



DATA NOTE

The genome sequence of the Silver-washed Fritillary, *Argynnis paphia* (Linnaeus, 1758) (Lepidoptera: Nymphalidae)

[version 1; peer review: 2 approved]

Yannick Chittaro ¹, Andreas Sanchez¹, Kay Lucek ², Charlotte J. Wright ³, Joana I. Meier³, 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

¹Info fauna, Neuchâtel, Switzerland

²Biodiversity Genomics Laboratory, Institute of Biology, University of Neuchâtel, Neuchâtel, Switzerland

³Tree of Life, Wellcome Sanger Institute, Hinxton, England, UK

V1 First published: 31 Jul 2025, 10:399
<https://doi.org/10.12688/wellcomeopenres.24635.1>
Latest published: 31 Jul 2025, 10:399
<https://doi.org/10.12688/wellcomeopenres.24635.1>

Abstract

We present a genome assembly from a female specimen of *Argynnis paphia* (Silver-washed Fritillary; Arthropoda; Insecta; Lepidoptera; Nymphalidae). The assembly contains two haplotypes with total lengths of 520.88 megabases and 444.74 megabases. Most of haplotype 1 (97.96%) is scaffolded into 30 chromosomal pseudomolecules, including the W, Z₁, and Z₂ sex chromosomes. Haplotype 2 was assembled to scaffold level. The mitochondrial genome has also been assembled, with a length of 15.22 kilobases

Keywords

Argynnis paphia, Silver-washed Fritillary, genome sequence, chromosomal, Lepidoptera



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

Open Peer Review

Approval Status

	1	2
version 1		
31 Jul 2025	view	view

1. Jason Charamis , Foundation for Research and Technology - Hellas, Irákleion, Greece
2. Jaakko Pohjoismäki , University of Eastern Finland, Joensuu, Finland

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 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.

Copyright: © 2025 Chittaro Y *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Chittaro Y, Sanchez A, Lucek K *et al.* **The genome sequence of the Silver-washed Fritillary, *Argynnis paphia* (Linnaeus, 1758) (Lepidoptera: Nymphalidae) [version 1; peer review: 2 approved]** Wellcome Open Research 2025, 10:399 <https://doi.org/10.12688/wellcomeopenres.24635.1>

First published: 31 Jul 2025, 10:399 <https://doi.org/10.12688/wellcomeopenres.24635.1>

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; Heliconiinae; Argynnini; *Argynnis*; *Argynnis paphia* (Linnaeus, 1758) (NCBI:txid171802)

Background

The Silver-washed Fritillary (*Argynnis paphia*) is a widely distributed species, occurring from North Africa (Algeria) across Europe to the temperate regions of Asia to Japan (Tshikolovets, 2011). This is a large Fritillary species that can be easily distinguished from others by examining the underside of the hindwings, which are metallic green and streaked with pearly striae. The male has four broad, parallel androconial striae, which release pheromones, on the upper surface of the forewings, while the slightly larger, duller-coloured female has none.

Argynnis paphia is a mesophilous forest species, occurring in most types of open forest, along forest roads, edges and clearings, including near wet meadows. To thrive, the species needs a mosaic of structures combining both areas of clear undergrowth rich in violets (*Viola* spp.), shadier forest surfaces for egg-laying, but also, in the immediate vicinity, places with lush vegetation where grow the nectar-bearing plants on which the adults feed (Lafranchis *et al.*, 2015; LSPN, 1987). Although isolated individuals are sometimes observed at elevations of over 2 000 m in the Alps, the species develops mainly in the colline and montane levels and rarely exceeds 1 500 m of elevation.

Caterpillars develop mainly on various species of violets (*Viola* spp.), but meadowsweet (*Filipendula ulmaria*) and raspberry (*Rubus idaeus*) have also been cited as hosts (Clarke, 2024; Ebert & Rennwald, 1991). Females do not usually lay eggs directly on the host plants used by the caterpillars, but more often in crevices in the bark of various trees, sometimes several metres high. According to Thomas and Lewington (2014), females choose particularly shady locations to lay their eggs. The species has only one generation a year, spread out over the course of the year, with adults living for several weeks. Widespread and undemanding, the species is not threatened in Europe (LC according to Van Swaay *et al.*, 2010). Locally, however, it may be threatened by the destruction of certain forest areas and also by over-intensive management of edges and roadsides. The reference genome presented here will allow to investigate the biogeography of this species and provide an invaluable resource for understanding the evolution of Lepidoptera. The sequence data was derived from a female specimen (Figure 1) collected from Barthélemy NE, Switzerland.

Methods

Sample acquisition and DNA barcoding

The specimen used for genome sequencing was an adult female *Argynnis paphia* (specimen ID SAN28000137, ToLID



Figure 1. Voucher photograph of the *Argynnis paphia* (ilArgPaph1) specimen used for genome sequencing.

ilArgPaph1; Figure 1), collected from Barthélemy NE, Switzerland (latitude 46.969, longitude 6.5462; elevation 1 050 m) on 10/08/2023. The specimen was collected and identified by Yannick Chittaro (Info Fauna, 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 ilArgPaph1 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.

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 47.99 ng/μL and a yield of 2 255.53 ng, with a fragment size of 14.0 kb. The 260/280 spectrophotometric ratio was 1.88, and the 260/230 ratio was 2.67.

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 ilArgPaph1 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).

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.

Table 1. Specimen and sequencing data for BioProject PRJEB78670.

Platform	PacBio HiFi	Hi-C
ToLID	ilArgPaph1	ilArgPaph1
Specimen ID	SAN28000137	SAN28000137
BioSample (source individual)	SAMEA115110003	SAMEA115110003
BioSample (tissue)	SAMEA115110022	SAMEA115110021
Tissue	thorax	head
Sequencing platform and model	Revio	Illumina NovaSeq X
Run accessions	ERR13485706	ERR13474171
Read count total	1.69 million	909.98 million
Base count total	16.57 Gb	137.41 Gb

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 mis-assemblies. 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 50 joins. The curation process is documented at <https://gitlab.com/wtsi-grit/rapid-curation>. PretextViewSnapshot was used to generate a Hi-C contact map of the final assembly.

Assembly quality assessment

Chromosomal painting was performed using lep_buscoPainter using Merian elements, which represent the 32 ancestral linkage groups in Lepidoptera (Wright *et al.*, 2024). Painting was based on gene locations from the lepidoptera_odb10 BUSCO analysis and chromosome lengths from the genome index produced using SAMtools faidx (Danecek *et al.*, 2021). Each complete BUSCO (including both single-copy and duplicated BUSCOs) was assigned to a Merian element using a reference database, and coloured positions were plotted along chromosomes drawn to scale.

The Merqury.FK tool (Rhie *et al.*, 2020), run in a Singularity container (Kurtzer *et al.*, 2017), was used to evaluate *k*-mer completeness and assembly quality for both haplotypes using the *k*-mer databases (*k* = 31) computed prior to genome assembly. The analysis outputs included assembly QV scores and completeness statistics.

The genome was analysed using the BlobToolKit pipeline, a Nextflow implementation of the earlier Snakemake BlobToolKit pipeline (Challis *et al.*, 2020). The pipeline aligns PacBio reads using minimap2 (Li, 2018) and SAMtools (Danecek *et al.*, 2021) to generate coverage tracks. Simultaneously, it queries the GoT database (Challis *et al.*, 2023) to identify relevant BUSCO lineages and runs BUSCO

(Manni *et al.*, 2021). For the three domain-level BUSCO lineages, BUSCO genes are aligned to the UniProt Reference Proteomes database (Bateman *et al.*, 2023) using DIAMOND blastp (Buchfink *et al.*, 2021). The genome is divided into chunks based on the density of BUSCO genes from the closest taxonomic lineage, and each chunk is aligned to the UniProt Reference Proteomes database with DIAMOND blastx. Sequences without hits are chunked using seqtk and aligned to the NT database with blastn (Altschul *et al.*, 1990). The BlobToolKit suite consolidates all outputs into a blobdir for visualisation. The BlobToolKit pipeline was developed using nf-core tooling (Ewels *et al.*, 2020) and MultiQC (Ewels *et al.*, 2016), with package management via Conda and Bioconda (Grüning *et al.*, 2018), and containerisation through Docker (Merkel, 2014) and Singularity (Kurtzer *et al.*, 2017).

Genome sequence report

Sequence data

The genome of a specimen of *Argynnis paphia* was sequenced using Pacific Biosciences single-molecule HiFi long reads, generating 16.57 Gb (gigabases) from 1.69 million reads, which were used to assemble the genome. GenomeScope2.0 analysis estimated the haploid genome size at 479.84 Mb, with a heterozygosity of 1.41% and repeat content of 29.59%. These estimates guided expectations for the assembly. Based on the estimated genome size, the sequencing data provided approximately 33× coverage. Hi-C sequencing produced 137.41 Gb from 909.98 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 520.88 Mb in 140 scaffolds, with 169 gaps, and a scaffold N50 of 17.51 Mb (Table 2).

Most of the assembly sequence (97.96%) was assigned to 30 chromosomal-level scaffolds, representing 27 autosomes and the W, Z₁ and Z₂ sex chromosomes. 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 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 59.9, and for haplotype 2, 60.0. When the two haplotypes are combined, the assembly

Table 2. Genome assembly statistics.

Assembly name	ilArgPaph1.hap1.1	ilArgPaph1.hap2.1
Assembly accession	GCA_964264515.1	GCA_964264475.1
Assembly level	chromosome	scaffold
Span (Mb)	520.88	444.74
Number of chromosomes	30	N/A
Number of contigs	309	199
Contig N50	5.47 Mb	6.22 Mb
Number of scaffolds	140	95
Scaffold N50	17.51 Mb	16.5 Mb
Longest scaffold length (Mb)	37.79	N/A
Sex chromosomes	W; Z ₁ and Z ₂	N/A
Organelles	Mitochondrial genome: 15.22 kb	N/A

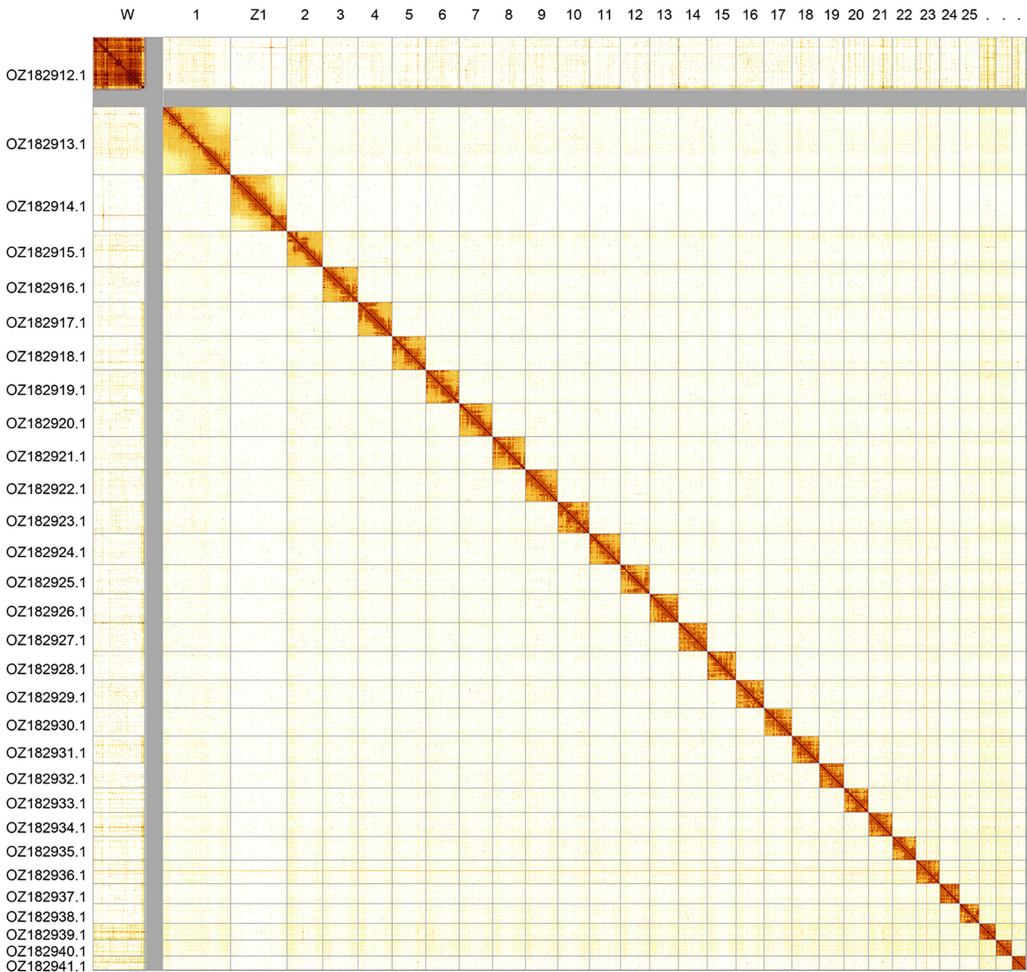


Figure 2. Hi-C contact map of the *Argynnis paphia* genome assembly. Assembled chromosomes are shown in order of size. The map was produced in HiGlass.

Table 3. Chromosomal pseudomolecules in the haplotype 1 genome assembly of *Argynnis paphia* ilArgPaph1.

INSDC accession	Molecule	Length (Mb)	GC%	Assigned Merian elements
OZ182913.1	1	37.79	32.50	M5;M6
OZ182915.1	2	19.96	33	M17;M20
OZ182916.1	3	19.58	32.50	M2
OZ182917.1	4	19.04	33	M1
OZ182918.1	5	18.81	33	M3
OZ182919.1	6	18.59	32.50	M9
OZ182920.1	7	18.57	32.50	M8
OZ182921.1	8	18.35	32.50	M12
OZ182922.1	9	18.10	32.50	M7
OZ182923.1	10	17.51	32.50	M18
OZ182924.1	11	17.39	32.50	M16
OZ182925.1	12	16.28	32.50	M4
OZ182926.1	13	16.08	33	M10
OZ182927.1	14	16.08	33	M22
OZ182928.1	15	16.08	32.50	M21
OZ182929.1	16	15.58	33	M11
OZ182930.1	17	15.51	32.50	M15
OZ182931.1	18	15.22	33	M14
OZ182932.1	19	13.84	32.50	M13
OZ182933.1	20	13.54	33	M23
OZ182934.1	21	13.53	33.50	M19
OZ182935.1	22	13.13	32.50	M26
OZ182936.1	23	13.11	33	M24
OZ182937.1	24	11.13	32.50	M28
OZ182938.1	25	10.99	32.50	M27
OZ182939.1	26	9.19	35	M30
OZ182940.1	27	9.04	34	M29
OZ182912.1	W	29	37	N/A
OZ182914.1	Z ₁	31.56	33	M25;MZ
OZ182941.1	Z ₂	7.65	34.50	M31
OZ182942.1	MT	0.02	18.50	N/A

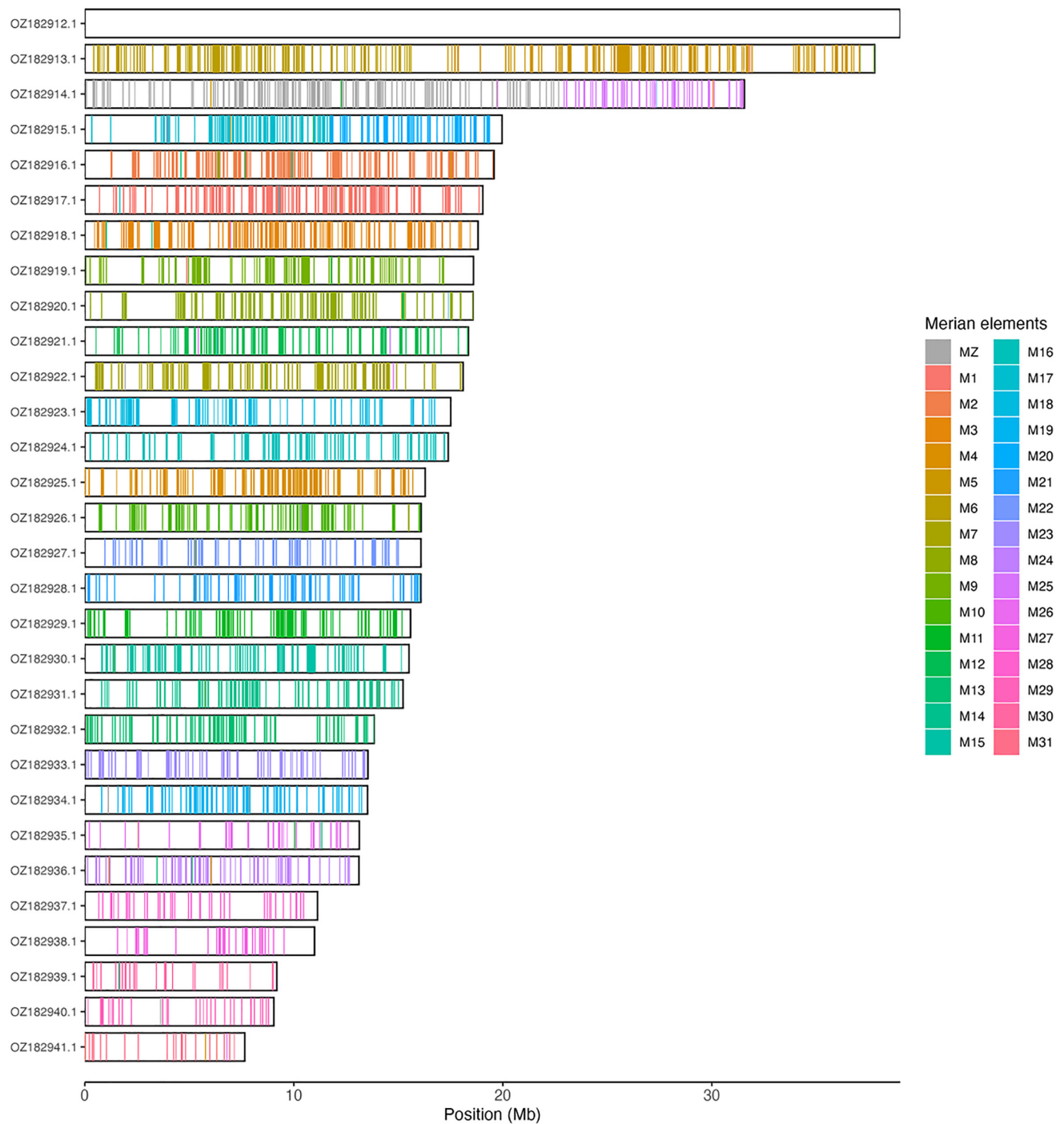


Figure 3. Merian elements painted across chromosomes in the *ilArgPaph1.hap1.1* assembly of *Argynnis paphia*. Chromosomes are drawn to scale, with the positions of orthologues shown as coloured bars. Each orthologue is coloured by the Merian element that it belongs to. All orthologues which could be assigned to Merian elements are shown.

achieves an estimated QV of 60.0. The *k*-mer completeness is 76.69% for haplotype 1, 70.22% for haplotype 2, and 99.64% for the combined haplotypes (Figure 4). BUSCO analysis using the *lepidoptera_odb10* reference set ($n = 5\,286$)

(Kriventseva *et al.*, 2019) identified 98.8% of the expected gene set (single = 98.0%, duplicated = 0.8%) for haplotype 1. The snail plot in Figure 5 summarises the scaffold length distribution and other assembly statistics for haplotype 1.

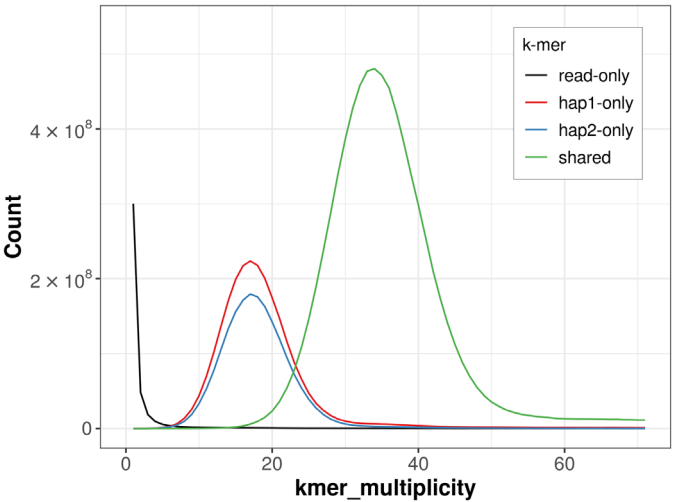


Figure 4. Evaluation of *k*-mer completeness using MerquyFK. This plot illustrates the recovery of *k*-mers from the original read data in the final assemblies. The horizontal axis represents *k*-mer multiplicity, and the vertical axis shows the number of *k*-mers. The black curve represents *k*-mers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

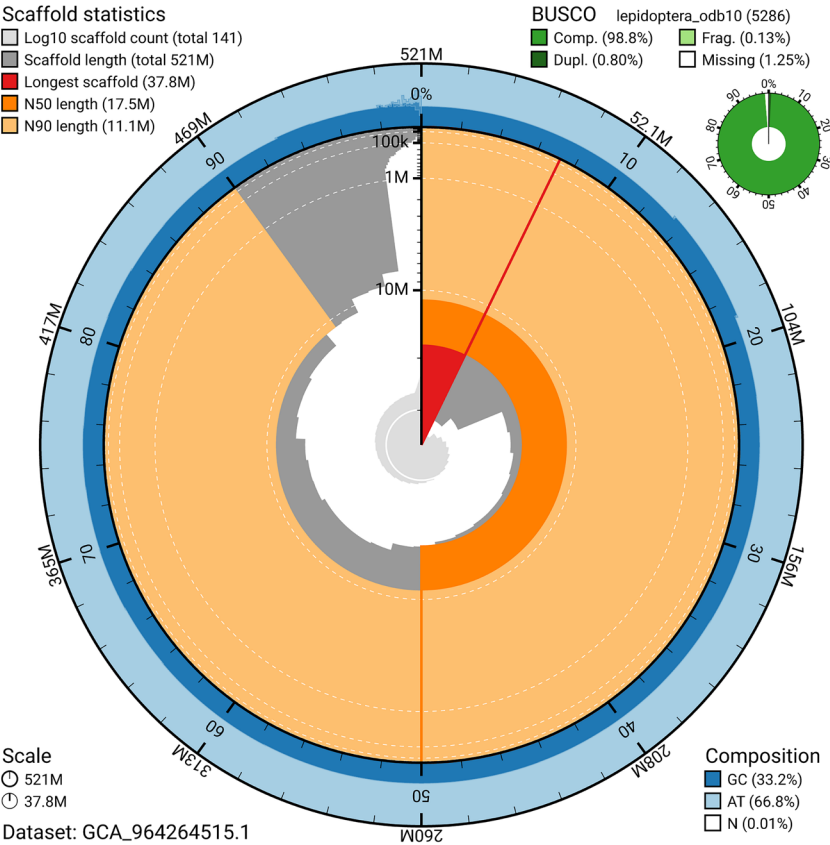


Figure 5. Assembly metrics for *ilArgPaph1.hap1.1*. The BlobToolKit snail plot provides an overview of assembly metrics and BUSCO gene completeness. The circumference represents the length of the whole genome sequence, and the main plot is divided into 1,000 bins around the circumference. The outermost 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](#).

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

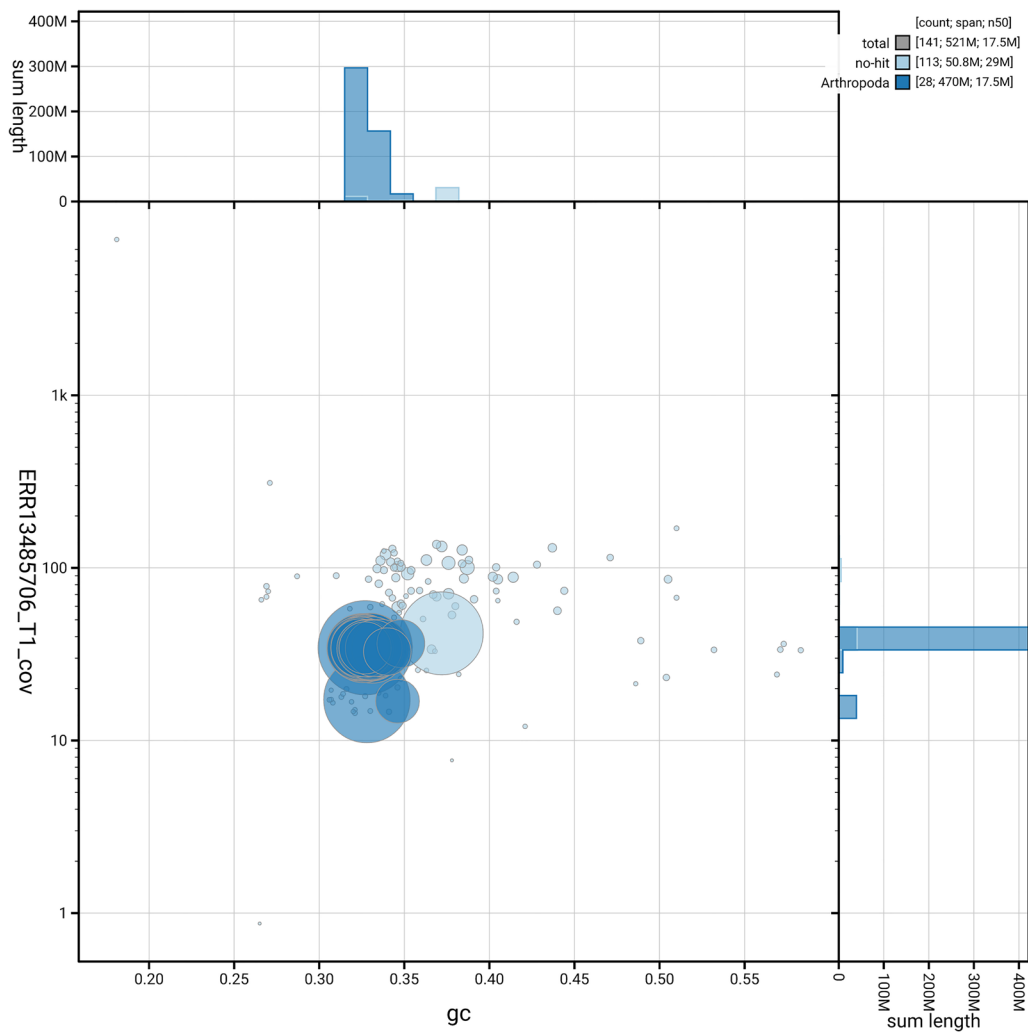


Figure 6. BlobToolKit GC-coverage plot for ilArgPaph1.hap1.1. Blob plot showing sequence coverage (vertical axis) and GC content (horizontal axis). The circles represent scaffolds, with the size proportional to scaffold length and the colour representing phylum membership. The histograms along the axes 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 *Argynnis paphia* assembly.

Measure (Benchmark)	Value
EBP summary (haplotype 1)	6.C.Q59
Contig N50 length (≥ 1 Mb)	5.47 Mb
Scaffold N50 length (= chromosome N50)	17.51 Mb
Consensus quality (QV) (≥ 40)	Haplotype 1: 59.9; haplotype 2: 60.0; combined: 60.0
<i>k</i> -mer completeness ($\geq 95\%$)	Haplotype 1: 76.69%; Haplotype 2: 70.22%; combined: 99.64%
BUSCO (S > 90%; D < 5%)	C:98.8%[S:98.0%,D:0.8%],F:0.1%,M:1.1%,n:5286
Percentage of assembly assigned to chromosomes ($\geq 90\%$)	97.96%

Assembly Standards [September 2024](#). The EBP metric, calculated for the haplotype 1, is **6.C.Q59**, 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: *Argynnis paphia* (silver-washed fritillary). Accession number [PRJEB78670](#). The genome sequence is released openly for reuse. The *Argynnis paphia* 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/chhy123/hifiasm
HiGlass	1.13.4	https://github.com/higlass/higlass
lep_buscoPainter	1.0.0	https://github.com/charlottewright/lep_buscoPainter
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
PretextSnapshot	N/A	https://github.com/sanger-tol/PretextSnapshot

Software	Version	Source
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](#)

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, Rennwald E: **Die Schmetterlinge Baden-Württembergs. Band 1, Tagfalter 1.** Stuttgart: Ulmer-Verlag, 1991. [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](#)

Grüning B, Dale R, Sjödin A, et al.: **Bioconda: sustainable and comprehensive software distribution for the life sciences.** *Nat Methods.* 2018; **15**(7): 475–476. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Howard C, Denton A, Jackson B, et al.: **On the path to reference genomes for all biodiversity: lessons learned and laboratory protocols created in the Sanger Tree of Life core laboratory over the first 2000 species.** *bioRxiv.* 2025. [Publisher Full Text](#)

Howe K, Chow W, Collins J, et al.: **Significantly improving the quality of genome assemblies through curation.** *GigaScience.* 2021; **10**(1): g1aa153. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Kerpedjiev P, Abdennur N, Lekschas F, et al.: **HiGlass: web-based visual exploration and analysis of genome interaction maps.** *Genome Biol.* 2018; **19**(1): 125. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Kriventseva EV, Kuznetsov D, Tegenfeldt F, et al.: **OrthoDB v10: sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs.** *Nucleic Acids Res.* 2019; **47**(D1): D807–D811. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Kurtzer GM, Sochat V, Bauer MW: **Singularity: scientific containers for**

mobility of compute. *PLoS One*. 2017; **12**(5): e0177459.

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

Lafranchis T, Jutzeler D, Guillosson JY, *et al.*: **La vie des papillons : Écologie, biologie 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**. Volume 1. Bâle: Ligue Suisse pour la Protection de la Nature, 1987.

[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](#)

Thomas J, Lewington R: **The Butterflies of Britain and Ireland, new revised edition**. Oxford: British Wildlife Publishing Ltd, 2014.

[Reference Source](#)

Tshikolovets VV: **Butterflies of Europe & the Mediterranean area**. Padubrice, Czech Republic: Tshikolovets Publications, 2011.

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.

[Reference Source](#)

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](#)

Wright CJ, Stevens L, Mackintosh A, *et al.*: **Comparative genomics reveals the dynamics of chromosome evolution in Lepidoptera**. *Nat Ecol Evol*. 2024; **8**(4): 777–790.

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

Zhou C, McCarthy SA, Durbin R: **YaHS: Yet another Hi-C Scaffolding tool**. *Bioinformatics*. 2023; **39**(1): btac808.

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

Open Peer Review

Current Peer Review Status:  

Version 1

Reviewer Report 11 September 2025

<https://doi.org/10.21956/wellcomeopenres.27149.r130579>

© 2025 Pohjoismäki J. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Jaakko Pohjoismäki 

University of Eastern Finland, Joensuu,, Finland

Yannick Chittaro, Andreas Sanchez, Kay Lucek, Charlotte J. Wright, Joana I. Meier, Mark L. Blaxter, the Tree of Life initiative, and the project Psyche community provide a genome report of *Argynnis paphia*. The genome meets Earth BioGenome Project (EBP) standards and represents a highly continuous, chromosome-level reference genome assembly with two assembled haplotypes.

The introduction to the species is well written and informative, providing nicely also ecological context for the species. The sequencing, Hi-C-based chromosomal scaffolding, assembly, and annotation are carried out following the Darwin Tree of Life pipeline and appear technically sound and of high quality.

I find little to improve in this report. One point of additional interest could have been a commentary on the two types of Z sex chromosomes, but this may be better addressed in a broader comparative framework with related species are available.

Overall, this is a nice genome report and represents growing number of genomic resources for future research on Lepidoptera.

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: molecular biology, biodiversity genetics, taxonomy

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 03 September 2025

<https://doi.org/10.21956/wellcomeopenres.27149.r130583>

© 2025 Charamis J. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Jason Charamis 

Foundation for Research and Technology - Hellas, Irákleion, Greece

This manuscript presents a reference genome for the silver-washed fritillary butterfly, *Argynnis paphia*. The work is technically sound and I have only a few text corrections to suggest:

1. Instead of "Silver-washed Fritillary" write "silver-washed fritillary butterfly" in the first sentence of Background.
2. Similarly in the second sentence of Background replace "Fritillary species" with "fritillary butterfly species"
3. Correct this sentence and make the relevance of this study more clear

"The reference genome presented here will allow to investigate the biogeography of this species and provide an invaluable resource for understanding the evolution of Lepidoptera."

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: Arthropod Comparative 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.
