GENOME NOTE

The genome sequence of the Knapweed Fritillary, *Melitaea phoebe* (Goeze, 1779) (Lepidoptera: Nymphalidae)

[version 1; peer review: 1 approved]

Paula Escuer[1], Mathieu Joron[2], Kay Lucek[1], 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]Centre d'Ecologie Fonctionnelle et Evolutive, Univ Montpellier, CNRS, EPHE, IRD, Montpellier, France
[3]Tree of Life, Wellcome Sanger Institute, Hinxton, England, UK

v1 **First published:** 04 Sep 2026, **11**:595
https://doi.org/10.12688/wellcomeopenres.27708.1
Latest published: 04 Sep 2026, **11**:595
https://doi.org/10.12688/wellcomeopenres.27708.1

Abstract

We present a genome assembly from a female specimen of *Melitaea phoebe* (Knapweed Fritillary; Arthropoda; Insecta; Lepidoptera; Nymphalidae). The assembly contains two haplotypes with total lengths of 505.71 megabases and 466.56 megabases. Most of the haplotype 1 assembly (99.73%) was scaffolded into 32 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 15.21 kilobases. Gene annotation of this assembly by Ensembl identified 13,375 protein-coding genes. This work is part of Project Psyche, a collaborative programme generating genomes for European butterflies and moths.

Keywords

Melitaea phoebe, Knapweed Fritillary, genome sequence, chromosomal, Lepidoptera

This article is included in the Tree of Life gateway.

Open Peer Review

Approval Status ✓

	1
version 1 04 Sep 2026	✓ view

1. **Automated Benchmarking**, Wellcome Sanger Institute, Hinxton, 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: Escuer P: Investigation, Resources; **Joron M**: Writing – Original Draft Preparation; **Lucek K**: Resources, 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 and PE received funding from the 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: Escuer P, Joron M, Lucek K *et al.* **The genome sequence of the Knapweed Fritillary, *Melitaea phoebe* (Goeze, 1779) (Lepidoptera: Nymphalidae) [version 1; peer review: 1 approved]** Wellcome Open Research 2026, **11**:595 https://doi.org/10.12688/wellcomeopenres.27708.1

First published: 04 Sep 2026, **11**:595 https://doi.org/10.12688/wellcomeopenres.27708.1

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; Nymphalinae; Melitaeini; Melitaeina; *Melitaea*; *Melitaea phoebe* (Goeze, 1779) (NCBI:txid113342)

Background

Melitaea phoebe (Denis & Schiffermüller, 1775), the Knapweed Fritillary, is a medium-sized butterfly in the family Nymphalidae. *M. phoebe* has a widespread Palaearctic distribution, ranging from western Europe to central and eastern Asia. *M. phoebe* also has a wide altitudinal range, being found from sea level up to 2,500 m, depending on latitude. The species has a Least Concern (LC) status on the European IUCN Red List (van Swaay *et al.*, 2025).

M. phoebe thrives in open habitats and dry meadows of deciduous woodland biomes, including subalpine prairies as well as agricultural landscapes, where its larval host plants grow. Larvae feed on various knapweed species (genus *Centaurea*) as well as many thistles in the genera *Cirsium* or *Carduus* (Lafranchis *et al.*, 2015; Tóth *et al.*, 2015). *M. phoebe* is multivoltine, typically producing two or sometimes three generations per year, with adults on the wing from mid-spring to late summer, depending on latitude.

Among the Melitaeini, *M. phoebe* stands as a large and robust species, with broad wings displaying a striking orange and black wing pattern. Although its size, shape and wing pattern make it relatively easy to distinguish from other co-occurring *Melitaea* species, this butterfly exhibits morphological variation across its range, making it a subject for studies on phenotypic plasticity and local variation, especially with regard to the effect of hybridisation with closely-related species (Hinojosa *et al.*, 2024; Pateman *et al.*, 2022; Tóth *et al.*, 2014). Taxonomically, *M. phoebe* is indeed part of a species- rich group containing several close allies, such as *M. ornata*, *M. pseudornata*, *M. telona* or *M. punica*, which undergo sustained gene flow with *M. phoebe* and whose relationships with *M. phoebe* are the subject of active investigation (Hinojosa *et al.*, 2022; Russell *et al.*, 2022; Russell & Tennent, 2016; Tóth *et al.*, 2017). More generally, the genus *Melitaea* has been used extensively as a model for understanding metapopulation dynamics, dispersal behaviour, and the impact of habitat fragmentation on genetic diversity.

This reference genome for *M. phoebe* will serve as a key resource for advancing our understanding of the genetic mechanisms underlying its ecological adaptations, phenotypic variation, and speciation. We present a chromosome-level genome sequence of the Knapweed Fritillary, *Melitaea phoebe*, sequenced as part of Project Psyche (Wright *et al.*, 2025). The sequence data were derived from a specimen (Figure 1) collected from Zwischbergen, Switzerland.


FS49170416
0
1
2
3
4

Figure 1. Voucher photograph of the *Melitaea phoebe* (ilMelPhoe1) specimen used for genome sequencing.

Methods

Sample acquisition

The specimen used for genome sequencing was an adult female *Melitaea phoebe* (specimen ID SAN28000093, ToLID ilMelPhoe1; Figure 1), collected from Zwischbergen, Switzerland (latitude 46.1694, longitude 8.1179; elevation 1,356 m) on 2023-07-11. The specimen was collected and identified by Paula Escuer (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 thorax of ilMelPhoe1 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 Femto Pulse system. For this sample, the final post-shearing DNA had a Qubit concentration of 61.26 ng/μL and a yield of 2,879.22 ng, with a fragment size of 16.2 kb. The 260/280 spectrophotometric ratio was 1.87, and the 260/230 ratio was 1.89. The Genomic Quality Number (GQN) was 6.7.

PacBio HiFi library preparation and sequencing

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

Library PSYCHE14901159 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 PSYCHE14901159 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 ilMelPhoe1 head 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 in Hi-C phasing mode (Cheng *et al.*, 2021, 2022), producing two haplotypes. Hi-C reads (Rao *et al.*, 2014) were mapped to 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 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 *Melitaea bellona* (NC_070111.1) 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). The two haplotype assemblies were combined for curation.

Manual corrections included ten breaks and 19 joins. This reduced the scaffold count by 10.0% and increased the scaffold N50 by 2.0%. The curation process is described at sanger-tol/curation-resources. A Hi-C contact map of the haplotype 1 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 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 haplotype 1 assembly was assessed using BlobToolKit (Challis *et al.*, 2020). PacBio reads were mapped to the haplotype 1 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 *M. phoebe* specimen generated 26.27 Gb (gigabases) from 2.97 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 479.55 Mb, with heterozygosity of 2.20% and repeat content of 28.99%; the full-model fit was 97.54% (Figure 2). The GenomeScope estimates guided expectations for the assembly. The sequencing data provided approximately 52× coverage of the haplotype 1 assembly. Hi-C sequencing produced 117.75 Gb from 389.89 million reads, which were used to scaffold the haplotype 1 assembly. Table 1 summarises the specimen and sequencing details.

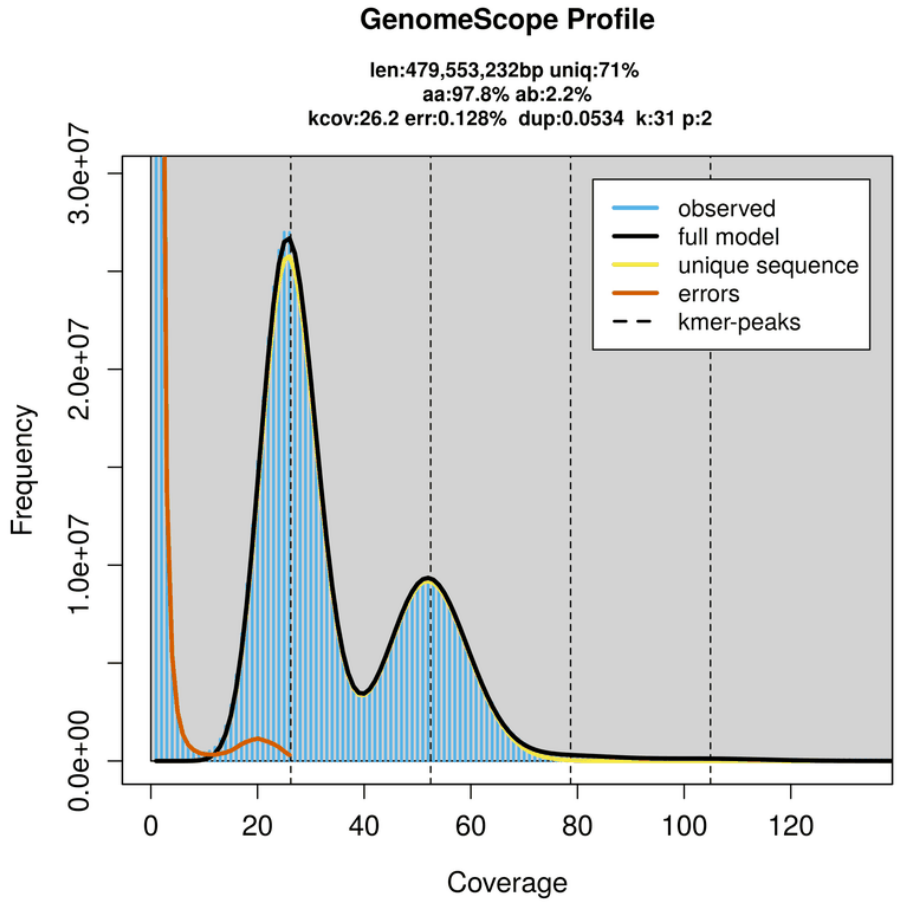

Figure 2. GenomeScope model fitted to the *k*-mer frequency distribution. The plot shows the observed and modelled *k*-mer spectra used to estimate genome size, heterozygosity and repeat content from unassembled sequencing reads.

Table 1. Specimen and sequencing data for *M. phoebe* (BioProject PRJEB79202)

Platform	PacBio HiFi	Hi-C
ToLID	ilMelPhoe1	ilMelPhoe1
Specimen ID	SAN28000093	SAN28000093
BioSample (source individual)	SAMEA115109877	SAMEA115109877
BioSample (tissue)	SAMEA115109915	SAMEA115109916
Tissue	thorax	head

Platform	PacBio HiFi	Hi-C
Instrument	Revio (1 run)	Illumina NovaSeq X
Library ID & run accession(s)	PSYCHE14901159: ERR13605514	ERR13602177
Read count total	2.97 million reads	389.89 million read pairs
Base count total	26.27 Gb	117.75 Gb

Assembly statistics

The genome was assembled into two haplotypes using Hi-C phasing. The haplotype 1 assembly was curated to chromosome level, while haplotype 2 was assembled to scaffold level. The haplotype 1 assembly, which is proposed as the reference assembly for this species, has a total length of 505.71 Mb in 80 scaffolds (excluding organelle sequences), with 68 gaps, and a scaffold N50 of 17.39 Mb (Table 2). The haplotype 1 assembly had a scaffold auN of 16.6 Mb.

Table 2. Genome assembly statistics for *M. phoebe*

Genome assembly	Haplotype 1	Haplotype 2
Assembly name	ilMelPhoe1.hap1.1	ilMelPhoe1.hap2.1
Assembly accession	GCA_964274925.1	GCA_964274915.1
Assembly level	chromosome	scaffold
Span (Mb)	505.71	466.56
Number of chromosomes	32	-
Number of contigs	148	122
Contig N50 (Mb)	8.94	10.45
Number of scaffolds (excluding organelle sequences)	80	71
Scaffold N50 (Mb)	17.39	17.28
Scaffold auN (Mb)	16.6	16.1
Longest scaffold length (Mb)	20.52	20.36
Sex chromosomes	W and Z	-
Organelles	Mitochondrial genome: 15.21 kb	-

Most of the haplotype 1 assembly sequence (99.73%) was assigned to 32 chromosome-level scaffolds, representing 30 autosomes and the W and Z sex chromosomes. These chromosome-level scaffolds, confirmed by Hi-C data, are named according to size (Figure 3; Table 3). Chromosome painting of the haplotype 1 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 haplotype 1 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, OZ188065.1). This sequence is included as a contig in the multifasta file of the haplotype 1 genome submission and as a standalone record.

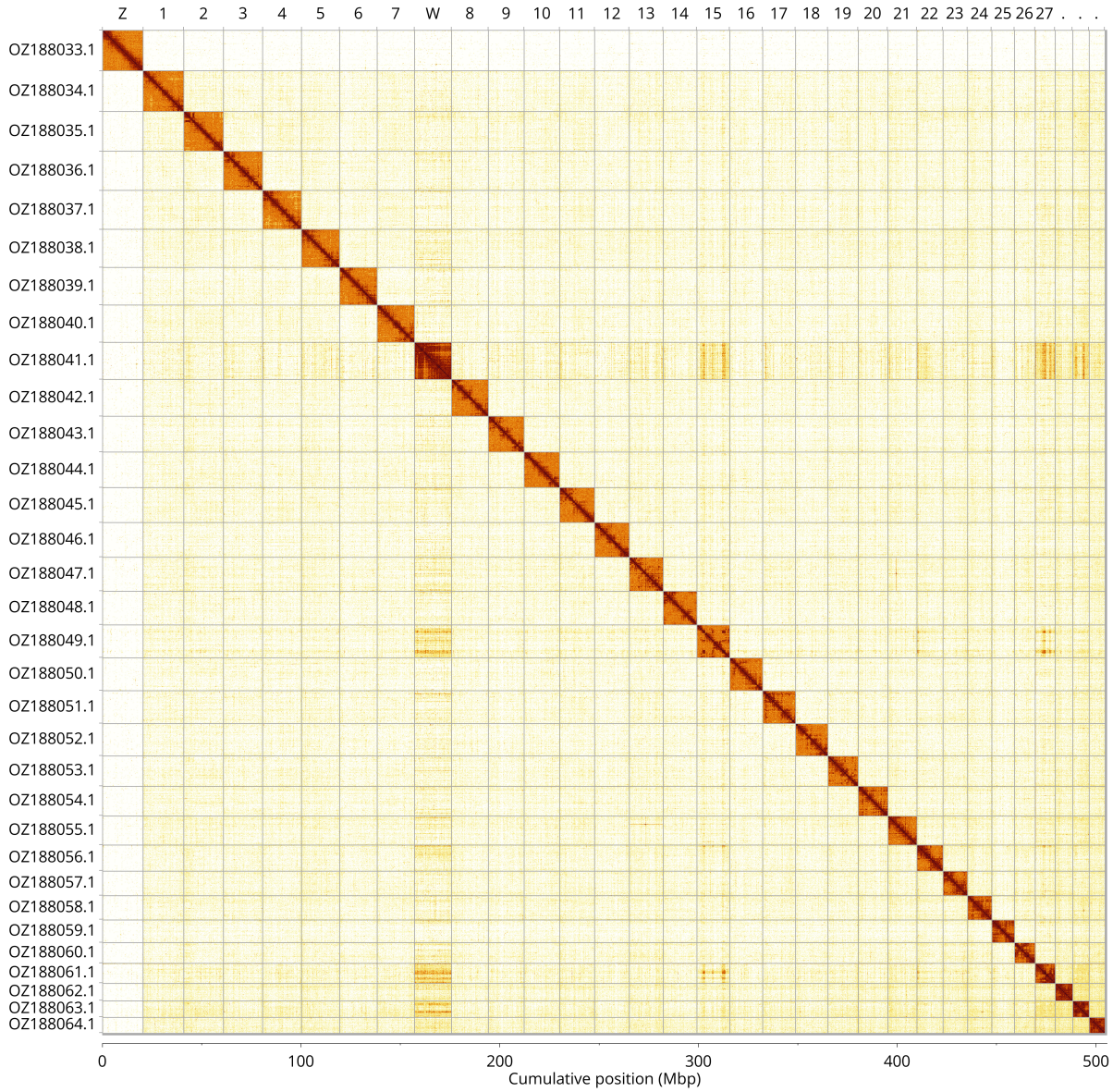

Figure 3. Hi-C contact map of the *M. phoebe* haplotype 1 assembly. Assembled chromosomes are shown in order of size and labelled along the axes. The plot was generated using PretextSnapshot.

Table 3. Chromosomal pseudomolecules in the haplotype 1 genome assembly of *M. phoebe* ilMelPhoe1

INSDC accession	Molecule	Length (Mb)	GC%	Assigned Merian elements
OZ188034.1	1	20.46	34	M2
OZ188035.1	2	20.01	34	M17; M20
OZ188036.1	3	19.67	34.50	M8
OZ188037.1	4	19.61	34.50	M1
OZ188038.1	5	19.13	34.50	M7
OZ188039.1	6	18.86	34	M9

INSDC accession	Molecule	Length (Mb)	GC%	Assigned Merian elements
OZ188040.1	7	18.76	34	M3
OZ188042.1	8	18.44	34	M12
OZ188043.1	9	18.05	34	M18
OZ188044.1	10	17.86	34	M5
OZ188045.1	11	17.58	34	M6
OZ188046.1	12	17.39	34	M16
OZ188047.1	13	17.04	34.50	M22
OZ188048.1	14	16.99	34	M21
OZ188049.1	15	16.56	35	M23
OZ188050.1	16	16.55	34	M4
OZ188051.1	17	16.54	34.50	M10
OZ188052.1	18	16.20	34.50	M15
OZ188053.1	19	15.35	34.50	M11
OZ188054.1	20	14.83	34.50	M13
OZ188055.1	21	14.64	35	M14
OZ188056.1	22	13.08	34.50	M26
OZ188057.1	23	12.39	34	M24
OZ188058.1	24	12.19	35	M19
OZ188059.1	25	11.44	35	M28
OZ188060.1	26	10.38	34.50	M27
OZ188061.1	27	10.05	37	M30
OZ188062.1	28	8.75	35.50	M25
OZ188063.1	29	8.35	37.50	M31
OZ188064.1	30	7.98	36.50	M29
OZ188041.1	W	18.69	41.50	-
OZ188033.1	Z	20.52	33.50	MZ

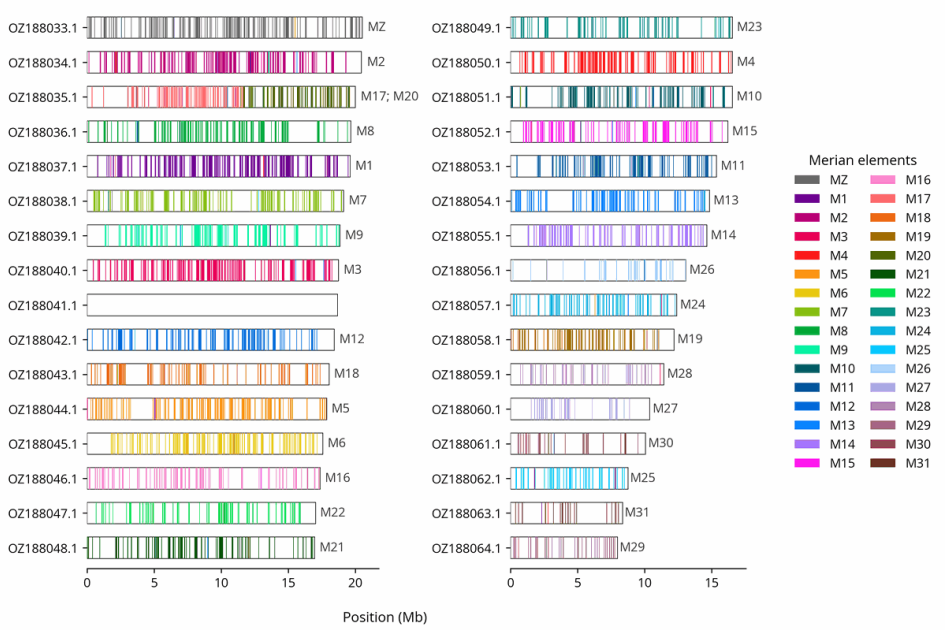

Figure 4. Merian elements painted across chromosomes in the ilMelPhoe1.hap1.1 haplotype 1 assembly of *M. phoebe*. Chromosomes are drawn to scale, with the positions of orthologues shown as coloured bars. Each orthologue is coloured by the Merian element that it belongs to. All orthologues which could be assigned to Merian elements are shown.

Assembly quality metrics

For the haplotype 1 assembly, the estimated QV is 64.9, and for the haplotype 2 assembly, 64.0. When the two haplotypes are combined, the assembly achieves an estimated QV of 64.4. The *k*-mer completeness is 68.77% for the haplotype 1 assembly, 65.05% for the haplotype 2 assembly, and 99.69% for the combined haplotypes (Figure 5).

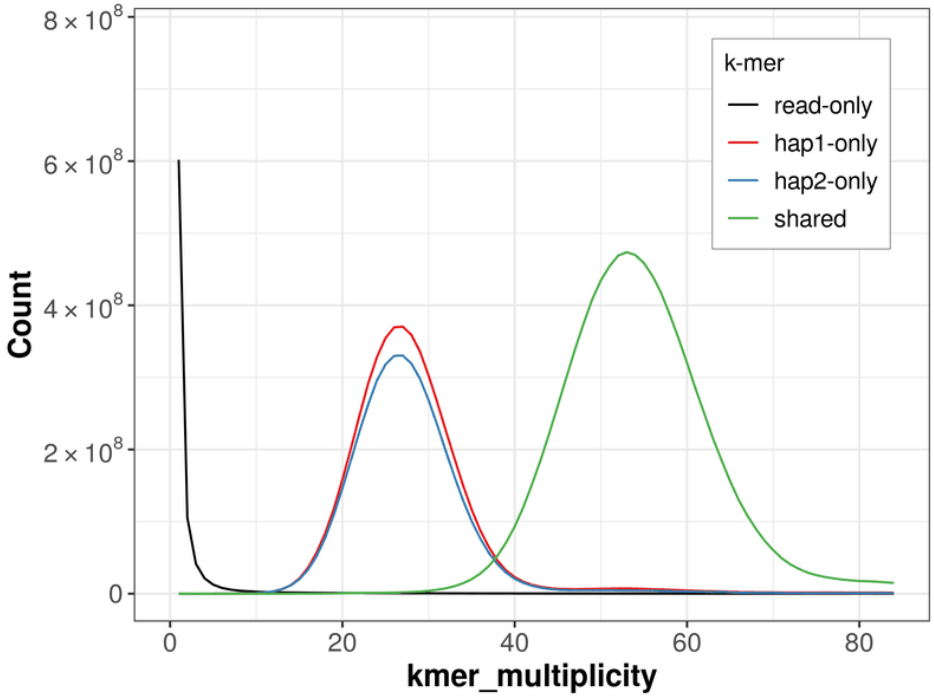

Figure 5. Evaluation of *k*-mer completeness using MerquryFK. This plot illustrates the recovery of *k*-mers from the original read data in the haplotype 1 assembly, the haplotype 2 assembly and the combined haplotypes. The horizontal axis represents *k*-mer multiplicity, and the vertical axis shows 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 analysis using the lepidoptera_odb10 reference set (n = 5,286) identified 98.7% of the expected gene set (single = 98.2%, duplicated = 0.5%) in the haplotype 1 assembly. For the haplotype 2 assembly, BUSCO analysis identified 94.4% of the expected gene set (single = 93.7%, duplicated = 0.6%).

The snail plot in Figure 6 provides a summary of assembly contiguity, cumulative scaffold count, sequence composition and BUSCO completeness for the haplotype 1 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 haplotype 1 assembly by GC proportion and base coverage. Marker area is proportional to scaffold length, and colour indicates taxonomic assignment (Challis *et al.*, 2020).

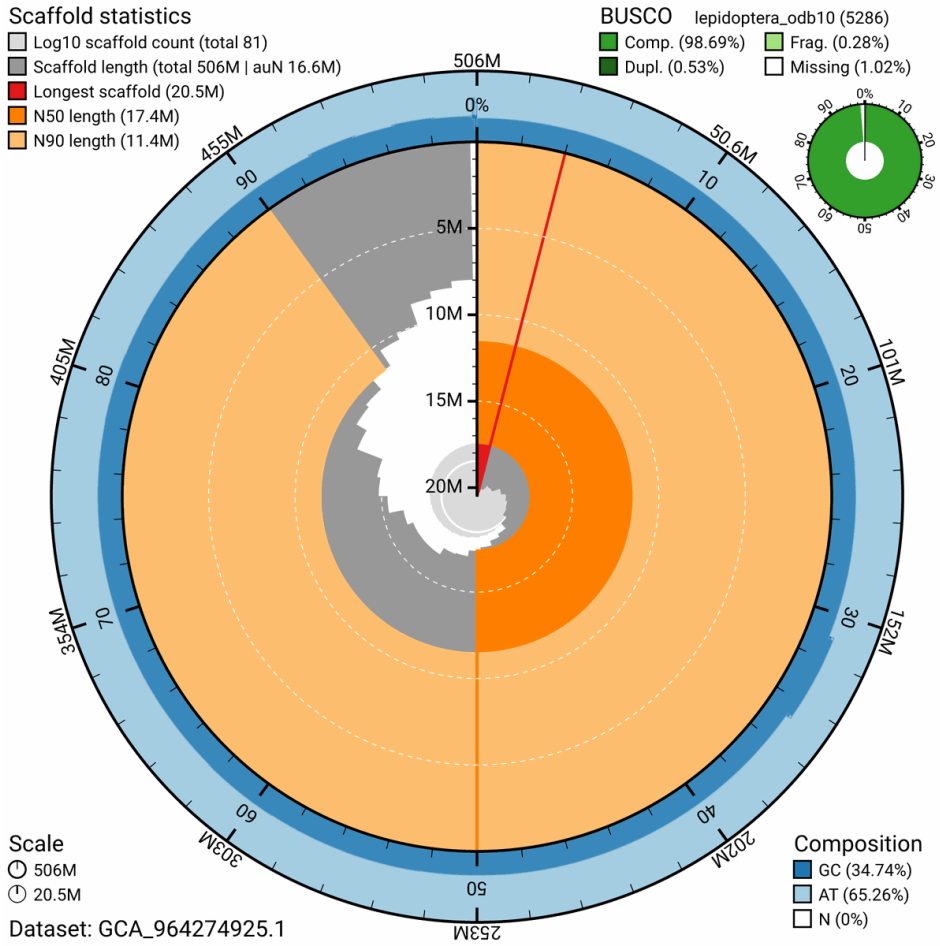

Figure 6. Assembly metrics for ilMelPhoe1.hap1.1, the haplotype 1 assembly. The BlobToolKit snail plot provides an overview of assembly metrics and BUSCO gene completeness. The circumference represents the length of the haplotype 1 assembly, and the main plot is divided into 1,000 bins around the circumference. 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.

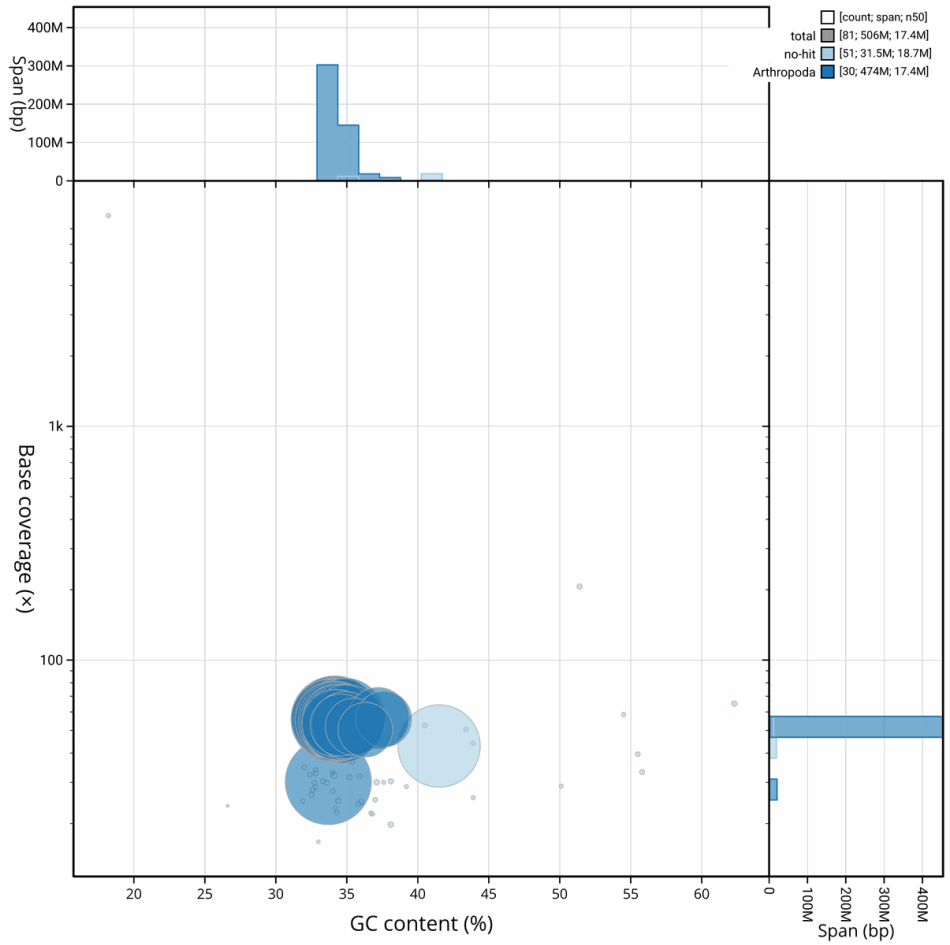

Figure 7. BlobToolKit blob plot for ilMelPhoe1.hap1.1, the haplotype 1 assembly. The plot shows base coverage (vertical axis) and GC content (horizontal axis). Circles represent scaffolds; marker area is proportional to scaffold length, and colour indicates taxonomic assignment. The histograms along 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 haplotype 1 assembly, is **6.C.Q64**.

Table 4. Earth BioGenome Project summary metrics for the *M. phoebe* assembly

Measure	Value	*Benchmark*
EBP summary (haplotype 1)	6.C.Q64	*6.C.Q40*
Contig N50 length	8.94 Mb	*> 1 Mb*
Scaffold N50 length	17.39 Mb	*= chromosome N50*
Consensus quality (QV)	Haplotype 1: 64.9; haplotype 2: 64.0; combined: 64.4	*# 40*
k-mer completeness	Haplotype 1: 68.77%; Haplotype 2: 65.05%; combined: 99.69%	*> 90%*
BUSCO	Haplotype 1: C:98.7% [S:98.2%, D:0.5%], F:0.3%, M:1.0%, n:5,286;	*S > 90%; D < 5%*

Measure	Value	*Benchmark*
	Haplotype 2: C:94.4% [S:93.7%, D:0.6%], F:0.3%, M:5.4%, n:5,286	
Percentage of assembly assigned to chromosomes	99.73%	*# 90%*

Notes: The EBP summary uses log10(Contig N50); chromosome-level (C) or log10(Scaffold N50); Q (Merqury QV). BUSCO v5.5.0 using the lepidoptera_odb10 lineage dataset: C=complete; S=single-copy; D=duplicated; F=fragmented; M=missing; n=orthologues.

Genome annotation report

The *M. phoebe* genome assembly (GCA_964274925.1) was annotated by Ensembl at the European Bioinformatics Institute (EBI). The annotation comprises 28,946 transcripts from 13,375 protein-coding genes and 5,381 non-coding genes. The average transcript length is 16,790.81 bp, with an average of 1.54 coding transcripts per gene and 6.39 exons per transcript. For further information, please refer to the Ensembl annotation page.

Data availability

European Nucleotide Archive: Melitaea phoebe (knapweed fritillary). Accession number: PRJEB79202; URL: https://identifiers.org/ena.embl/PRJEB79202. The genome sequence is released openly for reuse. The *M. phoebe* genome sequencing initiative is part of the Sanger Institute Tree of Life Programme (PRJEB43745) and Project Psyche (PRJEB71705). All raw sequence data and the haplotype 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 *M. phoebe*

Software	Version	Source
BLAST	2.14.0	https://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/
BlobToolKit	4.3.9	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	5.5.0	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

Software	Version	Source
MerquryFK	1.1.0-c1	https://github.com/thegenemyers/MERQURY.FK
Minimap2	2.24-r1122	https://github.com/lh3/minimap2
MitoHiFi	3.2.2	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	0.0.6	https://github.com/sanger-tol/PretextSnapshot
PretextView	1.0.3	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
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	1.1.1	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:

- 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 |

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., Jarvis E. D., Fedrigo O.: **Haplotype-resolved assembly of diploid genomes without parental data**. *Nature Biotechnology*; 2022; **40**(9): 1332–1335. 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 |

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 |

Hinojosa J. C., Tóth J. P., Monasterio Y.: **Integrative taxonomy reveals a new *Melitaea* (Lepidoptera: Nymphalidae) species widely distributed in the Iberian Peninsula**. *Insect Systematics and Diversity*; 2022; **6**(2): ixac004. Publisher Full Text |

Hinojosa J. C., Marques V., Mesa L. S.: **Can species endure massive introgression? Genomic evidence of asymmetric gene flow in *Melitaea* butterflies**. *Systematic Entomology*; 2024; **49**(4): 583–595. 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 |

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 |

Lafranchis T., Jutzeler D., Guillosson J.-Y.: *La vie des papillons: écologie, biologie et comportement des Rhopalocères de France*. Paris: Diathéo; 2015.

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 |

Pateman J. A., Russell P. J. C., Tennent W. J.: **Some observations on hybridisation between closely related species of Rhopalocera (Lepidoptera: Lycaenidae, Nymphalidae)**. *SHILAP Revista de Lepidopterología*; 2022; **50**(198): 313–320. 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 |

Russell P., Lukhtanov V. A., Tennent W. J.: **Reassessment of the status of some European and Asian *Melitaea* taxa described as subspecies of *Melitaea phoebe* ([Denis & Schiffermüller], 1775), with designations of lectotypes where appropriate (Lepidoptera: Nymphalidae)**. *Zootaxa*; 2022; **5141**(1): 25–38. Publisher Full Text |

Russell P., Tennent W. J.: **A synonymic list of names associated with Western Palaearctic *Melitaea phoebe* (Denis & Schiffermüller, 1775) species group taxa (*M. phoebe*, *M. punica* Oberthür, 1876, *M. ornata* Christoph, 1893) (Lepidoptera, Nymphalidae)**. *Nota Lepidopterologica*; 2016; **39**(1): 27–56. Publisher Full Text |

Tóth J. P., Bereczki J., Varga Z.: **Relationships within the *Melitaea phoebe* species group (Lepidoptera: Nymphalidae): New insights from molecular and morphometric information**. *Systematic Entomology*; 2014; **39**(4): 749–757. Publisher Full Text |

Tóth J. P., Bereczki J., Végvári Z.: **Different host plant utilization ability of two closely related *Melitaea* species (Lepidoptera: Nymphalidae)**. *European Journal of Entomology*; 2015; **112**(1): 120–125. Publisher Full Text |

Tóth J. P., Varga Z., Verovnik R.: **Mito-nuclear discordance helps to reveal the phylogeographic patterns of *Melitaea ornata* (Lepidoptera: Nymphalidae)**. *Biological Journal of the Linnean Society*; 2017; **121**(2): 267–281. Publisher Full Text |

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.: ***Melitaea phoebe* (Europe assessment)**. *The IUCN Red List of Threatened Species*; 2025; **2025**: e.T274914748A211396008. 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 04 September 2026

https://doi.org/

✔ **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 Melitaea phoebe (ilMelPhoe1), all three checks passed.

Data available: pass. ENA study PRJEB79202, raw-read accessions ERR13605514 and ERR13602177, and assembly accessions GCA_964274925.1 and GCA_964274915.1 are public.

Earth BioGenome Project three-part metric: pass. Haplotype 1 is reported as 6.C.Q64, meeting the 6.C.Q40 reference standard.

BUSCO: pass. Haplotype 1, which is proposed as the reference assembly for this species, has 98.7% complete BUSCOs (98.2% single-copy and 0.5% 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.