DATA NOTE

The genome sequence of the Purple-edged Copper, *Lycaena hippothoe* (Lepidoptera: Lycaenidae)

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

Hans-Peter Wymann[1], Alena Suchackova Bartonova[2], Kay Lucek[3], Charlotte J. Wright[4], Joana I. Meier[4], Mark L. Blaxter[4], 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]Independent researcher, Münchringen, Bern, Switzerland
[2]Biology Centre Czech Academy of Sciences, Ceske Budejovice, South Bohemia, Czech Republic
[3]University of Neuchâtel, Neuchâtel, Switzerland
[4]Tree of Life, Wellcome Sanger Institute, Hinxton, England, UK

v1 **First published:** 22 Aug 2025, **10**:458
https://doi.org/10.12688/wellcomeopenres.24747.1
Latest published: 22 Aug 2025, **10**:458
https://doi.org/10.12688/wellcomeopenres.24747.1

Abstract

We present a genome assembly from a female specimen of *Lycaena hippothoe* (Purple-edged Copper; Arthropoda; Insecta; Lepidoptera; Lycaenidae). The assembly contains two haplotypes with total lengths of 484.25 megabases and 431.80 megabases. Most of haplotype 1 (99.6%) is scaffolded into 25 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 15.43 kilobases.

Keywords

Lycaena hippothoe, Purple-edged 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 22 Aug 2025	? view	✓ view	? view

1. **Ozren Polašek**, University of Split, Split, Croatia
2. **David Lohman**, City College of New York AND National Museum of Natural History, New York, USA
The City College of New York Department of Biology (Ringgold ID: 171428), New York, USA
City University of New York (Ringgold ID: 2009), New York, USA
National Museum of the Philippines (Ringgold ID: 198024), Manila, Philippines
3. **Megan Barkdull**, Cornell University, Ithaca, USA

Natural History Museum of Los Angeles County (Ringgold ID: 5119), Los Angeles, USA

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: Wymann HP: Investigation, Resources; **Suchackova Bartonova A**: Writing – Original Draft Preparation; **Lucek K**: Investigation, Resources; **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). AS was supported by VEDA FELLOWSHIPS (CZ.02.01.01/00/22_010/0008117) and 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: Wymann HP, Suchackova Bartonova A, Lucek K *et al.* **The genome sequence of the Purple-edged Copper, *Lycaena hippothoe* (Lepidoptera: Lycaenidae) [version 1; peer review: 1 approved, 2 approved with reservations]** Wellcome Open Research 2025, **10**:458 https://doi.org/10.12688/wellcomeopenres.24747.1

First published: 22 Aug 2025, **10**:458 https://doi.org/10.12688/wellcomeopenres.24747.1

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; *Lycaena*; *Lycaena hippothoe* (NCBI:txid580924)

Background

Lycaena hippothoe (Linnaeus, 1760), the Purple-edged Copper, is a Palaearctic species, occurring from Northern Spain, Italy, northern Fennoscandia, and throughout Siberia to the Far East. It is absent from the Great Britain. The forewing length is 14–18.5 mm (Middleton-Welling *et al.*, 2020) The males have coppery uppersides with iridescent violet margins, females are brown with orange submarginal markings. In the form described as subspecies *L. hippothoe eurydame* (Hoffmannsegg, 1806), occurring in the Alps, no iridescent colouration is present, and the wing margins are black (Tolman & Lewington, 2009). The species is replaced by *Lycaena candens* (Herrich-Schäffer, 1844) in the southern Balkans, Turkey, Caucasus, Transcaucasia, and northern Iran; *L. candens* differs from *L. hippothoe* in genitalia structures (Wagner, 2005). DNA barcoding studies showed the highest genetic diversity of *L. hippothoe* in the Alps, where several lineages meet (Dapporto *et al.*, 2022; Sucháčková Bartoňová *et al.*, 2024).

Lycaena hippothoe inhabits wet meadows and fens, up to 2 200 m a. s. l. (Settele *et al.*, 2008; Tolman & Lewington, 2009), or drier grasslands rich in forbs (Sawchik *et al.*, 2003). It is a sedentary species (Trense *et al.*, 2022a; Trense *et al.*, 2022b), requiring a dense network of habitats (Fischer & Fiedler, 2001). The main host plant is *Rumex acetosa* L, but other *Rumex* spp., and *Polygonum bistorta* L. are also used (Tolman & Lewington, 2009). Females preferentially oviposit on host plants protruding surrounding vegetation, where the warmer microclimate allows for quicker development of larvae (Turlure & Van Dyck, 2009). Larva overwinters at the base of the host plant stem. The species forms one adult generation in June/ July, and locally two generations in May and August (Beneš *et al.*, 2002; Tolman & Lewington, 2009).

According to the IUCN European Red List of butterflies, *L. hippothoe* is considered of least concern in Europe, and near threatened in the European Union (Maes *et al.*, 2019; Van Swaay *et al.*, 2010). However, it has declined in Central Europe (Settele *et al.*, 2008). Populations from the Alps have demonstrated a vulnerability to habitat fragmentation (Trense *et al.*, 2022a).

We present a chromosome-level, haplotype-resolved genome sequence of *Lycaena hippothoe*, sequenced as part of Project Psyche. The sequence data was derived from a female specimen (Figure 1) collected from St-Imier, Bern, Switzerland.

Figure 1. Voucher photograph of the *Lycaena hippothoe* (ilLycHipp1) specimen used for genome sequencing.

Methods

Sample acquisition and DNA barcoding

The specimen used for genome sequencing was an adult female *Lycaena hippothoe* (specimen ID SAN28000071, ToLID ilLycHipp1; Figure 1), collected from St-Imier, Bern, Switzerland (latitude 47.1282, longitude 6.9989; elevation 1 095 m). The specimen was collected and identified by Hans-Peter Wymann (Natural History Museum of Bern).

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 ilLycHipp1 sample was weighed and triaged to determine the appropriate extraction protocol. Tissue from the thorax was homogenised by powermashing using a PowerMasher II tissue disruptor.

HMW DNA was extracted in the WSI Scientific Operations core using the Automated MagAttract v2 protocol. DNA was sheared into an average fragment size of 12–20 kb following the Megaruptor®3 for LI PacBio protocol. Sheared DNA was purified by automated SPRI (solid-phase reversible immobilisation). The concentration of the sheared and purified DNA was assessed using a Nanodrop spectrophotometer and Qubit Fluorometer using the Qubit dsDNA High Sensitivity Assay kit. Fragment size distribution was evaluated by running the sample on the FemtoPulse system. For this sample, the final post-shearing DNA had a Qubit concentration of 47.05 ng/μL and a yield of 2 211.35 ng, with a fragment size of 14.7 kb. The 260/280 spectrophotometric ratio was 1.88, and the 260/230 ratio was 2.73.

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 ilLycHipp1 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

Table 1. Specimen and sequencing data for BioProject.

Platform	PacBio HiFi	Hi-C
ToLID	ilLycHipp1	ilLycHipp1
Specimen ID	SAN28000071	SAN28000071
BioSample (source individual)	SAMEA115109746	SAMEA115109746
BioSample (tissue)	SAMEA115109855	SAMEA115109792
Tissue	thorax	head
Sequencing platform and model	Revio	Illumina NovaSeq X
Run accessions	ERR13605510	ERR13602173
Read count total	2.04 million	906.18 million
Base count total	22.50 Gb	136.83 Gb

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 102 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 *Lycaena hippothoe* was sequenced using Pacific Biosciences single-molecule HiFi long reads, generating 22.50 Gb (gigabases) from 2.04 million reads, which were used to assemble the genome. GenomeScope2.0 analysis estimated the haploid genome size at 452.63 Mb, with a heterozygosity of 0.96% and repeat content of 28.66%. These estimates guided expectations for the assembly. Based on the estimated genome size, the sequencing data provided approximately 48× coverage. Hi-C sequencing produced 136.83 Gb from 906.18 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 484.25 Mb in 54 scaffolds, with 152 gaps, and a scaffold N50 of 20.36 Mb (Table 2).

Table 2. Genome assembly statistics.

Assembly name	ilLycHipp1.hap1.1	ilLycHipp1.hap2.1
Assembly accession	GCA_964264015.1	GCA_964264095.1
Assembly level	chromosome	scaffold
Span (Mb)	484.25	431.80
Number of chromosomes	25	N/A
Number of contigs	206	130
Contig N50	10.16 Mb	10.09 Mb
Number of scaffolds	54	92
Scaffold N50	20.36 Mb	19.25 Mb
Longest scaffold length (Mb)	29.58	N/A
Sex chromosomes	W and Z	N/A
Organelles	Mitochondrial genome: 15.43 kb	N/A

Most of the assembly sequence (99.6%) was assigned to 25 chromosomal-level scaffolds, representing 23 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).

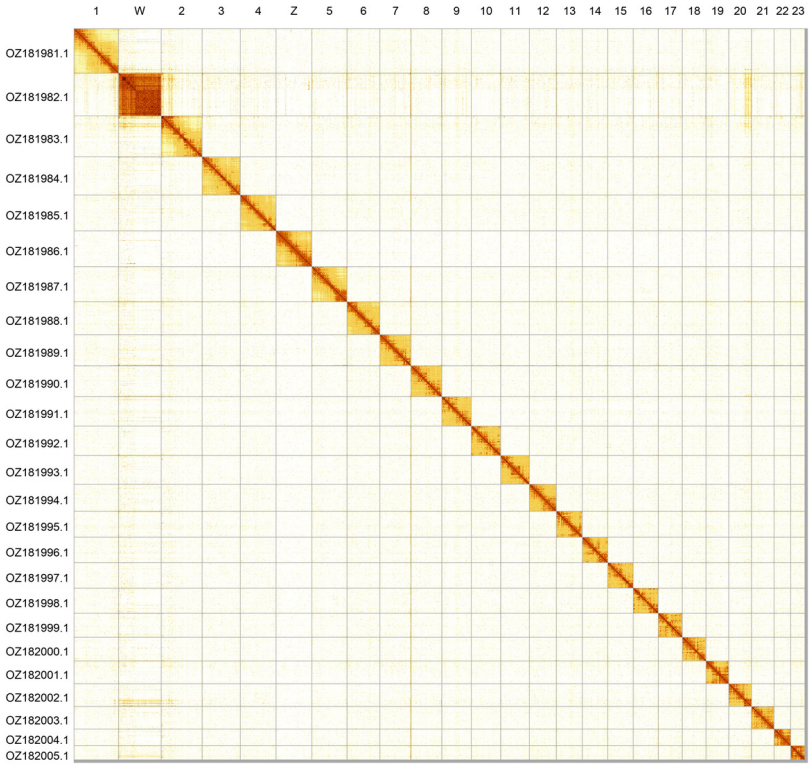

Figure 2. Hi-C contact map of the *Lycaena hippothoe* genome assembly. Assembled chromosomes are shown in order of size and labelled along the axes. The plot was generated using PretextSnapshot.

Table 3. Chromosomal pseudomolecules in the haplotype 1 genome assembly of *Lycaena hippothoe* ilLycHipp1.

INSDC accession	Molecule	Length (Mb)	GC%	Assigned Merian elements
OZ181981.1	1	29.58	36	M1;M19
OZ181983.1	2	27.09	36.50	M18;M30
OZ181984.1	3	25.16	35.50	M11;M23
OZ181985.1	4	23.60	35.50	M12;M29
OZ181987.1	5	23.23	35.50	M25;M5
OZ181988.1	6	21.75	36.50	M14;M26
OZ181989.1	7	20.43	36	M17;M20
OZ181990.1	8	20.36	36	M2
OZ181991.1	9	19.45	36	M3
OZ181992.1	10	19.37	35.50	M9
OZ181993.1	11	18.91	36	M8
OZ181994.1	12	17.83	36.50	M7
OZ181995.1	13	17.17	36.50	M4
OZ181996.1	14	16.77	36	M6
OZ181997.1	15	16.73	35.50	M16
OZ181998.1	16	16.58	36	M21
OZ181999.1	17	15.83	36.50	M22
OZ182000.1	18	15.59	36.50	M15
OZ182001.1	19	15.12	37	M28;M31
OZ182002.1	20	14.95	36.50	M13
OZ182003.1	21	14.88	37	M10
OZ182004.1	22	10.97	37.50	M24
OZ182005.1	23	9.94	38	M27
OZ181982.1	W	28.08	38.50	N/A
OZ181986.1	Z	23.54	35	MZ

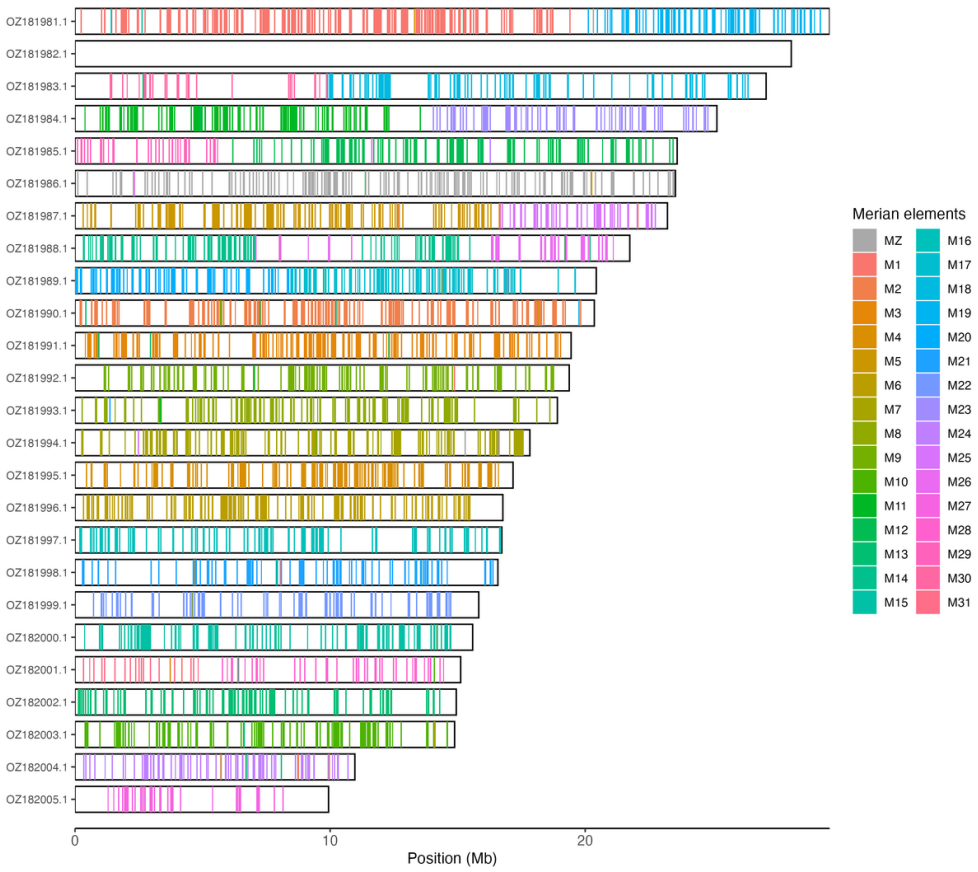

Figure 3. Merian elements painted across chromosomes in the ilLycHipp1.hap1.1 assembly of *Lycaena hippothoe*. Chromosomes are drawn to scale, with the positions of orthologues shown as coloured bars. Each orthologue is coloured by the Merian element that it belongs to. All orthologues which could be 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 62.8, and for haplotype 2, 63.0. When the two haplotypes are combined, the assembly achieves an estimated QV of 62.9. The *k*-mer completeness is 81.47% for haplotype 1, 77.10% for haplotype 2, and 99.56% 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 98.0%, 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 **7.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 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

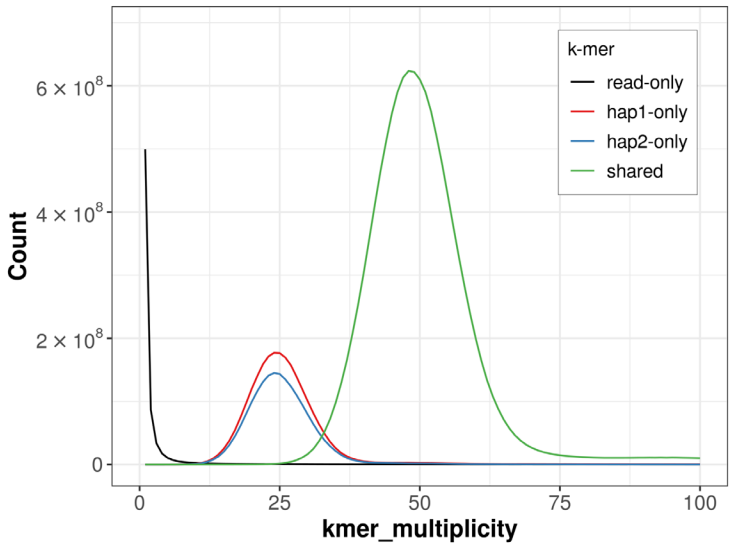

Figure 4. Evaluation of *k*-mer completeness using MerquryFK. This plot illustrates the recovery of *k*-mers from the original read data in the final assemblies. The horizontal axis represents *k*-mer multiplicity, and the vertical axis shows the number of *k*-mers. The black curve represents *k*-mers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

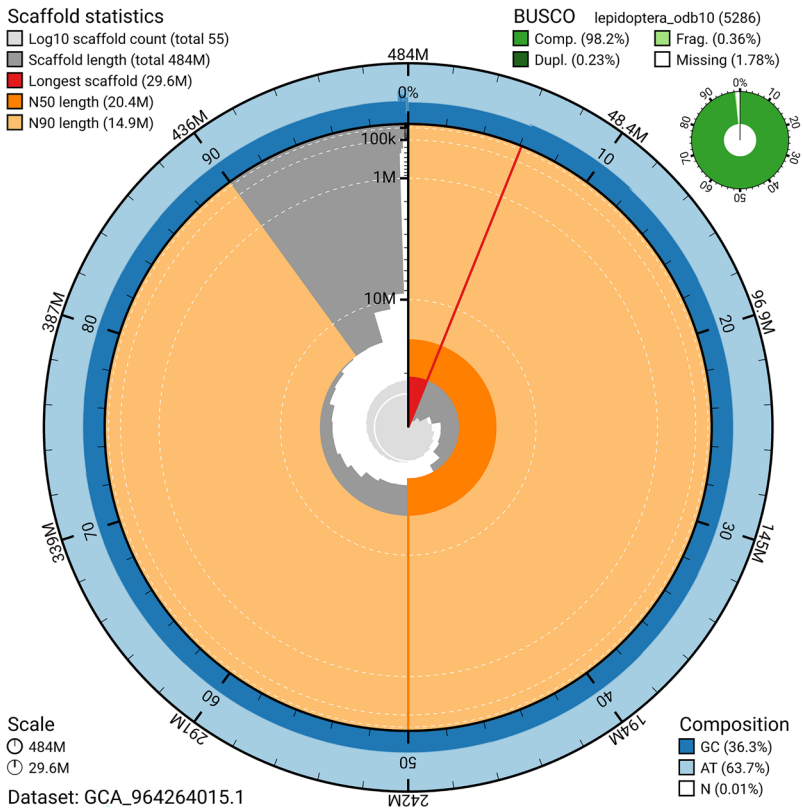

Figure 5. Assembly metrics for ilLycHipp1.hap1.1. The BlobToolKit snail plot provides an overview of assembly metrics and BUSCO gene completeness. The circumference represents the length of the whole genome sequence, and the main plot is divided into 1,000 bins around the circumference. The outermost 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.

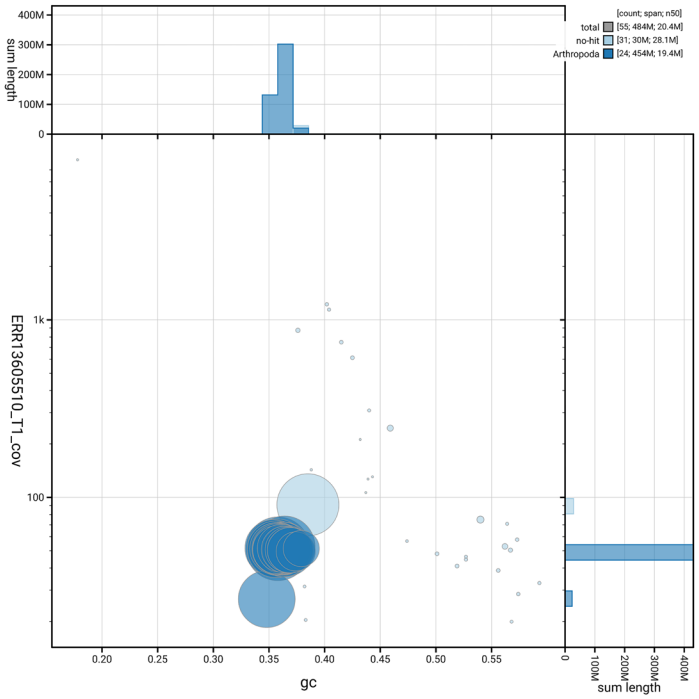

Figure 6. BlobToolKit GC-coverage plot for ilLycHipp1.hap1.1. Blob plot showing sequence coverage (vertical axis) and GC content (horizontal axis). The circles represent scaffolds, with the size proportional to scaffold length and the colour representing phylum membership. The histograms along the axes 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 *Lycaena hippothoe* assembly.

Measure	Value	Benchmark
EBP summary (haplotype 1)	7.7.Q62	6.C.Q40
Contig N50 length	10.16 Mb	≥ 1 Mb
Scaffold N50 length	20.36 Mb	= chromosome N50
Consensus quality (QV)	Haplotype 1: 62.8; haplotype 2: 63.0; combined: 62.9	≥ 40
k-mer completeness	Haplotype 1: 81.47%; Haplotype 2: 77.10%; combined: 99.56%	≥ 95%
BUSCO	C:98.2%[S:98.0%,D:0.2%], F:0.4%,M:1.4%,n:5 286	S > 90%; D < 5%
Percentage of assembly assigned to chromosomes	99.6%	≥ 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: Lycaena hippothoe (purple-edged copper). Accession number PRJEB79194. The genome sequence is released openly for reuse. The *Lycaena hippothoe* 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
MerquryFK	1.1.2	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
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

Beneš J, Konvička M, Dvořák J, *et al.*: **Motýli České Republiky: Rozšíření a ochrana I, II.** Společnost pro ochranu motýlů, 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

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

Dapporto L, Menchetti M, Vodă R, *et al.*: **The atlas of mitochondrial genetic diversity for Western Palaearctic butterflies.** *Glob Ecol Biogeogr.* 2022; **31**(11): 2184–2190.
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

Fischer K, Fiedler K: **Egg weight variation in the butterfly *Lycaena hippothoe*: more small or fewer large eggs?** *Popul Ecol.* 2001; **43**(1): 105–9.
Publisher Full Text

Formenti G, Abueg L, Brajuka A, *et al.*: **Gfastats: conversion, evaluation and manipulation of genome sequences using assembly graphs.** *Bioinformatics.* 2022; **38**(17): 4214–4216.
PubMed Abstract | Publisher Full Text | Free Full Text

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

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

Howe K, Chow W, Collins J, *et al.*: **Significantly improving the quality of genome assemblies through curation.** *GigaScience.* 2021; **10**(1): 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

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

Maes D, Verovnik R, Wiemers M, *et al.*: **Integrating national Red Lists for prioritising conservation actions for European butterflies.** *J Insect Conserv.* 2019; **23**(2): 301–30.
Publisher 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

Middleton-Welling J, Dapporto L, García-Barros E, *et al.*: **A new comprehensive trait database of European and Maghreb butterflies, Papilionoidea.** *Sci Data.* 2020; **7**(1): 351.
PubMed Abstract | Publisher Full Text | Free 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

Sawchik J, Dufrêne M, Lebrun P: **Estimation of habitat quality based on plant community, and effects of isolation in a network of butterfly habitat patches.** *Acta Oecologica.* 2003; **24**(1): 25–33.
Publisher Full Text

Settele J, Kudrna O, Harpke A, *et al.*: **Climatic risk Atlas of European butterflies.** *BioRisk.* 2008; **1**: 1–712.
Publisher Full Text

Sucháčková Bartoňová A, Škopek P, Konvička M, *et al.*: **Czech Republic butterfly barcoding reveals that distribution of genetic lineages depends on species traits.** *J Biogeogr.* 2024; **51**(8): 1575–86.
Publisher Full Text

Tolman T, Lewington R: **Collins Butterfly Guide.** HarperCollins Publishers, 2009.

Trense D, Habel JC, Finger A, *et al.*: **Contrasting genetic response to habitat fragmentation for two Lycaenid butterfly species.** *Insect Conserv Divers.* 2022a; **15**(3): 337–47.
Publisher Full Text

Trense D, Jager L, Fischer K: **The central Alps comprise a major dispersal barrier between western and eastern populations of two butterfly species.** *J Biogeogr.* 2022b; **49**(8): 1508–20.
Publisher Full Text

Turlure C, Van Dyck H: **On the consequences of aggressive male mate-locating behaviour and micro-climate for female host plant use in the butterfly *Lycaena hippothoe*.** *Behav Ecol Sociobiol.* 2009; **63**: 1581–91.
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, Cuttelod A, Collins S, *et al.*: **European red list of butterflies.** Luxembourg: Publications Office of the European Union, 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

Wagner W: ***Lycaena hippothoe* (Linnaeus, 1761).** 2005.
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

Open Peer Review

Current Peer Review Status: ? ✓ ?

Version 1

Reviewer Report 22 October 2025

https://doi.org/10.21956/wellcomeopenres.27264.r131672

?

Megan Barkdull
[1] Cornell University, Ithaca, New York, USA
[2] Entomology, Natural History Museum of Los Angeles County (Ringgold ID: 5119), Los Angeles, California, USA

The authors present a genome assembly for the Lycaenid butterfly Lycaena hippothoe.

My most substantial criticism is that the authors have not indicate where the voucher specimen pictured in Fig. 1 has been deposited. It should be deposited in a recognized natural history collection to permit future study.

I indicated that the authors did not convey a rationale for creating the dataset; although they did not, in my view simply providing additional genomic resources to the research community is sufficient rationale.

It would be helpful if the authors could add collecting permit information or indicate that permits were not necessary.

The colors in Fig. 3 are extremely difficult to distinguish, and I urge the authors to adopt a different color scheme that allows the 31 Merian elements to be more clearly identified.

Is the rationale for creating the dataset(s) clearly described?
No

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: Comparative genomics, entomology.

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 23 September 2025

https://doi.org/10.21956/wellcomeopenres.27264.r131668

David Lohman
[1] City College of New York AND National Museum of Natural History, New York, New York, USA
[2] The City College of New York Department of Biology (Ringgold ID: 171428), New York, USA
[3] Graduate Center, PhD Program in Biology, City University of New York (Ringgold ID: 2009), New York, USA
[4] Entomology Section, National Museum of Natural History, National Museum of the Philippines (Ringgold ID: 198024), Manila, Philippines

This concise genome note summarizes the biology of the widespread European copper butterfly *Lycaena hippothoe* before explaining the methods that resulted in a highly complete genome assembled at the chromosome level. Since the primary aim of this note is to describe and publicize the genome, the brevity of the introduction and lack of hypotheses is not a problem.
Nevertheless, I suggest the addition of a few sentences of additional text. Host ant associations of lycaenid larvae (or the absence of such associations) is nearly as ecologically important as data on host plant association. Therefore, I would suggest that the following text be added after the sentences on host plants and before the text on voltanism.

"Unlike some of its close relatives, larvae of *L. hippothoe* have never been recorded in association with ants. In contrast, larvae of Palearctic *Lycaena dispar* facultatively associate with ants from two different subfamilies (Fiedler 2021). Larvae of *Tharsalea xanthoides, T. edita, T. heteronea,* and *T. rubidus* have also occasionally been recorded with ants (Ballmer & Pratt 1988; Fiedler 2021). These latter four species were previously placed in the genus *Lycaena* before the recent subfamily revision of Lycaeninae by Zhang et al. (2023)."

By citing the Zhang et al. paper, you are also acknowledging the existence of this revision published in a somewhat obscure journal, and tacitly confirm that *hippothoe* is still in *Lycaena.*

Aside from that addition of text, I would suggest the following rather minor changes

- I would write "northern Spain", not "Northern Spain" because the area is not a formal place name. Later in the text "northern Iran" is used.
-Add that the host plant family is Polygonaceae
-Change "Larva overwinters" to "Larvae overwinter"
-Change "The species forms one adult..." to "The species is univoltine with one adult..."
-Remove "considered" in the phrase "*L. hippothoe* is considered of least concern"
-After "..sequenced as part of Project Psyche" cite the pre-print and/or website
-Figure 1 - Is it possible to show the wing uppersides, too?
-Perhaps cite an appropriate reference to explain what Hi-C is.
-In this phrase "enrichment" does not need to be capitalized: "...DNA remains bound to the Enrichment beads..."
-Provide a citation explaining what "Chromosome painting" is.
-The GC content seems slightly higher in the 2nd contig, which seems to be the W chromosome. This might merit some discussion or speculation. If this is common in lepidopteran W chromosomes, this should be mentioned and cited.
-Fig. 6 - Dichotomizing the reads into Arthropod/non-Arthropod seems to be lost opportunity to look for *Wolbachia* and other endosymbiotic bacteria.
-In the reference Trense et al 2022a, "lycaenid" should not be capitalized because it is not a proper noun; this is an adjective.

Reference:
3. Zhang, J., Q. Cong, J. Shen, L. Song, and N. V. Grishin. 2023. Butterfly classification and species discovery using genomics. The Taxonomic Report of the International Lepidoptera Survey 11:1-94. https://digitalcommons.unl.edu/taxrpt/88/
https://doi.org/10.5281/zenodo.8400884

References
1. Ballmer G, Pratt G: A survey of the last instar larvae of the Lycaenidae (Lepidoptera) of California. *The Journal of Research on the Lepidoptera*. 1988; **27** (1): 1-81 Publisher Full Text
2. Fiedler K: The ant associates of Lycaenidae butterfly caterpillars – revisited. *Nota Lepidopterologica*. 2021; **44**: 159-174 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: Ecology, evolution, systematics, and biogeography of butterflies

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 22 September 2025

https://doi.org/10.21956/wellcomeopenres.27264.r133132

?

Ozren Polašek
University of Split, Split, Croatia

Dear authors, I recently reviewed another of your submissions, and I would suggest to do a similar protocol in here as well. The molecular work does not seem to have any major limitations or problems, but I am again inclined to double check the taxonomic situation to be sure. The photo that you provided does not show sufficient detail for me to be convinced about the species, so please try to provide two more things. The first is to show the wing underside, if you still have it. The second is to consider using mtDNA/COI to validate the species against BOLD. That would reduce the chance of an error and strengthen the overall message.

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: Taxonomy, genetics

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.