DATA NOTE

The genome sequence of the Dusky Meadow Brown, *Cercyonis lycaon* (Lepidoptera: Nymphalidae)

[version 1; peer review: 3 approved]

Yannick Chittaro[1], Eric Toro-Delgado[2], Kay Lucek[3], Charlotte J. Wright[4], Joana I. Meier[4], Mark L. Blaxter[4],
Wellcome Sanger Institute Tree of Life Management, Samples and Laboratory team,
Wellcome Sanger Institute Scientific Operations: Sequencing Operations,
Wellcome Sanger Institute Tree of Life Core Informatics team,
Tree of Life Core Informatics collective, Project Psyche Community

[1]Info fauna, Neuchâtel, Switzerland
[2]Institut de Biologia Evolutiva, CSIC - Universitat Pompeu Fabra, Barcelona, Spain
[3]University of Neuchâtel, Neuchâtel, Switzerland
[4]Tree of Life, Wellcome Sanger Institute, Hinxton, England, UK

v1 **First published:** 31 Jul 2025, **10**:395
https://doi.org/10.12688/wellcomeopenres.24656.1
Latest published: 31 Jul 2025, **10**:395
https://doi.org/10.12688/wellcomeopenres.24656.1

Abstract

We present a genome assembly from a female specimen of *Cercyonis lycaon* (Dusky Meadow Brown; Arthropoda; Insecta; Lepidoptera; Nymphalidae). The assembly contains two haplotypes with total lengths of 601.00 megabases and 548.79 megabases. Most of haplotype 1 (94.99%) is scaffolded into 30 chromosomal pseudomolecules, including the W and Z sex chromosomes. Haplotype 2 was assembled to scaffold level. The mitochondrial genome has also been assembled, with a length of 15.2 kilobases.

Keywords

Cercyonis lycaon, Dusky Meadow Brown, genome sequence, chromosomal, Lepidoptera

This article is included in the Tree of Life gateway.

Open Peer Review

Approval Status ✓ ✓ ✓

	1	2	3
version 1 31 Jul 2025	✓ view	✓ view	✓ view

1. **Annabel Whibley**, The University of Auckland, Auckland, New Zealand
Bragato Research Institute, Lincoln, New Zealand

2. **Arun Arumugaperumal**, Rajalakshmi Engineering College, Thandalam 602105, Chennai, India

3. **Jackson Wolfe**, San Diego State University College of Sciences (Ringgold ID: 115104), San Diego, USA
Arun Sethuraman, San Diego State University, 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; **Toro-Delgado E**: 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). E. T-D. is supported by an FPU grant from the Spanish Ministry of Science, Innovation and Universities (MCIU; grant number FPU22/02358). KL was supported by a grant (Grant ID 202869) from the Swiss National Science Foundation (SNSF).
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

How to cite this article: Chittaro Y, Toro-Delgado E, Lucek K *et al.* **The genome sequence of the Dusky Meadow Brown, *Cercyonis lycaon* (Lepidoptera: Nymphalidae) [version 1; peer review: 3 approved]** Wellcome Open Research 2025, **10**:395 https://doi.org/10.12688/wellcomeopenres.24656.1

First published: 31 Jul 2025, **10**:395 https://doi.org/10.12688/wellcomeopenres.24656.1

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; Satyrini; Maniolina; *Cercyonis*; *Hyponephele*; *Cercyonis lycaon* (NCBI:txid3038947)

Background

The Dusky Meadow Brown (*Cercyonis lycaon*) is a species of butterfly in the family Nymphalidae, subfamily Satyrinae. Its wing dorsal side is dark brown, while the ventral side is orange with a grey margin on the forewing and grey with black marbling on the hindwing, which is also festooned (Tolman & Lewington, 2011). This species is sexually dimorphic, with dorsal forewings presenting subapical black ocelli and androconia in males and large black ocelli framed in orange in females. It can be distinguished from the oriental meadow brown (*C. lupina*) by the male's androconia (narrower in *C. lycaon*) and the extent of orange colouration on the forewing dorsal side in females (more orange in *C. lycaon*).

Cercyonis lycaon is found from the Iberian Peninsula throughout most of Europe (except the plains of western France up to eastern Germany, the British and Irish Isles and most of Scandinavia), as well as Anatolia, the Caucasus, Iran, and southern Russia up to lake Baikal in Central Asia (Kudrna *et al.*, 2015; Tolman & Lewington, 2011). It is a cryo-xerophilous species inhabiting dry montane and steppic habitats between 800 and 2 000 m (García-Barros *et al.*, 2013). It is an univoltine species, on flight between July and September, and overwinters as 1st or 2nd instar larva (Lafranchis *et al.*, 2015; LSPN, 1987; Vila *et al.*, 2018). Females lay eggs individually on the grasses of Poaceae, such as *Festuca*, *Bromus* and *Stipa* (Vila *et al.*, 2018). Males exhibit territorial behaviour. This butterfly is classified as Least Concern (LC) by the European Red List of Butterflies (Van Swaay *et al.*, 2010). However, it is local and scarce in some regions, such as Catalonia and Switzerland, where it is considered Vulnerable (Vila *et al.*, 2018; Wermeille *et al.*, 2014).

Cytogenetic studies determined a haploid chromosome number of n = 29 based on specimens from France and Turkey (De Lesse, 1960). Mitochondrial data (DNA barcodes) suggests a high genetic diversity and pronounced genetic structure, with some lineages in the east Mediterranean that are highly divergent from the main European one (Dapporto *et al.*, 2022). However, some of these lineages may correspond to other closely related taxa, as the *lycaon* species complex is poorly studied and suspected to contain more species in West Asia (Lukhtanov & Novikova, 2015; Tshikolovets, 2011). The generic classification is also debated, as historically *C. lycaon* had been classified as *Hyponephele*, which some authors consider a subgenus of *Cercyonis* (Zhang *et al.*, 2020). Although we follow the NCBI taxonomy, given the reciprocal monophyly of *Hyponephele* and *Cercyonis sensu stricto* and the lack of precedent for using *Cercyonis* for Old World taxa, we favour the use of *Hyponephele* as genus to maintain taxonomic stability, at least until further evidence with a more extensive taxon sampling is gathered. The reference genome presented here will prove invaluable in clarifying the taxonomy of the genus and the diversity within the *lycaon* complex.

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

Methods

Sample acquisition and DNA barcoding

The specimen used for genome sequencing was an adult female *Cercyonis lycaon* (specimen ID SAN28000143, ToLID ilCerLyca1; Figure 1), collected from Conthey, Valais, Switzerland (latitude 46.2872, longitude 7.3116; elevation 1 650 m) on 02/08/2023. 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 ilCerLyca1 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.

Figure 1. Voucher photograph of the *Cercyonis lycaon* (ilCerLyca1) specimen used for genome sequencing.

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 33.03 ng/μL and a yield of 1 552.41 ng, with a fragment size of 14.4 kb. The 260/280 spectrophotometric ratio was 2.01, and the 260/230 ratio was 3.25.

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 head tissue of the ilCerLyca1 sample using the Arima-HiC v2 kit (Arima Genomics). Following the manufacturer's instructions, tissue was fixed and DNA crosslinked using TC buffer to a final formaldehyde concentration of 2%. The tissue was homogenised using the Diagnocine Power Masher-II. Crosslinked DNA was digested with a restriction enzyme master mix, biotinylated, and ligated. Clean-up was performed with SPRISelect beads before library preparation. DNA concentration was measured with the Qubit Fluorometer (Thermo Fisher Scientific) and Qubit HS Assay Kit. The biotinylation percentage was estimated using the Arima-HiC v2 QC beads.

Hi-C library preparation and sequencing

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

Table 1. Specimen and sequencing data for BioProject PRJEB78762.

Platform	PacBio HiFi	Hi-C
ToLID	ilCerLyca1	ilCerLyca1
Specimen ID	SAN28000143	SAN28000143
BioSample (source individual)	SAMEA115117751	SAMEA115117751
BioSample (tissue)	SAMEA115117754	SAMEA115117753
Tissue	thorax	head
Sequencing platform and model	Revio	Illumina NovaSeq X
Run accessions	ERR13485727	ERR13493984
Read count total	2.04 million	782.12 million
Base count total	21.56 Gb	118.10 Gb

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 7 breaks and 12 joins. The curation process is documented at https://gitlab.com/wtsi-grit/rapid-curation. PretextSnapshot was used to generate a Hi-C contact map of the final assembly.

Assembly quality assessment

Chromosomal painting was performed using lep_busco_painter 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 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 *Cercyonis lycaon* was sequenced using Pacific Biosciences single-molecule HiFi long reads, generating 21.56 Gb (gigabases) from 2.04 million reads, which were used to assemble the genome. GenomeScope2.0 analysis estimated the haploid genome size at 560.00 Mb, with a heterozygosity of 1.82% and repeat content of 42.08%. These estimates guided expectations for the assembly. Based on the estimated genome size, the sequencing data provided approximately 37× coverage. Hi-C sequencing produced 118.10 Gb from 782.12 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 601.00 Mb in 209 scaffolds, with 121 gaps, and a scaffold N50 of 20.64 Mb (Table 2).

Most of the assembly sequence (94.99%) was assigned to 30 chromosomal-level scaffolds, representing 28 autosomes and

the W and Z sex chromosomes. 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).

Table 2. Genome assembly statistics.

Assembly name	ilCerLyca1.hap1.1	ilCerLyca1.hap2.1
Assembly accession	GCA_964270755.1	GCA_964270765.1
Assembly level	chromosome	scaffold
Span (Mb)	601.00	548.79
Number of chromosomes	30	N/A
Number of contigs	330	302
Contig N50	7.85 Mb	7.66 Mb
Number of scaffolds	209	189
Scaffold N50	20.64 Mb	20.16 Mb
Longest scaffold length (Mb)	26.43	N/A
Sex chromosomes	W and Z	N/A
Organelles	Mitochondrial genome: 15.2 kb	N/A

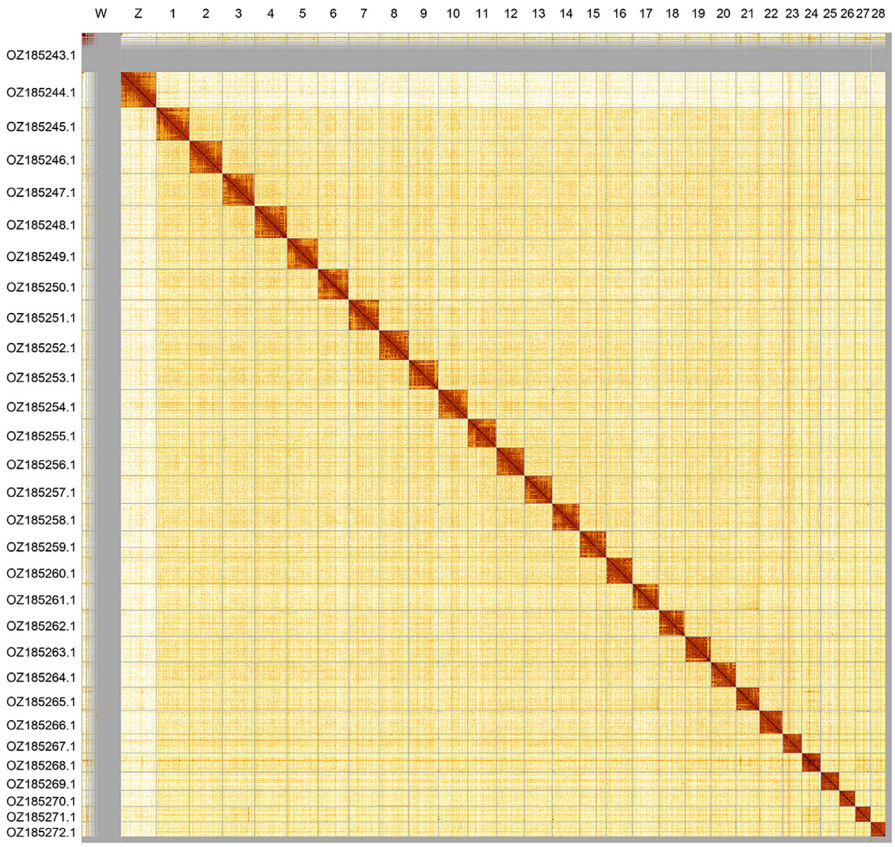

Figure 2. Hi-C contact map of the *Cercyonis lycaon* genome assembly. Assembled chromosomes are shown in order of size and labelled along the axes.

Table 3. Chromosomal pseudomolecules in the haplotype 1 genome assembly of *Cercyonis lycaon* ilCerLyca1.

INSDC accession	Molecule	Length (Mb)	GC%	Assigned Merian elements
OZ185245.1	1	24.61	37	M1
OZ185246.1	2	24.41	36.50	M2
OZ185247.1	3	24.01	37	M17;M20
OZ185248.1	4	23.99	36.50	M8
OZ185249.1	5	22.99	36.50	M19;M26
OZ185250.1	6	22.66	36.50	M3
OZ185251.1	7	22.65	36.50	M9
OZ185252.1	8	22.07	36.50	M5
OZ185253.1	9	21.91	37	M12
OZ185254.1	10	21.78	37	M14;M29
OZ185255.1	11	21.36	36.50	M18
OZ185256.1	12	20.77	36.50	M7
OZ185257.1	13	20.64	36.50	M16
OZ185258.1	14	20.33	36.50	M6
OZ185259.1	15	19.77	36.50	M21
OZ185260.1	16	19.52	37	M22
OZ185261.1	17	19.51	36.50	M4
OZ185262.1	18	19.44	37	M15
OZ185263.1	19	19.16	37	M10
OZ185264.1	20	18.52	37	M11
OZ185265.1	21	17.44	37	M23
OZ185266.1	22	17.14	36.50	M13
OZ185267.1	23	14.36	37	M24
OZ185268.1	24	13.89	38.50	M30
OZ185269.1	25	13.66	36.50	M28
OZ185270.1	26	11.76	37	M27
OZ185271.1	27	11.47	38	M31
OZ185272.1	28	11.08	37	M25
OZ185243.1	W	3.58	38	N/A
OZ185244.1	Z	26.43	37	MZ

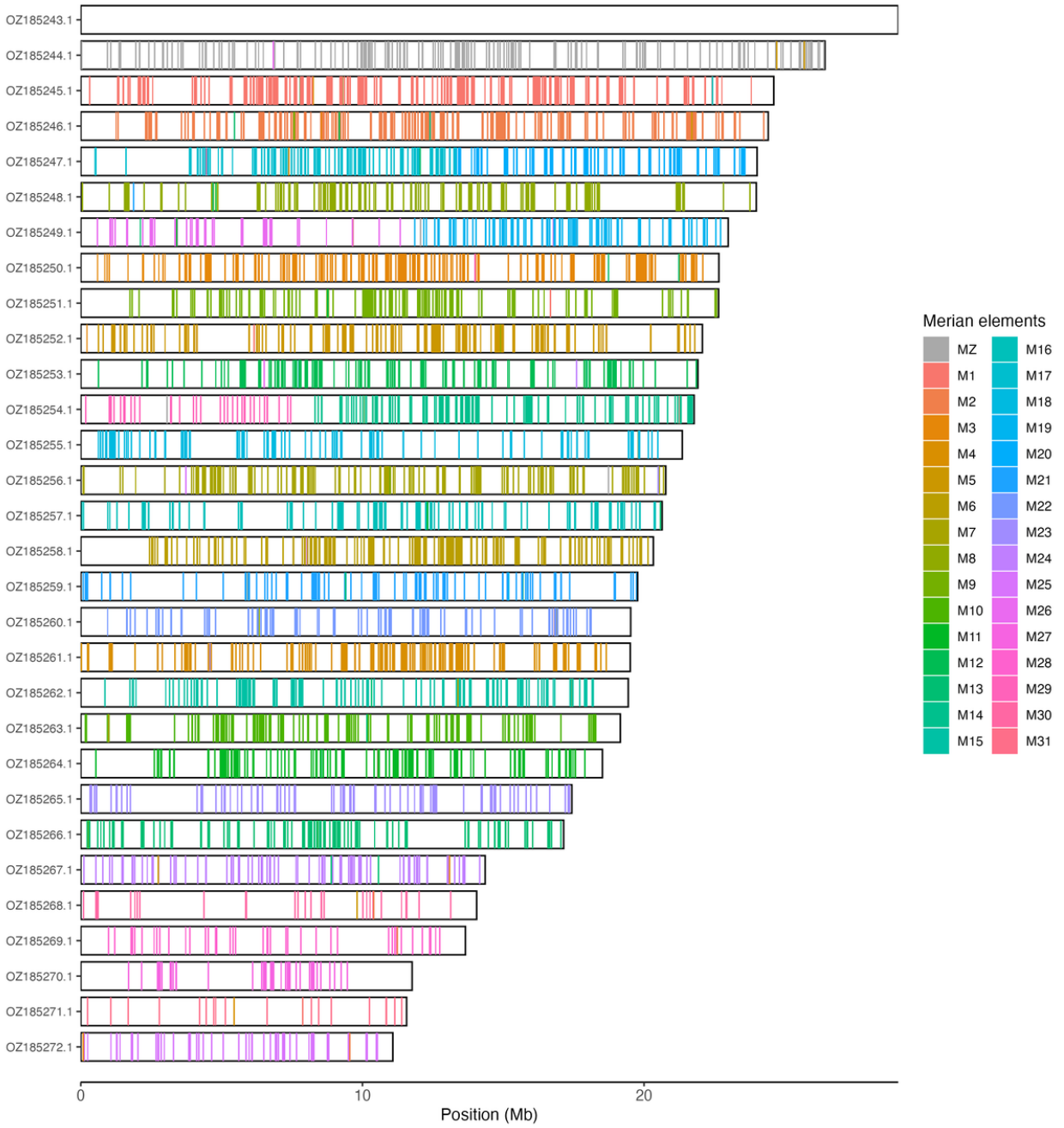

Figure 3. Merian elements painted across chromosomes in the ilCerLyca1.hap1.1 assembly of *Cercyonis lycaon*. Chromosomes are drawn to scale, with the positions of orthologues shown as coloured bars. Each orthologue is coloured by the Merian element that it belongs to. All orthologues which could be assigned to Merian elements are shown.

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 64.8, and for haplotype 2, 65.8. When the two haplotypes are combined, the assembly achieves an estimated QV of 65.2. The *k*-mer completeness is 71.32% for haplotype 1, 67.87% for haplotype 2, and 99.38% for the combined haplotypes (Figure 4). BUSCO analysis using the lepidoptera_odb10 reference set (n = 5 286) (Kriventseva *et al.*, 2019) identified 98.5% of the expected gene set (single = 97.8%, duplicated = 0.6%) for haplotype 1. The snail plot in Figure 5 summarises the scaffold length distribution

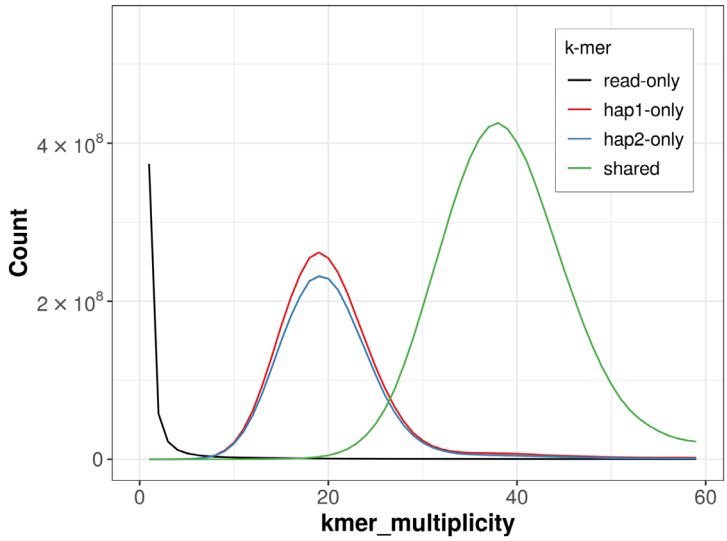

Figure 4. Evaluation of *k*-mer completeness using MerquryFK. This plot illustrates the recovery of *k*-mers from the original read data in the final assemblies. The horizontal axis represents *k*-mer multiplicity, and the vertical axis shows the number of *k*-mers. The black curve represents *k*-mers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

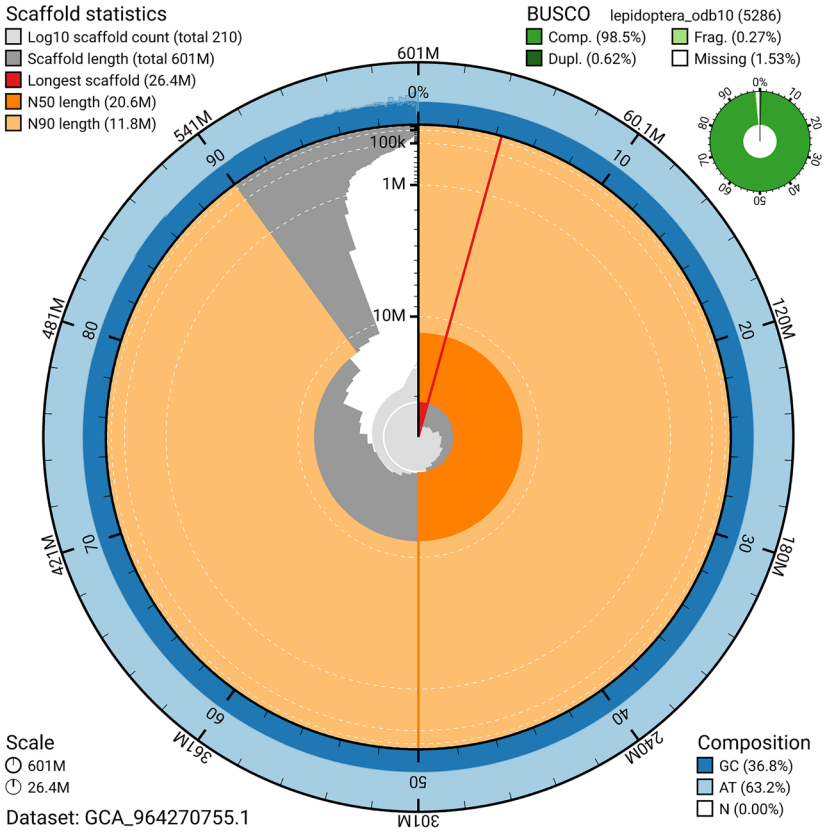

Figure 5. Assembly metrics for ilCerLyca1.hap1.1. The BlobToolKit snail plot provides an overview of assembly metrics and BUSCO gene completeness. The circumference represents the length of the whole genome sequence, and the main plot is divided into 1,000 bins around the circumference. The outermost 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.

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

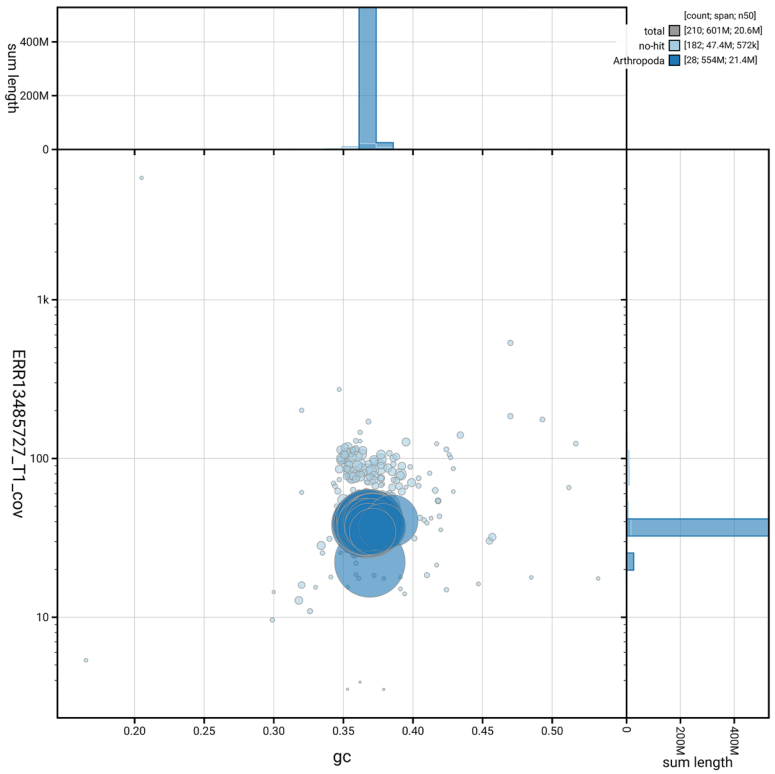

Figure 6. BlobToolKit GC-coverage plot for ilCerLyca1.hap1.1. Blob plot showing sequence coverage (vertical axis) and GC content (horizontal axis). The circles represent scaffolds, with the size proportional to scaffold length and the colour representing phylum membership. The histograms along the axes 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 *Cercyonis lycaon* assembly.

Measure	Value	Benchmark
EBP summary (haplotype 1)	6.7.Q64	6.C.Q40
Contig N50 length	7.85 Mb	≥ 1 Mb
Scaffold N50 length	20.64 Mb	= chromosome N50
Consensus quality (QV)	Haplotype 1: 64.8; haplotype 2: 65.8; combined: 65.2	≥ 40
k-mer completeness	Haplotype 1: 71.32%; Haplotype 2: 67.87%; combined: 99.38%	≥ 95%
BUSCO	C:98.5%[S:97.8%,D:0.6%], F:0.3%,M:1.3%,n:5 286	S > 90%; D < 5%
Percentage of assembly assigned to chromosomes	94.99%	≥ 90%

for the haplotype 1, is **6.C.Q64**, 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.

Data availability

European Nucleotide Archive: Cercyonis lycaon (dusky meadow brown). Accession number PRJEB78762. The genome sequence is released openly for reuse. The *Cercyonis lycaon* 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
lep_busco_painter	1.0.0	https://github.com/charlottewright/lep_busco_painter
MerquryFK	1.1	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

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

Dapporto L, Menchetti M, Vodă R, *et al.*: **The atlas of mitochondrial genetic diversity for Western Palaearctic butterflies.** *Glob Ecol Biogeogr.* 2022; **31**(11): 2184–90.
Publisher Full Text

De Lesse H: **Spéciation et variation chromosomique chez les Lépidoptères Rhopalocères.** *Annales des Sciences Naturelles, Zoologie.* 1960; **12**(2): 1–223.
Reference Source

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

García-Barros E, Munguira ML, Stefanescu C, *et al.*: **Fauna Ibérica, Vol. 37. Lepidoptera. Papilionoidea.** Madrid: Museo Nacional de Ciencias Naturales, 2013.
Reference Source

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): giaa153.
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

Kudrna O, Pennerstorfer J, Lux K: **Distribution atlas of European butterflies and skippers.** Wissenschaftlicher Verlag PEKS, 2015.

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

Lafranchis T, Jutzeler D, Guillosson JY, *et al.*: **La vie des papillons : éologie, biologie et comportement des Rhopalocères de France.** *Et Comportement Des Rhopalocères de France.* Diathéo, 2015.
Reference Source

Li H: **Minimap2: pairwise alignment for nucleotide sequences.** *Bioinformatics.* 2018; **34**(18): 3094–3100.
PubMed Abstract | Publisher Full Text | Free Full Text

LSPN: **Les papillons de jour et leurs biotopes: espèces, dangers qui les menacent, protection.** Fotorotar AG, 1987; **1**.
Reference Source

Lukhtanov VA, Novikova AV: **Interpretation of mitochondrial diversity in terms of taxonomy: a case study of *Hyponephele lycaon* species complex in Israel (*Lepidoptera, Nymphalidae, Satyrinae*).** *Zookeys.* 2015; (538): 21–34.
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

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

Tolman T, Lewington R: **Mariposas de España y Europa.** Barcelona: Lynx Edicions, 2011.
Reference Source

Tshikolovets VV: **Butterflies of Europe and the Mediterranean area.** Tshikolovets Publications, 2011.
Reference Source

Uliano-Silva M, Ferreira JGRN, Krasheninnikova K, *et al.*: **MitoHiFi: a python pipeline for mitochondrial genome assembly from PacBio high fidelity reads.** *BMC Bioinformatics.* 2023; **24**(1): 288.
PubMed Abstract | Publisher Full Text | Free Full Text

Van Swaay C, Cuttelod A, Collins S, *et al.*: **European red list of butterflies.** Luxembourg: Publications Office of the European Union, 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 R, Stefanescu C, Sesma JM: **Guia de Les Papallones Diürnes de Catalunya.** Barcelona: Lynx Edicions, 2018.
Reference Source

Wermeille E, Chittaro Y, Gonseth Y: **Liste rouge Papillons diurnes et Zygènes.** Espèces menacées en Suisse, état 2012. *L'environnement pratique n° 1403.* 2014.
Reference Source

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

Zhang J, Cong Q, Shen J, *et al.*: **Genomic evidence suggests further changes of butterfly names.** *Taxon Rep Int Lepid Surv.* 2020; **8**: 7.
PubMed Abstract | 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.27170.r128202

✔ **Jackson Wolfe**
Biology, San Diego State University College of Sciences (Ringgold ID: 115104), San Diego, California, USA

Arun Sethuraman
San Diego State University, San Diego, California, USA

Summary: The authors report a high quality nuclear genome (~600Mb) of the dusky meadow brown, with 95% of the genome assembled into 30 chromosomal scaffolds, and a mitochondrial genome of 15.2 kb.

The authors do a good job of describing the methods used in generating the assembly. The introduction section included sufficient biological information regarding the rationale for the project.

No major concerns were noted by the reviewers, all methods are accessible and reproducible, with the software and its parameters. The genome itself available and open for reuse via EBI.

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, population genetics, evolution

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

Reviewer Report 04 September 2025

https://doi.org/10.21956/wellcomeopenres.27170.r128205

Arun Arumugaperumal
Department of Biotechnology, Rajalakshmi Engineering College, Thandalam 602105, Chennai, Tamil Nadu, 602105, India

The data note describes the genome sequence of *Cercyonis lycaon*, commonly known as the Dusky Meadow Brown. The photograph of the specimen remains in Figure 1 is of very good quality. This is the first report of the high quality genome of the butterfly species. The assembly reported is of size 601 Mb distributed among 30 chromosomes. The mitogenome has also been assembled of size 15.2 kb. The genomic DNA was sequenced using high-end sequencing machines. The assembly was said to be decontaminated by ASCC pipeline. What was the result? Does the presented genome size after removal of considerable amount of bases? The blob plot also shows emergence of a different blob.
The links provided in the data note pointing to databases were working fine. A BUSCO completeness of 98.5% indicates that the genome is a near-complete one. The data note can be indexed.

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: Bioinformatics; Genomics

I confirm that I have read this submission and believe that I have an appropriate level of

expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 30 August 2025

https://doi.org/10.21956/wellcomeopenres.27170.r128203

Annabel Whibley
[1] The University of Auckland, Auckland, New Zealand
[2] Grapevine Improvement, Bragato Research Institute, Lincoln, New Zealand

Chittaro and colleagues report the genome assembly of the Dusky Meadow Brown butterfly *(Cercyonis lyacon).* This species was sequenced as part of Project Psyche, a multi-stakeholder project with an overarching goal to sequence all European Lepidopteran species.

The background section sets out very effectively the natural history, population distribution, taxonomic context and karyotype evidence of *C. lyracon* and the voucher specimen photograph is of good quality. The sample processing methods section contains hyperlinks to protocols which have been given DOIs. I assume that a new DOI would be minted when the protocol version is updated, but I wonder whether it would be helpful to also note the protocol versions in the text itself.

Combining Revio HiFi data and Arima Hi-C, the genome assembly is of exceptional quality (formally 6.C.Q64.). Close to 95% of scaffolds have been assigned to chromosomes, and BUSCO and kmer completeness are >98% and >99% respectively. formally 6.C.Q64. The use of a female individual means that both Z and W chromosomes have been captured, which adds to the utility of the reference. The inclusion of analyses of Merian elements, reconstructing ancestral linkage groups, is a really nice addition to standard elements of the DToL reporting. As usual for the data notes, the report follows templating and the methodology, reporting of sample metadata, protocols and assembly protocols are comprehensively and clearly documented.

Please note there is a typo in Table 4 (for the EBP summary of haplotype 1 Value; it should be 6.C.Q64 not 6.7.Q64).

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

Are the protocols appropriate and is the work technically sound?
Yes

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

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

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Genomics, Bioinformatics

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