Check for updates

DATA NOTE

The genome sequence of the Violet Copper, *Helleia helle* (Denis & Schiffermüller), 1776 (Lepidoptera: Lycaenidae)

[version 1; peer review: 1 approved, 2 approved with reservations]

Paula Escuer[1], Kay Lucek[1], Charlotte J. Wright[2], Joana I. Meier[2], Mark L. Blaxter[2],
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]Tree of Life, Wellcome Sanger Institute, Hinxton, England, UK

v1 **First published:** 11 Aug 2025, **10**:429
https://doi.org/10.12688/wellcomeopenres.24699.1
Latest published: 11 Aug 2025, **10**:429
https://doi.org/10.12688/wellcomeopenres.24699.1

Abstract

We present a genome assembly from a male specimen of *Helleia helle* (Violet Copper; Arthropoda; Insecta; Lepidoptera; Lycaenidae). The assembly contains two haplotypes with total lengths of 540.27 megabases and 540.28 megabases. Most of haplotype 1 (99.66%) is scaffolded into 24 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 15.51 kilobases.

Keywords

Helleia helle, Violet Copper, genome sequence, chromosomal, Lepidoptera

This article is included in the Tree of Life gateway.

Open Peer Review

Approval Status ✓ ? ?

	1	2	3
version 1 11 Aug 2025	✓ view	? view	? view

1. **Bin Zhang**, Qingdao Agricultural University, Qingdao, China
2. **Laurence Despres**, Universite Savoie Mont Blanc, Chambéry, France
Laboratory of Alpine Ecology (Ringgold ID: 56837), Grenoble, France
3. **Arun Arumugaperumal**, Rajalakshmi Engineering College, Thandalam 602105, Chennai, India

Any reports and responses or comments on the article can be found at the end of the article.

Corresponding author: Charlotte J. Wright (cw22@sanger.ac.uk)

Author roles: Escuer P: Investigation, Resources, 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). PE and KL were supported by the Swiss National Science Foundation Grant PCEFP3_202869.
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2025 Escuer P *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: Escuer P, Lucek K, Wright CJ *et al.* **The genome sequence of the Violet Copper, *Helleia helle* (Denis & Schiffermüller), 1776 (Lepidoptera: Lycaenidae) [version 1; peer review: 1 approved, 2 approved with reservations]** Wellcome Open Research 2025, **10**:429 https://doi.org/10.12688/wellcomeopenres.24699.1

First published: 11 Aug 2025, **10**:429 https://doi.org/10.12688/wellcomeopenres.24699.1

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; Lycaenidae; Lycaeninae; *Helleia*; *Helleia helle* (Denis & Schiffermüller), 1776 (NCBI:txid2795559)

Background

The Violet Copper or *Helleia helle* (Denis & Schiffermüller, 1776) is a species of butterfly within the Lycaenidae family. It is present in Central and Northern Europe, Central and South Siberia, Mongolia, Transbaikal and Amur. This species can be found in very dispersed and restricted localities (Fischer *et al.*, 1999) in altitudes of 100–1 800m.

Helleia helle shows clear sexual dimorphism, with males displaying small forewings with intense violet reflections overlaying orange and black, and a marginal band. The hindwing has a prominent orange stripe, with connected black spots and white borders. Females are larger, with reduced or absent violet fading in comparison to the males.

Helleia helle prefers moist habitats such as abandoned floodplain meadows with flowers, close to rivers and lakes, and swamps with moss (*Sphagnum*), often associated with *Vaccinium* spp. and open forests. Larvae feed on *Bistorta officinalis* and *Polygonum viviparum*. The species lays its eggs on the underside of the leaf, and the young arva consume the lower cuticle, creating characteristic translucent "windows". It then overwinters as pupae and emerges as adults from May to July, depending on the latitude and altitude of the population (Tolman & Lewington, 2011).

During the last ice age and early postglacial period, the Violet Copper was widely distributed across Central Europe (Habel *et al.*, 2011). The species underwent a strong decline during the postglacial warming, causing a shift to northern regions, and only a few isolated populations at higher altitudes persisted across Central Europe (Habel *et al.*, 2011).

Close populations exhibit high genetic differentiation, reflecting strong habitat fragmentation and reduced gene flow (Habel *et al.*, 2011). Threats such as habitat deterioration from drainage, secondary succession and forestation have caused a marked population decline in recent decades (Trense *et al.*, 2022). In addition, climate change may reduce connectivity among populations, further threatening the species' conservation (Habel *et al.*, 2011). Therefore, *H. helle* is listed as Endangered on the IUCN Red List of Threatened Species (van Swaay *et al.*, 2010) and in the Spanish Red List (Jubete *et al.*, 2019).

The chromosome-level assembly of *H. helle* represents an important resource to assess more precisely the genetic diversity of this endangered species through population genomics studies. This will be important for developing effective conservation strategies and predicting the long-term viability of populations. The sequence data were derived from a male specimen (Figure 1) collected from St-Imier, Bern, Switzerland.

Figure 1. Voucher photograph of the *Helleia helle* (ilHelHell3) specimen used for genome sequencing.

Methods

Sample acquisition

The specimen used for genome sequencing was an adult male *Helleia helle* (specimen ID SAN28000063, ToLID ilHelHell3; Figure 1), collected from St-Imier, Bern, Switzerland (latitude 47.128, longitude 6.9992; elevation 1 091 m) on 06/06/2023. The specimen was collected and identified by Paula Escuer (University of Neuchâtel).

Nucleic acid extraction

Protocols for high molecular weight (HMW) DNA extraction developed at the Wellcome Sanger Institute (WSI) Tree of Life Core Laboratory are available on protocols.io (Howard *et al.*, 2025). The ilHelHell3 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 39.08 ng/μL and a yield of 1 836.76 ng, with a fragment size of 17.1 kb. The 260/280 spectrophotometric ratio was 1.84, and the 260/230 ratio was 2.5.

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 ilHelHell3 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.

Platform	PacBio HiFi	Hi-C
ToLID	ilHelHell3	ilHelHell3
Specimen ID	SAN28000063	SAN28000063
BioSample (source individual)	SAMEA115109739	SAMEA115109739
BioSample (tissue)	SAMEA115109778	SAMEA115109777
Tissue	thorax	head
Sequencing platform and model	Revio	Illumina NovaSeq X
Run accessions	ERR13605515	ERR13602178
Read count total	1.49 million	566.10 million
Base count total	17.30 Gb	85.48 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 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 9 breaks and 5 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 *Helleia helle* was sequenced using Pacific Biosciences single-molecule HiFi long reads, generating 17.30 Gb (gigabases) from 1.49 million reads, which were used to assemble the genome. GenomeScope2.0 analysis estimated the haploid genome size at 539.53 Mb, with a heterozygosity of 0.47% and repeat content of 30.79%. These estimates guided expectations for the assembly. Based on the estimated genome size, the sequencing data provided approximately 31× coverage. Hi-C sequencing produced 85.48 Gb from 566.10 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 540.27 Mb in 45 scaffolds, with 82 gaps, and a scaffold N50 of 23.19 Mb (Table 2).

Most of the assembly sequence (99.66%) was assigned to 24 chromosomal-level scaffolds, representing 23 autosomes and the Z sex chromosome. The Z chromosome was identified based on synteny with the genome of *Lycaena phlaeas* (GCA_905333005.2). 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.

Table 2. Genome assembly statistics.

Assembly name	ilHelHell3.hap1.1	ilHelHell3.hap2.1
Assembly accession	GCA_964264505.1	GCA_964264415.1
Assembly level	chromosome	scaffold
Span (Mb)	540.27	540.28
Number of chromosomes	24	N/A
Number of contigs	127	116
Contig N50	10.72 Mb	9.37 Mb
Number of scaffolds	45	42
Scaffold N50	23.19 Mb	23.07 Mb
Longest scaffold length (Mb)	33.98	N/A
Sex chromosomes	Z	N/A
Organelles	Mitochondrial genome: 15.51 kb	N/A

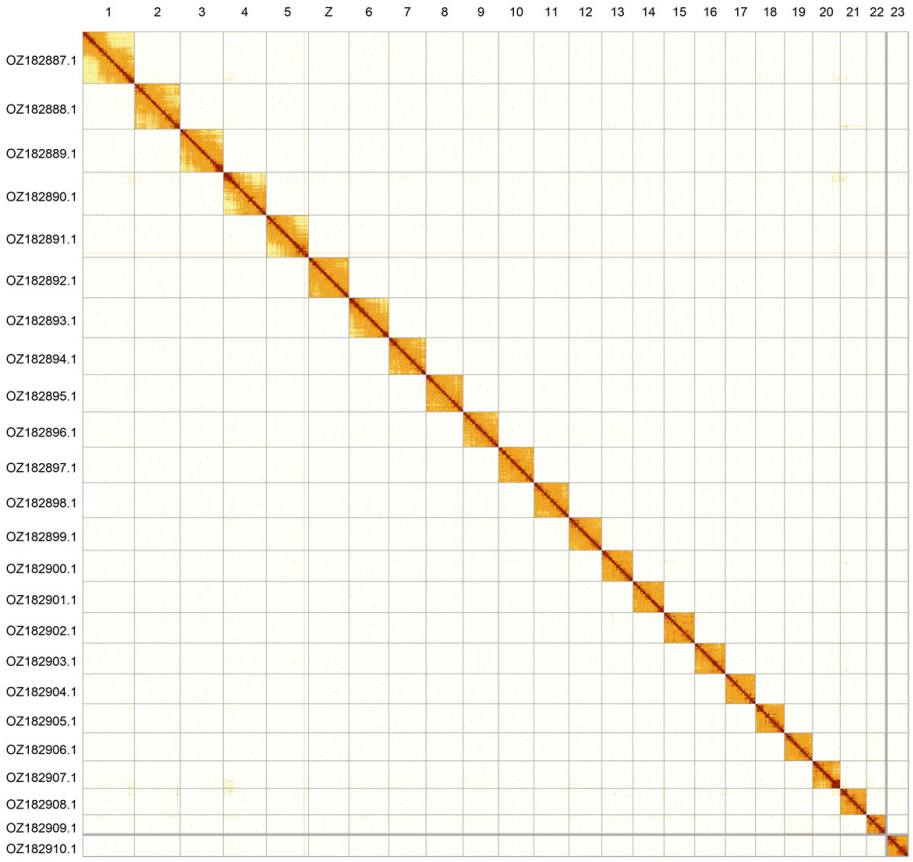

Figure 2. Hi-C contact map of the *Helleia helle* genome assembly. Assembled chromosomes are shown in order of size and labelled along the axes. The plot was generated using PretextSnapshot.

Table 3. Chromosomal pseudomolecules in the haplotype 1 genome assembly of *Helleia helle* ilHelHell3.

INSDC accession	Molecule	Length (Mb)	GC%	Assigned Merian elements
OZ182887.1	1	33.98	37	M1;M19
OZ182888.1	2	30.02	37.50	M11;M23
OZ182889.1	3	28.14	37.50	M12;M29
OZ182890.1	4	28.10	37	M18;M30
OZ182891.1	5	27.64	37.50	M25;M5
OZ182893.1	6	26.15	38	M14;M26
OZ182894.1	7	24.32	37.50	M17;M20
OZ182895.1	8	24.19	37.50	M2
OZ182896.1	9	23.19	37.50	M3
OZ182897.1	10	23.08	37.50	M8
OZ182898.1	11	22.90	37.50	M9
OZ182899.1	12	21.51	38	M7
OZ182900.1	13	20.41	38	M6
OZ182901.1	14	20.25	38	M16
OZ182902.1	15	20.07	38.50	M22
OZ182903.1	16	20.05	38	M4
OZ182904.1	17	19.73	38	M21
OZ182905.1	18	18.84	38	M15
OZ182906.1	19	18.37	38	M10
OZ182907.1	20	18.12	38	M28;M31
OZ182908.1	21	17.12	38	M13
OZ182909.1	22	13.66	39	M27
OZ182910.1	23	13.60	39	M24
OZ182892.1	Z	26.33	36.50	MZ

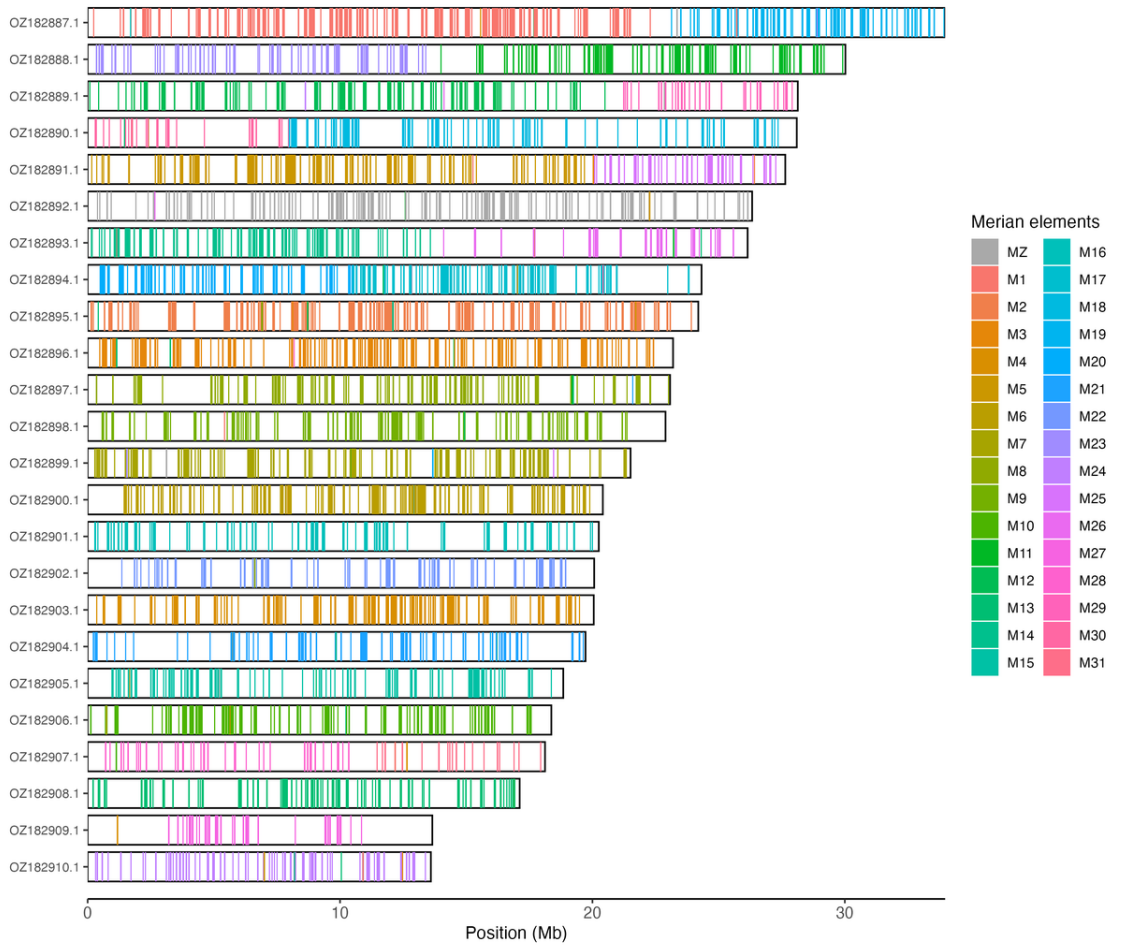

Figure 3. Merian elements painted across chromosomes in the ilHelHell3.hap1.1 assembly of *Helleia helle*. Chromosomes are drawn to scale, with the positions of orthologues shown as coloured bars. Each orthologue is coloured by the Merian element that it belongs to. All orthologues which could be assigned to Merian elements are shown.

Assembly quality metrics

For haplotype 1, the estimated QV is 64.9, and for haplotype 2, 65.1. When the two haplotypes are combined, the assembly achieves an estimated QV of 65.0. The *k*-mer completeness is 88.29% for haplotype 1, 88.27% for haplotype 2, and 99.49% for the combined haplotypes (Figure 4). BUSCO analysis using the lepidoptera_odb10 reference set (n = 5 286) (Kriventseva *et al.*, 2019) identified 98.4% of the expected gene set (single = 98.1%, duplicated = 0.3%) 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 **7.C.Q64**, 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

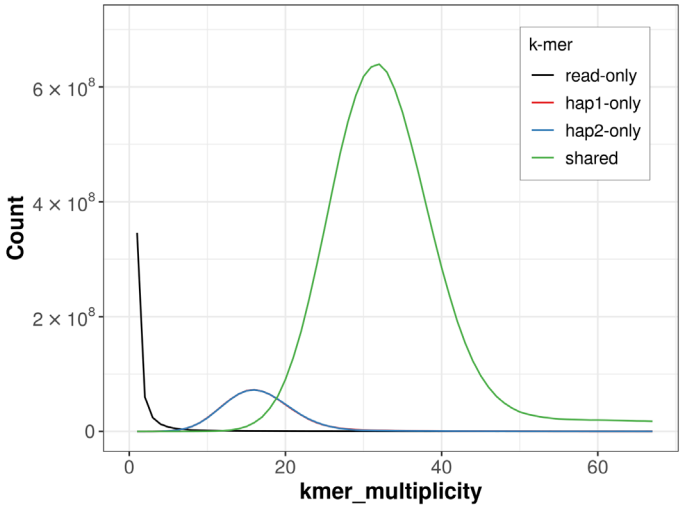

Figure 4. Evaluation of *k*-mer completeness using MerquryFK. This plot illustrates the recovery of *k*-mers from the original read data in the final assemblies. The horizontal axis represents *k*-mer multiplicity, and the vertical axis shows the number of *k*-mers. The black curve represents *k*-mers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

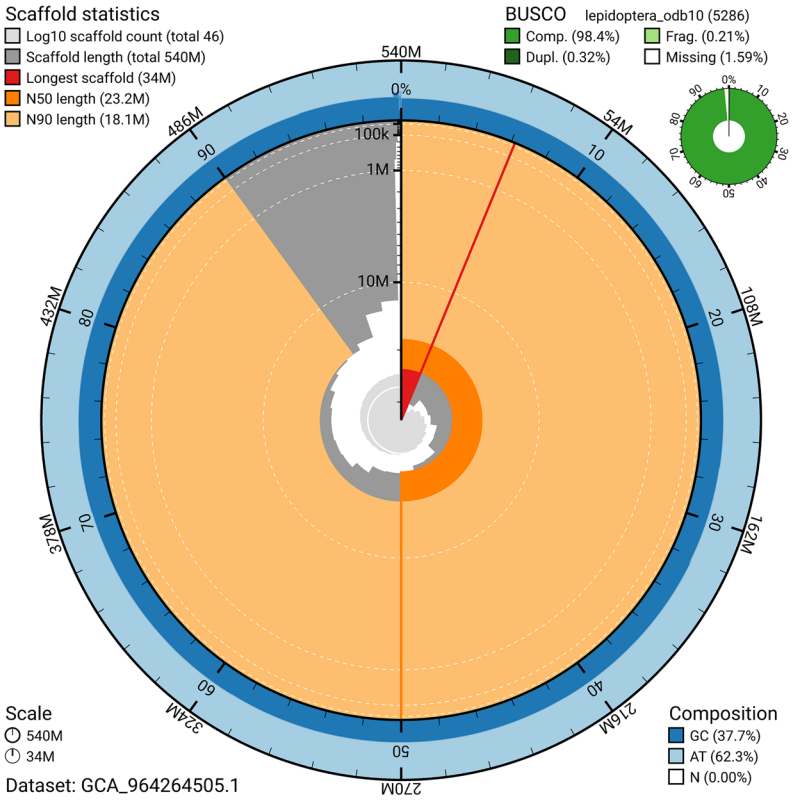

Figure 5. Assembly metrics for ilHelHell3.hap1.1. The BlobToolKit snail plot provides an overview of assembly metrics and BUSCO gene completeness. The circumference represents the length of the whole genome sequence, and the main plot is divided into 1,000 bins around the circumference. The outermost 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.

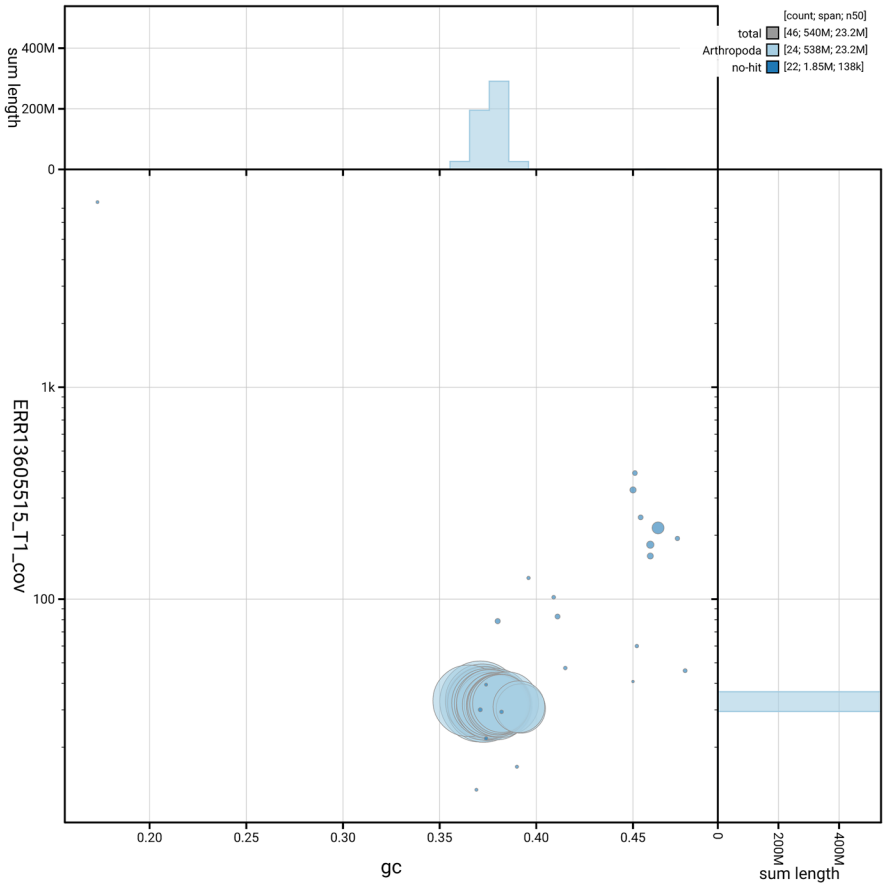

Figure 6. BlobToolKit GC-coverage plot for ilHelHell3.hap1.1. Blob plot showing sequence coverage (vertical axis) and GC content (horizontal axis). The circles represent scaffolds, with the size proportional to scaffold length and the colour representing phylum membership. The histograms along the axes 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 *Helleia helle* assembly.

Measure	Value	Benchmark
EBP summary (haplotype 1)	7.7.Q64	6.C.Q40
Contig N50 length	10.72 Mb	≥ 1 Mb
Scaffold N50 length	23.19 Mb	= chromosome N50
Consensus quality (QV)	Haplotype 1: 64.9; haplotype 2: 65.1; combined: 65.0	≥ 40
k-mer completeness	Haplotype 1: 88.29%; Haplotype 2: 88.27%; combined: 99.49%	≥ 95%
BUSCO	C:98.4%[S:98.1%,D:0.3%], F:0.2%,M:1.4%,n:5 286	S > 90%; D < 5%
Percentage of assembly assigned to chromosomes	99.66%	≥ 90%

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: Helleia helle (violet copper). Accession number PRJEB79204. The genome sequence is released openly for reuse. The *Helleia helle* 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
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

Software	Version	Source
sanger-tol/blobtoolkit	0.6.0	https://github.com/sanger-tol/blobtoolkit
Seqtk	1.3	https://github.com/lh3/seqtk
Singularity	3.9.0	https://github.com/sylabs/singularity
TreeVal	1.2.0	https://github.com/sanger-tol/treeval
YaHS	1.2a.2	https://github.com/c-zhou/yahs

References

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

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

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

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

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

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

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

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

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

Fischer K, Beinlich B, Plachter H: **Population structure, mobility and habitat preferences of the violet copper *Lycaena helle* (Lepidoptera: Lycaenidae) in Western Germany: implications for conservation.** *Journal of Insect Conservation.* 1999; **3**: 43–52.
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

Habel JC, Roedder D, Schmitt T, *et al.*: **Global warming will affect the genetic diversity and uniqueness of *Lycaena helle* populations.** *Glob Change Biol.* 2011; **17**(1): 194–205.
Publisher 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

Jubete F, Barea Azcón J, Jiménez R, *et al.*: **Bases Técnicas Para La Conservación de Los Lepidópteros Amenazados En España.** Asociación de Naturalistas Palentinos, 2019.

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

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

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

Tolman T, Lewington R: **Mariposas de España y Europa.** Lynx Edicions, 2011.
Reference Source

Trense D, Habel JC, Finger A, *et al.*: **Contrasting genetic responses to habitat fragmentation for two Lycaenid butterfly species.** *Insect Conserv Divers.* 2022; **15**(3): 337–47.
Publisher 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, Wynhoff I, Verovnik R, *et al.*: ***Lycaena helle* (Europe assessment).** 2010.
Publisher Full Text

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

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 12 September 2025

https://doi.org/10.21956/wellcomeopenres.27215.r129655

© **2025 Arumugaperumal A.** 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.

?

Arun Arumugaperumal
Department of Biotechnology, Rajalakshmi Engineering College, Thandalam 602105, Chennai, Tamil Nadu, India

The authors have presented the genome sequencing of a butterfly species *Helleia helle*. This is the second high-quality genome of the organism to be sequenced. The assembly provided is of size 540.27 Mb. The genome assembly consists of 24 chromosomes. Pladevall *et al* had already sequenced the genome of the same organism but the specimen used was a female and the assembly consisted of 25 chromosomes including Z and W chromosomes [1]. This data note has used a male specimen. The genome assembly in both the cases is nearly of the same quality. The genome was sequenced using high-end platforms and assembled using standard procedures. However, the previously published article was not cited. The authors should include the Pladvell *et al* paper in the reference. The voucher photograph in figure 1 is very clear but lacking the morphological details of the organism. Also, the organism is having confusion in name. So it is advisable to include a photograph of alive butterfly as an inset to figure 1. It will help the readers to recognize the species immediately without any confusion.

References

1. Pladevall C, Caritg R, Cámara F, Carbonell S, et al.: The genome sequence of the Violet Copper, Lycaena helle (Denis & Schiffermüller, 1775). *F1000Research*. 2025; **14**. Publisher Full Text

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: Genomics; Bioinformatics

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, however I have significant reservations, as outlined above.

Reviewer Report 30 August 2025

https://doi.org/10.21956/wellcomeopenres.27215.r129650

© 2025 Despres L. 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.

?

Laurence Despres
[1] Universite Savoie Mont Blanc, Chambéry, France
[2] Université Grenoble Alpes, Laboratory of Alpine Ecology (Ringgold ID: 56837), Grenoble, Auvergne-Rhône-Alpes, France

The authors report on the chromosome-level genome assembly of a male specimen of the violet copper Lycaena (Helleia) helle. The genome assembly is of good quality and was constructed using appropriate methods : Pac Bio HiFi long reads and Hi-C data to confirm scaffolding, following the high standard of the Darwin Tree of Life Project for results presentation.
I have two main concerns regarding this data note.
First, the taxonomy used : Helleia helle (Denis & Schiffermüller, 1776), is not the traditionally used taxonomy for the violet copper : Lycaena helle (Denis & Schiffermüller, 1775)-the type species being *Papilio helle* Denis & Schiffermüller, 1775. All the cited references are about Lycaena helle (not Helleia helle). The splitting of the Lycaeninae and Lycaena genus into several genera including the monotypic genus Helleia has been recently proposed (Zhang et al 2020 https://pmc.ncbi.nlm.nih.gov/articles/PMC8794283/#S10) based on a phylogeny of the Lycaeninae sub-family using protein coding genes which involved a unique specimen of L. helle from Siberia, but its phylogenetic placement was inconsistent with the phylogeny obtained based on the Z chromosome (Zhang et al. 2022 https://pmc.ncbi.nlm.nih.gov/articles/PMC9645532/). This new taxonomy is far from being accepted/known by the lepidopterists' community and I suggest either to keep the name Lycaena helle or explain the new genus name 'Helleia (Verity 1947)' and cite Zhang et al 2020, 2022.
Second, there is another chromosome level assembly of a female genome of Lycaena helle https://pmc.ncbi.nlm.nih.gov/articles/PMC12174908/ released in January 2025 which should be cited. The present genome is an assembly of a male from Switzerland, while the female genome is from Andorra, and together with the specimen from Siberia, the three genomic datasets could shed light on the taxonomic status of Lycaena/Hellia helle species and its placement in the phylogeny of Lycaeninae.

Is the rationale for creating the dataset(s) clearly described?
Partly

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?
Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: evolutionary biology

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, however I have significant reservations, as outlined above.

Reviewer Report 28 August 2025

https://doi.org/10.21956/wellcomeopenres.27215.r129657

© 2025 Zhang B. 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.

Bin Zhang
Qingdao Agricultural University, Qingdao, China

This Data Note presents a high-quality, haplotype-resolved chromosome-level genome assembly of the endangered butterfly Helleia helle (Violet Copper), including PacBio HiFi (17.3 Gb) and Hi-C (85.48 Gb) sequencing data. Key achievements include: Assembly of two haplotypes (540.27/540.28 Mb), with 99.66% of haplotype 1 scaffolded into 24 chromosomes (including Z sex chromosome). BUSCO completeness of 98.4% (lepidoptera_odb10), QV > 64, and mitochondrial genome (15.51 kb). Contextualization of the genome's utility for conservation genomics of this IUCN-endangered species. I would like to give the decision with "Accept after minor revisions".
1. Table 3: State that chromosomes are ordered by size (match Fig 2's Hi-C map).
2. Table 4: Decode "7.7.Q64" (e.g., "Scaffold N50 = 7.7 Mb, QV = 64") and benchmark "6.C.Q40" ("C" = chromosome-level).
3. Discuss specific applications (e.g., identifying inbreeding depression loci, designing genetic corridors). Update IUCN citation (van Swaay et al. 2010 Latest assessment).
4. Define "Merian elements" upon first use (Page 5: "ancestral lepidopteran linkage groups").
5. Provide ENA hyperlinks for raw data (e.g., HiFi: ERR13605515; Hi-C: ERR13602178). Specify Ensembl annotation release timeline.

6. Use "kb" not "kilobases" (e.g., "15.51 kb"). Add commas to numbers (e.g., "1,836.76 ng" not "1 836.76 ng").
7. Correct errors: "Bateman" (not "Baerman"); "Altschul" (not "Alicciato"); Nature (not "Not").
8. Table 5: Replace "N/A" for *PretextSnapshot* with a version/git commit hash.

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: Entomology, Bioinformatics

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.