DATA NOTE

The genome sequence of the Alpine Heath, *Coenonympha gardetta* Hübner, 1819 (Lepidoptera: Nymphalidae)

[version 1; peer review: awaiting peer review]

Camille Cornet[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, Switzerland
[2]Centre for Evolutionary and Functional Ecology (CEFE CNRS), Montpellier, France
[3]Tree of Life, Wellcome Sanger Institute, Hinxton, England, UK

v1 **First published:** 31 Jul 2025, **10**:403
https://doi.org/10.12688/wellcomeopenres.24645.1
Latest published: 31 Jul 2025, **10**:403
https://doi.org/10.12688/wellcomeopenres.24645.1

Abstract

We present a genome assembly from a female specimen of *Coenonympha gardetta* (Alpine Heath; Arthropoda; Insecta; Lepidoptera; Nymphalidae). The assembly contains two haplotypes with total lengths of 500.33 megabases and 474.74 megabases. Most of haplotype 1 (99.65%) is scaffolded into 29 chromosomal pseudomolecules, including the Z sex chromosome. Haplotype 2 was assembled to scaffold level. The mitochondrial genome has also been assembled, with a length of 15.29 kilobases.

Keywords

Coenonympha gardetta, Alpine Heath, 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: Cornet C: Investigation, Resources; **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). CC and KL were supported by Swiss National Science Foundation (SNSF) Grant ID 202869.
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

How to cite this article: Cornet C, Joron M, Lucek K *et al.* **The genome sequence of the Alpine Heath, *Coenonympha gardetta* Hübner, 1819 (Lepidoptera: Nymphalidae) [version 1; peer review: awaiting peer review]** Wellcome Open Research 2025, **10**:403 https://doi.org/10.12688/wellcomeopenres.24645.1

First published: 31 Jul 2025, **10**:403 https://doi.org/10.12688/wellcomeopenres.24645.1

Species taxonomy

Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Protostomia; Ecdysozoa; Panarthropoda; Arthropoda; Mandibulata; Pancrustacea; Hexapoda; Insecta; Dicondylia; Pterygota; Neoptera; Endopterygota; Amphiesmenoptera; Lepidoptera; Glossata; Neolepidoptera; Heteroneura; Ditrysia; Obtectomera; Papilionoidea; Nymphalidae; Satyrinae; Satyrini; Coenonymphina; *Coenonympha*; *Coenonympha gardetta* Hübner, 1819 (NCBI: txid554481)

Background

Coenonympha gardetta (de Prunner, 1798), known as the Alpine Heath, is a small nymphalid butterfly from a diverse genus with a Holarctic distribution (Bozano, 2002). *C. gardetta* is part of a clade that has diversified in the Western Palaearctic (Kodandaramaiah *et al.*, 2009). *C. gardetta* is a high-altitude specialist restricted to the European Alps, and the Massif Central of France (Bozano, 2002). This species is listed as Least Concern in the IUCN Red List (Europe) with a decreasing trend (Van Swaay *et al.*, 2010). Subspecies *C. g. lecerfi* de Lesse 1949, the only non-Alpine population with a disjunct distribution on the hilltops of the Forez massif of France, is listed as Vulnerable (UICN-France, 2014).

Coenonympha gardetta inhabits open grasslands around the treeline and above into alpine meadows from 1 400 to 2 800 metres above sea level. It depends on grasses for larval development, laying eggs singly in the shade of herbs or shrubs. Adults fly from late June to late July, depending on geography, and may be found perching or sun basking on plants and rocks. The wing pattern includes a dark brown upper side and a grey and buff colour on the underside, featuring a regular series of black eyespots with a white centre, which contrast prominently against a crescent of cream-coloured scales. It resembles *C. orientalis*, a closely related high-altitude species in the Balkan Peninsula (Bozano, 2002).

Coenonympha gardetta is part of a clade that displays significant signatures of interspecific gene flow (Greenwood *et al.*, 2025), and is therefore worthy of further investigation to reveal the role of gene flow in local adaptation and speciation. *C. gardetta* diverged from its lowland relative *C. arcania* approximately 1.7 million years ago, replacing it on the altitudinal gradient and suggesting specific altitudinal adaptations. For instance, genomic signals of selection have been found in genes related to anoxia (Capblancq *et al.*, 2024), and experiments have uncovered thermal adaptation to adult flight, larval cold hardiness, and egg heat tolerance (Doniol-Valcroze *et al.*, 2024; Nève & Després, 2020; Vrba *et al.*, 2022). *C. gardetta* is the subject of ongoing research into hybrid speciation, with population resequencing of genomes from diverse Alpine locations and from the contact zones with its close allies (Capblancq *et al.*, 2024; Doniol-Valcroze, 2024). Although *C. gardetta* exhibits strong premating isolation with *C. arcania*, involving chemical divergence and genetic incompatibilities (Capblancq *et al.*, 2019; Doniol-Valcroze *et al.*, 2025), hybridisation around 300 thousand years ago has formed two local hybrid lineages in the Alps, to which *C. gardetta* contributes about 23% of its genome (Capblancq *et al.*, 2015; Capblancq *et al.*, 2024). Understanding how this genome fraction contributes to adaptation and to barriers to gene flow with related species is crucial to our understanding of the evolutionary history of this clade. More generally, this case of hybrid speciation is an excellent model for investigating the modalities of the emergence of new species (Capblancq *et al.*, 2015; Capblancq *et al.*, 2024).

A reference genome for *C. gardetta* is an invaluable resource for advancing our understanding of clade diversification, hybrid speciation, and altitudinal adaptation. The sequence data were derived from a female specimen (Figure 1) collected from Brochwald, Meiringen, Switzerland.

Methods

Sample acquisition

The specimen used for genome sequencing was an adult female *Coenonympha gardetta* (specimen ID SAN28000081, ToLID ilCoeGard1; Figure 1), collected from Brochwald, Meiringen, Switzerland (latitude 46.6743, longitude 8.1334; elevation 1 470 m) on 09/07/2023. The specimen was collected and identified by Camille Cornet (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). The ilCoeGard1 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

Figure 1. Voucher photograph of the *Coenonympha gardetta* (ilCoeGard1) specimen used for genome sequencing.

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 32.23 ng/μL and a yield of 1,514.81 ng, with a fragment size of 13.6 kb. The 260/280 spectrophotometric ratio was 1.87, and the 260/230 ratio was 2.72.

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.

Table 1. Specimen and sequencing data for BioProject PRJEB78768.

Platform	PacBio HiFi	Hi-C
ToLID	ilCoeGard1	ilCoeGard1
Specimen ID	SAN28000081	SAN28000081
BioSample (source individual)	SAMEA115109803	SAMEA115109803
BioSample (tissue)	SAMEA115109828	SAMEA115109827
Tissue	thorax	head
Sequencing platform and model	Revio	Illumina NovaSeq X
Run accessions	ERR13485730	ERR13493987
Read count total	1.82 million	774.20 million
Base count total	20.32 Gb	116.90 Gb

Hi-C

Sample preparation and crosslinking

The Hi-C sample was prepared from 20–50 mg of frozen head tissue of the ilCoeGard1 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.

The HiFi reads were assembled using Hifiasm in Hi-C phasing mode (Cheng *et al.*, 2021, Cheng *et al.*, 2022), producing two haplotypes. Hi-C reads (Rao *et al.*, 2014) were mapped to the primary contigs using bwa-mem2 (Vasimuddin *et al.*, 2019). Contigs were further scaffolded with Hi-C data in YaHS

(Zhou *et al.*, 2023), using the --break option for handling potential misassemblies. The scaffolded assemblies were evaluated using Gfastats (Formenti *et al.*, 2022), BUSCO (Manni *et al.*, 2021) and MERQURY.FK (Rhie *et al.*, 2020).

The mitochondrial genome was assembled using MitoHiFi (Uliano-Silva *et al.*, 2023), which runs MitoFinder (Allio *et al.*, 2020) and uses these annotations to select the final mitochondrial contig and to ensure the general quality of the sequence.

Assembly curation

The assembly was decontaminated using the Assembly Screen for Cobionts and Contaminants (ASCC) pipeline. TreeVal was used to generate the flat files and maps for use in curation. Manual curation was conducted primarily in PretextView and HiGlass (Kerpedjiev *et al.*, 2018). Scaffolds were visually inspected and corrected as described by Howe *et al.* (2021). Manual corrections included 8 breaks and 17 joins. The curation process is documented at https://gitlab.com/wtsi-grit/rapid-curation. PretextSnapshot was used to generate a Hi-C contact map of the final assembly.

Assembly quality assessment

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

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

The genome was analysed using the BlobToolKit pipeline, a Nextflow implementation of the earlier Snakemake BlobToolKit pipeline (Challis *et al.*, 2020). The pipeline aligns PacBio reads using minimap2 (Li, 2018) and SAMtools (Danecek *et al.*, 2021) to generate coverage tracks. Simultaneously, it queries the GoaT database (Challis *et al.*, 2023) to identify relevant BUSCO lineages and runs BUSCO (Manni *et al.*, 2021). For the three domain-level BUSCO lineages, BUSCO genes are aligned to the UniProt Reference Proteomes database (Bateman *et al.*, 2023) using DIAMOND blastp (Buchfink *et al.*, 2021). The genome is divided into chunks based on the density of BUSCO genes from the closest taxonomic lineage, and each chunk is aligned to the UniProt Reference Proteomes database with DIAMOND blastx. Sequences without hits are chunked using seqtk and aligned to the NT database with blastn (Altschul *et al.*, 1990). The BlobToolKit suite consolidates all outputs into a blobdir for visualisation. The BlobToolKit pipeline was developed using nf-core tooling (Ewels *et al.*, 2020) and MultiQC (Ewels *et al.*, 2016), with package management via Conda and Bioconda (Grüning *et al.*, 2018), and containerisation through Docker (Merkel, 2014) and Singularity (Kurtzer *et al.*, 2017).

Genome sequence report

Sequence data

The genome of a specimen of *Coenonympha gardetta* was sequenced using Pacific Biosciences single-molecule HiFi long reads, generating 20.32 Gb (gigabases) from 1.82 million reads, which were used to assemble the genome. GenomeScope2.0 analysis estimated the haploid genome size at 480.04 Mb, with a heterozygosity of 1.81% and repeat content of 38.66%. These estimates guided expectations for the assembly. Based on the estimated genome size, the sequencing data provided approximately 41× coverage. Hi-C sequencing produced 116.90 Gb from 774.20 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 500.33 Mb in 74 scaffolds, with 178 gaps, and a scaffold N50 of 18.14 Mb (Table 2).

Most of the assembly sequence (99.65%) was assigned to 29 chromosomal-level scaffolds, representing 28 autosomes and the Z sex chromosome. The specimen appears to be a ZO female. These chromosome-level scaffolds, confirmed by Hi-C data, are named according to size (Figure 2; Table 3). Chromosome painting with Merian elements illustrates the distribution of orthologues along chromosomes and highlights patterns of chromosomal evolution relative to Lepidopteran ancestral linkage groups (Figure 3).

Table 2. Genome assembly statistics.

Assembly name	ilCoeGard1.hap1.1	ilCoeGard1.hap2.1
Assembly accession	GCA_964258855.1	GCA_964258745.1
Assembly level	chromosome	scaffold
Span (Mb)	500.33	474.74
Number of chromosomes	29	N/A
Number of contigs	252	264
Contig N50	4.46 Mb	4.91 Mb
Number of scaffolds	74	122
Scaffold N50	18.14 Mb	17.71 Mb
Longest scaffold length (Mb)	26.01	N/A
Sex chromosomes	Z	N/A
Organelles	Mitochondrial genome: 15.29 kb	N/A

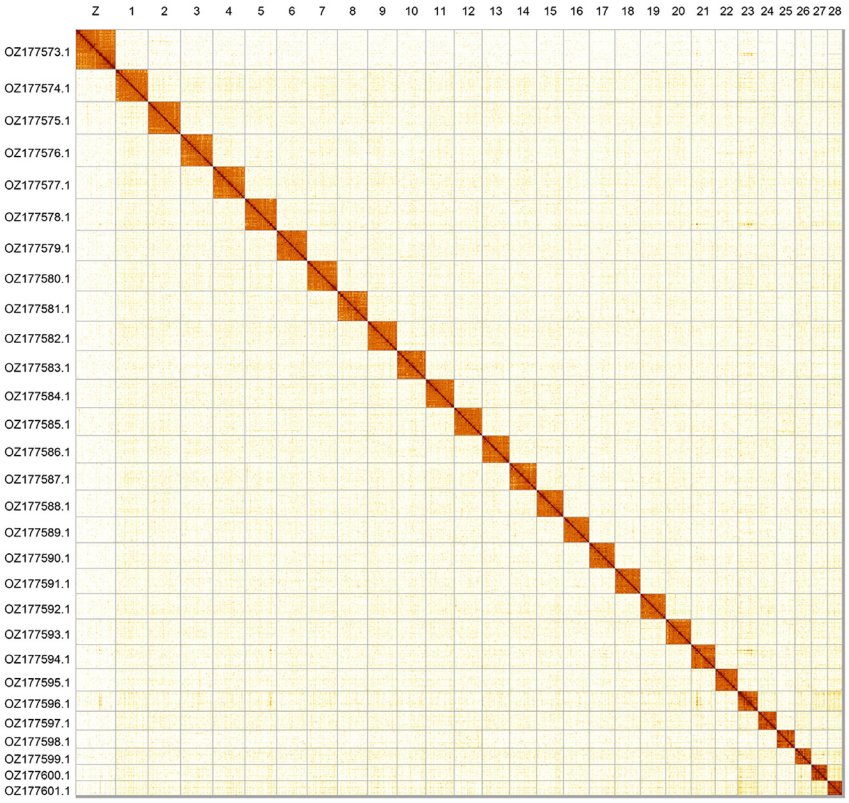

Figure 2. Hi-C contact map of the *Coenonympha gardetta* genome assembly. Assembled chromosomes are shown in order of size and labelled along the axes. The plot was generated using PretextSnapshot.

Table 3. Chromosomal pseudomolecules in the haplotype 1 genome assembly of *Coenonympha gardetta* ilCoeGard1.

INSDC accession	Molecule	Length (Mb)	GC%	Assigned Merian elements
OZ177574.1	1	21.17	38.50	M17;M20
OZ177575.1	2	21.08	38	M2
OZ177576.1	3	21.05	38	M1
OZ177577.1	4	20.85	38	M3
OZ177578.1	5	20.64	38	M19;M26
OZ177579.1	6	19.82	37.50	M9
OZ177580.1	7	19.73	39	M14;M29
OZ177581.1	8	19.58	38	M8
OZ177582.1	9	19.17	38	M5
OZ177583.1	10	18.71	38	M18
OZ177584.1	11	18.43	38.50	M7
OZ177585.1	12	18.14	38	M12
OZ177586.1	13	17.86	38.50	M6
OZ177587.1	14	17.67	38	M4
OZ177588.1	15	17.46	37.50	M16
OZ177589.1	16	16.76	37.50	M21
OZ177590.1	17	16.66	38.50	M15
OZ177591.1	18	16.58	38	M22
OZ177592.1	19	16.49	39	M11
OZ177593.1	20	16.47	38.50	M10
OZ177594.1	21	15.73	39	M23
OZ177595.1	22	14.60	38	M13
OZ177596.1	23	13.05	41	M30
OZ177597.1	24	12.28	39	M24
OZ177598.1	25	11.86	39	M28
OZ177599.1	26	10.63	38.50	M27
OZ177600.1	27	10.56	40	M25
OZ177601.1	28	9.55	41.50	M31
OZ177573.1	Z	26.01	37	MZ

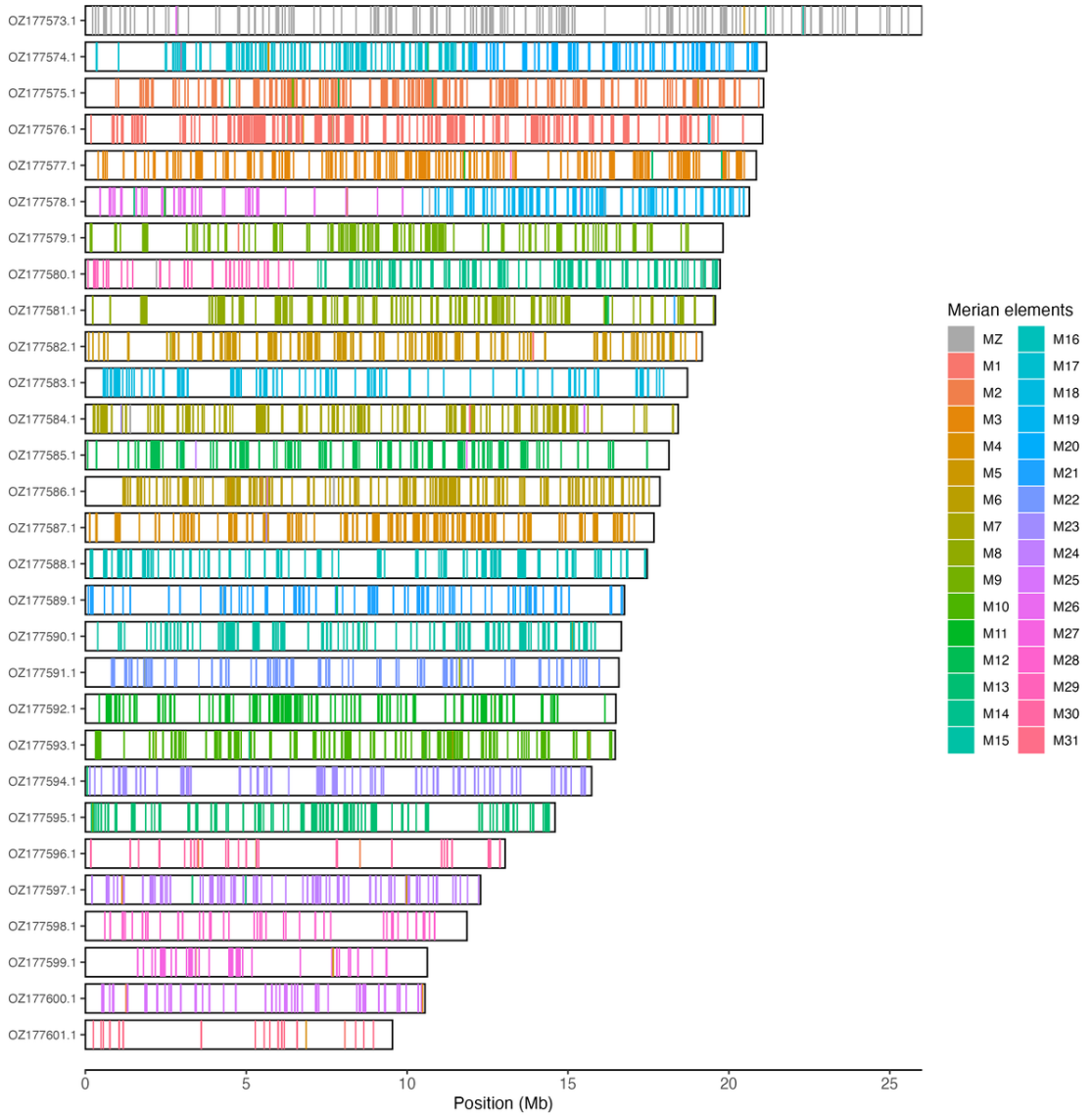

Figure 3. Merian elements painted across chromosomes in the ilCoeGard1.hap1.1 assembly of *Coenonympha gardetta*. Chromosomes are drawn to scale, with the positions of orthologues shown as coloured bars. Each orthologue is coloured by the Merian element that it belongs to. All orthologues which could be assigned to Merian elements are shown.

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

Assembly quality metrics

For haplotype 1, the estimated QV is 63.7, and for haplotype 2, 64.2. When the two haplotypes are combined, the assembly achieves an estimated QV of 63.9. The *k*-mer completeness is 72.16% for haplotype 1, 68.67% for haplotype 2, and 99.33% for the combined haplotypes (Figure 4). BUSCO analysis using the lepidoptera_odb10 reference set (n = 5 286) (Kriventseva *et al.*, 2019) identified 98.2% of the expected gene set (single = 97.5%, duplicated = 0.7%) 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.Q63**, 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

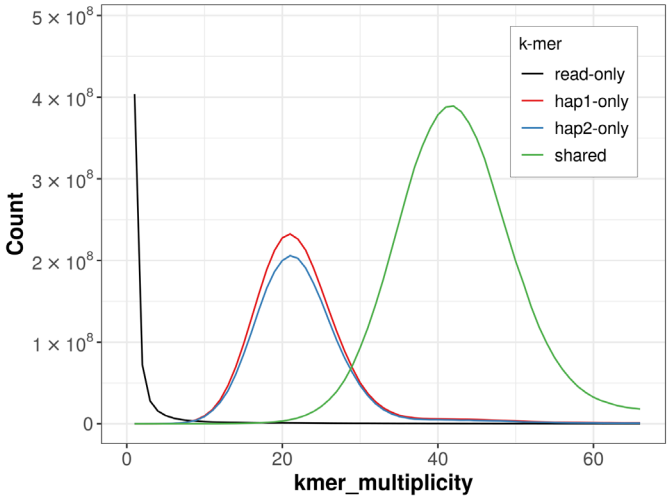

Figure 4. Evaluation of *k*-mer completeness using MerquryFK. This plot illustrates the recovery of *k*-mers from the original read data in the final assemblies. The horizontal axis represents *k*-mer multiplicity, and the vertical axis shows the number of *k*-mers. The black curve represents *k*-mers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

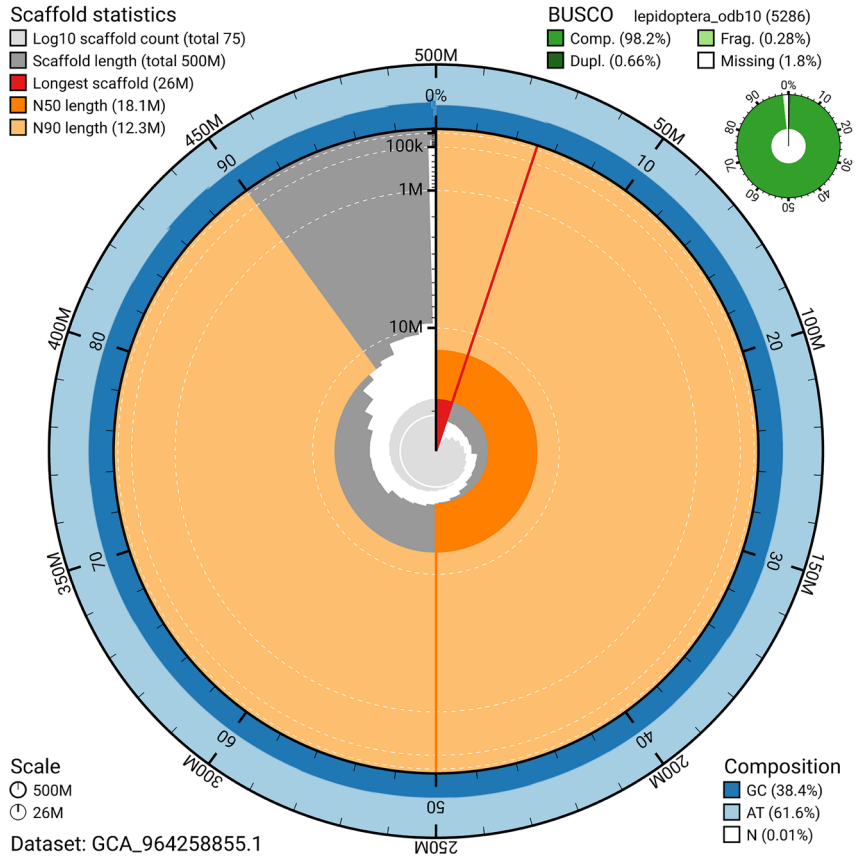

Figure 5. Assembly metrics for ilCoeGard1.hap1.1. The BlobToolKit snail plot provides an overview of assembly metrics and BUSCO gene completeness. The circumference represents the length of the whole genome sequence, and the main plot is divided into 1,000 bins around the circumference. The outermost 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.

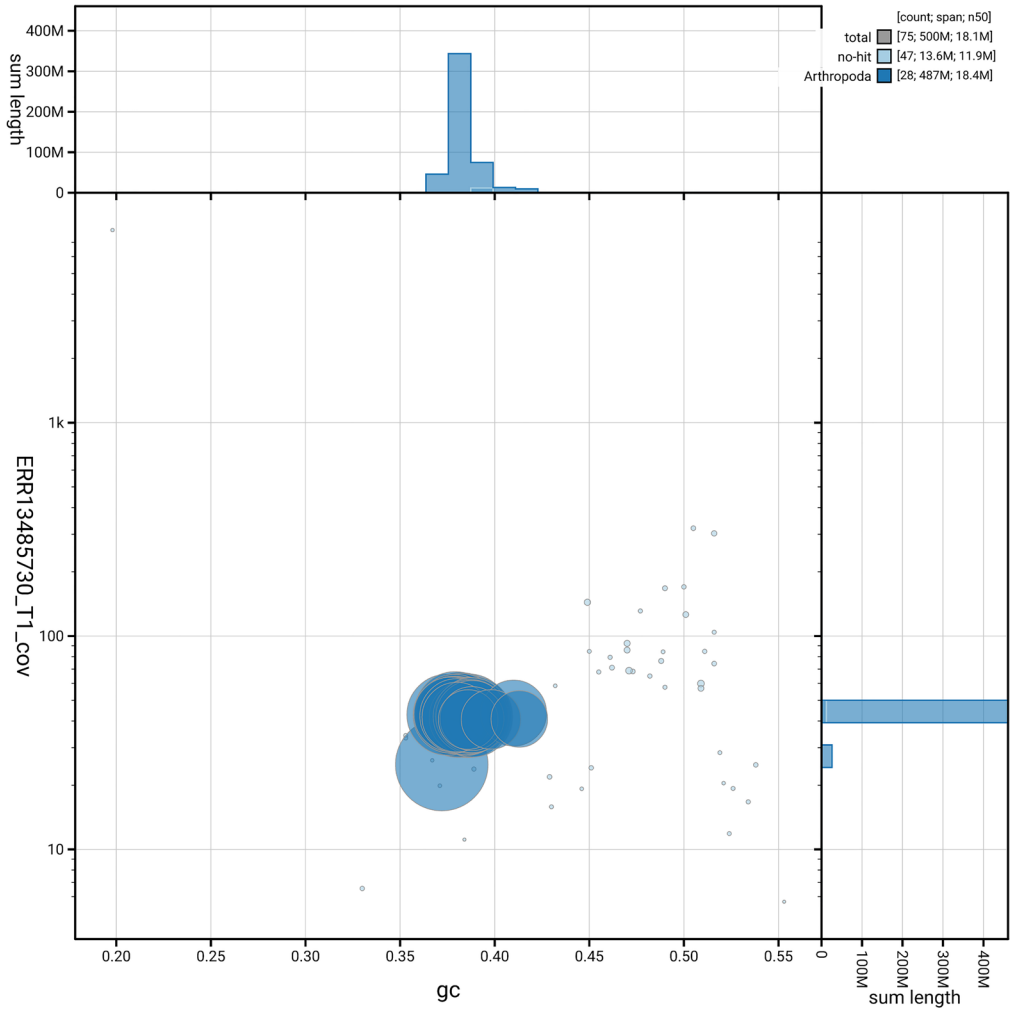

Figure 6. BlobToolKit GC-coverage plot for ilCoeGard1.hap1.1. Blob plot showing sequence coverage (vertical axis) and GC content (horizontal axis). The circles represent scaffolds, with the size proportional to scaffold length and the colour representing phylum membership. The histograms along the axes 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 *Coenonympha gardetta* assembly.

Measure	Value	Benchmark
EBP summary (haplotype 1)	6.C.Q63	6.C.Q40
Contig N50 length	4.46 Mb	≥ 1 Mb
Scaffold N50 length	18.14 Mb	= chromosome N50
Consensus quality (QV)	Haplotype 1: 63.7; haplotype 2: 64.2; combined: 63.9	≥ 40
k-mer completeness	Haplotype 1: 72.16%; Haplotype 2: 68.67%; combined: 99.33%	≥ 95%
BUSCO	C:98.2%[S:97.5%,D:0.7%], F:0.3%,M:1.5%,n:5286	S > 90%; D < 5%
Percentage of assembly assigned to chromosomes	99.65%	≥ 90%

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: Coenonympha gardetta (alpine heath). Accession number PRJEB78768. The genome sequence is released openly for reuse. The *Coenonympha gardetta* 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

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

References

Allio R, Schomaker-Bastos A, Romiguier J, *et al.*: **MitoFinder: efficient automated large-scale extraction of mitogenomic data in target enrichment phylogenomics.** *Mol Ecol Resour.* 2020; **20**(4): 892–905.
PubMed Abstract | Publisher Full Text | Free Full Text

Altschul SF, Gish W, Miller W, *et al.*: **Basic Local Alignment Search Tool.** *J Mol Biol.* 1990; **215**(3): 403–410.
PubMed Abstract | Publisher Full Text

Bateman A, Martin MJ, Orchard S, *et al.*: **UniProt: the Universal Protein Knowledgebase in 2023.** *Nucleic Acids Res.* 2023; **51**(D1): D523–D531.
PubMed Abstract | Publisher Full Text | Free Full Text

Bozano GC: **Guide to the Butterflies of the Palearctic Region. Satyrinae, Part III. Tribe Satyrini, subtribes Melanargiina and Coenonymphina, Melanargia, Coenonympha, Sinonympha, Triphysa.** Milano: Omnes Partes, 2002.

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

Capblancq T, Després L, Rioux D, *et al.*: **Hybridization promotes speciation in *Coenonympha* butterflies.** *Mol Ecol.* 2015; **24**(24): 6209–22.
PubMed Abstract | Publisher Full Text

Capblancq T, Mavarez J, Rioux D, *et al.*: **Speciation with gene flow: evidence from a complex of alpine butterflies (*Coenonympha*, Satyridae).** *Ecol Evol.* 2019; **9**(11): 6444–6457.
PubMed Abstract | Publisher Full Text | Free Full Text

Capblancq T, Roux C, Boyer F, *et al.*: **Untangling the contribution of adaptive versus non-adaptive processes in the evolution of reproductive isolation between *Coenonympha* butterflies.** 2024.
Publisher 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

Doniol-Valcroze P: **Hybrid speciation and adaptation in alpine butterflies in the genus Coenonympha.** PhD thesis: Université de Montpellier, 2024.

Doniol-Valcroze P, Després L, Joron M: **Strong differences in egg thermal tolerance between closely related lowland and alpine butterfly species.** *Ecol Entomol.* 2024; **49**(4): 598–603.
Publisher Full Text

Doniol-Valcroze P, Develay Nguyen L, Buatois B, *et al.*: **Non-random sorting of parental chemical compounds during hybrid speciation.** *J Evol Biol.* 2025; **38**(5): 559–571.
PubMed Abstract | Publisher 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

Greenwood MP, Capblancq T, Wahlberg N, *et al.*: **Whole genome data confirm pervasive gene discordance in the evolutionary history of *Coenonympha* (Nymphalidae) butterflies.** *Mol Phylogenet Evol.* 2025; **202**: 108222.
PubMed Abstract | Publisher 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): giaa153.
PubMed Abstract | Publisher Full Text | Free Full Text

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

Kodandaramaiah U, Wahlberg N: **Phylogeny and biogeography of *Coenonympha* butterflies (Nymphalidae: Satyrinae) - patterns of colonization in the holarctic.** *Syst Entomol.* 2009; **34**(2): 315–23.
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.
Reference Source

Nève G, Després L: **Cold adaptation across the elevation gradient in an alpine butterfly species complex.** *Ecol Entomol.* 2020; **45**(5): 997–1003.
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

UICN-France: **Chapitre papillons de jour de france métropolitaine.** 2014.
Reference Source

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

Van Swaay CAM, Wynhoff I, Verovnik R, *et al.*: **Coenonympha gardetta.** *The IUCN Red List of Threatened Species.* 2010; **2010**: e.T173245A6979358.
Publisher Full Text

Vasimuddin M, Misra S, Li H, *et al.*: **Efficient architecture-aware acceleration of BWA-MEM for multicore systems.** In: *2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS).* IEEE, 2019; 314–324.
Publisher Full Text

Vrba P, Sucháčková Bartoňová A, Andres M, *et al.*: **Exploring cold hardiness within a butterfly clade: supercooling ability and polyol profiles in European Satyrinae.** *Insects.* 2022; **13**(4): 369.
PubMed Abstract | Publisher Full Text | Free 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