Check for updates

GENOME NOTE

The genome sequence of the Blind Ringlet, *Erebia pharte* (Hubner, 1803-1804) (Lepidoptera: Nymphalidae)

[version 1; peer review: 2 approved]

Kay Lucek[1], Irena Klečková[2], Charlotte J. Wright[3], Joana I. Meier[3], 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

[1]University of Neuchâtel, Neuchâtel, Neuchâtel, Switzerland
[2]Biology Centre Czech Academy of Sciences, České Budějovice, South Bohemia, Czech Republic
[3]Tree of Life, Wellcome Sanger Institute, Hinxton, England, UK

v1 **First published:** 04 Sep 2026, **11**:592
https://doi.org/10.12688/wellcomeopenres.27705.1
Latest published: 04 Sep 2026, **11**:592
https://doi.org/10.12688/wellcomeopenres.27705.1

Abstract

We present a genome assembly from a female specimen of *Erebia pharte* (Blind Ringlet; Arthropoda; Insecta; Lepidoptera; Nymphalidae). The genome sequence has a total length of 616.83 megabases. Most of the assembly (98.95%) is scaffolded into 20 chromosomal pseudomolecules, including the Z sex chromosome. The mitochondrial genome has also been assembled, with a length of 15.21 kilobases. Gene annotation of this assembly by Ensembl identified 15,135 protein-coding genes. This work is part of Project Psyche, a collaborative programme generating genomes for European butterflies and moths.

Keywords

Erebia pharte, Blind Ringlet, genome sequence, chromosomal, Lepidoptera

This article is included in the Tree of Life gateway.

Open Peer Review

Approval Status ✓ ✓

	1	2
version 1 04 Sep 2026	✓ view	✓ view

1. **Automated Benchmarking**, Wellcome Sanger Institute, Hinxton, UK
2. **Ricardo Rodriguez-de-la-vega**, Université Paris-Saclay, Gif-sur-Yvette, France

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

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

Author roles: Lucek K: Investigation, Resources, Supervision, Writing – Review & Editing; **Klečková I**: Supervision, Writing – Original Draft Preparation, 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]. KL received funding from the Swiss National Science Foundation (SNSF) Grant ID 202869. IK was funded by the 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: © 2026 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 the Blind Ringlet, *Erebia pharte* (Hubner, 1803-1804) (Lepidoptera: Nymphalidae) [version 1; peer review: 2 approved]** Wellcome Open Research 2026, **11**:592 https://doi.org/10.12688/wellcomeopenres.27705.1

First published: 04 Sep 2026, **11**:592 https://doi.org/10.12688/wellcomeopenres.27705.1

Species taxonomy

Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Protostomia; Ecdysozoa; Panarthropoda; Arthropoda; Mandibulata; Pancrustacea; Altocrustacea; Allotriocarida; Hexapoda; Insecta; Dicondylia; Pterygota; Neoptera; Eumetabola; Endopterygota; Aparaglossata; Panorpida; Amphiesmenoptera; Lepidoptera; Glossata; Neolepidoptera; Heteroneura; Ditrysia; Obtectomera; Papilionoidea; Nymphalidae; Satyrinae; Erebiini; *Erebia*; *Erebia pharte* (Hubner, 1803-1804) (NCBI:txid111906)

Background

Erebia pharte, the Blind Ringlet, is a small, brownish *Erebia* species with orange spots on the upper and lower sides of the wings. The orange spots lack central black dots, which is why the species is called the Blind Ringlet. The species may resemble individuals of *Erebia melampus*, which can also lack black centres in the ocelli (Sonderegger, 2005). *Erebia pharte* differs from *E. manto* and *E. eriphyle* in the shape of the orange ocelli.

Erebia pharte is distributed in the Alps and the Slovak Carpathians (Tatra Mountains) (Panigaj *et al.*, 2015; Tolman & Lewington, 2009). The species is also reported from Romania (see Sonderegger, 2005), but probably does not occur there (Cuvelier & Dincă, 2007). Similarly, the species has been reported from the Vosges (Tolman & Lewington, 2009), but it probably does not occur there either (Arnscheid & Roos, 1981; Sonderegger, 2005). *Erebia pharte* belongs to the *Erebia epiphron* species group and is related to *Erebia epiphron*, *E. orientalis*, *E. flavofasciata* and *E. melampus* (Augustijnen *et al.*, 2024; Klečková *et al.*, 2023).

Larvae feed on grasses and sedges (Poaceae and Cyperaceae), including *Carex ferruginea*, *C. flacca*, *Festuca ovina*, *F. quadriflora*, *F. rubra* and *Nardus stricta* (Clarke, 2024). The habitat of the species includes mountain meadows and pastures, as well as subalpine and alpine meadows with taller grasses. The species has a two-year life cycle with four larval stages, and it overwinters in the first and third larval instars (Sonderegger, 2005). *Erebia pharte* is currently assessed as Least Concern by the IUCN, with a stable population trend (van Swaay *et al.*, 2025).

We present a chromosome-level, haplotype-resolved genome sequence of the Blind Ringlet, *Erebia pharte*, sequenced as part of Project Psyche (Wright *et al.*, 2025). The sequence data were derived from a specimen (Figure 1) collected from Gadmen, Bern, Switzerland.



Figure 1. Voucher photograph of the *Erebia pharte* (ilErePhar2) specimen used for genome sequencing.

Methods

Sample acquisition

The specimen used for genome sequencing was an adult female *Erebia pharte* (specimen ID SAN20001732, ToLID ilErePhar2; Figure 1), collected from Gadmen, Bern, Switzerland (latitude 46.7317, longitude 8.4244; elevation 1,850 m) on 2022-07-07. 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). For HMW DNA extraction, tissue from the whole organism of ilErePhar2 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 manual 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 Femto Pulse system. For this sample, the final post-shearing DNA had a Qubit concentration of 24.2 ng/μL and a yield of 1,403.60 ng, with a fragment size of 13.4 kb. The 260/280 spectrophotometric ratio was 2.01, and the 260/230 ratio was 2.47. The Genomic Quality Number (GQN) was 5.6.

PacBio HiFi library preparation and sequencing

Library preparation and sequencing took place in the WSI Scientific Operations core.

Library 6771STDY14439721 was prepared with the SMRTbell Prep Kit 3.0 (Pacific Biosciences) 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 with a Qubit Fluorometer v4.0 (Thermo Fisher 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.

Library 6771STDY14439721 was sequenced on a Revio instrument (Pacific Biosciences, California, USA). The prepared libraries selected for multiplexing were pooled based on genome size, library molarity and the intended plex level. Primers were annealed and polymerases were bound to generate circularised complexes following the manufacturer's instructions. Complexes were purified using SMRTbell beads, diluted to the Revio loading concentration, and spiked with a Revio sequencing internal control. The pooled libraries were sequenced on a Revio 25M Plex SMRT cell in a 4-plex run. SMRT Link software (Pacific Biosciences) was used to configure and monitor the run and to perform primary and secondary data analysis.

Hi-C

Sample preparation and crosslinking

Hi-C data were generated from the ilErePhar2 whole organism 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. Libraries were amplified with 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. Libraries were sequenced on the Illumina NovaSeq X, generating paired-end 150-bp reads.

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. The *k*-mer frequency distribution was analysed with GenomeScope (version 2.1.0) (Ranallo-Benavidez *et al.*, 2020), providing fitted point estimates of genome size, heterozygosity and repeat content.

The HiFi reads were assembled using Hifiasm (Cheng *et al.*, 2021) with the --primary option. Haplotypic duplications were identified and removed using purge_dups (Guan *et al.*, 2020). The Hi-C reads (Rao *et al.*, 2014) were mapped to contigs using bwa-mem2 (Vasimuddin *et al.*, 2019), and the contigs were scaffolded in YaHS (Zhou *et al.*, 2023) with the --break option for handling potential misassemblies. The scaffolded assemblies were evaluated using Gfastats (Formenti *et al.*, 2022), BUSCO (Manni *et al.*, 2021) and MerquryFK (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, using the mitochondrial genome of *Minois paupera* (NC_085606.1) as the reference.

Assembly curation

The assembly was decontaminated using the Assembly Screen for Cobionts and Contaminants (ASCC) pipeline. The flat files and maps for curation were generated with TreeVal. 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 21 breaks and 30 joins, which reduced the scaffold count by 16.5%. The curation process is described at sanger-tol/curation-resources. A Hi-C contact map of the primary assembly was generated with PretextSnapshot.

Assembly quality assessment

The MerquryFK 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 the primary and alternate haplotypes using the *k*-mer databases ($k = 31$) computed prior to genome assembly. The analysis outputs included assembly QV scores and completeness statistics.

The primary assembly was assessed using BlobToolKit (Challis *et al.*, 2020). PacBio reads were mapped to the primary assembly to derive sequence coverage, and BUSCO (Manni *et al.*, 2021) was used to assess gene-set completeness. DIAMOND BLASTp searches of extracted BUSCO proteins and DIAMOND BLASTx searches of assembly sequences against UniProt Reference Proteomes (Buchfink *et al.*, 2021), followed by BLASTn searches of sequences without DIAMOND BLASTx hits against the NCBI nt database (Altschul *et al.*, 1990), supported taxonomic assignment. Coverage, sequence composition, taxonomic assignment and BUSCO results were integrated into a BlobDir for interactive inspection and visualisation. The snail and blob plots were rendered locally from this hosted BlobDir using BlobTk v0.8.3. The scaffold-length radial scale function for the snail plot was set to linear (Challis & Blaxter, 2026). The blob plot used circular markers with area proportional to scaffold length (Challis *et al.*, 2020).

We used *busco-alg-painter* v0.1.0, a Python reimplementation and extension of *lep_busco_painter*, to paint Merian elements along chromosomes. Merian elements represent the 32 ancestral linkage groups reconstructed for Lepidoptera (Wright *et al.*, 2024). The BUSCO-to-Merian reference assignments were derived from *lep_busco_painter*. Complete single-copy and duplicated BUSCO genes from the lepidoptera_odb10 results were assigned to Merian elements. Chromosome identifiers and lengths were obtained from NCBI Datasets, and BUSCO positions were plotted on chromosomes drawn to scale.

Genome sequence report

Sequence data

PacBio sequencing of the *E. pharte* specimen generated 21.27 Gb (gigabases) from 2.48 million reads. We assembled the genome from these reads. GenomeScope analysis using a $p = 2$ model gave a *k*-mer-based haploid genome size estimate of 594.20 Mb, with heterozygosity of 0.84% and repeat content of 44.11%; the full-model fit was 98.55% (Figure 2). The GenomeScope estimates guided expectations for the assembly. The sequencing data provided approximately 35× coverage of the primary assembly. Hi-C

sequencing produced 91.16 Gb from 301.87 million reads, which were used to scaffold the primary assembly. Table 1 summarises the specimen and sequencing details.

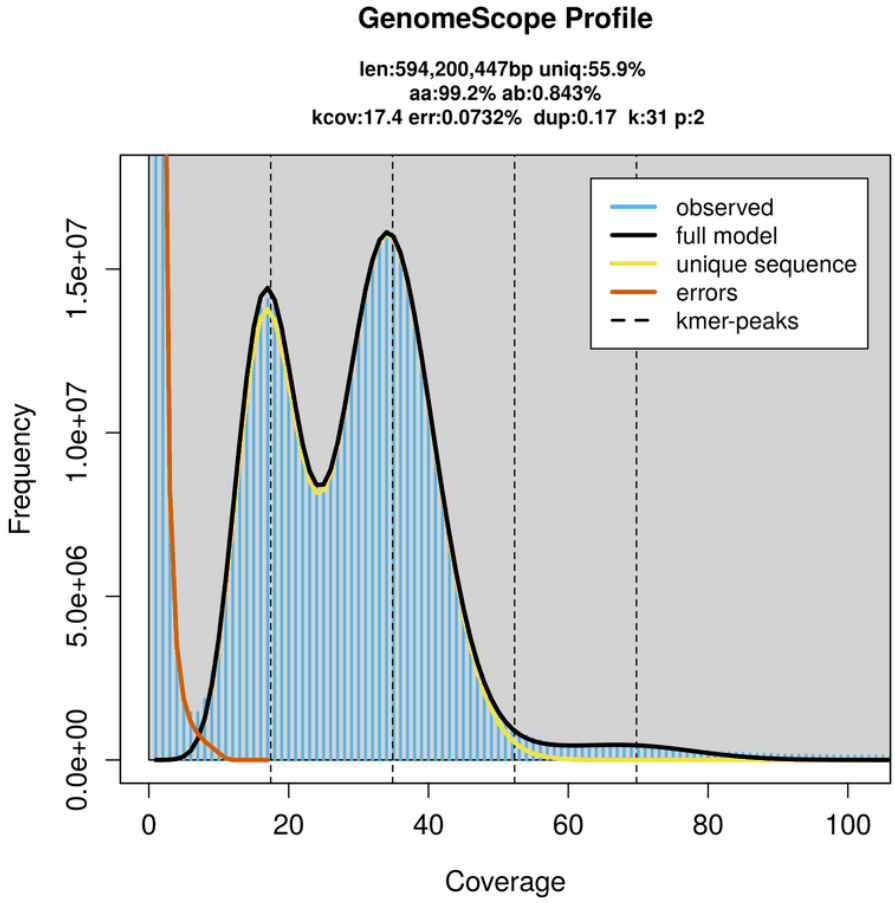


Figure 2. GenomeScope model fitted to the *k*-mer frequency distribution. The plot shows the observed and modelled *k*-mer spectra used to estimate genome size, heterozygosity and repeat content from unassembled sequencing reads.

Table 1. Specimen and sequencing data for *E. pharte* (BioProject PRJEB75866)

Platform	PacBio HiFi	Hi-C
ToLID	ilErePhar2	ilErePhar2
Specimen ID	SAN20001732	SAN20001732
BioSample (source individual)	SAMEA114286227	SAMEA114286227
BioSample (tissue)	SAMEA114286288	SAMEA114286288
Tissue	whole organism	whole organism
Instrument	Revio (1 run)	Illumina NovaSeq X
Library ID & run accession(s)	6771STDY14439721: ERR13158709	ERR13148270
Read count total	2.48 million reads	301.87 million read pairs
Base count total	21.27 Gb	91.16 Gb

Assembly statistics

The primary assembly was generated and an alternate haplotype assembly was also deposited at scaffold level in INSDC databases. The primary assembly is proposed as the reference assembly for this species. It has a total length of 616.83 Mb in 70 scaffolds (excluding organelle sequences), with 134 gaps, and a scaffold N50 of 31.3 Mb (Table 2). The primary assembly had a scaffold auN of 34.0 Mb.

Table 2. Genome assembly statistics for *E. pharte*

Genome assembly	Primary assembly	Alternate haplotype assembly
Assembly name	ilErePhar2.1	ilErePhar2.1 alternate haplotype
Assembly accession	GCA_965112705.1	GCA_965112695.1
Assembly level	chromosome	scaffold
Span (Mb)	616.83	582.07
Number of chromosomes	20	-
Number of contigs	204	3,036
Contig N50 (Mb)	8.12	0.64
Number of scaffolds (excluding organelle sequences)	70	3,008
Scaffold N50 (Mb)	31.3	0.65
Scaffold auN (Mb)	34.0	-
Longest scaffold length (Mb)	52.41	3.66
Sex chromosomes	Z	-
Organelles	Mitochondrial genome: 15.21 kb	-

Most of the primary assembly sequence (98.95%) was assigned to 20 chromosome-level scaffolds, representing 19 autosomes and the Z sex chromosome. These chromosome-level scaffolds, confirmed by Hi-C data, are named according to size (Figure 3; Table 3). Chromosome painting of the primary assembly with Merian elements illustrates the distribution of orthologues along chromosomes and highlights patterns of chromosomal evolution relative to Lepidopteran ancestral linkage groups (Figure 4). The Z chromosome of the primary assembly was assigned based on the presence of the ancestral Merian element MZ (Wright *et al.*, 2024).

The mitochondrial genome was also assembled (length 15.21 kb, OZ218833.1). This sequence is included as a contig in the multifasta file of the primary genome submission and as a standalone record.

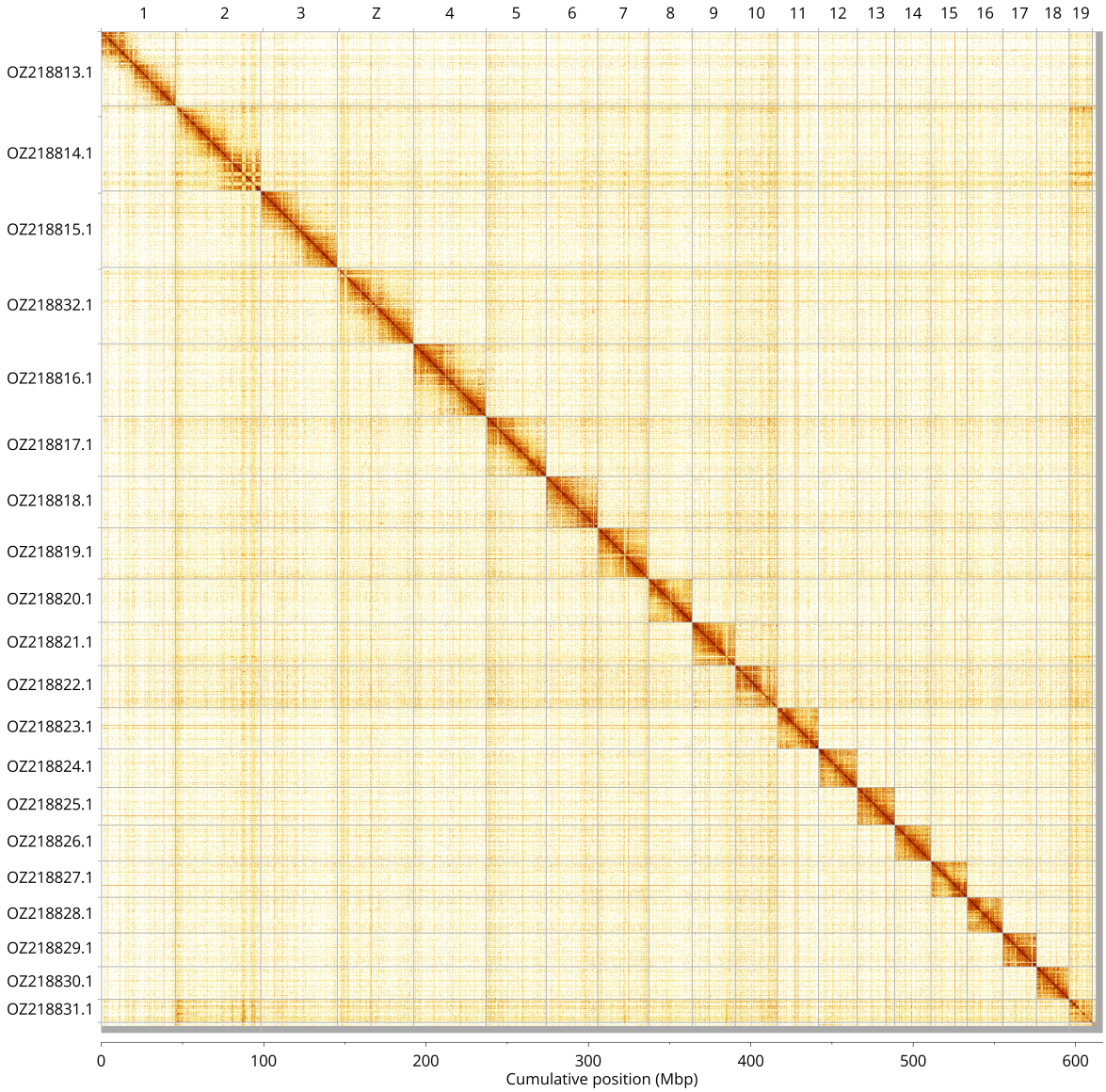


Figure 3. Hi-C contact map of the *E. pharte* primary assembly. Assembled chromosomes are shown in order of size and labelled along the axes. The plot was generated using PretextSnapshot.

Table 3. Chromosomal pseudomolecules in the primary genome assembly of *E. pharte* ilErePhar2

INSDC accession	Molecule	Length (Mb)	GC%	Assigned Merian elements
OZ218813.1	1	52.49	37.50	M2; M23; M30
OZ218814.1	2	47.12	37.50	M14; M16; M29
OZ218815.1	3	46.90	37.50	M10; M12
OZ218816.1	4	44.69	37.50	M21; M8
OZ218817.1	5	36.96	37.50	M17; M20; M25
OZ218818.1	6	31.79	37.50	M22; M27

INSDC accession	Molecule	Length (Mb)	GC%	Assigned Merian elements
OZ218819.1	7	31.30	37.50	M13; M28
OZ218820.1	8	26.80	37.50	M1
OZ218821.1	9	26.54	37.50	M19; M26
OZ218822.1	10	26.04	37.50	M24; M31
OZ218823.1	11	25.31	37.50	M3
OZ218824.1	12	23.82	37.50	M5
OZ218825.1	13	23.07	37	M9
OZ218826.1	14	22.28	37.50	M7
OZ218827.1	15	22.24	37.50	M4
OZ218828.1	16	21.95	37.50	M6
OZ218829.1	17	20.89	38	M11
OZ218830.1	18	20.04	37.50	M15
OZ218831.1	19	14.23	38	M30
OZ218832.1	Z	45.88	37	M18; MZ

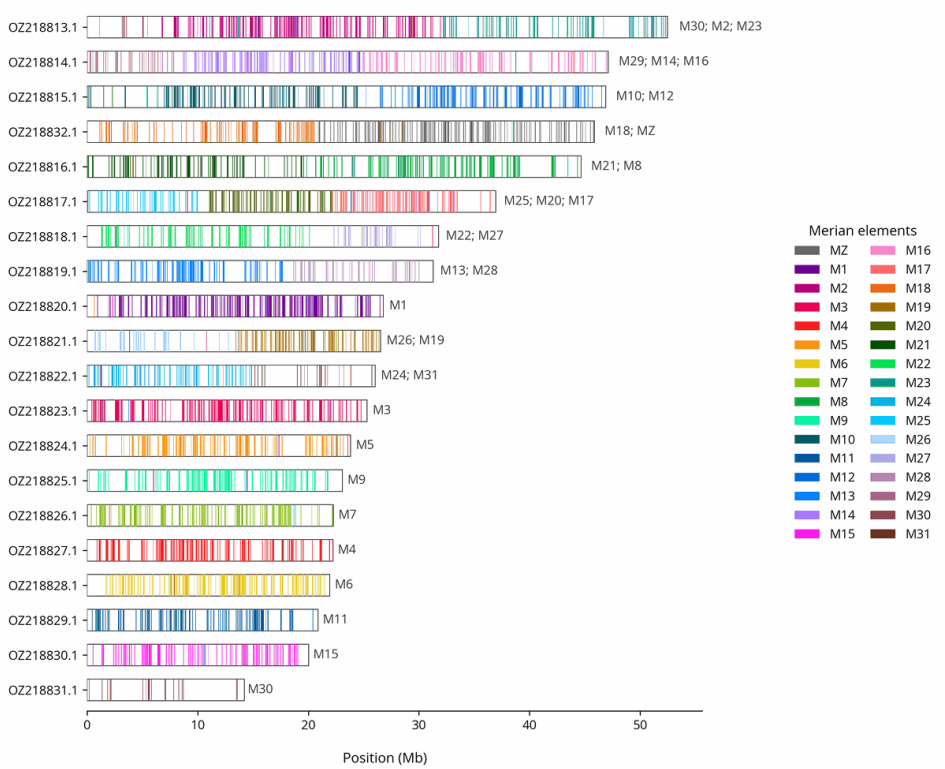


Figure 4. Merian elements painted across chromosomes in the ilErePhar2.1 primary assembly of *E. pharte*. Chromosomes are drawn to scale, with the positions of orthologues shown as coloured bars. Each orthologue is coloured by the Merian element that it belongs to. All orthologues which could be assigned to Merian elements are shown.

Assembly quality metrics

The combined primary and alternate assemblies achieve an estimated QV of 63.0. The *k*-mer completeness is 85.24% for the primary assembly, 75.25% for the alternate haplotype, and 98.78% for the combined assemblies (Figure 5).

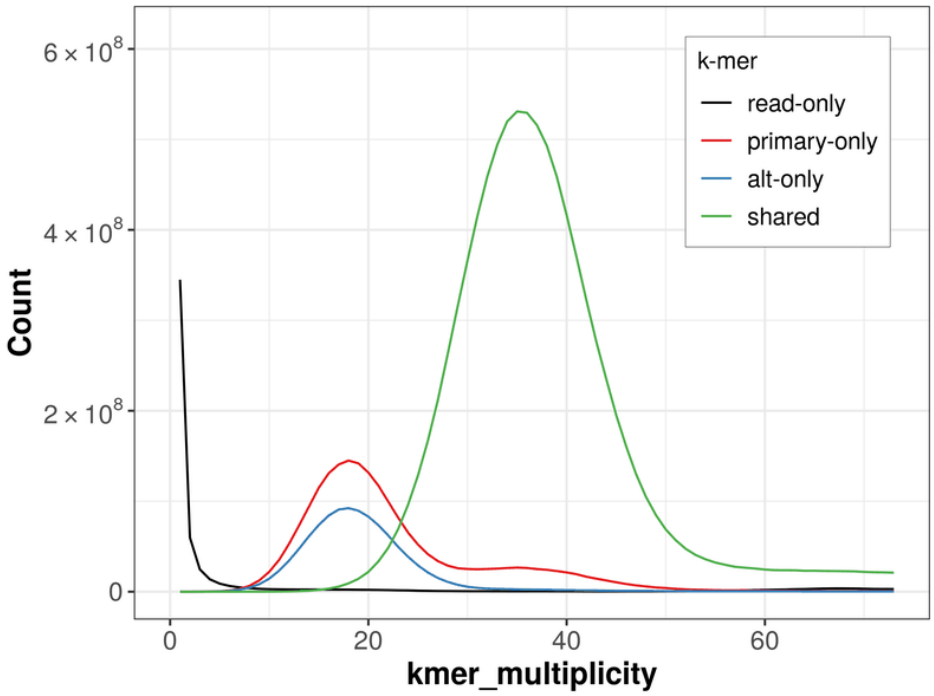


Figure 5. Evaluation of *k*-mer completeness using MerquryFK. This plot illustrates the recovery of *k*-mers from the original read data in the primary assembly, the alternate haplotype assembly and the combined assemblies. The horizontal axis represents *k*-mer multiplicity, and the vertical axis shows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

BUSCO v.5.7.1 analysis using the lepidoptera_odb10 reference set (n = 5,286) (Kriventseva *et al.*, 2019) identified 98.5% of the expected gene set (single = 97.6%, duplicated = 0.9%).

The snail plot in Figure 6 provides a summary of assembly contiguity, cumulative scaffold count, sequence composition and BUSCO completeness for the primary assembly (Challis & Blaxter, 2026). The relative sizes of the longest scaffold, scaffold N50 and N90, together with the dark-grey scaffold-length distribution, show how much of the assembly is contained in long scaffolds. The outer composition track shows variation in GC, AT and N content, while the central track shows the cumulative scaffold count and the BUSCO inset summarises gene-set completeness.

The blob plot in Figure 7 displays scaffolds from the primary assembly by GC proportion and base coverage. Marker area is proportional to scaffold length, and colour indicates taxonomic assignment (Challis *et al.*, 2020).

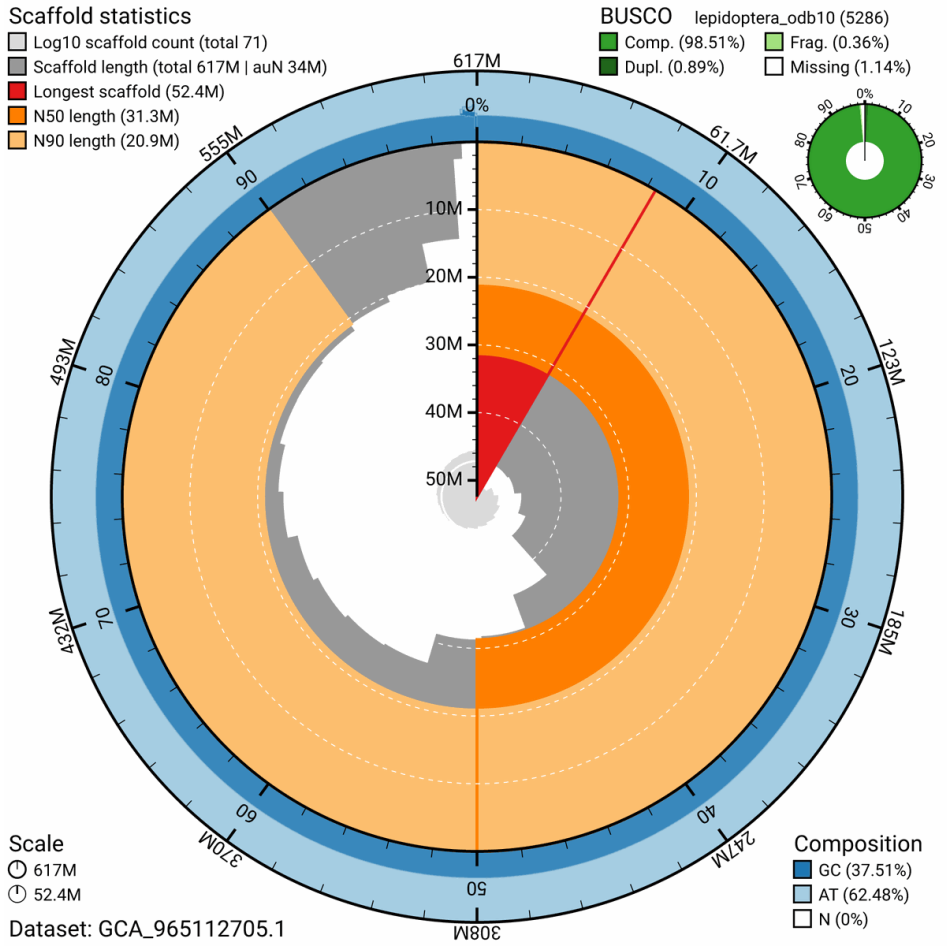


Figure 6. Assembly metrics for ilErePhar2.1, the primary assembly. The BlobToolKit snail plot provides an overview of assembly metrics and BUSCO gene completeness. The circumference represents the length of the primary assembly, and the main plot is divided into 1,000 bins around the circumference. The 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. Scaffold length is shown on a linear radial scale, scaled to the longest scaffold. Red shading indicates the longest scaffold, and the deeper orange and pale orange shading represent the N50 and N90 lengths. The summary statistics include scaffold auN, the area under the scaffold Nx curve. The 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 lepidoptera_odb10 set is presented at the top right. The assembly can be explored interactively in the BlobToolKit viewer.

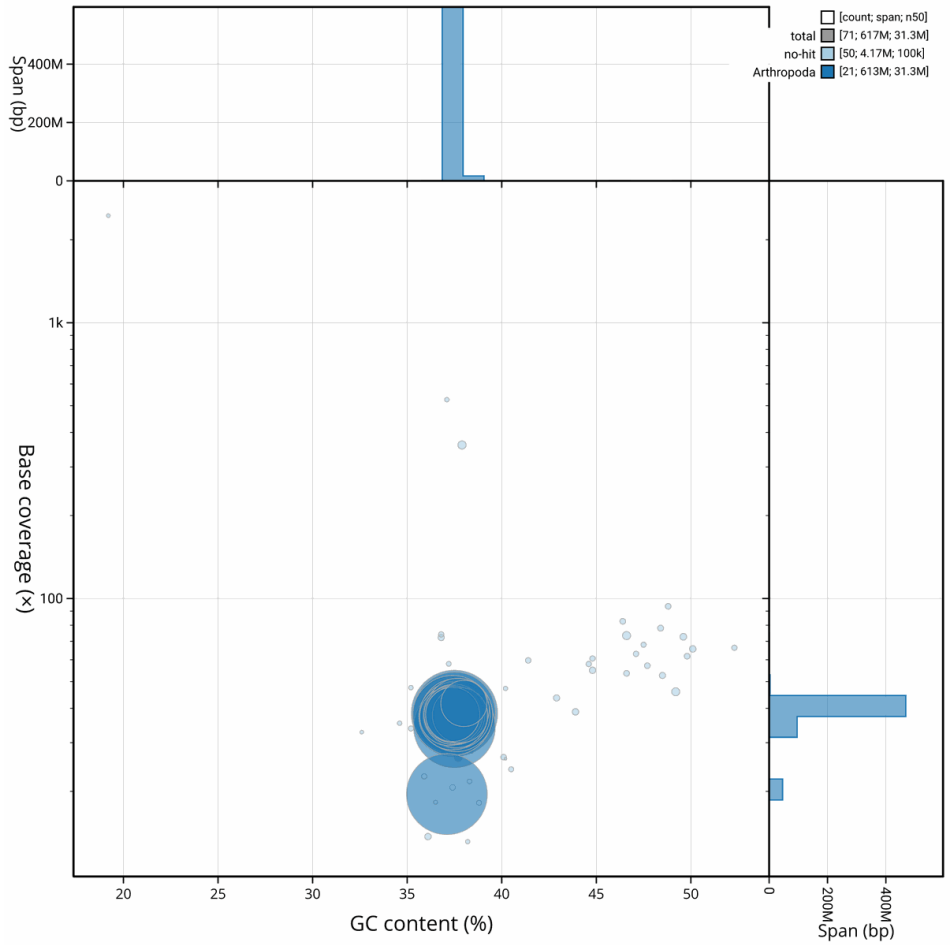


Figure 7. BlobToolKit blob plot for ilErePhar2.1. The plot shows base coverage (vertical axis) and GC content (horizontal axis). Circles represent scaffolds; marker area is proportional to scaffold length, and colour indicates taxonomic assignment. The histograms along the axes display the total length of sequences distributed across different levels of coverage and GC content. The assembly can be explored interactively in the BlobToolKit viewer.

Table 4 lists the assembly metric benchmarks adapted from Rhie *et al.*. (2021) the Earth BioGenome Project (EBP) Report on Assembly Standards. The three-part EBP metric, calculated for the primary assembly, is **6.C.Q64**.

Table 4. Earth BioGenome Project summary metrics for the *E. pharte* assembly

Measure	Value	*Benchmark*
EBP summary (primary)	6.C.Q64	*6.C.Q40*
Contig N50 length	8.12 Mb	*> 1 Mb*
Scaffold N50 length	31.30 Mb	*= chromosome N50*
Consensus quality (QV)	Primary: 64.2; alternate: 62.5; combined: 63.0	*# 40*
k-mer completeness	Primary: 85.24%; alternate: 75.25%; combined: 98.78%	*> 90%*
BUSCO	C:98.5% [S:97.6%, D:0.9%], F:0.4%, M:1.1%, n:5,286	*S > 90%; D < 5%*

Measure	Value	*Benchmark*
Percentage of assembly assigned to chromosomes	98.95%	*# 90%*

Notes: The EBP summary uses log10(Contig N50); chromosome-level (C) or log10(Scaffold N50); Q (Merqury QV). BUSCO v5.7.1 using the lepidoptera_odb10 lineage dataset: C=complete; S=single-copy; D=duplicated; F=fragmented; M=missing; n=orthologues.

Genome annotation report

The *E. pharte* genome assembly (GCA_965112705.1) was annotated by Ensembl at the European Bioinformatics Institute (EBI). The annotation comprises 33,980 transcripts from 15,135 protein-coding genes and 6,348 non-coding genes. The average transcript length is 17,381.29 bp, with an average of 1.58 coding transcripts per gene and 6.51 exons per transcript. For further information, please refer to the Ensembl annotation page.

Data availability

European Nucleotide Archive: Erebia pharte (blind ringlet). Accession number: PRJEB75866; URL: https://identifiers.org/ena.embl/PRJEB75866. The genome sequence is released openly for reuse. The *E. pharte* genome sequencing initiative is part of the Sanger Institute Tree of Life Programme (PRJEB43745) and Project Psyche (PRJEB71705). All raw sequence data and the primary and alternate assemblies 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.

Table 5. Software versions and sources used for *E. pharte*

Software	Version	Source
BLAST	2.14.0	https://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/
BlobToolKit	4.4.5	https://github.com/genomehubs/blobtoolkit
BlobTk (hosted BlobDir)	0.5.1	https://github.com/genomehubs/blobtk
BlobTk (snail and blob plot renderer)	0.8.3	https://github.com/genomehubs/blobtk
BUSCO	5.7.1	https://gitlab.com/ezlab/busco
busco-alg-painter	0.1.0	https://github.com/Karenvn/busco-alg-painter
bwa-mem2	2.2.1	https://github.com/bwa-mem2/bwa-mem2
DIAMOND	2.1.8	https://github.com/bbuchfink/diamond
fasta_windows	0.2.4	https://github.com/tolkit/fasta_windows
FastK	1.2	https://github.com/thegenemyers/FASTK
GenomeScope	2.1.0	https://github.com/tbenavi1/genomescope2.0
Gfastats	1.3.6	https://github.com/vgl-hub/gfastats
Hifiasm	0.19.8-r603	https://github.com/chhylp123/hifiasm
HiGlass	1.13.4	https://github.com/higlass/higlass
MerquryFK	1.1.0-c1	https://github.com/thegenemyers/MERQURY.FK
Minimap2	2.28-r1209	https://github.com/lh3/minimap2

Software	Version	Source
MitoHiFi	3.2.2	https://github.com/marcelauliano/MitoHiFi
MultiQC	1.14; 1.17 and 1.18	https://github.com/MultiQC/MultiQC
Nextflow	24.10.4	https://github.com/nextflow-io/nextflow
PretextSnapshot	0.0.6	https://github.com/sanger-tol/PretextSnapshot
PretextView	1.0.3	https://github.com/sanger-tol/PretextView
purge_dups	1.2.3	https://github.com/dfguan/purge_dups
samtools	1.21	https://github.com/samtools/samtools
sanger-tol/ascc	0.1.0	https://github.com/sanger-tol/ascc
sanger-tol/blobtoolkit	v0.7.1	https://github.com/sanger-tol/blobtoolkit
sanger-tol/curationpretext	1.4.2	https://github.com/sanger-tol/curationpretext
Seqtk	1.4-r122	https://github.com/lh3/seqtk
Singularity	3.9.0	https://github.com/sylabs/singularity
TreeVal	1.1.0	https://github.com/sanger-tol/treeval
YaHS	1.1a.2	https://github.com/c-zhou/yahs

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

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.

References

Allio R., Schomaker-Bastos A., Romiguier J.: **MitoFinder: Efficient automated large-scale extraction of mitogenomic data in target enrichment phylogenomics**. *Molecular Ecology Resources*; 2020; **20**(4): 892–905. PubMed Abstract | Publisher Full Text |

Altschul S. F., Gish W., Miller W.: **Basic Local Alignment Search Tool**. *Journal of Molecular Biology*; 1990; **215**(3): 403–410. PubMed Abstract | Publisher Full Text |

Arnscheid W., Roos P.: **Ein beitrag zum vorkommen von *Erebia pharte* (Hübner 1804) in den Vogesen (Lep.: Satyridae). Beiträge zur kenntnis der erebien, XIII**. *Entomologische Zeitschrift*; 1981; **91**(6): 66–69.

Augustijnen Hannah, Bätscher Livio, Cesanek Martin: **A macroevolutionary role for chromosomal fusion and fission in *Erebia* butterflies**. *Science Advances*; 2024; **10**(16): eadl0989. PubMed Abstract | Publisher Full Text |

Buchfink B., Reuter K., Drost H.-G.: **Sensitive protein alignments at tree-of-life scale using DIAMOND**. *Nature Methods*; 2021; **18**(4): 366–368. PubMed Abstract | Publisher Full Text |

Challis R., Richards E., Rajan J.: **BlobToolKit – interactive quality assessment of genome assemblies**. *G3: Genes, Genomes, Genetics*; 2020; **10**(4): 1361–1374. PubMed Abstract | Publisher Full Text |

Challis Richard, Blaxter Mark: **Snail plots are badges of genome assembly quality**. *G3 Genes|Genomes|Genetics*; 2026; **16**(6): jkag074. PubMed Abstract | Publisher Full Text |

Cheng H., Concepcion G. T., Feng X.: **Haplotype-resolved *de novo* assembly using phased assembly graphs with Hifiasm**. *Nature Methods*; 2021; **18**(2): 170–175. PubMed Abstract | Publisher Full Text |

Clarke Harry E.: **A checklist of European butterfly larval foodplants**. *Ecology and Evolution*; 2024; **14**(1): e10834. PubMed Abstract | Publisher Full Text |

Cuvelier S., Dincă V.: **New data regarding the butterflies (Lepidoptera: Rhopalocera) of Romania, with additional comments (general distribution in Romania, habitat preferences, threats and protection) for ten localized Romanian species**. *Phegea*; 2007; **35**(3): 93–115.

Formenti G., Abueg L., Brajuka A.: **Gfastats: Conversion, evaluation and manipulation of genome sequences using assembly graphs**. *Bioinformatics*; 2022; **38**(17): 4214–4216. PubMed Abstract | Publisher Full Text |

Guan Desheng, McCarthy Shane A., Wood Jonathan: **Identifying and removing haplotypic duplication in primary genome assemblies**. *Bioinformatics*; 2020; **36**(9): 2896–2898. PubMed Abstract | Publisher Full Text |

Howard Caroline, Denton Amy, Jackson Benjamin W.: **On the path to reference genomes for all biodiversity: Laboratory protocols and lessons learned from processing over 2,000 species in the Sanger Tree of Life**. *GigaScience*; 2025; **14** giaf119. Reference Source Publisher Full Text |

Howe K., Chow W., Collins J.: **Significantly improving the quality of genome assemblies through curation**. *GigaScience*; 2021; **10**(1): giaa153. PubMed Abstract | Publisher Full Text |

Kerpedjiev P., Abdennur N., Lekschas F.: **HiGlass: Web-based visual exploration and analysis of genome interaction maps**. *Genome Biology*; 2018; **19**(1): 125. PubMed Abstract | Publisher Full Text |

Klečková Irena, Klečka Jan, Fric Zdeněk Faltýnek: **Climatic niche conservatism and ecological diversification in the Holarctic cold-dwelling butterfly genus *Erebia***. *Insect Systematics and Diversity*; 2023; **7**(1): 2. Publisher Full Text |

Kriventseva Evgenia V., Kuznetsov Denis, Tegenfeldt Fredrik: **OrthoDB v10: Sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs**. *Nucleic Acids Research*; 2019; **47**(D1): D807–D811. PubMed Abstract | Publisher Full Text |

Kurtzer G. M., Sochat V., Bauer M. W.: **Singularity: Scientific containers for mobility of compute**. *PLOS ONE*; 2017; **12**(5): e0177459. PubMed Abstract | Publisher Full Text |

Manni Mosè, Berkeley Matthew R, Seppey Mathieu: **BUSCO update: Novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes**. *Molecular Biology and Evolution*; 2021; **38**(10): 4647–4654. PubMed Abstract | Publisher Full Text |

Panigaj Ľubomír, Šemeláková Martina, Pristaš Peter: **The phylogenetic relationship of Western Carpathians *Erebia* spp. (Lepidoptera: Nymphalidae) based on mitochondrial cytochrome oxidase subunit I sequence analysis**. *Biologia*; 2015; **70**(3): 393–399. Publisher Full Text |

Ranallo-Benavidez Timothy R., Jaron Kai S., Schatz Michael C.: **GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes**. *Nature Communications*; 2020; **11**(1): 1432. PubMed Abstract | Publisher Full Text |

Rao S. S. P., Huntley M. H., Durand N. C.: **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 |

Rhie A., Walenz B. P., Koren S.: **Merqury: Reference-free quality, completeness, and phasing assessment for genome assemblies**. *Genome Biology*; 2020; **21**(1): 245. PubMed Abstract | Publisher Full Text |

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

Sonderegger Peter: Die erebien der Schweiz (Lepidoptera: Satyrinae, genus *Erebia*). Brügg bei Biel: Peter Sonderegger; 2005. Publisher Full Text |

Tolman Tom, Lewington Richard: *Collins butterfly guide: The most complete guide to the butterflies of Britain and Europe*. London: HarperCollins Publishers; 2009.

Uliano-Silva M., Ferreira J. G. R. N., Krasheninnikova K.: **MitoHiFi: A Python pipeline for mitochondrial genome assembly from PacBio high fidelity reads**. *BMC Bioinformatics*; 2023; **24**(1): 288. PubMed Abstract | Publisher Full Text |

van Swaay C., Ellis S., Warren M.: ***Erebia pharte***. International Union for Conservation of Nature. 2025. Publisher Full Text |

Vasimuddin Md., Misra S., Li H.: **Efficient architecture-aware acceleration of BWA-MEM for multicore systems**. *2019 IEEE international parallel and distributed processing symposium (IPDPS)*. IEEE. 2019. 314–324. Publisher Full Text |

Wright Charlotte J., Stevens Lewis, Mackintosh Alexander: **Comparative genomics reveals the dynamics of chromosome evolution in Lepidoptera**. *Nature Ecology & Evolution*; 2024; **8**(4): 777–790. PubMed Abstract | Publisher Full Text |

Wright Charlotte J., Wahlberg Niklas, Vila Roger: **Project Psyche: Reference genomes for all Lepidoptera in Europe**. *Trends in Ecology & Evolution*; 2025; **40**(12): 1234–1250. Publisher Full Text |

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

Open Peer Review

Current Peer Review Status: ✔ ✔

Version 1

Reviewer Report 15 October 2026

https://doi.org/10.21956/wellcomeopenres.30496.r169050

© **2026 Rodriguez-de-la-vega R.** 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.

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

Comments on Lucek et al., "The genome sequence of the Blind Ringlet..."

In this Psyche project Data Note, the authors present an Earth BioGenome Project (EBP) reference-level, haplotype-resolved assembly of a female specimen of E*rebia pharte*, a nymphalid moth commonly known as the blind ringlet. Unlike most other Psyche project assemblies, the alternative haplotype produced is highly fragmented (>3,000 scaffolds, EBP metric: 5.5.62).

Nucleic acid extraction, library preparation (HiFi and Hi-C), and sequencing were carried out at the Wellcome Sanger Institute (WSI) Scientific Operations core or followed protocols developed by the WSI Tree of Life Core Laboratory. The assemblies were generated using hifiasm on HiFi reads scaffolded with Hi-C reads. The primary assembly was manually curated in PretextView and HiGlass using Hi-C reads mapped onto the assembly. As is now common for Psyche project Data Notes, the supporting materials include tables for:

- "Specimen and sequencing data",
- "Genome assembly statistics",
- "Chromosomal pseudomolecules in the primary assembly",
- EBP "summary statistics",
- "Software versions and sources".

Supporting figures include:

- GenomeScope2 FastK 31-mer frequency distribution,
- PretextSnapshot of Hi-C contacts,
- Merian elements on length-scaled chromosome pseudomolecules of the primary assembly,
- MerquryFK read-only, haplotype-specific, and combined assemblies 31-mer distributions,
- Snail and blob plots of the primary assembly scaffolds.

I detected no departures from the well-established protocols associated with Psyche project genome assemblies. Beyond the incremental nature of the resources presented here (as of

September 5th, this is the twenty-second *Erebia* reference-level assembly released by the Psyche or Darwin Tree of Life projects in Wellcome Open Research, and the 16th obtained from a female specimen in the genus), their most interesting contribution will emerge once researchers begin analyzing the data in a comparative genomics framework. For instance, among the *Erebia* assemblies released by DToL or Psyche projects, W sex chromosomes are conspicuously absent from several female assemblies (e.g., WOR_11-650; WOR_11-588), while in others, the W chromosome is the largest (WOR_11-589; WOR_10-425) or the smallest assembled chromosome (WOR_11-75).

For convenience, I am including links to the DToL quality check and Genomes on a Tree page for this specimen/species.
https://tolqc.cog.sanger.ac.uk/tol/insects/Erebia_pharte/
https://goat.genomehubs.org/record?recordId=111906&result=taxon&taxonomy=ncbi#Erebia%20pharte

References

1. Lucek K, Kleckova I, Wright C, Meier J, et al.: The genome sequence of the Sudeten Ringlet, Erebia sudetica Dalman, 1816 (Lepidoptera: Nymphalidae). *Wellcome Open Research*. 2026; **11**.
Publisher Full Text
2. Chittaro Y, Vila M, Lucek K, Wright C, et al.: The genome sequence of the de Prunner's Ringlet, Erebia triaria von Prunner, 1798 (Lepidoptera: Nymphalidae). *Wellcome Open Research*. 2026; **10**.
Publisher Full Text

Are the rationale for sequencing the genome and the species significance clearly described?

Yes

Are the rationale for sequencing the genome and the species significance clearly described?

Yes

Are the protocols appropriate and is the work technically sound?

Yes

Are the protocols appropriate and is the work technically sound?

Yes

Are sufficient details of the sequencing and extraction, software used, and materials provided to allow replication by others?

Yes

Are sufficient details of the sequencing and extraction, software used, and materials provided to allow replication by others?

Yes

Are the datasets clearly presented in a usable and accessible format, and the assembly and annotation available in an appropriate subject-specific repository?

Yes

Are the datasets clearly presented in a usable and accessible format, and the assembly and

annotation available in an appropriate subject-specific repository?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Evolutionary Genomics; Bioinformatics

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

Reviewer Report 04 September 2026

https://doi.org/

© 2026 Benchmarking A. 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.

Automated Benchmarking
Wellcome Sanger Institute, Hinxton, UK

As this article is a Genome Note, Automated Benchmarking has been used to determine whether the reported assembly passes three defined checks established by the Earth BioGenome Project: the raw and assembled data are publicly available, the Earth BioGenome Project three-part metric is met, and the BUSCO score passes the stated threshold.

For Erebia pharte (ilErePhar2), all three checks passed.

Data available: pass. ENA study PRJEB75866, raw-read accessions ERR13158709 and ERR13148270, and assembly accessions GCA_965112705.1 and GCA_965112695.1 are public.

Earth BioGenome Project three-part metric: pass. The primary assembly is reported as 6.C.Q64, meeting the 6.C.Q40 reference standard.

BUSCO: pass. The primary assembly, which is proposed as the reference assembly for this species, has 98.5% complete BUSCOs (97.6% single-copy and 0.9% duplicated). It meets the target of greater than 90% complete and single-copy BUSCOs.

This decision uses values reported in the Genome Note and public accession records.

Competing Interests: This Automated Benchmarking peer-review report was generated automatically and relies on analyses performed by the Tree of Life programme, which also produced the Genome Note under review.

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.