DATA NOTE

The genome sequence of the Nickerl's Fritillary, *Melitaea aurelia* Nickerl, 1850 (Lepidoptera: Nymphalidae)

[version 1; peer review: 2 approved]

Yannick Chittaro[1], Andreas Sanchez[1], Kay Lucek[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]Info fauna, Neuchâtel, Switzerland
[2]University of Neuchâtel, Neuchâtel, Switzerland
[3]Tree of Life, Wellcome Sanger Institute, Hinxton, England, UK

v1 **First published:** 22 Aug 2025, **10**:459
https://doi.org/10.12688/wellcomeopenres.24749.1
Latest published: 22 Aug 2025, **10**:459
https://doi.org/10.12688/wellcomeopenres.24749.1

Abstract

We present a genome assembly from a female specimen of *Melitaea aurelia* (Nickerl's Fritillary; Arthropoda; Insecta; Lepidoptera; Nymphalidae). The assembly contains two haplotypes with total lengths of 595.21 megabases and 565.99 megabases. Most of haplotype 1 (99.72%) is scaffolded into 31 chromosomal pseudomolecules, including the Z sex chromosome. Haplotype 2 was assembled to scaffold level. The mitochondrial genome has also been assembled, with a length of 15.17 kilobases.

Keywords

Melitaea aurelia, Nickerl's Fritillary, genome sequence, chromosomal, Lepidoptera

This article is included in the Tree of Life gateway.

Open Peer Review

Approval Status ✓ ✓

	1	2
version 1 22 Aug 2025	✓ view	✓ view

1. **Laurence Despres**, Universite Savoie Mont Blanc, Chambéry, France
Université Grenoble Alpes, Grenoble, France

2. **Sina Beier**, MRC Toxicology Unit, Cambridge, UK
Gurdon Institute (Ringgold ID: 47693), Cambridge, UK

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; **Sanchez A**: Writing – Original Draft Preparation; **Lucek K**: Supervision, Writing – Review & Editing; **Wright CJ**: Funding Acquisition, Project Administration, Supervision; **Meier JI**: Funding Acquisition, Project Administration, Supervision; **Blaxter ML**: Funding Acquisition, Project Administration, Supervision;

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by Wellcome through core funding to the Wellcome Sanger Institute (220540). KL was supported by Swiss National Science Foundation (SNSF) (Grant ID 202869).
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

How to cite this article: Chittaro Y, Sanchez A, Lucek K *et al.* **The genome sequence of the Nickerl's Fritillary, *Melitaea aurelia* Nickerl, 1850 (Lepidoptera: Nymphalidae) [version 1; peer review: 2 approved]** Wellcome Open Research 2025, **10**:459 https://doi.org/10.12688/wellcomeopenres.24749.1

First published: 22 Aug 2025, **10**:459 https://doi.org/10.12688/wellcomeopenres.24749.1

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; Nymphalinae; Melitaeini; Melitaeina; Melitaea; *Mellicta*; *Melitaea aurelia* Nickerl, 1850 (NCBI:txid2716722)

Background

Melitaea aurelia is found throughout central and eastern Europe, from France (western limit) to western Siberia (eastern limit), but is absent from the British and Irish Isles, Spain, Corsica, Sardinia and a few other islands in the Mediterranean basin. To the north, it has been recorded in southern Sweden (Ebert, 2011; Lafranchis *et al.*, 2015; Tshikolovets, 2011). There are several subspecies in its range, detailed by Tshikolovets (2011).

The upper side of the wings is orange, squared off with black lines of varying thickness, giving a generally darker appearance than other species in the genus. The species is also generally smaller in size than its congeners. The species is not easy to identify, however, and reliable identification requires observation of the male genitalia (especially the tips of the valves, LSPN, 1987). There is little sexual dimorphism in this species.

Melitaea aurelia is a species associated with calcareous grasslands (Van Swaay, 2002), and *Arrhenateretum* with meadow sage on dry soils (*Arrhenateretum salvietosum*) in the colline and montane levels (optimum around 1 500 m) but can occur singly up to almost 2 000 m in the warmest locations. It also likes flowering edges and bushy pastures (Ebert, 2011; Essayan *et al.*, 2013; Lafranchis *et al.*, 2015). In theory, *M. aurelia* has only one generation per year (but a second generation in August cannot be ruled out, according to Lafranchis *et al.*, 2015), from late April to late September. After mating, the female lays her eggs in small clutches (one or two clutches containing between 100 and 300 eggs) under the leaves of host plants (mainly *Plantago* spp., although they may consume other plants such as *Veronica* spp., *Digitalis* spp., *Cirsium* spp. or *Melampyrum* spp.) (Clarke, 2024; Lafranchis *et al.*, 2015; Tshikolovets, 2011). From the second to the third larval instar, caterpillars live grouped in a web woven between two host plant leaves and feed on the entire leaves of their host plant, leaving only the veins. Around a month after hatching, the caterpillars undergo their third moult and stop feeding, taking refuge in small groups under dead leaves in the litter. They then enter winter diapause, weaving a silken mat to enclose themselves. The following spring, they emerge from diapause very emaciated and feed little, spending most of their time immobile in the sun. At the end of the 7th instar, they pupate. The pupa hangs in low vegetation (Lafranchis *et al.*, 2015).

On a European scale, it is potentially threatened (NT in Europe and LC in EU27 according to Van Swaay *et al.*, 2010). Nevertheless, regionally, it may be more threatened, as is the case in Switzerland where it is endangered (EN according to Wermeille *et al.*, 2014) or in Austria where it is vulnerable (VU according to Höttinger & Pennerstorfer, 2005). It appears, however, that *M. aurelia* is steadily expanding its range, at least locally, as has been observed in northwest Germany (Fartmann, 2004) and the Czech Republic (Spitzer & Beneš, 2010). Despite its restriction to calcareous grasslands, *M. aurelia* seems to avoid a strong differentiation between isolated populations and adverse effects of genetic bottlenecks (study based on allozymes, Habel *et al.*, 2009). However, it is a sedentary species, confined to the meadows where its larvae have developed, and is therefore particularly sensitive to changes in its habitat (Eichel & Fartmann, 2008), in particular to the intensification of agricultural practices (intensive grazing, mowing too early or too frequent manuring and watering). The abandonment of farming and the urbanisation of sunny hillsides also threaten certain populations.

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

Methods

Sample acquisition

The specimen used for genome sequencing was an adult female *Melitaea aurelia* (specimen ID SAN28000126, ToLID ilMelAure1), collected from Euseigne VS, Switzerland (latitude 46.1773, longitude 7.4159; elevation 800 m) on 01/04/2023. 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 ilMelAure1 sample was weighed and

Figure 1. Voucher image of the *Melitaea aurelia* specimen (SAN28000126, ToLID ilMelAure1) used for genome sequencing.

triaged to determine the appropriate extraction protocol. Tissue from the thorax was homogenised by powermashing using a PowerMasher II tissue disruptor.

HMW DNA was extracted in the WSI Scientific Operations core using the Automated MagAttract v2 protocol. DNA was sheared into an average fragment size of 12–20 kb following the Megaruptor®3 for LI PacBio protocol. Sheared DNA was purified by automated SPRI (solid-phase reversible immobilisation). The concentration of the sheared and purified DNA was assessed using a Nanodrop spectrophotometer and Qubit Fluorometer using the Qubit dsDNA High Sensitivity Assay kit. Fragment size distribution was evaluated by running the sample on the FemtoPulse system. For this sample, the final post-shearing DNA had a Qubit concentration of 46.73 ng/μL and a yield of 2 196.31 ng, with a fragment size of 14.8 kb. The 260/280 spectrophotometric ratio was 1.92, and the 260/230 ratio was 2.58.

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 ilMelAure1 sample using the Arima-HiC v2 kit (Arima Genomics). Following the manufacturer's instructions, tissue was fixed and DNA crosslinked using TC buffer to a final formaldehyde concentration of 2%. The tissue was homogenised using the Diagnocine Power Masher-II. Crosslinked DNA was digested with a restriction enzyme master mix, biotinylated, and ligated. Clean-up was performed with SPRISelect beads before library preparation. DNA concentration was measured with the Qubit Fluorometer (Thermo Fisher Scientific) and Qubit HS Assay Kit. The biotinylation percentage was estimated using the Arima-HiC v2 QC beads.

Hi-C library preparation and sequencing

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

Table 1. Specimen and sequencing data for BioProject.

Platform	PacBio HiFi	Hi-C
ToLID	ilMelAure1	ilMelAure1
Specimen ID	SAN28000126	SAN28000126
BioSample (source individual)	SAMEA115110016	SAMEA115110016
BioSample (tissue)	SAMEA115110061	SAMEA115110062
Tissue	thorax	head
Sequencing platform and model	Revio	Illumina NovaSeq X
Run accessions	ERR13605516	ERR13602179
Read count total	2.13 million	684.49 million
Base count total	21.19 Gb	103.36 Gb

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 3 breaks and 10 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 *Melitaea aurelia* was sequenced using Pacific Biosciences single-molecule HiFi long reads, generating 21.19 Gb (gigabases) from 2.13 million reads, which were used to assemble the genome. GenomeScope2.0 analysis estimated the haploid genome size at 575.01 Mb, with a heterozygosity of 1.99% and repeat content of 39.89%. These estimates guided expectations for the assembly. Based on the estimated genome size, the sequencing data provided approximately 35× coverage. Hi-C sequencing produced 103.36 Gb from 684.49 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 595.21 Mb in 78 scaffolds, with 100 gaps, and a scaffold N50 of 20.52 Mb (Table 2).

Most of the assembly sequence (99.72%) was assigned to 31 chromosomal-level scaffolds, representing 30 autosomes and the Z sex chromosome. Chromsome Z was assigned based on the Hi-C signal. This appears to be a Z0 female. The 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

Table 2. Genome assembly statistics.

Assembly name	ilMelAure1.hap1.1	ilMelAure1.hap2.1
Assembly accession	GCA_964265045.1	GCA_964265085.1
Assembly level	chromosome	scaffold
Span (Mb)	595.21	565.99
Number of chromosomes	31	N/A
Number of contigs	178	149
Contig N50	8.6 Mb	8.46 Mb
Number of scaffolds	78	65
Scaffold N50	20.52 Mb	20.43 Mb
Longest scaffold length (Mb)	27.45	N/A
Sex chromosomes	Z	N/A
Organelles	Mitochondrial genome: 15.17 kb	N/A

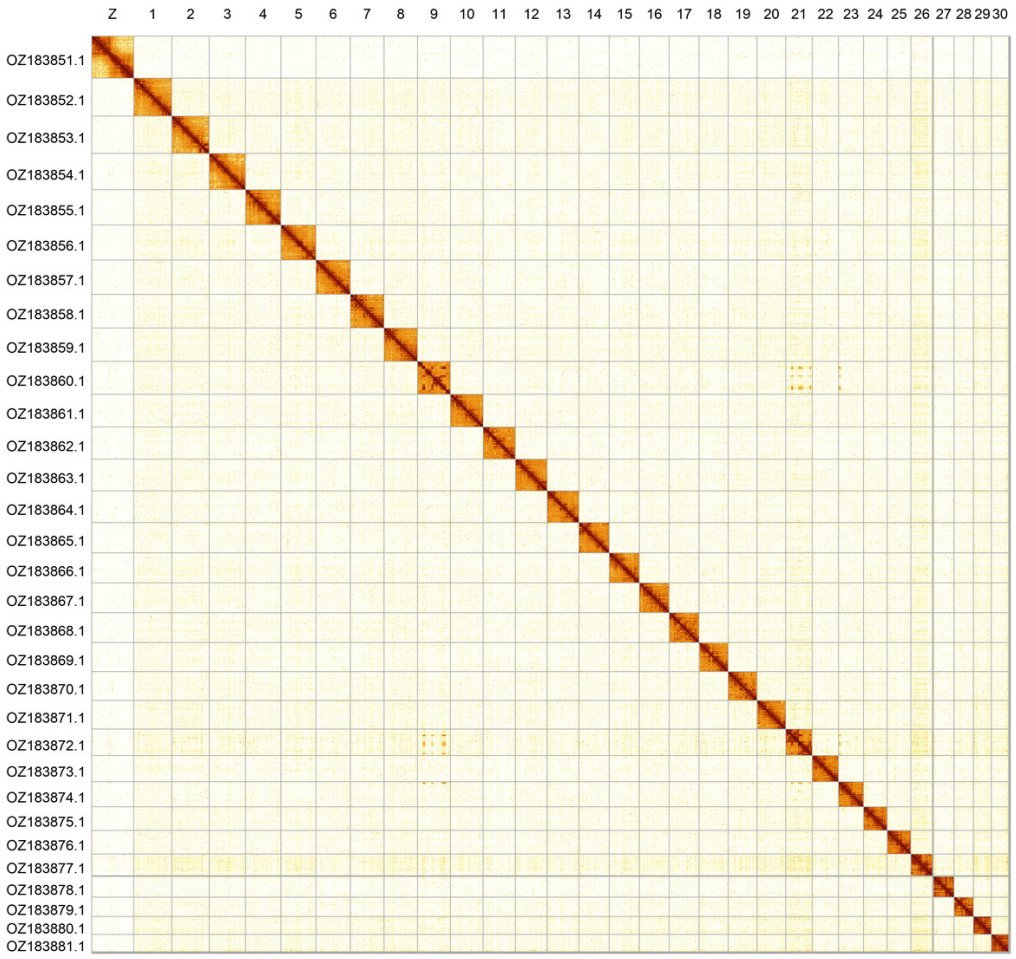

Figure 2. Hi-C contact map of the *Melitaea aurelia* genome assembly. Assembled chromosomes are shown in order of size and labelled along the axes. The plot was generated using PretextSnapshot.

Table 3. Chromosomal pseudomolecules in the haplotype 1 genome assembly of *Melitaea aurelia* ilMelAure1.

INSDC accession	Molecule	Length (Mb)	GC%	Assigned Merian elements
OZ183852.1	1	24.56	34.50	M2
OZ183853.1	2	24.31	34.50	M17;M20
OZ183854.1	3	23.47	34.50	M1
OZ183855.1	4	22.87	34	M8
OZ183856.1	5	22.80	34.50	M3
OZ183857.1	6	22.14	34	M9
OZ183858.1	7	21.87	34.50	M6
OZ183859.1	8	21.58	34.50	M7
OZ183860.1	9	21.42	34.50	M23
OZ183861.1	10	21.19	34	M12
OZ183862.1	11	20.80	34	M18
OZ183863.1	12	20.55	34	M5
OZ183864.1	13	20.52	34	M16
OZ183865.1	14	19.75	34	M21
OZ183866.1	15	19.43	34	M4
OZ183867.1	16	19.37	34.50	M10
OZ183868.1	17	19.32	34	M22
OZ183869.1	18	18.88	34	M15
OZ183870.1	19	18.63	35	M11
OZ183871.1	20	18.62	35	M14
OZ183872.1	21	17.14	37	M30
OZ183873.1	22	16.98	34.50	M13
OZ183874.1	23	16.27	34.50	M26
OZ183875.1	24	15.31	35	M19
OZ183876.1	25	15.19	34.50	M24
OZ183877.1	26	14.58	37	M31
OZ183878.1	27	13.54	34.50	M28
OZ183879.1	28	12.53	34.50	M27
OZ183880.1	29	11.47	35.50	M29
OZ183881.1	30	11.41	35.50	M25
OZ183851.1	Z	27.45	34	MZ

chromosomal evolution relative to Lepidopteran ancestral linkage groups (Figure 3).

The mitochondrial genome was also assembled. This sequence is included as a contig in the multifasta file of the genome submission and as a standalone record.

Assembly quality metrics

For haplotype 1, the estimated QV is 65.3, and for haplotype 2, 64.7. When the two haplotypes are combined, the assembly achieves an estimated QV of 65.0. The *k*-mer completeness is 69.90% for haplotype 1, 66.69% for haplotype 2, and 99.31% for the combined haplotypes (Figure 4). BUSCO

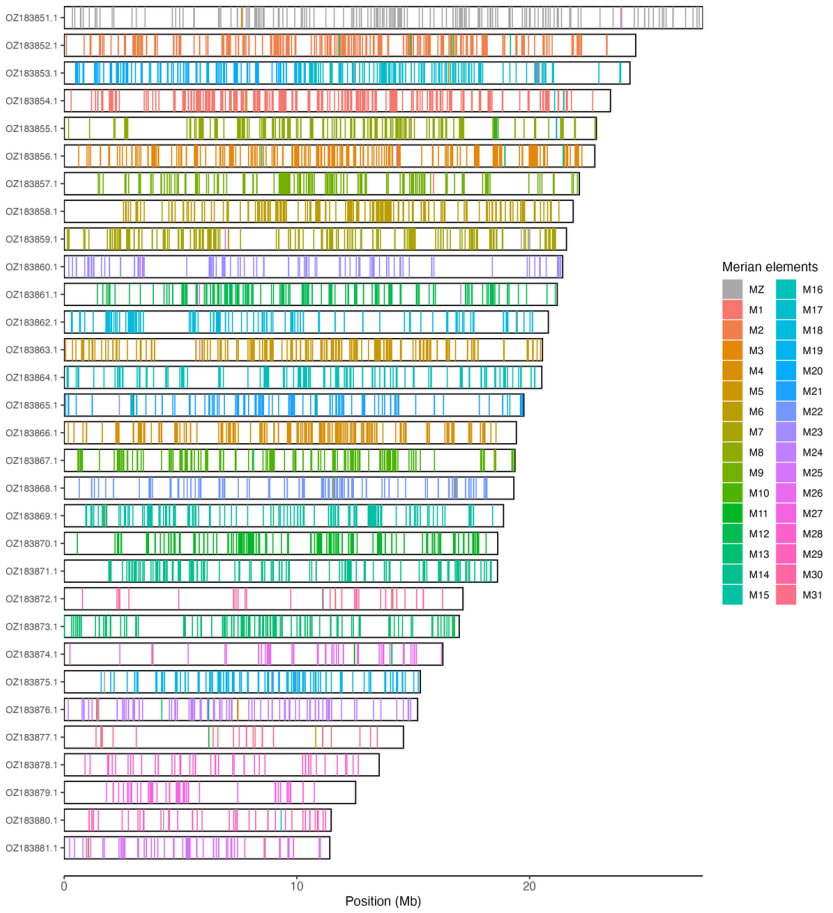

Figure 3. Merian elements painted across chromosomes in the ilMelAure1.hap1.1 assembly of *Melitaea aurelia*. Chromosomes are drawn to scale, with the positions of orthologues shown as coloured bars. Each orthologue is coloured by the Merian element that it belongs to. All orthologues which could be assigned to Merian elements are shown.

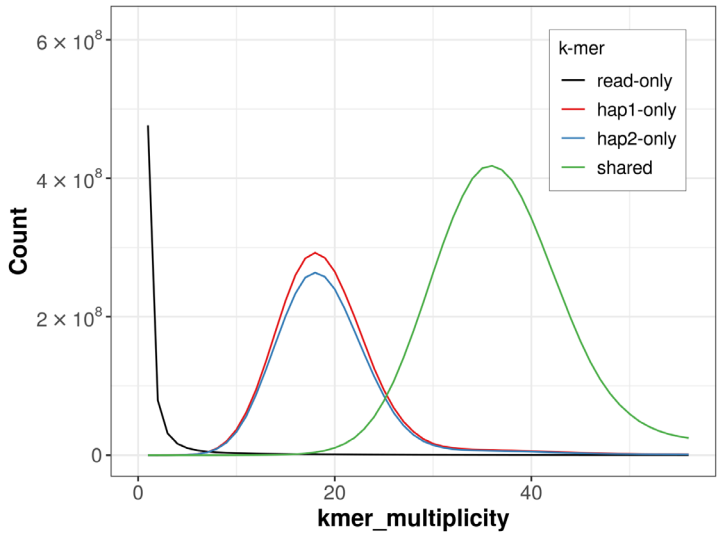

Figure 4. Evaluation of *k*-mer completeness using MerquryFK. This plot illustrates the recovery of *k*-mers from the original read data in the final assemblies. The horizontal axis represents *k*-mer multiplicity, and the vertical axis shows the number of *k*-mers. The black curve represents *k*-mers that appear in the reads but are not assembled. The green curve (the homozygous peak) corresponds to *k*-mers shared by both haplotypes and the red and blue curves (the heterozygous peaks) show *k*-mers found only in one of the haplotypes.

analysis using the lepidoptera_odb10 reference set (n = 5 286) (Kriventseva *et al.*, 2019) identified 98.4% of the expected gene set (single = 97.6%, duplicated = 0.8%) for haplotype 1. The snail plot in Figure 5 summarises the scaffold length distribution and other assembly statistics for haplotype 1. The blob plot in Figure 6 shows the distribution of scaffolds by GC proportion and coverage for haplotype 1.

Table 4 lists the assembly metric benchmarks adapted from Rhie *et al.* (2021) the Earth BioGenome Project Report on

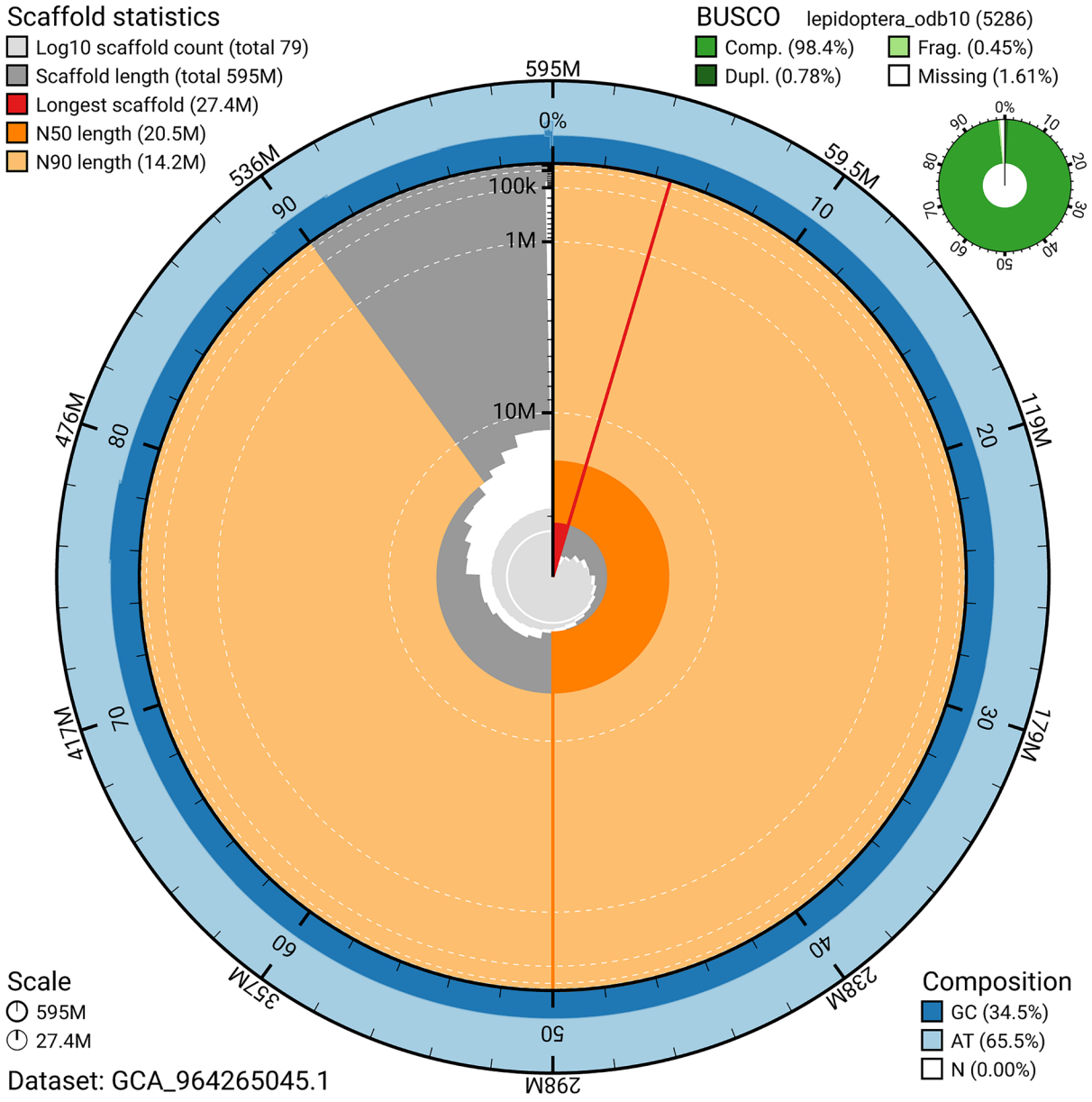

Figure 5. Assembly metrics for ilMelAure1.hap1.1. The BlobToolKit snail plot provides an overview of assembly metrics and BUSCO gene completeness. The circumference represents the length of the whole genome sequence, and the main plot is divided into 1,000 bins around the circumference. The outermost 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.

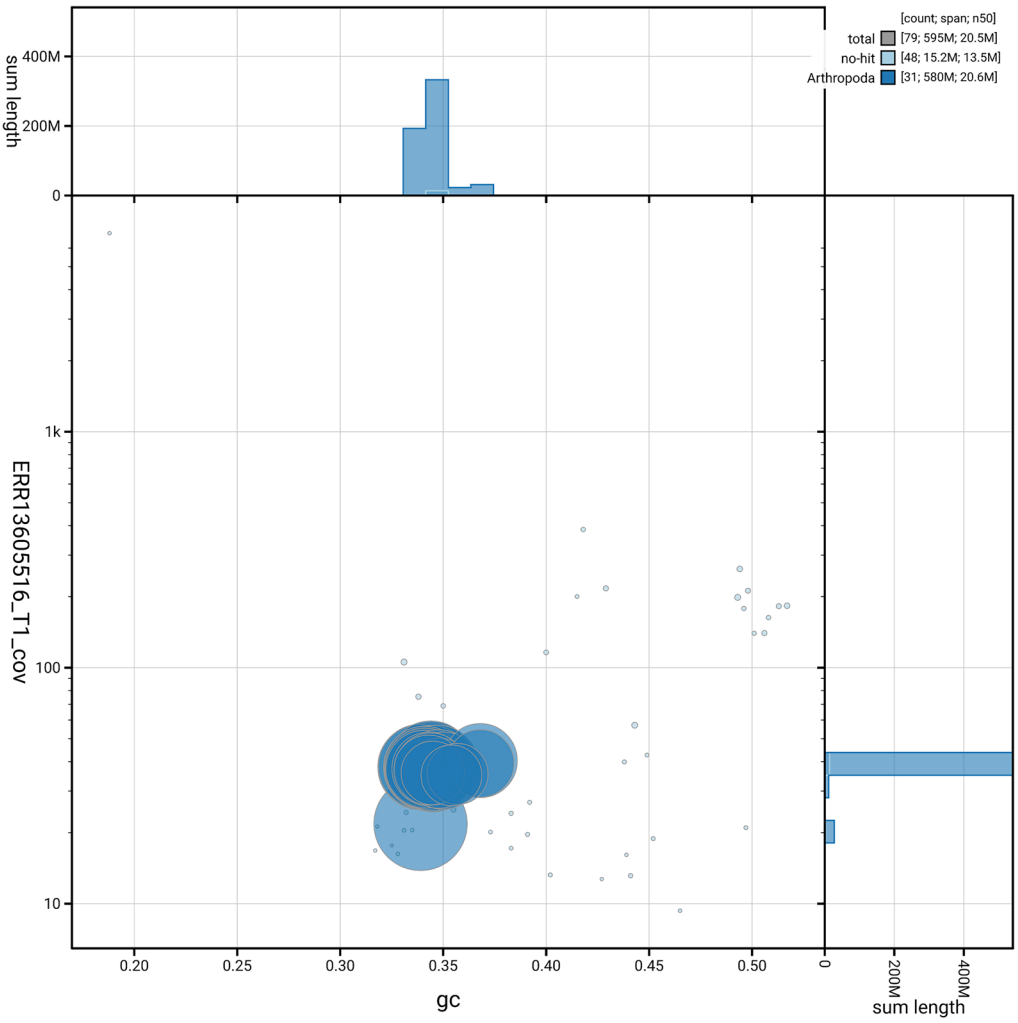

Figure 6. BlobToolKit GC-coverage plot for ilMelAure1.hap1.1. Blob plot showing sequence coverage (vertical axis) and GC content (horizontal axis). The circles represent scaffolds, with the size proportional to scaffold length and the colour representing phylum membership. The histograms along the axes 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 *Melitaea aurelia* assembly.

Measure	Value	Benchmark
EBP summary (haplotype 1)	6.7.Q65	6.C.Q40
Contig N50 length	8.60 Mb	≥ 1 Mb
Scaffold N50 length	20.52 Mb	= chromosome N50
Consensus quality (QV)	Haplotype 1: 65.3; haplotype 2: 64.7; combined: 65.0	≥ 40
k-mer completeness	Haplotype 1: 69.90%; Haplotype 2: 66.69%; combined: 99.31%	≥ 95%
BUSCO	C:98.4%[S:97.6%,D:0.8%],F:0.5%,M:1.2%,n:5286	S > 90%; D < 5%
Percentage of assembly assigned to chromosomes	99.72%	≥ 90%

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

Wellcome Sanger Institute – Legal and Governance

The materials that have contributed to this genome note have been supplied by a Tree of Life collaborator. The Wellcome Sanger Institute employs a process whereby due diligence is carried out proportionate to the nature of the materials themselves, and the circumstances under which they have been/are to be collected and provided for use. The purpose of this is to address and mitigate any potential legal and/or ethical implications of receipt and use of the materials as part of the research project, and to ensure that in doing so, we align with best practice wherever possible. The overarching areas of consideration are:

- Ethical review of provenance and sourcing of the material
- Legality of collection, transfer and use (national and international).

Each transfer of samples is undertaken according to a Research Collaboration Agreement or Material Transfer Agreement entered into by the Tree of Life collaborator, Genome Research Limited (operating as the Wellcome Sanger Institute), and in some circumstances, other Tree of Life collaborators.

Data availability

European Nucleotide Archive: Mellicta aurelia (Nickerl's fritillary). Accession number PRJEB79206. The genome sequence is released openly for reuse. The *Melitaea aurelia* 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

Software	Version	Source
MerquryFK	1.1.2	https://github.com/thegenemyers/MERQURY.FK
Minimap2	2.24-r1122	https://github.com/lh3/minimap2
MitoHiFi	3	https://github.com/marcelauliano/MitoHiFi
MultiQC	1.14; 1.17 and 1.18	https://github.com/MultiQC/MultiQC
Nextflow	23.10.0	https://github.com/nextflow-io/nextflow
PretextSnapshot	N/A	https://github.com/sanger-tol/PretextSnapshot
PretextView	0.2.5	https://github.com/sanger-tol/PretextView
samtools	1.19.2	https://github.com/samtools/samtools
sanger-tol/ascc	0.1.0	https://github.com/sanger-tol/ascc
sanger-tol/blobtoolkit	0.6.0	https://github.com/sanger-tol/blobtoolkit
Seqtk	1.3	https://github.com/lh3/seqtk
Singularity	3.9.0	https://github.com/sylabs/singularity
TreeVal	1.2.0	https://github.com/sanger-tol/treeval
YaHS	1.2a.2	https://github.com/c-zhou/yahs

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

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

Ebert G: **Die schmetterlinge Baden-Württembergs. Band 1.** Ulmer, 2011.

Eichel S, Fartmann T: **Management of calcareous grasslands for Nickerl's fritillary *Melitaea aurelia* has to consider habitat requirements of the immature stages, isolation, and patch area.** *J Insect Conserv.* 2008; **12**: 677–688.
Publisher Full Text

Essayan R, Jugan D, Mora F, *et al.*: **Atlas des papillons de jour de Bourgogne et de Franche-Comté (Rhopalocères et Zygènes).** Rev. Sci. Bourgogne-Nature Hors-série, 2013; **13**.
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

Fartmann T: **Die schmetterlingsgemeinschaften der halbtrockenrasen-komplexe des diemeltales. Biozönologie von tagfaltern und widderchen in einer alten hudelandschaft.** *Abhandlungen aus dem Westfälischen Museum für Naturkunde.* 2004; **66**.
Reference Source

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

Habel JC, Meyer M, Schmitt T: **The genetic consequence of differing ecological demands of a generalist and a specialist butterfly species.** *Biodivers Conserv.* 2009; **18**(7): 1895–1908.
Publisher Full Text

Höttinger H, Pennerstorfer J: **Rote liste der tagschmetterlinge Österreichs (Lepidoptera: Papilionoidea & Hesperioidea).** In: K. P. Zulka (ed), *Rote Listen Gefährdeter Tiere Österreichs. Teil 1.* Wien: Böhlau, 2005; 313–54.

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**

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: Écologie, biologie et comportement des Rhopalocères de France.** Paris: Diatheo, 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.** Bâle: Ligue Suisse pour la Protection de la Nature, 1987; **1**.
Reference Source

Manni M, Berkeley MR, Seppey M, *et al.*: **BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes.** *Mol Biol Evol.* 2021; **38**(10): 4647–4654.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Merkel D: **Docker: lightweight Linux containers for consistent development and deployment.** *Linux J.* 2014; **2014**(239): 2.
Reference Source

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

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

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

Rhie A, Walenz BP, Koren S, *et al.*: **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**

Spitzer L, Beneš J: **Nové a významné nálezy denních motýlů a vřetenuškovitých (Lepidoptera) na Valašsku (okres Vsetín, Česká Republika).** *Acta Carp.* 2010; **1**: 19–39.
Publisher Full Text

Tshikolovets VV: **Butterflies of Europe & the Mediterranean area.** Padubrice, Czech Republic: 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: **The importance of calcareous grasslands for butterflies in Europe.** *Biol Conserv.* 2002; **104**(3): 315–318.
Publisher 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

Wermeille E, Chittaro Y, Gonseth Y: **Liste rouge papillons diurnes et zygènes. Espèces menacées en Suisse, état 2012**. 2014; 97.
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**

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 03 October 2025

https://doi.org/10.21956/wellcomeopenres.27266.r133558

✔ **Sina Beier**
[1] MRC Toxicology Unit, Cambridge, UK
[2] Gurdon Institute (Ringgold ID: 47693), Cambridge, England, UK

The authors present a high quality genome of the Nickerl's Fritillary. It has been assembled to full chromosomal resolution using long read PacBio Sequencing paired with Hi-C sequencing for scaffolding and haplotyping.
As the specimen was female, only the W sex chromosome is missing from the data, with the Z chromosome and mitochondrium included.

The data is finished to high quality and will provide a valuable resource for studies of this endangered but highly locally adapted species.

It would be beneficial to have a supplementary map of the breaks and joins from manual curation to improve reproducibility.

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?
Partly

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 23 September 2025

https://doi.org/10.21956/wellcomeopenres.27266.r133556

Laurence Despres
[1] Universite Savoie Mont Blanc, Chambéry, Auvergne-Rhône-Alpes, France
[2] Laboratoire d'Ecologie Alpine (Ringgold ID: 56837), Université Grenoble Alpes, Grenoble, Auvergne-Rhône-Alpes, France

The authors report on the chromosome-level, haplotype-resolved genome sequence from a female specimen of Nickerl's Fritillary *Melitaea aurelia* (Nymphalidae). The final assembly includes 31 chromosomes : 30 autosomes and the chromosome Z. The W chromosome was not recovered and the female is supposed to be Z0. The mitogenome was also assembled. The quality of the assembly was further assessed by the proportion of complete BUSCO genes recovered, which is very high (~98% of Lepidoptera database BUSCO genes).
The genome assembly is of good quality and was constructed using appropriate methods : Pac Bio HiFi long reads (35-fold coverage) and Hi-C data to confirm scaffolding, following the high standard of the Darwin Tree of Life Project for quality checking and results presentation.
This high quality genome assembly will provide a reference genome for further genomic studies on this diversified genus of butterflies.

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: evolutionary genomics

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

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