of 21.59 Mb (Table 2).

Table 2. Genome assembly statistics.

Assembly name	ilEreMont1.hap1.1	ilEreMont1.hap2.1
Assembly accession	GCA_964260675.1	GCA_964260635.1
Assembly level	chromosome	scaffold
Span (Mb)	597.21	499.88
Number of chromosomes	24	N/A
Number of contigs	223	184
Contig N50	8.87 Mb	8.7 Mb
Number of scaffolds	92	92
Scaffold N50	21.59 Mb	21.46 Mb
Longest scaffold length (Mb)	55.06	N/A
Sex chromosomes	W; Z ₁ and Z ₂	N/A
Organelles	Mitochondrial genome: 15.2 kb	N/A

Most of the assembly sequence (98.25%) was assigned to 24 chromosomal-level scaffolds, representing 21 autosomes and the W, Z₁, and Z₂ sex chromosomes. The contigs in chromosome W from 11.72 to 28.34 Mbp have uncertain order and orientation. These chromosome-level scaffolds, confirmed by Hi-C data, are named according to size (Figure 2; 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 3).

The mitochondrial genome was also assembled. 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 65.3, and for haplotype 2, 64.5. When the two haplotypes are combined, the assembly achieves an estimated QV of 64.9. The *k*-mer completeness is 84.20% for haplotype 1, 73.40% for haplotype 2, and 97.35% for the combined haplotypes (Figure 4). BUSCO analysis using the lepidoptera_odb10 reference set ($n = 5286$) (Kriventseva *et al.*, 2019) identified 98.6% of the expected gene set (single = 94.6%, duplicated = 4.0%) 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) the Earth BioGenome Project Report on

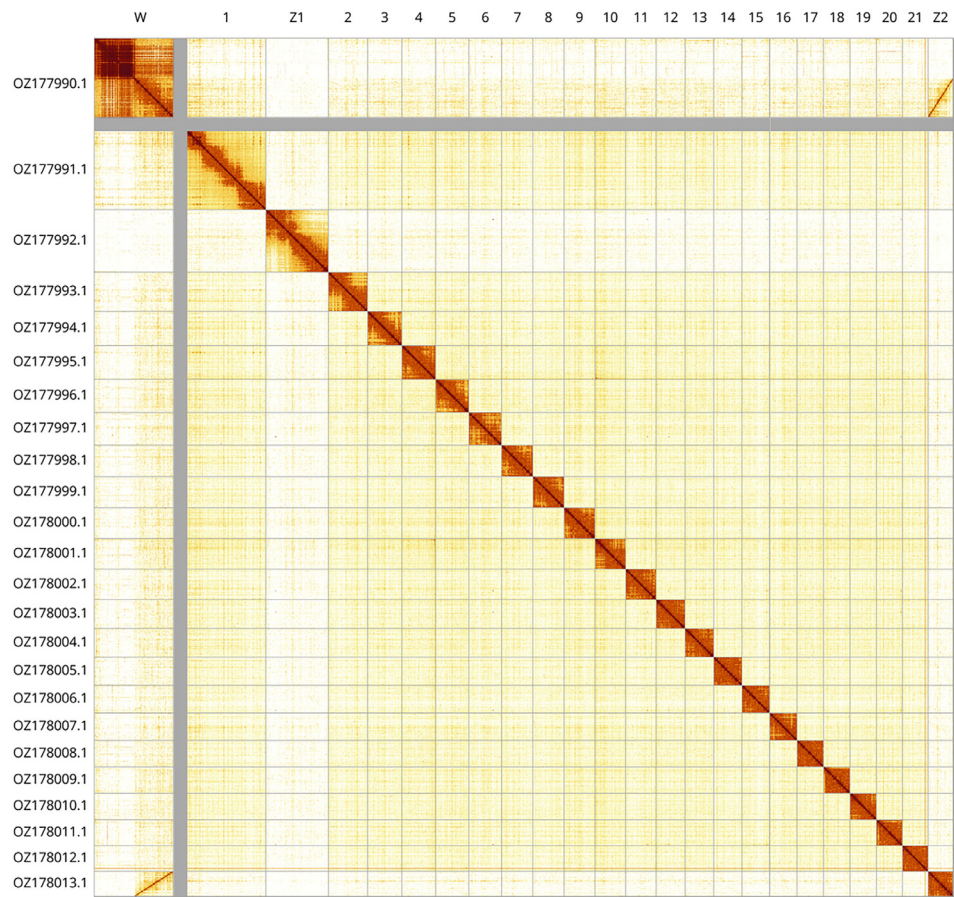


Figure 2. Hi-C contact map of the *Erebia montana* genome assembly. Assembled chromosomes are shown in order of size and labelled along the axes. The plot was generated using PretextViewSnapshot.

Table 3. Chromosomal pseudomolecules in the haplotype 1 genome assembly of *Erebia montana* ilEreMont1.

INSDC accession	Molecule	Length (Mb)	GC%	Assigned Merian elements
OZ177991.1	1	54.77	37	M2;M23;M30
OZ177993.1	2	27.32	37	M13;M28
OZ177994.1	3	23.75	37	M1
OZ177995.1	4	23.43	37	M19;M26
OZ177996.1	5	23.02	37	M17;M20
OZ177997.1	6	22.82	37	M3
OZ177998.1	7	21.71	37	M8
OZ177999.1	8	21.59	37	M9
OZ178000.1	9	21.43	37	M14;M29
OZ178001.1	10	21.37	37.50	M24;M31
OZ178002.1	11	21.17	37	M5
OZ178003.1	12	20.11	37.50	M7
OZ178004.1	13	19.80	37	M6
OZ178005.1	14	19.65	37	M12
OZ178006.1	15	19.21	37	M16
OZ178007.1	16	19.05	37	M4
OZ178008.1	17	18.39	37	M22
OZ178009.1	18	18.32	37	M15
OZ178010.1	19	18.31	37.50	M25;M27
OZ178011.1	20	18.15	37.50	M10
OZ178012.1	21	17.82	37	M21
OZ177990.1	W	64.57	37	N/A
OZ177992.1	Z1	43.35	36.50	M18;MZ
OZ178013.1	Z2	17.21	37.50	M11

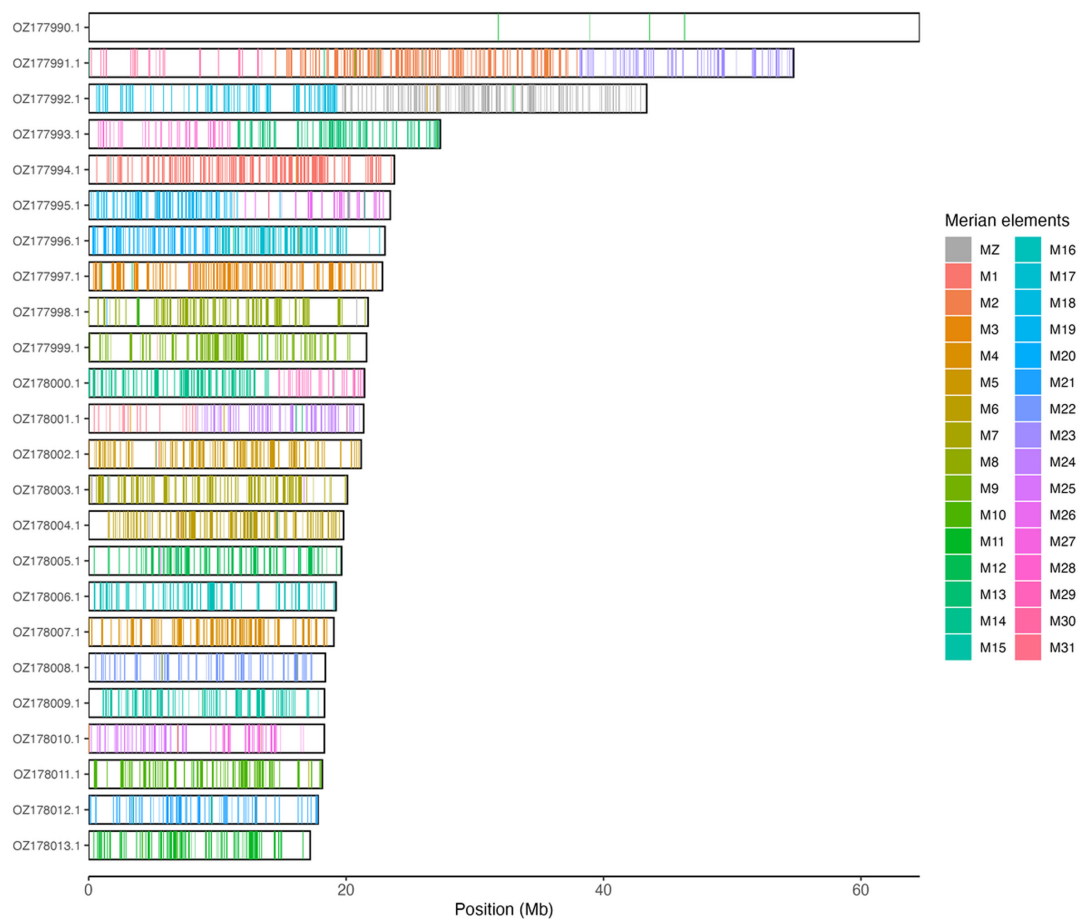


Figure 3. Merian elements painted across chromosomes in the iEreMont1.hap1.1 assembly of *Erebia montana*. Chromosomes are drawn to scale, with the positions of orthologues shown as coloured bars. Each orthologue is coloured by the Merian element that it belongs to. All orthologues which could be assigned to Merian elements are shown.

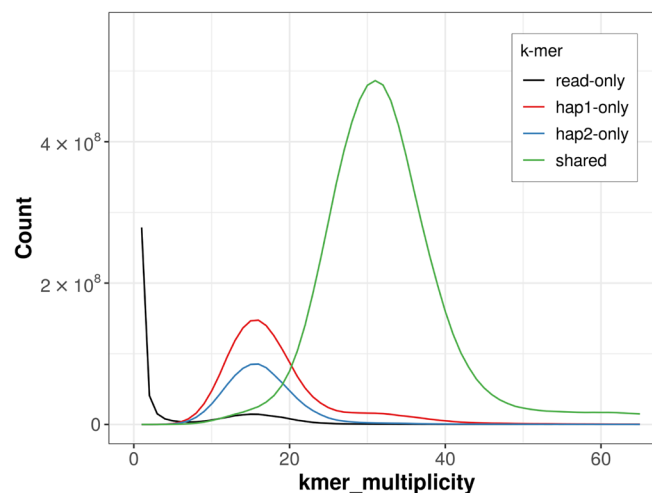


Figure 4. Evaluation of k -mer completeness using MerquryFK. This plot illustrates the recovery of k -mers from the original read data in the final assemblies. The horizontal axis represents k -mer multiplicity, and the vertical axis shows the number of k -mers. The black curve represents k -mers that appear in the reads but are not assembled. The green curve (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.

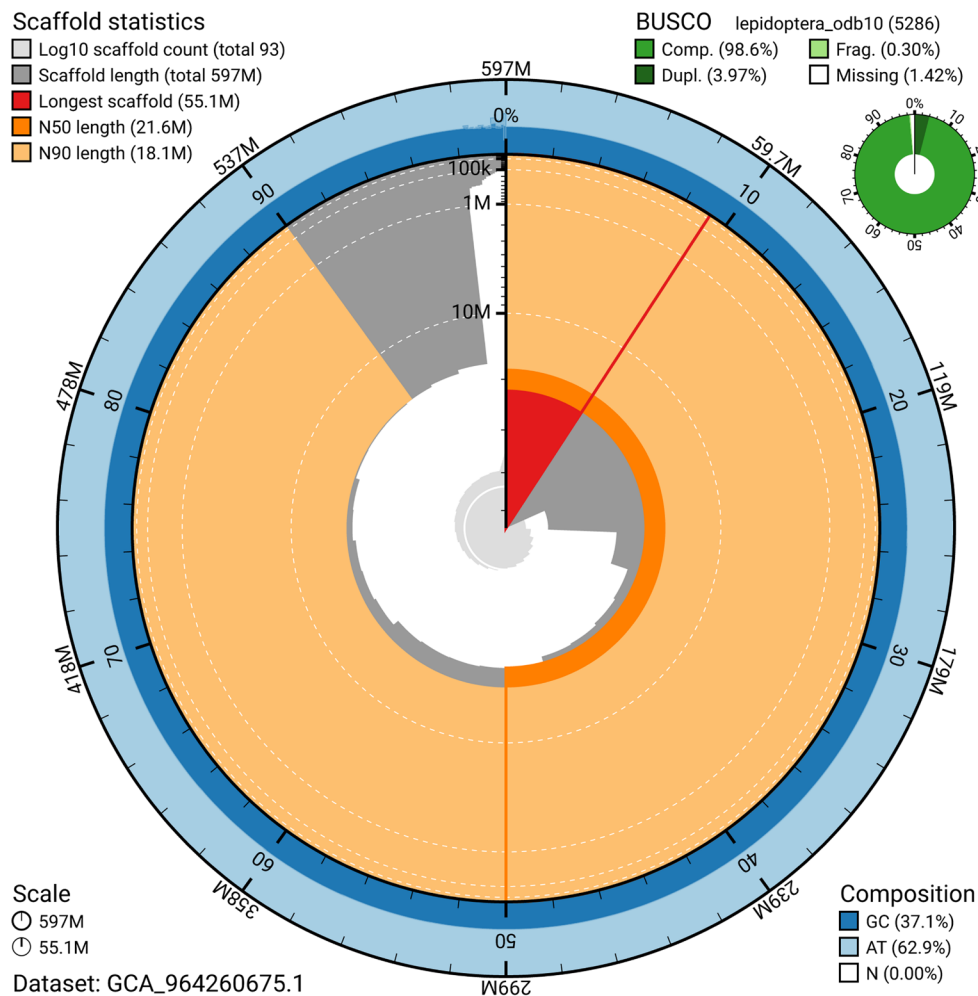


Figure 5. Assembly metrics for iEreMont1.hap1.1. The BlobToolKit snail plot provides an overview of assembly metrics and BUSCO gene completeness. The circumference represents the length of the whole genome sequence, and the main plot is divided into 1,000 bins around the circumference. The outermost 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](#).

Assembly Standards [September 2024](#). The EBP metric, calculated for the haplotype 1, is **6.C.Q65**, meeting the recommended reference standard.

Wellcome Sanger Institute – Legal and Governance

The materials that have contributed to this genome note have been supplied by a 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.

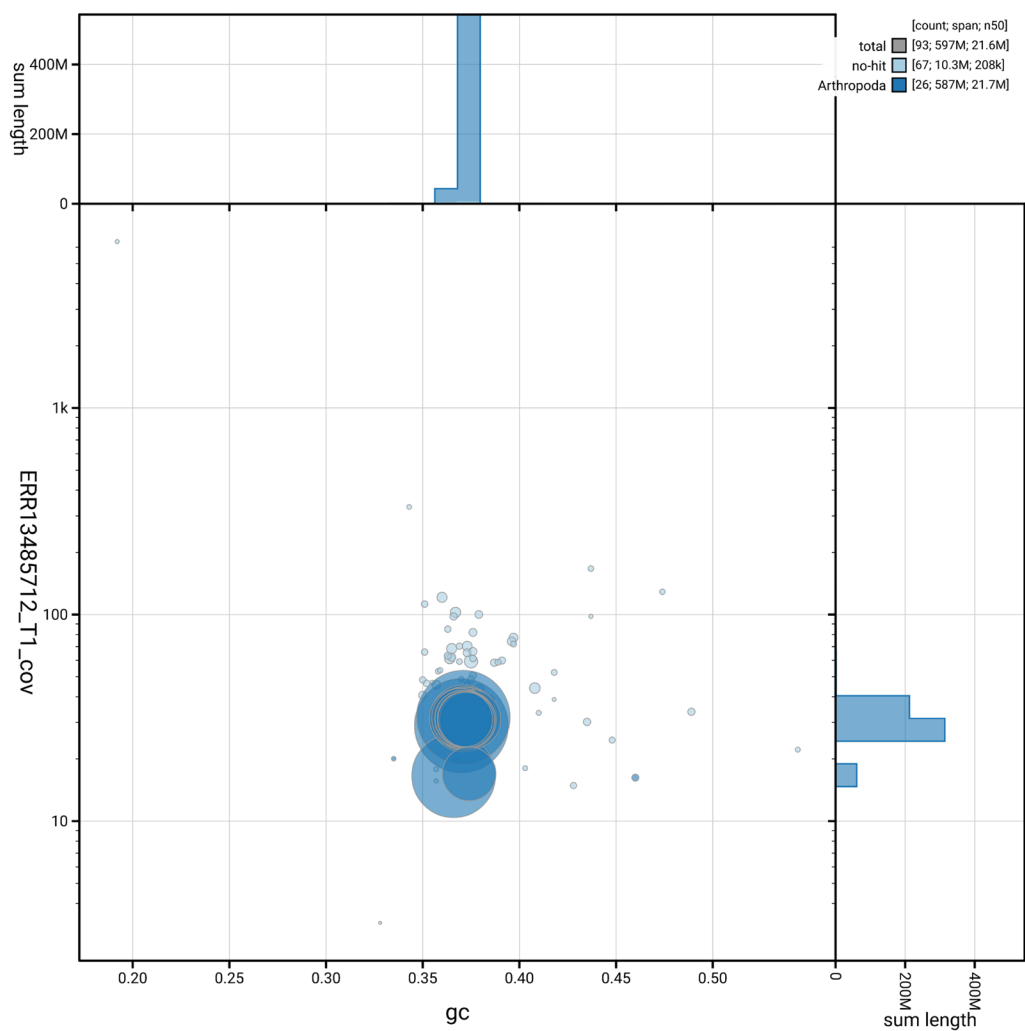


Figure 6. BlobToolKit GC-coverage plot for iEreMont1.hap1.1. Blob plot showing sequence coverage (vertical axis) and GC content (horizontal axis). The circles represent scaffolds, with the size proportional to scaffold length and the colour representing phylum membership. The histograms along the axes 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 *Erebia montana* assembly.

Measure	Value	Benchmark
EBP summary (haplotype 1)	6.7.Q65	6.C.Q40
Contig N50 length	8.87 Mb	≥ 1 Mb
Scaffold N50 length	21.59 Mb	= chromosome N50
Consensus quality (QV)	Haplotype 1: 65.3; haplotype 2: 64.5; combined: 64.9	≥ 40
k-mer completeness	Haplotype 1: 84.20%; Haplotype 2: 73.40%; combined: 97.35%	≥ 95%
BUSCO	C:98.6%[S:94.6%,D:4.0%], F:0.3%,M:1.1%,n:5 286	S > 90%; D < 5%
Percentage of assembly assigned to chromosomes	98.25%	≥ 90%

Data availability

European Nucleotide Archive: *Erebia montana* (marbled ringlet). Accession number [PRJEB78682](#). The genome sequence is released openly for reuse. The *Erebia montana* 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. The genome will be annotated using available RNA-Seq data and presented through [Ensembl](#) at the European Bioinformatics Institute. 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.

Table 5. Software versions and sources.

Software	Version	Source
BEDTools	2.30.0	https://github.com/arq5x/bedtools2
BLAST	2.14.0	ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/
BlobToolKit	4.3.9	https://github.com/blobtoolkit/blobtoolkit
BUSCO	5.5.0	https://gitlab.com/ezlab/busco
bwa-mem2	2.2.1	https://github.com/bwa-mem2/bwa-mem2
Cooler	0.8.11	https://github.com/open2c/cooler
DIAMOND	2.1.8	https://github.com/bbuchfink/diamond
fasta_windows	0.2.4	https://github.com/tolkit/fasta_windows
FastK	1.1	https://github.com/thegenemyers/FASTK
GenomeScope2.0	2.0.1	https://github.com/tbenavi1/genomescope2.0
Gfastats	1.3.6	https://github.com/vgl-hub/gfastats
Goat CLI	0.2.5	https://github.com/genomehubs/goat-cli
Hifiasm	0.19.8-r603	https://github.com/chhylp123/hifiasm
HiGlass	1.13.4	https://github.com/higlass/higlass
MerquryFK	1.1.2	https://github.com/thegenemyers/MERQURY.FK
Minimap2	2.24-r1122	https://github.com/lh3/minimap2
MitoHiFi	3	https://github.com/marcelauliano/MitoHiFi
MultiQC	1.14; 1.17 and 1.18	https://github.com/MultiQC/MultiQC
Nextflow	23.10.0	https://github.com/nextflow-io/nextflow
PretextSnapshot	N/A	https://github.com/sanger-tol/PretextSnapshot
PretextView	0.2.5	https://github.com/sanger-tol/PretextView
samtools	1.19.2	https://github.com/samtools/samtools
sanger-tol/ascc	0.1.0	https://github.com/sanger-tol/ascc
sanger-tol/blobtoolkit	0.6.0	https://github.com/sanger-tol/blobtoolkit
Seqtk	1.3	https://github.com/lh3/seqtk
Singularity	3.9.0	https://github.com/sylabs/singularity
TreeVal	1.2.0	https://github.com/sanger-tol/treeval
YaHS	1.2a.2	https://github.com/c-zhou/yahs

Author information

Contributors are listed at the following links:

- [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.](#)

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, Kumar S, Sotero-Caio C, *et al.*: **Genomes on a Tree (GoAT): a versatile, scalable search engine for genomic and sequencing project metadata across the eukaryotic Tree of Life [version 1; peer review: 2 approved].** *Wellcome Open Res.* 2023; **8**: 24. [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](#)
- Danecek P, Bonfield JK, Liddle J, *et al.*: **Twelve years of SAMtools and BCFtools.** *GigaScience.* 2021; **10**(2): giab008. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Ewels P, Magnusson M, Lundin S, *et al.*: **MultiQC: summarize analysis results for multiple tools and samples in a single report.** *Bioinformatics.* 2016; **32**(19): 3047–3048. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Ewels PA, Peltzer A, Fillinger S, *et al.*: **The nf-core framework for community-curated bioinformatics pipelines.** *Nat Biotechnol.* 2020; **38**(3): 276–278. [PubMed Abstract](#) | [Publisher Full Text](#)
- Formenti G, Abueg L, Brajuka A, *et al.*: **Gfastats: conversion, evaluation and manipulation of genome sequences using assembly graphs.** *Bioinformatics.* 2022; **38**(17): 4214–4216. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Grüning B, Dale R, Sjödin A, *et al.*: **Bioconda: sustainable and comprehensive software distribution for the life sciences.** *Nat Methods.* 2018; **15**(7): 475–476. [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): g1aa153. [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](#)
- Kleckova I, Konvicka M, Klecka J: **Thermoregulation and microhabitat use in mountain butterflies of the genus *Erebia*: importance of fine-scale habitat heterogeneity.** *J Therm Biol.* 2014; **41**: 50–58. [PubMed Abstract](#) | [Publisher 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](#)
- 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.*: **Merquy: reference-free quality, completeness, and phasing assessment for genome assemblies.** *Genome Biol.* 2020; **21**(1): 245. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Sonderegger P: **Die Erebieen Der Schweiz.** Brügg bei Biel: P. Sonderegger; 2005.
- Tolman T, Lewington R: **The most complete guide to the butterflies of Britain & Europe.** London: HarperCollins Publishers; 2008.
- Uliano-Silva M, Ferreira JGRN, Krashenninnikova 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 27 November 2025

<https://doi.org/10.21956/wellcomeopenres.27190.r138401>

© 2025 Wu L. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Li-Wei Wu 

Tunghai University, Taichung, Taichung City, Taiwan

This article provides genomic information for a female specimen of the nymphalid butterfly *Erebia montana*, and I believe it will serve as a valuable reference for subsequent research. Overall, the presentation of data is consistent with established genome-generation procedures. However, on page 3, within the section "Nucleic acid extraction," the phrase "a yield of 1 934.52 ng" appears to contain an unintended space between the digits "1" and "9." I am uncertain whether this spacing is purposeful or accidental, but it likely requires correction. In addition, this study reports two distinct genomic assemblies derived from the same female individual. I feel that further clarification of this aspect would be important. For example, whether this reflects potential chromosome-level chimerism, or whether it results from substantial heterozygosity between haplotypes. It may also be helpful to characterize the differences between the two genome haplotypes using the currently available assemblies. As the researchers who directly conducted the wet-lab assembly work, the authors are in the best position to clarify these points, and I believe that explicitly addressing this issue would greatly benefit researchers who may rely on this genome in future comparative or evolutionary studies.

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: butterfly conservation, phylogeography

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 25 November 2025

<https://doi.org/10.21956/wellcomeopenres.27190.r139222>

© 2025 Alqassar J. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Jasmine D Alqassar 

The George Washington University Washington, District of Columbia, USA

In this article Lucek et al. present a high-quality chromosome-level reference genome for *Erebia montana*. BUSCO analysis demonstrates the genome assembly is highly contiguous and complete (BUSCO score: C: 98.6%[S:94.6%,D:4%], F:0.3%, M:1.1%, n:5286). Additionally, this assembly is chromosome-level with 98.25% of the assembly assigned to chromosomes. This quality chromosome-level genome resource will serve as a valuable resource for the lepidopteran genomics community. The methods are well-explained and the genome resource is already accessible to the public. The inclusion of Merian element assignment will serve as an important resource for chromosome evolution studies.

The addition of an explanation for why scaffold OZ178013.1 is termed as a Z2 sex chromosome as well as a statement that compares this finding to existing *Erebiini* genomes will strengthen this article.

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 and evolutionary developmental biology

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 10 November 2025

<https://doi.org/10.21956/wellcomeopenres.27190.r138403>

© 2025 Mei Y. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Yang Mei 

¹ Zhejiang University, Hangzhou, China

² Jilin Agricultural University (Ringgold ID: 85112), Changchun, Jilin, China

The authors reported a high-quality genome of *Erebia montana* Berg-Mohrenfalter (Lepidoptera: Nymphalidae). But I have some questions:

1. Why is there a difference of 100MB between the two haplotypes? The difference is so significant
2. The Z chromosome is very conserved in Lepidoptera and is often the largest chromosome. Why was it not fully assembled in this study
3. Maybe providing the genome annotation is well

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

Yes

Are the protocols appropriate and is the work technically sound?

Yes

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

Yes

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

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: insect 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.
