DATA NOTE

The genome sequence of the Map, *Araschnia levana* (Linnaeus, 1758) (Lepidoptera: Nymphalidae)

[version 1; peer review: awaiting peer review]

Mathieu Joron[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]CEFE CNRS, Montpellier, France
[2]University of Neuchâtel, Neuchâtel, Switzerland
[3]Tree of Life, Wellcome Sanger Institute, Hinxton, England, UK

v1 **First published:** 31 Jul 2025, **10**:398
https://doi.org/10.12688/wellcomeopenres.24633.1
Latest published: 31 Jul 2025, **10**:398
https://doi.org/10.12688/wellcomeopenres.24633.1

Abstract

We present a genome assembly from a female specimen of *Araschnia levana* (Map; Arthropoda; Insecta; Lepidoptera; Nymphalidae). The assembly contains two haplotypes with total lengths of 362.54 megabases and 332.31 megabases. Most of haplotype 1 (99.09%) is 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 15.96 kilobases.

Keywords

Araschnia levana, Map, genome sequence, chromosomal, Lepidoptera

This article is included in the Tree of Life gateway.

Open Peer Review

Approval Status *AWAITING PEER REVIEW*

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: Joron M: Writing – Original Draft Preparation; **Lucek K**: Resources, Supervision; **Wright CJ**: Funding Acquisition, Project Administration, Supervision; **Meier JI**: Funding Acquisition, Project Administration, Supervision; **Blaxter ML**: Funding Acquisition, Project Administration, Supervision;

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by Wellcome through core funding to the Wellcome Sanger Institute (220540). 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.

How to cite this article: Joron M, Lucek K, Wright CJ *et al.* **The genome sequence of the Map, *Araschnia levana* (Linnaeus, 1758) (Lepidoptera: Nymphalidae) [version 1; peer review: awaiting peer review]** Wellcome Open Research 2025, **10**:398 https://doi.org/10.12688/wellcomeopenres.24633.1

First published: 31 Jul 2025, **10**:398 https://doi.org/10.12688/wellcomeopenres.24633.1

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; Nymphalini; *Araschnia*; *Araschnia levana* (Linnaeus, 1758) (NCBI:txid171575).

Background

Araschnia levana Linnaeus, 1758, known as the map butterfly due to its underside wing pattern, is a nymphalid butterfly in a small predominantly East Asian genus (Fric *et al.*, 2004). *A. levana* has a range extending across the Palaearctic from North Japan to the Atlantic (Bozano & Floriani, 2012). In Europe, it thrives in regions where summers are warm but not dry, effectively excluding the Mediterranean area in the south, and most of the British and Irish Isles and Atlantic Scandinavia in the northwest. However, it has significantly expanded northward recently (Audusseau *et al.*, 2017; Widenfalk *et al.*, 2014) and is listed as Least Concern with an increasing trend by the IUCN (Van Swaay *et al.*, 2010).

Araschnia levana thrives in lowland, wooded countryside with an abundance of stinging nettle (*Urtica dioica* L.), its main food plant. An interesting behaviour is the unique mode of egg-laying of females, who stack eggs on top of each other in a neat pile under the host leaf (García-Barros & Fartmann, 2009). Caterpillars are gregarious but do not spin a silk nest. They are black with red-brown spines (scoli) and two prominent spines on their head capsule.

Individuals overwinter as pupae. *A. levana* produces two to three generations per year, and is famous for its striking seasonal polyphenism. Spring adults, emerging from overwintering pupae, display orange wings with black spots, while summer adults, emerging from direct development of pupae, are black with a white band. This makes it an iconic case of seasonal polyphenism (Baudach & Vilcinskas, 2021; Windig & Lammar, 1999). While the hormonal response to photoperiod that triggers this phenotypic variation is well studied (Koch & Buckmann, 1987), the evolutionary forces driving the wing pattern polyphenism are not entirely clear. Hypotheses include developmental and nutritional trade-offs (Friberg & Karlsson, 2010; Morehouse *et al.*, 2013; Windig, 1999), crypsis or mimicry (Ihalainen & Lindstedt, 2012; Joiris *et al.*, 2010), thermal regulation (Windig & Lammar, 1999), or links with immunity (Baudach & Vilcinskas, 2021; Freitak *et al.*, 2019; Vilcinskas & Vogel, 2016). Recent research into the genetic mechanisms underlying polyphenism has documented the differences in gene and miRNA expression between long-day and short-day forms (Mukherjee *et al.*, 2020; Vilcinskas & Vogel, 2016), providing new insight into functional differences between the two forms.

This reference genome for *A. levana* will be invaluable for furthering our understanding of these genetic and evolutionary processes underlying seasonal polyphenism, wing patterning, and voltinism. The sequence data was derived from a female specimen (Figure 1) collected from Forêt de Jorat, Twann-Tüscherz, Bern, Switzerland.

Methods

Sample acquisition and DNA barcoding

The specimen used for genome sequencing was an adult female *Araschnia levana* (specimen ID SAN28000060, ToLID ilAraLeva1; Figure 1), collected from Forêt de Jorat, Twann-Tüscherz, Bern, Switzerland (latitude 47.1211, longitude 7.2039; elevation 516 m) on 29/05/2023.

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 ilAraLeva1 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 31.41 ng/μL and a yield of 1 476.27 ng, with a fragment size of 13.4 kb. The 260/280 spectrophotometric ratio was 2.02, and the 260/230 ratio was 3.05.

Figure 1. Voucher photograph of the *Araschnia levana* (ilAraLeva1) specimen used for genome sequencing.

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 ilAraLeva1 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.

Table 1. Specimen and sequencing data for BioProject PRJEB78668.

Platform	PacBio HiFi	Hi-C
ToLID	ilAraLeva1	ilAraLeva1
Specimen ID	SAN28000060	SAN28000060
BioSample (source individual)	SAMEA115109736	SAMEA115109736
BioSample (tissue)	SAMEA115109761	SAMEA115109771
Tissue	thorax	head
Sequencing platform and model	Revio	Illumina NovaSeq X
Run accessions	ERR13485705	ERR13474170
Read count total	1.75 million	751.79 million
Base count total	18.32 Gb	113.52 Gb

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 flag to handle 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 17 breaks, 53 joins, and removal of 57 haplotypic duplications. 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 *Araschnia levana* was sequenced using Pacific Biosciences single-molecule HiFi long reads, generating 18.32 Gb (gigabases) from 1.75 million reads, which were used to assemble the genome. GenomeScope2.0 analysis estimated the haploid genome size at 350.24 Mb, with a heterozygosity of 0.48% and repeat content of 17.74%. These estimates guided expectations for the assembly. Based on the estimated genome size, the sequencing data provided approximately 51× coverage. Hi-C sequencing produced 113.52 Gb from 751.79 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 362.54 Mb in 111 scaffolds, with 85 gaps, and a scaffold N50 of 12.61 Mb (Table 2).

Most of the assembly sequence (99.09%) was assigned to 32 chromosomal-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 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.

Table 2. Genome assembly statistics.

Assembly name	ilAraLeva1.hap1.1	ilAraLeva1.hap2.1
Assembly accession	GCA_964258735.1	GCA_964258705.1
Assembly level	chromosome	scaffold
Span (Mb)	362.54	332.31
Number of chromosomes	32	N/A
Number of contigs	196	161
Contig N50	5.14 Mb	5.54 Mb
Number of scaffolds	111	95
Scaffold N50	12.61 Mb	12.33 Mb
Longest scaffold length (Mb)	17.03	N/A
Sex chromosomes	W and Z	N/A
Organelles	Mitochondrial genome: 15.96 kb	N/A

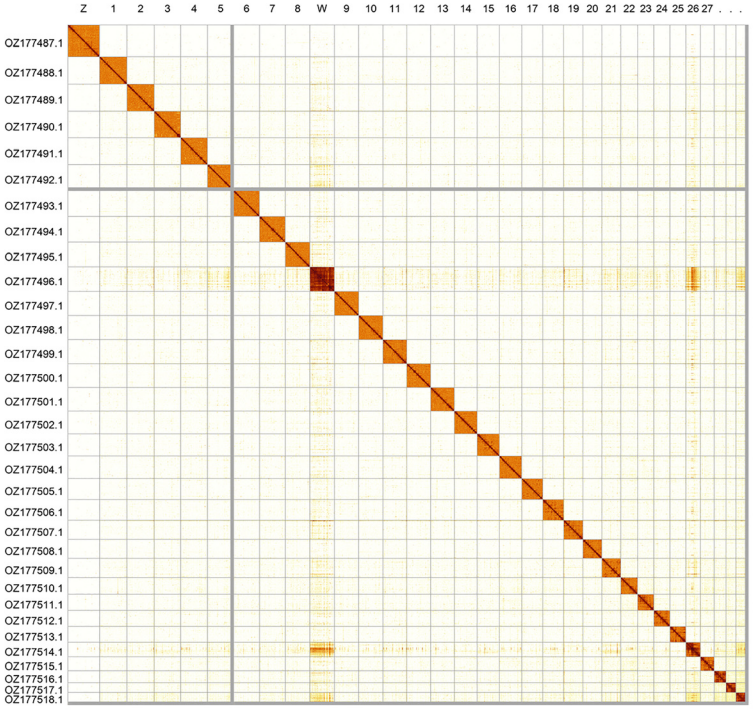

Figure 2. Hi-C contact map of the *Araschnia levana* genome assembly. Assembled chromosomes are shown in order of size and labelled along the axes. The plot was generated using PretextSnapshot.

Table 3. Chromosomal pseudomolecules in the haplotype 1 genome assembly of *Araschnia levana* ilAraLeva1.

INSDC accession	Molecule	Length (Mb)	GC%	Assigned Merian elements
OZ177488.1	1	14.57	32.50	M2
OZ177489.1	2	14.46	33	M1
OZ177490.1	3	14.20	33	M17;M20
OZ177491.1	4	14.16	33	M8
OZ177492.1	5	13.95	32.50	M21
OZ177493.1	6	13.76	32.50	M9
OZ177494.1	7	13.76	33	M3
OZ177495.1	8	13.07	33	M7
OZ177497.1	9	12.97	33	M12
OZ177498.1	10	12.86	33	M5
OZ177499.1	11	12.73	33.50	M6
OZ177500.1	12	12.73	33	M18
OZ177501.1	13	12.61	32.50	M16
OZ177502.1	14	12.16	33	M4
OZ177503.1	15	11.84	33	M15
OZ177504.1	16	11.83	33	M22
OZ177505.1	17	11.24	33	M10
OZ177506.1	18	11.14	33.50	M11
OZ177507.1	19	10.24	34	M14
OZ177508.1	20	10.09	33	M13
OZ177509.1	21	10.06	33.50	M23
OZ177510.1	22	8.94	33	M26
OZ177511.1	23	8.67	33.50	M24
OZ177512.1	24	8.63	34	M19
OZ177513.1	25	8.41	33.50	M28
OZ177514.1	26	7.72	37.50	M30
OZ177515.1	27	7.34	33.50	M27
OZ177516.1	28	6.33	34.50	M25
OZ177517.1	29	5.31	34.50	M29
OZ177518.1	30	4.94	35.50	M31
OZ177496.1	W	13.05	38	N/A
OZ177487.1	Z	17.03	32.50	MZ

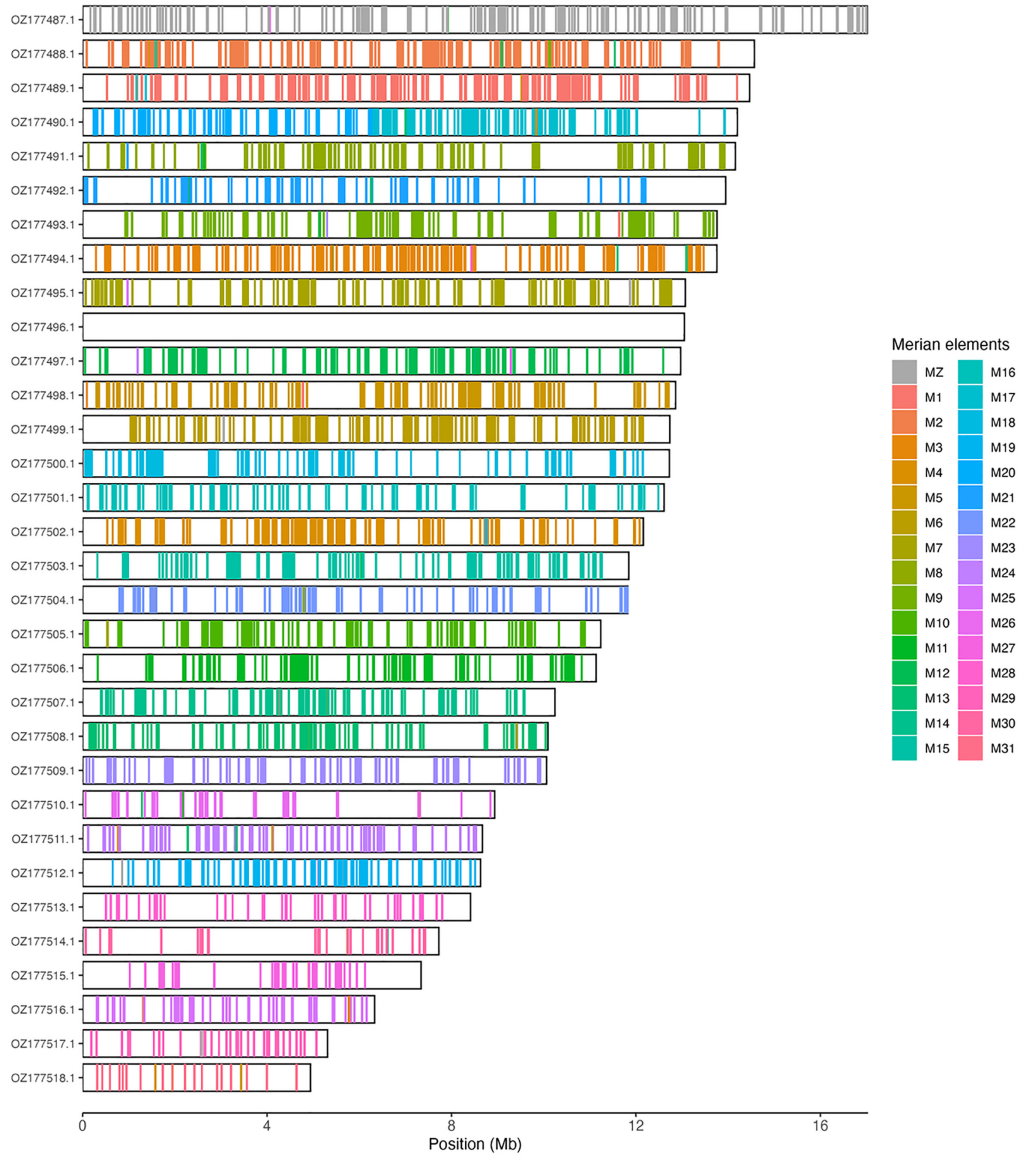

Figure 3. Merian elements painted across chromosomes in the ilAraLeva1.hap1.1 assembly of *Araschnia levana*. Chromosomes are drawn to scale, with the positions of orthologues shown as coloured bars. Each orthologue is coloured by the Merian element that it belongs to. All orthologues which could be assigned to Merian elements are shown.

Assembly quality metrics

For haplotype 1, the estimated QV is 62.2, and for haplotype 2, 62.6. When the two haplotypes are combined, the assembly achieves an estimated QV of 62.4. The *k*-mer completeness is 89.98% for haplotype 1, 85.12% for haplotype 2, and 99.69% 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 = 98.5%, duplicated = 0.2%) 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 Assembly Standards September 2024. The EBP metric, calculated for the haplotype 1, is **6.C.Q62**, 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

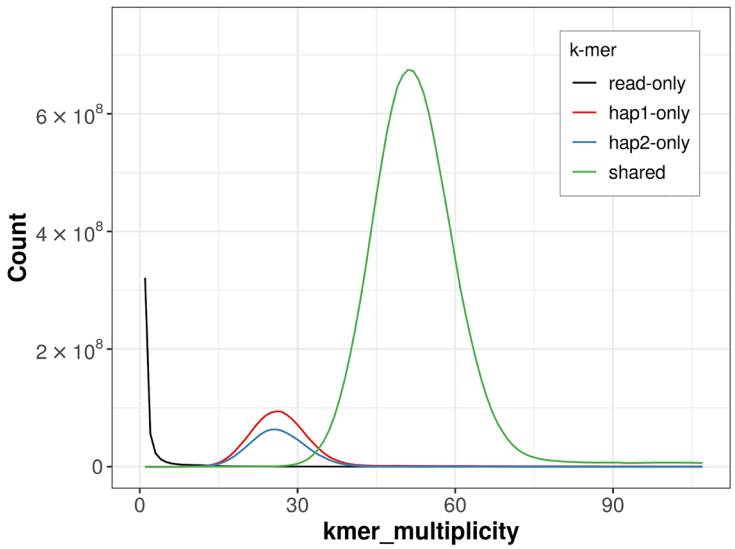

Figure 4. Evaluation of *k*-mer completeness using MerquryFK. This plot illustrates the recovery of *k*-mers from the original read data in the final assemblies. The horizontal axis represents *k*-mer multiplicity, and the vertical axis shows the number of *k*-mers. The black curve represents *k*-mers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

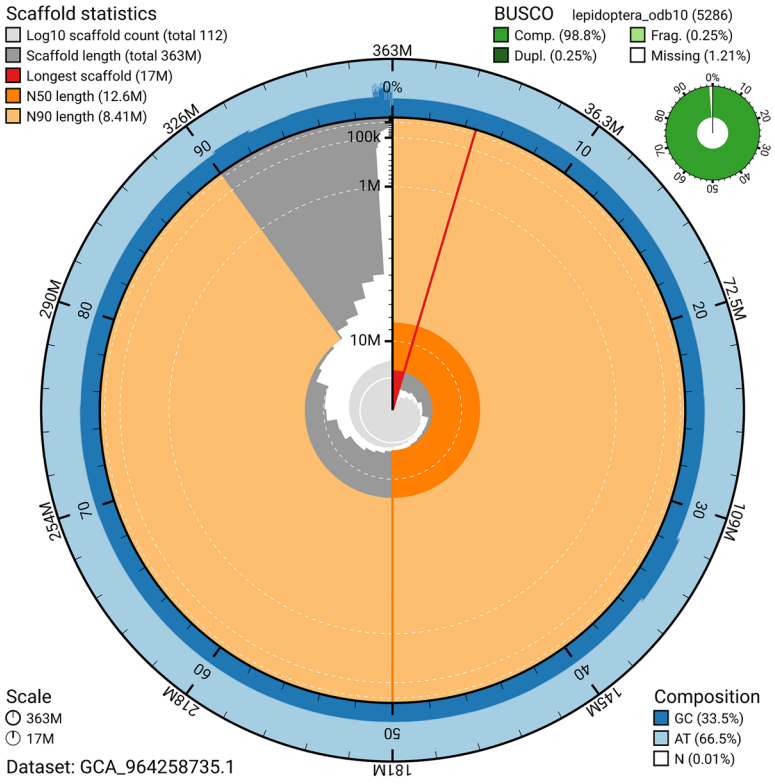

Figure 5. Assembly metrics for ilAraLeva1.hap1.1. The BlobToolKit snail plot provides an overview of assembly metrics and BUSCO gene completeness. The circumference represents the length of the whole genome sequence, and the main plot is divided into 1,000 bins around the circumference. The outermost 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.

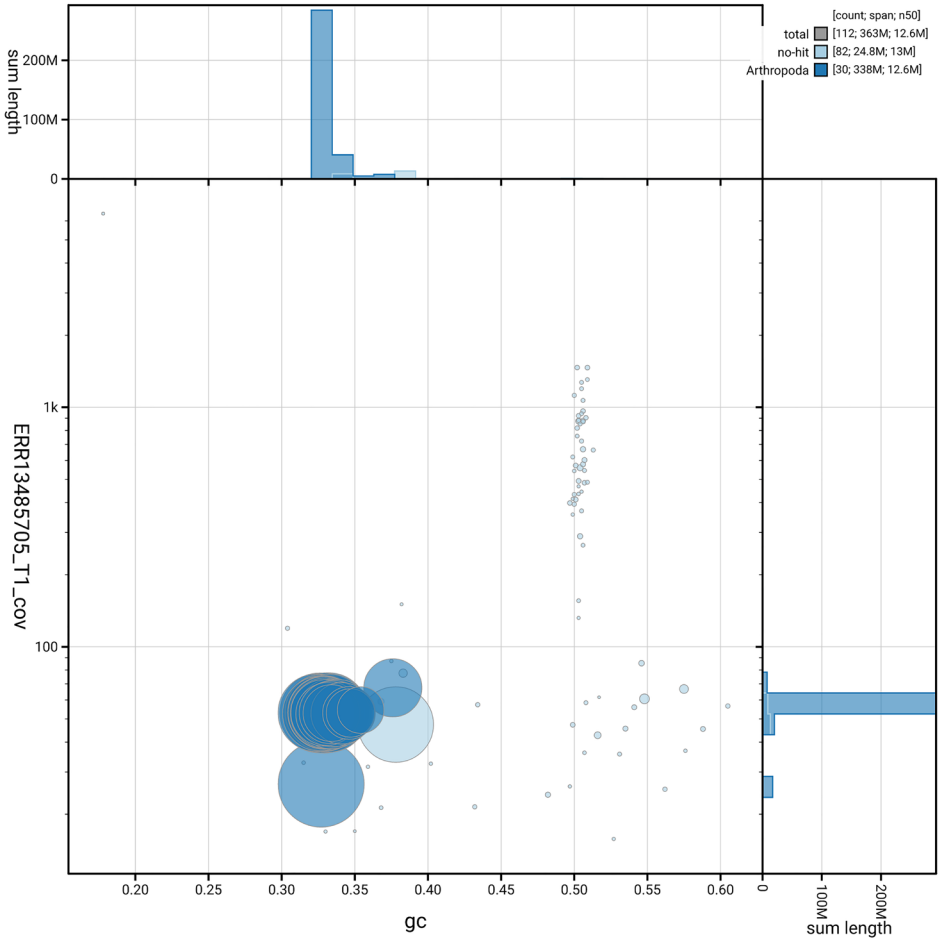

Figure 6. BlobToolKit GC-coverage plot for ilAraLeva1.hap1.1. Blob plot showing sequence coverage (vertical axis) and GC content (horizontal axis). The circles represent scaffolds, with the size proportional to scaffold length and the colour representing phylum membership. The histograms along the axes 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 *Araschnia levana* assembly.

Measure (Benchmark)	Value
EBP summary (haplotype 1)	6.C.Q62
Contig N50 length (≥ 1 Mb)	5.14 Mb
Scaffold N50 length (= chromosome N50)	12.61 Mb
Consensus quality (QV) (≥ 40)	Haplotype 1: 62.2; haplotype 2: 62.6; combined: 62.4
k-mer completeness (≥ 95%)	Haplotype 1: 89.98%; Haplotype 2: 85.12%; combined: 99.69%
BUSCO* (S > 90%; D < 5%)	C:98.8%[S:98.5%,D:0.2%],F:0.2%,M:1.0%,n:5286
Percentage of assembly assigned to chromosomes (≥ 90%)	99.09%

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: Araschnia levana (map butterfly). Accession number PRJEB78668; https://identifiers.org/ena.embl/PRJEB78668. The genome sequence is released openly for reuse. The *Araschnia levana* 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
lep_busco_painter	1.0.0	https://github.com/charlottewright/lep_busco_painter
MerquryFK	1.1.1	https://github.com/thegenemyers/MERQURY.FK
Minimap2	2.24-r1122	https://github.com/lh3/minimap2
MitoHiFi	3	https://github.com/marcelauliano/MitoHiFi

Software	Version	Source
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

Audusseau H, Le Vaillant M, Janz N, *et al.*: **Species range expansion constrains the ecological niches of resident butterflies.** *J Biogeogr.* 2017; **44**(1): 28–38.
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

Baudach A, Vilcinskas A: **The European map butterfly *Araschnia levana* as a model to study the molecular basis and evolutionary ecology of seasonal polyphenism.** *Insects.* 2021; **12**(4): 325.
PubMed Abstract | Publisher Full Text | Free Full Text

Bozano GC, Floriani A: **Guide to the butterflies of the Palearctic Region. Nymphalidae, Part v. Subfamily Nymphalinae, Tribes Nymphalini, Kallimini, Junoniini.** Milano: Omnes Partes, 2012.

Buchfink B, Reuter K, Drost HG: **Sensitive protein alignments at Tree-of-Life scale using DIAMOND.** *Nat Methods.* 2021; **18**(4): 366–368.
PubMed Abstract | Publisher Full Text | Free Full Text

Challis R, Kumar S, Sotero-Caio C, *et al.*: **Genomes on a Tree (GoaT): a versatile, scalable search engine for genomic and sequencing project metadata across the eukaryotic Tree of Life [version 1; peer review: 2 approved].** *Wellcome Open Res.* 2023; **8**: 24.
PubMed Abstract | Publisher Full Text | Free Full Text

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

Cheng H, Concepcion GT, Feng X, *et al.*: **Haplotype-resolved *de novo* assembly using phased assembly graphs with hifiasm.** *Nat Methods.* 2021; **18**(2): 170–175.
PubMed Abstract | Publisher Full Text | Free Full Text

Cheng H, Jarvis ED, Fedrigo O, *et al.*: **Haplotype-resolved assembly of diploid genomes without parental data.** *Nat Biotechnol.* 2022; **40**(9): 1332–1335.
PubMed Abstract | Publisher Full Text | Free Full Text

Danecek P, Bonfield JK, Liddle J, *et al.*: **Twelve years of SAMtools and BCFtools.** *GigaScience.* 2021; **10**(2): giab008.
PubMed Abstract | Publisher Full Text | Free Full Text

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

Freitak D, Tammaru T, Sandre SL, *et al.*: **Longer life span is associated with elevated immune activity in a seasonally polyphenic butterfly.** *J Evol Biol.* 2019; **32**(7): 653–665.
PubMed Abstract | Publisher Full Text | Free Full Text

Friberg M, Karlsson B: **Life-history polyphenism in the map butterfly (*Araschnia levana*): developmental constraints versus season-specific adaptations.** *Evol Ecol Res.* 2010; **12**(5): 603–615.
Reference Source

Fric Z, Konvicka M, Zrzavy J: **Red & black or black & white? Phylogeny of the *Araschnia* butterflies (Lepidoptera: Nymphalidae) and evolution of seasonal polyphenism.** *J Evol Biol.* 2004; **17**(2): 265–78.
PubMed Abstract | Publisher Full Text

García-Barros E, Fartmann T: **Butterfly oviposition: sites, behaviour and modes.** In: J. Settele, T. Shreeve, M. Konvička, *et al*. (eds), *Ecology of Butterflies in Europe.* Cambridge: Cambridge University Press, 2009; 29–42.
Reference Source

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

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

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

Ihalainen E, Lindstedt C: **Do avian predators select for seasonal polyphenism in the European Map butterfly *Araschnia levana* (Lepidoptera: Nymphalidae)?** *Biol J Linn Soc.* 2012; **106**(4): 737–748.
Publisher Full Text

Joiris A, Korati S, Van Dyck H: **Differential bird predator attack rate on seasonal forms of the Map butterfly (*Araschnia levana* L.): does the substrate matter?** *Ethology.* 2010; **116**(11): 1020–1026.
Publisher 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

Koch PB, Buckmann D: **Hormonal control of seasonal morphs by the timing of ecdysteroid release in *Araschnia levana L.* (Nymphalidae, Lepidoptera).** *J Insect Physiol.* 1987; **33**(11): 823–829.
Publisher 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

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

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

Merkel D: **Docker: lightweight Linux containers for consistent development and deployment.** *Linux J.* 2014; **2014**(239): 2, [Accessed 2 April 2024].
Reference Source

Morehouse NI, Mandon N, Christides JP, *et al.*: **Seasonal selection and resource dynamics in a seasonally polyphenic butterfly.** *J Evol Biol.* 2013; **26**(1): 175–85.
PubMed Abstract | Publisher Full Text

Mukherjee K, Baudach A, Vogel H, *et al.*: **Seasonal phenotype-specific expression of microRNAs during metamorphosis in the European map butterfly *Araschnia levana*.** *Arch Insect Biochem Physiol.* 2020; **104**(1): e21657.
PubMed Abstract | Publisher Full Text

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

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

Vilcinskas A, Vogel H: **Seasonal phenotype-specific transcriptional reprogramming during metamorphosis in the European Map butterfly *Araschnia levana*.** *Ecol Evol.* 2016; **6**(11): 3476–3485.
PubMed Abstract | Publisher Full Text | Free Full Text

Widenfalk LA, Ahrné K, Berggren Å: **Using citizen-reported data to predict distributions of two non-native insect species in Sweden.** *Ecosphere.* 2014; **5**(12): 156.
Publisher Full Text

Windig JJ: **Trade-offs between melanization, development time and adult size in *Inachis io* and *Araschnia levana* (Lepidoptera: Nymphalidae).** *Heredity.* 1999; **82**: 57–68.
Publisher Full Text

Windig JJ, Lammar P: **Evolutionary genetics of seasonal polyphenism in the Map butterfly *Araschnia levana* (Nymphalidae: Lepidoptera).** *Evol Ecol Res.* 1999; **1**(7): 875–894.
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