



DATA NOTE

The genome sequence of the de Prunner's Ringlet, *Erebia triaria* von Prunner, 1798 (Lepidoptera: Nymphalidae)

[version 1; peer review: 3 approved]

Yannick Chittaro ¹, Marta Vila ², Kay Lucek ³, Charlotte J. Wright ⁴,
Joana I. Meier⁴, 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

¹Info fauna, Neuchâtel, Switzerland

²Universidade da Coruña, A Coruña, Galicia, Spain

³University of Neuchâtel, Neuchâtel, Switzerland

⁴Tree of Life, Wellcome Sanger Institute, Hinxton, England, UK

V1 First published: 11 Aug 2025, 10:425
<https://doi.org/10.12688/wellcomeopenres.24693.1>

Latest published: 11 Aug 2025, 10:425
<https://doi.org/10.12688/wellcomeopenres.24693.1>

Abstract

We present a genome assembly from a female specimen of *Erebia triaria* (de Prunner's Ringlet; Arthropoda; Insecta; Lepidoptera; Nymphalidae). The assembly contains two haplotypes with total lengths of 521.30 megabases and 412.03 megabases. Most of haplotype 1 (99.7%) is scaffolded into 17 chromosomal pseudomolecules, including the Z₁, Z₂, and W sex chromosomes. Haplotype 2 was assembled to scaffold level. The mitochondrial genome has also been assembled, with a length of 15.26 kilobases.

Keywords

Erebia triaria, de Prunner's Ringlet, genome sequence, chromosomal, Lepidoptera



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

Open Peer Review

Approval Status

	1	2	3
version 1			
11 Aug 2025	view	view	view

1. **Doga Cedden** , University of Göttingen, Göttingen, Germany
2. **Bin Zhang** , Qingdao Agricultural University, Qingdao, China
3. **Noah H Rose** , University of California San Diego, San Diego, 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: Chittaro Y: Investigation, Resources; Vila M: Writing – Original Draft Preparation; Lucek K: Resources, Supervision; 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 the Swiss National Science Foundation Grant PCEFP3_202869. MV was supported by Xunta de Galicia (ED431B 2024/23) and the University of A Coruña.

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

Copyright: © 2025 Chittaro Y *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: Chittaro Y, Vila M, Lucek K *et al.* **The genome sequence of the de Prunner's Ringlet, *Erebia triaria* von Prunner, 1798 (Lepidoptera: Nymphalidae) [version 1; peer review: 3 approved]** Wellcome Open Research 2025, 10:425 <https://doi.org/10.12688/wellcomeopenres.24693.1>

First published: 11 Aug 2025, 10:425 <https://doi.org/10.12688/wellcomeopenres.24693.1>

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 triaria* von Prunner, 1798 (NCBI:txid242256)

Background

De Prunner's Ringlet, *Erebia triaria*, is a European montane species, distributed from the Iberian Peninsula through the western and central Alps to the Balkans, where it is extremely localised (Micevski, 2014; Tolman & Lewington, 2008). The scientific name frequently appears as *Erebia triarius* in the literature (Wiemers *et al.*, 2018) as, for instance, in the phylogenomic study of the genus *Erebia* that confirmed that de Prunner's Ringlet was a member of the *medusa* clade and revealed that it is the sister species of the monophyletic group formed by *E. polaris*, *E. medusa* and *E. epipsodea* (Augustijnen *et al.*, 2024).

Erebia triaria inhabits roadside verges, clearings, and edges of open pine forests and oak woodlands, favouring warm, dry, steep, and rocky locations with generally southern exposures. Its typical habitat includes arid stony slopes or limestone crags, often exposed to full midday sun. Typical habitats are regions such as the Valais, a characteristic inner-Alpine dry valley, with rocky heaths with xerophilous and steppe vegetation (Nardelli, 1988; Sonderegger, 2005; Tolman & Lewington, 2008). It has a very wide elevation range, from 200 m to 2 500 m, with most of the records occurring between 1 000 and 1 500 m in Spain and between 600 and 1 600 m in Switzerland.

This medium-sized butterfly (Figure 1) has a forewing length of 21–25 mm and exhibits the typical earthy and muted colouration of the genus *Erebia*. The dorsal side of the wing is dark brown with a postdiscal band that is reddish-orange in males and orange in females. The forewings display four or five ocelli with pupils. On the hindwings, three or four dorsal ocelli are visible, situated on a discontinuous orange band. The ventral surface is dark. The ventral ocelli of the hindwings are barely visible in males, whereas in females the underside is lighter

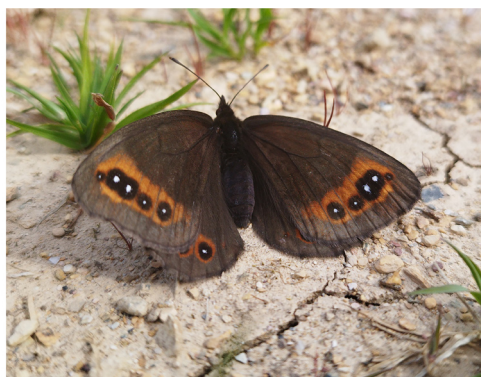


Figure 1. Photograph of *Erebia triaria* by Miguel Carballa (not the specimen used for genome sequencing).

and more contrasting. Final instar larvae are light green, with a brown head and dorsal and lateral black bands, the dorsal band being flanked by two white lines. The substantial geographical variability in the wing pattern of de Prunner's Ringlet has led to the description of at least 12 subspecies (reviewed by Vila, 2004). However, the only available phylogeographic study on this species focused on the genetic differentiation of the western-most one, *Erebia triaria pargapondalense*, whose allopatric isolation likely predates the last glaciation (Vila *et al.*, 2005).

Erebia triaria is a univoltine species, with adults mostly flying from May to the first half of July, depending on altitude. It may occur in sympatry with *E. meolans*, which typically flies during the summer season. This coexistence may lead to misidentification during late spring/early summer. Males of *E. triaria* are particularly active during their flight period, patrolling their territories in search of mates. The female lays eggs on various grasses, such as *Festuca* spp. and *Stipa pennata* L. (Clarke, 2024).

The species is classified as Least Concern (LC) at the European level (van Swaay *et al.*, 2010). However, it receives a higher level of protection in certain countries, being listed as Vulnerable (VU) in Switzerland, for example. Furthermore, specific lineages or populations may warrant particular attention. For instance, the aforementioned westernmost subspecies has been identified as a distinct conservation unit based on genetic analyses, as well as morphological and phenological observations (Vila *et al.*, 2006). Predictive models for the period 2061–2080 revealed that *E. triaria* will be one of the three species of *Erebia* likely to remain in Spain (almost entirely restricted to the Pyrenees and the Cantabrian Mountains), where the suitable areas for *E. triaria* will decrease by up to 77% (Romo *et al.*, 2023).

The first karyotypic analysis suggested *E. triaria* has an approximate haploid number of 16 chromosomes (de Lesse, 1961). As the chromosome numbers of the *medusa* clade range between 11 and 18, the chromosome-level reference genome of *E. triaria* provides an invaluable resource to unravel the chromosome evolution in *Erebia*, which has likely driven diversification in this highly speciose genus (Augustijnen *et al.*, 2024). Moreover, this reference genome provides the opportunity to resolve the cryptic diversity within *E. triaria*.

Methods

Sample acquisition

The specimen used for genome sequencing was an adult female *Erebia triaria* (specimen ID SAN20001744, ToLID iLEreTria2), collected from Hohtenn, Valais, Switzerland (latitude 46.33, longitude 7.763; elevation 1 443 m) on 06/06/2022. The specimen was collected and identified by Yannick Chittaro (Info Fauna, Neuchâtel, Switzerland).

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 iLEreTria2 sample was weighed and triaged to determine the appropriate extraction protocol. Tissue

from the whole organism was homogenised by [powermashing](#) using a PowerMasher II tissue disruptor.

HMW DNA was extracted 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 25.8 ng/μL and a yield of 1 496.40 ng, with a fragment size of 13.7 kb. The 260/280 spectrophotometric ratio was 1.92, and the 260/230 ratio was 2.21.

PacBio HiFi library preparation and sequencing

Library preparation and sequencing were performed at the WSI Scientific Operations core. Libraries were prepared using the SMRTbell Prep Kit 3.0 (Pacific Biosciences, California, USA), following the manufacturer’s instructions. The kit includes reagents for end repair/A-tailing, adapter ligation, post-ligation SMRTbell bead clean-up, and nuclease treatment. Size selection and clean-up were performed using diluted AMPure PB beads (Pacific Biosciences). DNA concentration was quantified using a Qubit Fluorometer v4.0 (ThermoFisher Scientific) and the Qubit 1X dsDNA HS assay kit. Final library fragment size was assessed with the Agilent Femto Pulse Automated Pulsed Field CE Instrument (Agilent Technologies) using the gDNA 55 kb BAC analysis kit.

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

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 whole organism tissue of the iEreTria2 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

Table 1. Specimen and sequencing data for BioProject.

Platform	PacBio HiFi	Hi-C
ToLID	iEreTria2	iEreTria2
Specimen ID	SAN20001744	SAN20001744
BioSample (source individual)	SAMEA114286221	SAMEA114286221
BioSample (tissue)	SAMEA114286282	SAMEA114286282
Tissue	whole organism	whole organism
Sequencing platform and model	Revio	Illumina NovaSeq X
Run accessions	ERR12954116	ERR12982554
Read count total	1.75 million	667.67 million
Base count total	19.49 Gb	100.82 Gb

(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 4 breaks and 205 joins. 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). 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 (Danecek *et al.*, 2021) to generate coverage tracks. Simultaneously, it queries the GoT 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 triaria* was sequenced using Pacific Biosciences single-molecule HiFi long reads, generating 19.49 Gb (gigabases) from 1.75 million reads, which were used to assemble the genome. GenomeScope2.0 analysis estimated the haploid genome size at 455.82 Mb, with a heterozygosity of 0.95% and repeat content of 28.80%. These estimates guided expectations for the assembly. Based on the estimated genome size, the sequencing data provided approximately 42× coverage. Hi-C sequencing produced 100.82 Gb from 667.67 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 521.30 Mb in 57 scaffolds, with 317 gaps, and a scaffold N50 of 33.97 Mb (Table 2).

Most of the assembly sequence (99.7%) was assigned to 17 chromosomal-level scaffolds, representing 14 autosomes and the Z₁, Z₂, and W sex chromosomes. Chromosomes Z and W

Table 2. Genome assembly statistics.

Assembly name	ilEreTria2.hap1.1	ilEreTria2.hap2.1
Assembly accession	GCA_964166005.1	GCA_964166405.1
Assembly level	chromosome	scaffold
Span (Mb)	521.30	412.03
Number of chromosomes	17	N/A
Number of contigs	374	149
Contig N50	5.93 Mb	5.95 Mb
Number of scaffolds	57	49
Scaffold N50	33.97 Mb	33.89 Mb
Longest scaffold length (Mb)	52.36	N/A
Sex chromosomes	W; Z1 and Z2	N/A
Organelles	Mitochondrial genome: 15.26 kb	N/A

were assigned based on read coverage and Hi-C data. Chromosome W is comprised of contigs of an 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).



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

Table 3. Chromosomal pseudomolecules in the haplotype 1 genome assembly of *Erebia triaria* iEreTria2.

INSDC accession	Molecule	Length (Mb)	GC%	Assigned Merian elements
OZ073195.1	1	43.41	37	M11;M13;M30
OZ073197.1	2	38.09	37	M14;M29;M8
OZ073198.1	3	35.54	37	M22;M3
OZ073199.1	4	33.99	37	M12;M4
OZ073200.1	5	33.97	37	M23;M9
OZ073201.1	6	33.39	37	M15;M5
OZ073202.1	7	31.18	37.50	M10;M27;M31

INSDC accession	Molecule	Length (Mb)	GC%	Assigned Merian elements
OZ073203.1	8	28.09	37	M28;M7
OZ073204.1	9	27.22	37	M21;M24
OZ073205.1	10	24.10	37	M25;M6
OZ073206.1	11	21.20	37.50	M19;M26
OZ073207.1	12	20.24	37	M1
OZ073208.1	13	20.17	37	M2
OZ073209.1	14	19.82	37	M17;M20
OZ073194.1	W	52.36	36	N/A
OZ073196.1	Z1	39.69	36.50	M18;MZ
OZ073210.1	Z2	17.29	37	M16

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 65.1, and for haplotype 2, 67.1. When the two haplotypes are combined, the assembly

achieves an estimated QV of 65.8. The *k*-mer completeness is 84.89% for haplotype 1, 74.12% for haplotype 2, and 99.33% for the combined haplotypes (Figure 4). BUSCO analysis using the lepidoptera_odb10 reference set (*n* = 5 286) (Kriventseva *et al.*, 2019) identified 98.4% of the expected gene set (single = 97.8%, duplicated = 0.5%) for haplotype 1. The snail plot in Figure 5 summarises the scaffold length distribution

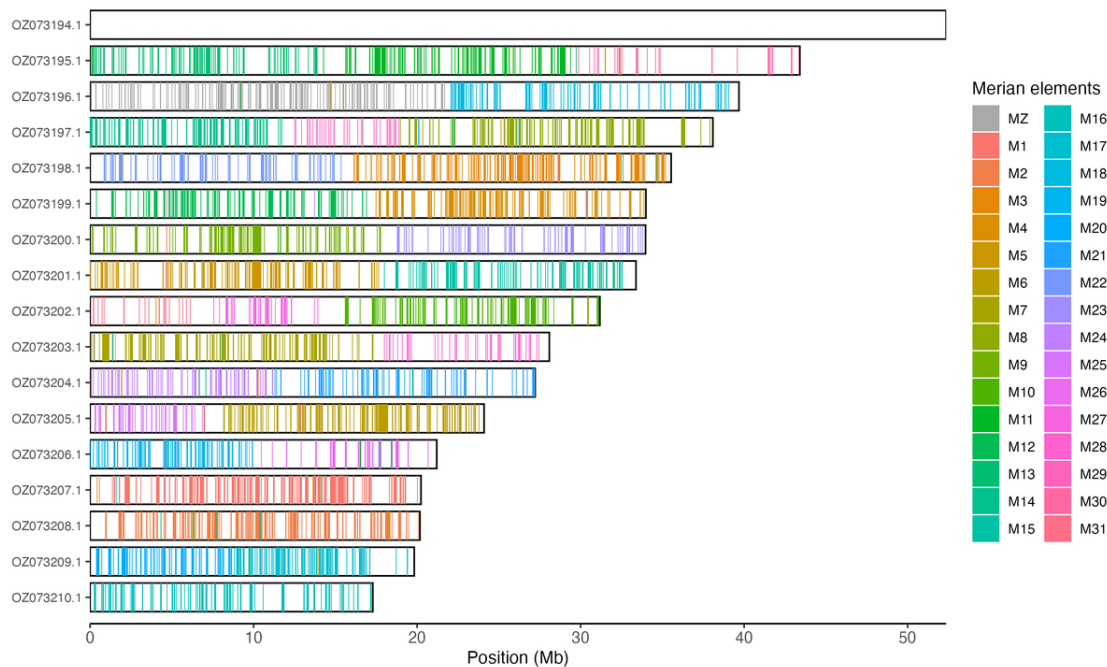


Figure 3. Merian elements painted across chromosomes in the iEreTria2.hap1.1 assembly of *Erebia triaria*. Chromosomes are drawn to scale, with the positions of orthologues shown as coloured bars. Each orthologue is coloured by the Merian element that it belongs to. All orthologues which could be assigned to Merian elements are shown.

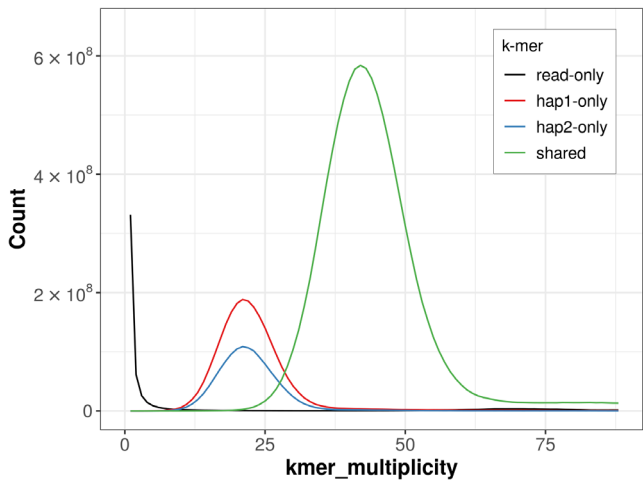


Figure 4. Evaluation of *k*-mer completeness using MerquyFK. This plot illustrates the recovery of *k*-mers from the original read data in the final assemblies. The horizontal axis represents *k*-mer multiplicity, and the vertical axis shows the number of *k*-mers. The black curve represents *k*-mers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

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 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

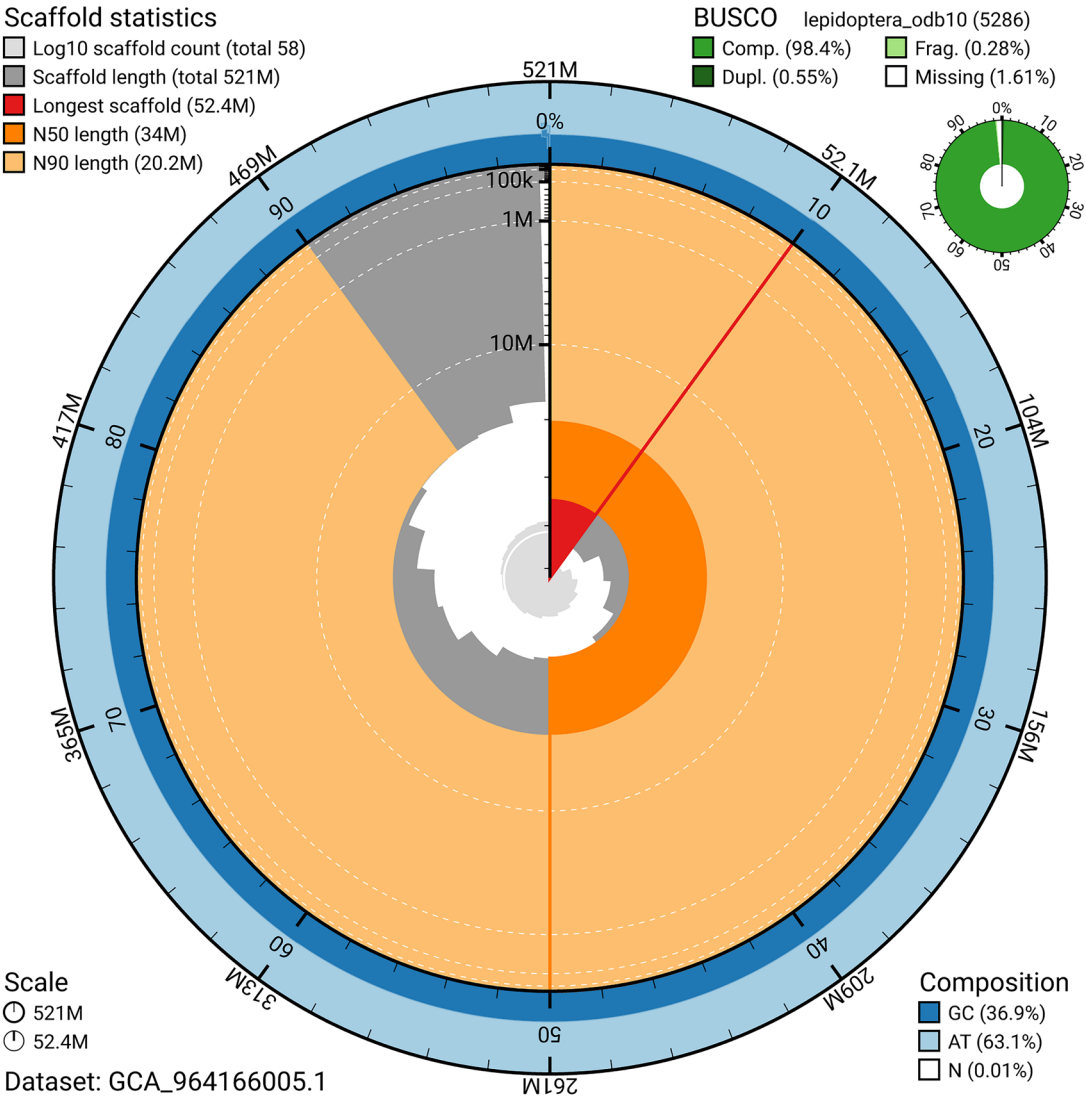


Figure 5. Assembly metrics for ilEreTria2.hap1.1. The BlobToolKit snail plot provides an overview of assembly metrics and BUSCO gene completeness. The circumference represents the length of the whole genome sequence, and the main plot is divided into 1,000 bins around the circumference. The outermost 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](#).

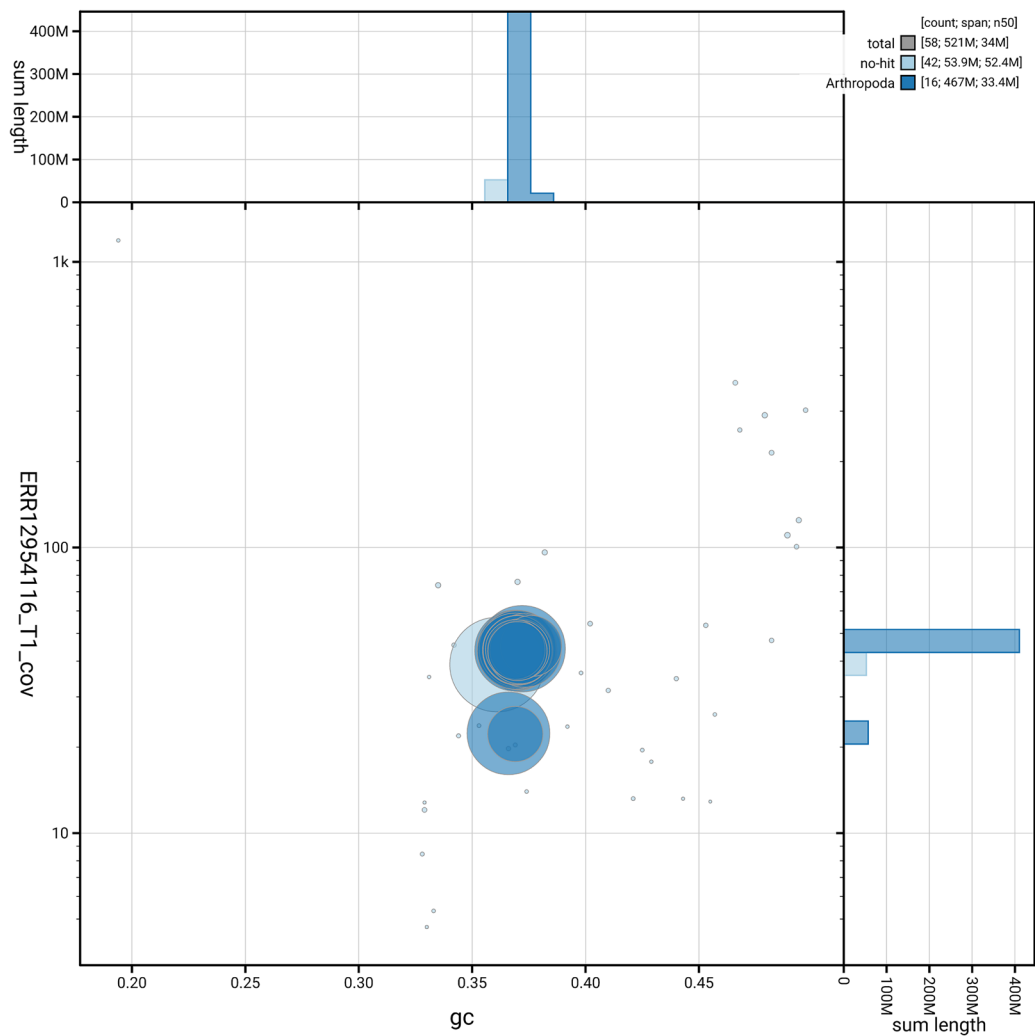


Figure 6. BlobToolKit GC-coverage plot for ilEreTria2.hap1.1. Blob plot showing sequence coverage (vertical axis) and GC content (horizontal axis). The circles represent scaffolds, with the size proportional to scaffold length and the colour representing phylum membership. The histograms along the axes 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 triaria* assembly.

Measure	Value	Benchmark
EBP summary (haplotype 1)	6.7.Q65	6.C.Q40
Contig N50 length	5.93 Mb	≥ 1 Mb
Scaffold N50 length	33.97 Mb	= chromosome N50
Consensus quality (QV)	Haplotype 1: 65.1; haplotype 2: 67.1; combined: 65.8	≥ 40
k-mer completeness	Haplotype 1: 84.89%; Haplotype 2: 74.12%; combined: 99.33%	≥ 95%
BUSCO	C:98.4%[S:97.8%,D:0.5%], F:0.3%,M:1.3%,n:5 286	S > 90%; D < 5%
Percentage of assembly assigned to chromosomes	0%	≥ 90%

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: *Erebia triaria* (de Prunner’s ringlet). Accession number [PRJEB75277](#). The genome sequence is released openly for reuse. The *Erebia triaria* 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.

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](#).

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
MerquyFK	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

Software	Version	Source
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

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](#)

Augustijn H, Batscher L, Cesanek M, et al.: **A macroevolutionary role for chromosomal fusion and fission in *Erebia* butterflies.** *Sci Adv.* 2024; **10**(16): ead10989.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free 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](#)

Clarke HE: **A checklist of European butterfly larval foodplants.** *Ecol Evol.* 2024; **14**(1): e10834.
[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](#)

de Lesse H: **Signification supraspécifique des formules chromosomiques chez les Lépidoptères.** *Bulletin de la Société entomologique de France.* 1961; **66**(3–4): 71–83.
[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](#)

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](#)

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, [Accessed 2 April 2024].
[Reference Source](#)

Micevski N: **De Prunner's Ringlet *Erebia triaria* (de Prunner, 1798) (Lepidoptera: Nymphalidae: Satyrinae) – a new species for the Republic of Macedonia.** *Natura Sloveniae.* 2014; **16**(1): 53–58.
[Publisher Full Text](#)

Nardelli U: ***Erebia triaria* De Prunner 1798 (Lepidoptera, Satyridae). Biologie und ökologie der population der Brenta-Gruppe in Italien.** *Entomologische Zeitschrift, Frankfurt am Main.* 1988; **98**(21): 305–12.
[Reference Source](#)

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

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

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

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

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

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

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Romo H, García-Barros E, Wilson RJ, *et al.*: **Modelling the scope to conserve an endemic-rich mountain butterfly taxon in a changing climate.** *Insect Conserv and Diver*. 2023; **16**(4): 451–467.

[Publisher Full Text](#)

Sonderegger P: **Die Erebie der Schweiz.** Biel/Bienne: Verlag Peter Sonderegger, 2005.

[Publisher Full Text](#)

Tolman T, Lewington R: **Field guide to the butterflies of Britain and Europe.** HarperCollins Publishers Ltd, 2008.

Uliano-Silva M, Ferreira JGRN, Krashennikova 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](#)

van Swaay C, Wynhoff I, Verovnik R, *et al.*: **Erebia triaria.** 2010.

[Publisher 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](#)

Vila M: **Effects of environmental changes during the quaternary on the distribution of extant species: the case of the butterflies *Erebia triaria* and *E. palarica*.** PhD thesis, Universidade da Coruña, 2004.

[Reference Source](#)

Vila M, Lundhagen AC, Thuman KA, *et al.*: **A new conservation unit in the butterfly *Erebia triaria* (Nymphalidae) as revealed by nuclear and mitochondrial markers.** *Ann Zool Fenn*. 2006; **43**(1): 72–79.

[Reference Source](#)

Vila M, Vidal-Romaní JR, Björklund M: **The importance of time scale and multiple refugia: incipient speciation and admixture of lineages in the butterfly *Erebia triaria* (Nymphalidae).** *Mol Phylogenet Evol*. 2005; **36**(2): 249–60.

[PubMed Abstract](#) | [Publisher Full Text](#)

Wiemers M, Balletto E, Dincă V, *et al.*: **An updated checklist of the European butterflies (Lepidoptera, Papilionoidea).** *ZooKeys*. 2018; (811): 9–45.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free 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 11 September 2025

<https://doi.org/10.21956/wellcomeopenres.27209.r129637>

© 2025 Rose N. 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.



Noah H Rose 

University of California San Diego, San Diego, California, USA

The present report describes the assembly of a reference genome for *Erebia triaria*. The authors do a good job providing context for the reference, including alternative names that appear in the literature, noting possibly distinct or vulnerable populations, and highlight the potential for interesting cryptic diversity in the genus. The methods are appropriate, the text is clear, and the assembly appears to be high quality.

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: Ecological genomics, evolutionary genetics, vector 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 September 2025

<https://doi.org/10.21956/wellcomeopenres.27209.r129639>

© 2025 Zhang B. 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.



Bin Zhang 

Qingdao Agricultural University, Qingdao, China

This manuscript presents a high-quality reference genome assembly of *Erebia triaria* (de Prunner's Ringlet), including chromosome-level assemblies of two haplotypes and the mitochondrial genome. The study is clearly structured, the methods are transparent and replicable, and the raw data and assemblies are openly available. The assembly quality metrics (BUSCO 98.4%, k-mer completeness 99.33%) indicate that the genome is highly reliable and of considerable value for comparative genomics, evolutionary biology, and conservation studies.

Points for Improvement.

1. While quality metrics are well reported, there is limited scientific interpretation. For example, it would strengthen the manuscript to briefly comment on potential chromosome rearrangements, or to compare the assembly with existing *Erebia* genomes.
2. Figures such as the Hi-C contact map, Merian element painting, and BUSCO snail plot are useful but under-discussed. A short description of what the plots reveal about genome structure would improve clarity.
3. Some sentences in the Background section are overly descriptive (e.g., habitat details) and could be streamlined.

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: Entomology, Bioinformatics

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

Reviewer Report 23 August 2025

<https://doi.org/10.21956/wellcomeopenres.27209.r129631>

© 2025 Cedden D. 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.



Doga Cedden 

University of Göttingen, Göttingen, Germany

Chittaro et al. report a high-quality genome assembly from *Erebia triaria*. The background information regarding the lepidopteran is adequate. Sufficient quality control of the genome assembly is provided with standard methodology. The assembly has a BUSCO score of 98.4% and *k*-mer completeness of 99.33% for the combined haplotypes. The raw data and genome are publicly available via the provided accession number. I have no further comments.

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: I work on coleopteran pests using RNAi and RNA-seq.

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.