DATA NOTE

The genome sequence of the Crepuscular Burnet, *Zygaena carniolica* (Scopoli, 1763) (Lepidoptera: Zygaenidae)

[version 1; peer review: 4 approved]

Marianne Espeland[1], Kay Lucek[2], Charlotte J. Wright[3], Joana I. Meier[3], Mark L. Blaxter[3],
Wellcome Sanger Institute Tree of Life Management, Samples and Laboratory team,
Wellcome Sanger Institute Scientific Operations: Sequencing Operations,
Wellcome Sanger Institute Tree of Life Core Informatics team,
Tree of Life Core Informatics collective, Project Psyche Community

[1]Leibniz Institute for the Analysis of Biodiversity Change, Bonn, Germany
[2]University of Neuchâtel, Neuchâtel, Switzerland
[3]Tree of Life, Wellcome Sanger Institute, Hinxton, England, UK

v1 **First published:** 28 Jul 2025, **10**:374
https://doi.org/10.12688/wellcomeopenres.24616.1
Latest published: 28 Jul 2025, **10**:374
https://doi.org/10.12688/wellcomeopenres.24616.1

Abstract

We present a genome assembly from a female specimen of *Zygaena carniolica* (Crepuscular Burnet; Arthropoda; Insecta; Lepidoptera; Zygaenidae). The assembly contains two haplotypes with total lengths of 353.39 megabases and 323.68 megabases. Most of haplotype 1 (99.94%) is scaffolded into 31 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 15.56 kilobases.

Keywords

Zygaena carniolica, Crepuscular Burnet, genome sequence, chromosomal, Lepidoptera

This article is included in the Tree of Life gateway.

Open Peer Review

Approval Status ✓ ✓ ✓ ✓

	1	2	3	4
version 1 28 Jul 2025	✓ view	✓ view	✓ view	✓ view

1. **Bin Zhang**, Qingdao Agricultural University, Shenzhen, China
2. **Annabel Whibley**, The University of Auckland, Auckland, New Zealand
Bragato Research Institute, Lincoln, New Zealand
3. **Matthew Merkin**, University of Liverpool, Liverpool, UK
4. **Luana Bataglia**, Universita Cattolica del Sacro Cuore - Campus di Piacenza e Cremona (Ringgold ID: 96985), Piacenza, Italy

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: Espeland M: Writing – Original Draft Preparation; **Lucek K**: Investigation, 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). The work was also funded by a Swiss National Science Foundation (SNSF) Grant ID 202869 to KL.
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: Espeland M, Lucek K, Wright CJ *et al.* **The genome sequence of the Crepuscular Burnet, *Zygaena carniolica* (Scopoli, 1763) (Lepidoptera: Zygaenidae) [version 1; peer review: 4 approved]** Wellcome Open Research 2025, **10**:374 https://doi.org/10.12688/wellcomeopenres.24616.1

First published: 28 Jul 2025, **10**:374 https://doi.org/10.12688/wellcomeopenres.24616.1

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; Apoditrysia; Zygaenoidea; Zygaenidae; Zygaeninae; *Zygaena*; *Zygaena carniolica* (Scopoli, 1763) (NCBI:txid287238)

Background

The Crepuscular Burnet moth (*Zygaena carniolica*) is a widespread species with extremely variable phenotype and ecology. It is distributed from eastern Spain through central Europe to southern Siberia and western Mongolia, at elevations from sea level to almost 4 000 m. As with all species in the genus *Zygaena*, *Z. carniolica* is highly aposematic due to its ability to synthesise and/or sequester cyanogenic glucosides from its host plants. Forewings are black with six red spots of variable size, with more or less cream colouration around the spots, while hindwings are red with a back border. In extreme cases the cream colouration completely takes over and the black colouration is hardly visible.

It occurs in a variety of habitats, mainly in calcareous regions. The species feeds on a variety of host plants mainly of the genera *Anthyllis*, *Dorycnium*, *Hippocrepis*, *Onobrychis*, *Lotus*, *Hedysarum* and *Astragalus*, all Fabaceae (Hofmann & Tremewan, 2020). It is often abundant where it occurs, and can benefit from planted *Lotus* and *Onobrychis* crops (Rennwald *et al.*, 2012). Individuals are rather lethargic, and many can be found roosting on a single flower head. Populations are univoltine or biovoltine depending on locality, and flight time is variable (Hofmann & Tremewan, 2020). *Z. carniolica* is listed as Vulnerable in Switzerland (Wermeille *et al.*, 2014) and threatened in some states in eastern Germany, where it reaches the northernmost boundary of its distribution (Fischer & Sobczyk, 2001; Wachlin *et al.*, 1997). It is not listed in the IUCN Red list Europe. An allozyme-based population genetic study of populations in Germany and adjoining areas in Luxembourg and France found the genetic diversity of *Z. carniolica* rather low and homogeneously distributed across all populations (Habel *et al.*, 2012). Different populations are known to have between 30 and 31 chromosomes (Burgeff & Haupt, 1967; Larsen, 1976; Lukhtanov & Kuznetsova, 1989).

We present a chromosome-level, haplotype-resolved genome sequence of the Crepuscular Burnet, *Zygaena carniolica*, sequenced as part of Project Psyche. The sequence data were derived from a female specimen (Figure 1) collected from Conthey, Valais, Switzerland.

Methods

Sample acquisition

The specimen used for genome sequencing was an adult female *Zygaena carniolica* (specimen ID SAN28000139, ToLID ilZygCarn1; Figure 1), collected from Conthey, Valais, Switzerland (latitude 46.2872, longitude 7.3116; elevation 1 650 m) on 02/08/2023. The specimen was collected and identified by Yannick Chittaro (Info Fauna, Neuchâtel, Switzerland).

Figure 1. Voucher photograph of the *Zygaena carniolica* (ilZygCarn1) specimen used for genome sequencing.

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 ilZygCarn1 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 39.4 ng/μL and a yield of 1 851.80 ng, with a fragment size of 16.2 kb. The 260/280 spectrophotometric ratio was 1.92, and the 260/230 ratio was 3.05.

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 ilZygCarn1 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 PRJEB78806.

Platform	PacBio HiFi	Hi-C
ToLID	ilZygCarn1	ilZygCarn1
Specimen ID	SAN28000139	SAN28000139
BioSample (source individual)	SAMEA115110002	SAMEA115110002
BioSample (tissue)	SAMEA115110086	SAMEA115110020
Tissue	thorax	head
Sequencing platform and model	Revio	Illumina NovaSeq X
Run accessions	ERR13485749	ERR13494006
Read count total	1.55 million	682.62 million
Base count total	16.26 Gb	103.08 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 17 breaks and 64 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 *Zygaena carniolica* was sequenced using Pacific Biosciences single-molecule HiFi long reads, generating 16.26 Gb (gigabases) from 1.55 million reads, which were used to assemble the genome. GenomeScope2.0 analysis estimated the haploid genome size at 336.97 Mb, with a heterozygosity of 1.27% and repeat content of 26.65%. These estimates guided expectations for the assembly. Based on the estimated genome size, the sequencing data provided approximately 47× coverage. Hi-C sequencing produced 103.08 Gb from 682.62 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 353.39 Mb in 38 scaffolds, with 99 gaps, and a scaffold N50 of 12.29 Mb (Table 2).

Most of the assembly sequence (99.94%) was assigned to 31 chromosomal-level scaffolds, representing 29 autosomes and the W and Z sex chromosomes. The Z and W sex chromosomes were identified by Hi-C signal and read coverage analysis. The chromosome-level scaffolds, confirmed by Hi-C data, are named according to size (Figure 2; Table 3). Chromosome painting with Merian elements illustrates the distribution of orthologues along chromosomes and highlights patterns of chromosomal evolution relative to Lepidopteran ancestral linkage groups (Figure 3).

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

Assembly quality metrics

For haplotype 1, the estimated QV is 64.3, and for haplotype 2, 64.5. When the two haplotypes are combined, the assembly achieves an estimated QV of 64.4. The *k*-mer completeness is 77.49% for haplotype 1, 73.39% for haplotype 2, and 99.65% for the combined haplotypes (Figure 4). BUSCO analysis using

Table 2. Genome assembly statistics.

Assembly name	ilZygCarn1.hap1.1	ilZygCarn1.hap2.1
Assembly accession	GCA_964261665.1	GCA_964261595.1
Assembly level	chromosome	scaffold
Span (Mb)	353.39	323.68
Number of chromosomes	31	N/A
Number of contigs	137	121
Contig N50	5.74 Mb	5.55 Mb
Number of scaffolds	38	37
Scaffold N50	12.29 Mb	12.24 Mb
Longest scaffold length (Mb)	15.56	N/A
Sex chromosomes	W and Z	N/A
Organelles	Mitochondrial genome: 15.56 kb	N/A

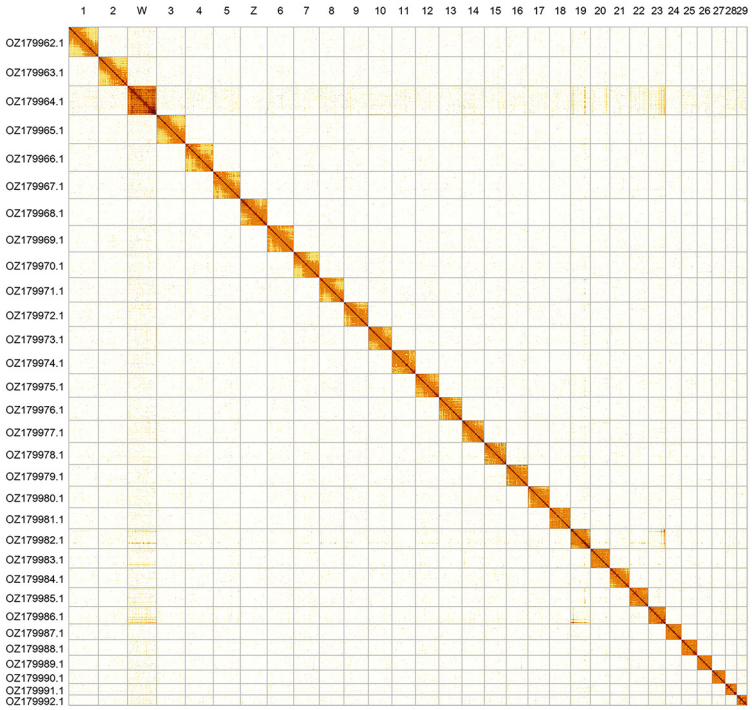

Figure 2. Hi-C contact map of the *Zygaena carniolica* genome assembly. Assembled chromosomes are shown in order of size and labelled along the axes. The plot was generated using PretextSnapshot.

Table 3. Chromosomal pseudomolecules in the haplotype 1 genome assembly of *Zygaena carniolica* ilZygCarn1.

INSDC accession	Molecule	Length (Mb)	GC%	Assigned Merian elements
OZ179962.1	1	15.56	37	M1
OZ179963.1	2	15.15	37	M17;M20
OZ179965.1	3	15	36.50	M2
OZ179966.1	4	14.47	36.50	M3
OZ179967.1	5	14.19	36.50	M8
OZ179969.1	6	13.79	36	M9
OZ179970.1	7	13.38	36.50	M5
OZ179971.1	8	12.72	36.50	M4
OZ179972.1	9	12.69	36	M12
OZ179973.1	10	12.34	37	M7
OZ179974.1	11	12.29	36	M18
OZ179975.1	12	12.16	36.50	M6
OZ179976.1	13	12.02	36.50	M16
OZ179977.1	14	11.64	37	M15
OZ179978.1	15	11.39	37.50	M11
OZ179979.1	16	11.35	36	M21
OZ179980.1	17	11.15	37	M10
OZ179981.1	18	10.97	36.50	M22
OZ179982.1	19	10.38	39.50	M30;M31
OZ179983.1	20	10.13	36.50	M13
OZ179984.1	21	10.13	37.50	M14
OZ179985.1	22	9.82	37	M23
OZ179986.1	23	9.14	38	M19
OZ179987.1	24	8.18	37	M24
OZ179988.1	25	8.12	37.50	M28
OZ179989.1	26	7.72	36.50	M26
OZ179990.1	27	6.99	37	M27
OZ179991.1	28	5.95	38	M25
OZ179992.1	29	5.35	38.50	M29
OZ179964.1	W	15.11	39	N/A
OZ179968.1	Z	13.93	36	MZ
OZ179993.1	MT	0.02	18.50	N/A

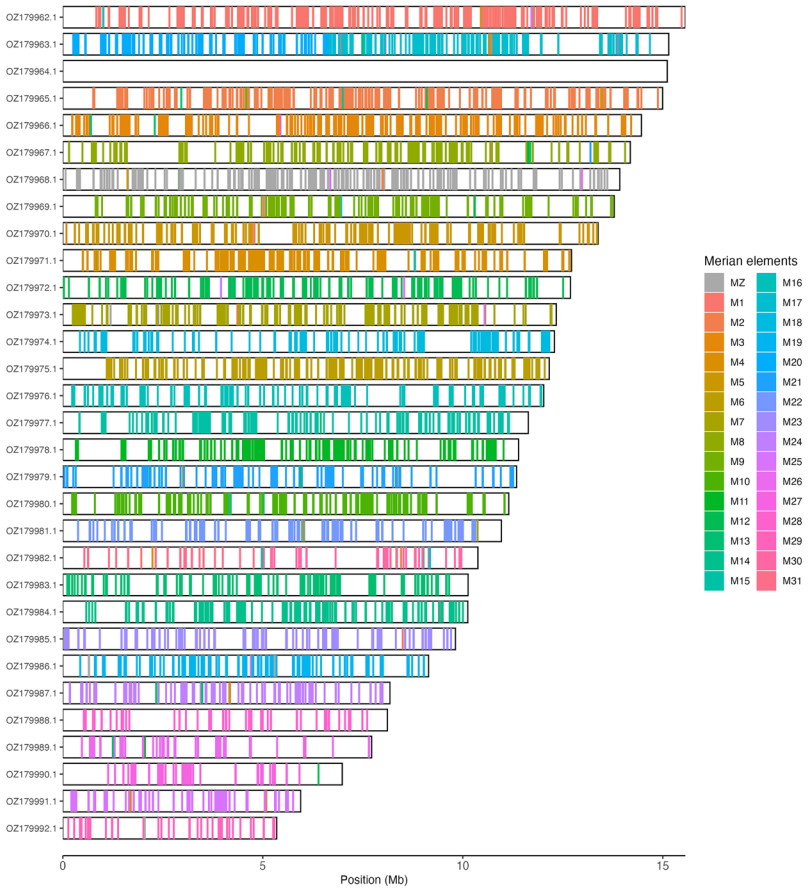

Figure 3. Merian elements painted across chromosomes in the ilZygCarn1.hap1.1 assembly of *Zygaena carniolica*. Chromosomes are drawn to scale, with the positions of orthologues shown as coloured bars. Each orthologue is coloured by the Merian element that it belongs to. All orthologues which could be assigned to Merian elements are shown.

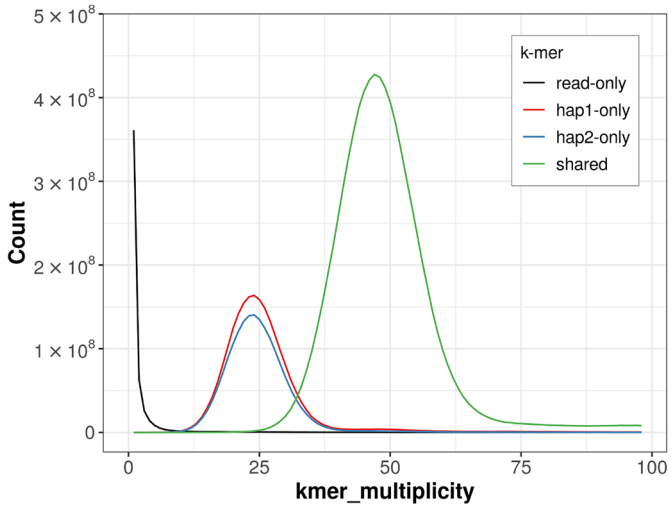

Figure 4. Evaluation of *k*-mer completeness using MerquryFK. This plot illustrates the recovery of *k*-mers from the original read data in the final assemblies. The horizontal axis represents *k*-mer multiplicity, and the vertical axis shows the number of *k*-mers. The black curve represents *k*-mers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

the lepidoptera_odb10 reference set (n = 5 286) (Kriventseva *et al.*, 2019) identified 98.0% of the expected gene set (single = 97.6%, duplicated = 0.4%) 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.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

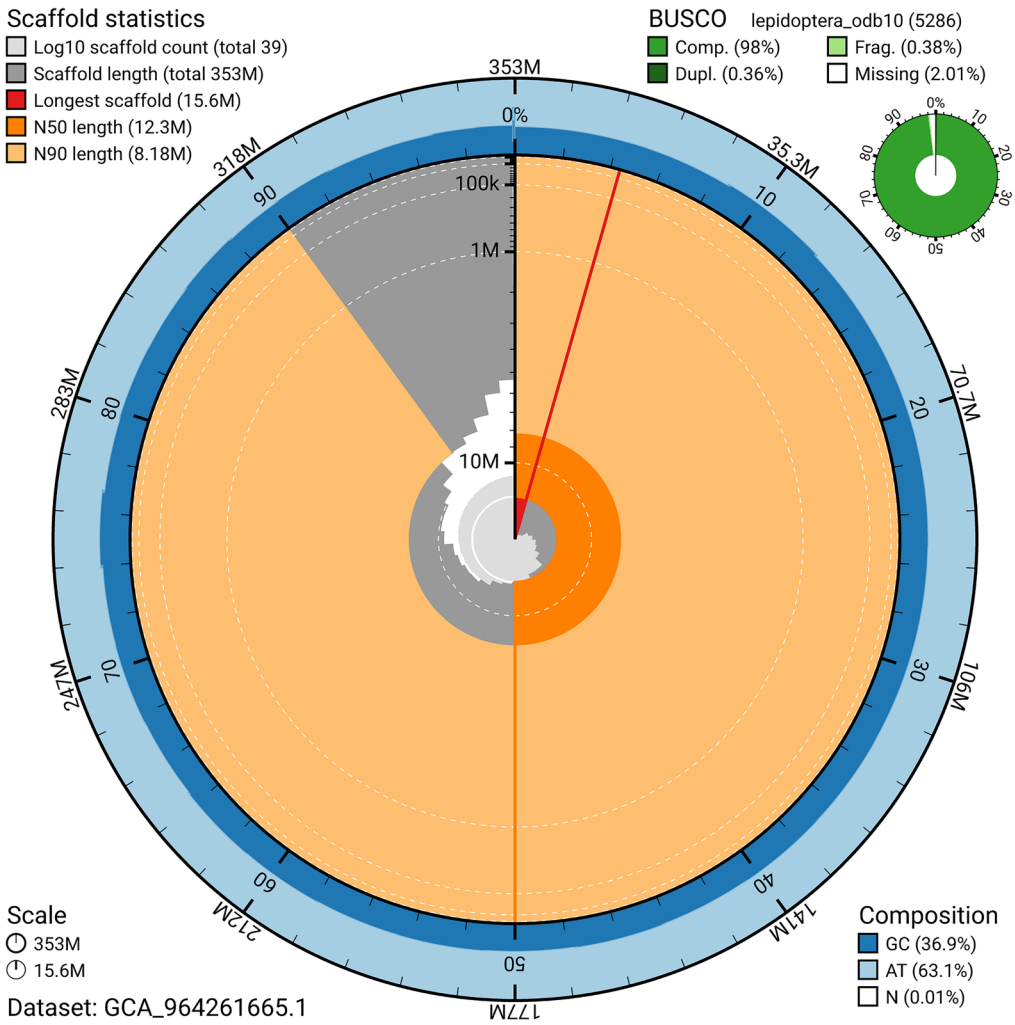

Figure 5. Assembly metrics for ilZygCarn1.hap1.1. The BlobToolKit snail plot provides an overview of assembly metrics and BUSCO gene completeness. The circumference represents the length of the whole genome sequence, and the main plot is divided into 1,000 bins around the circumference. The outermost 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.

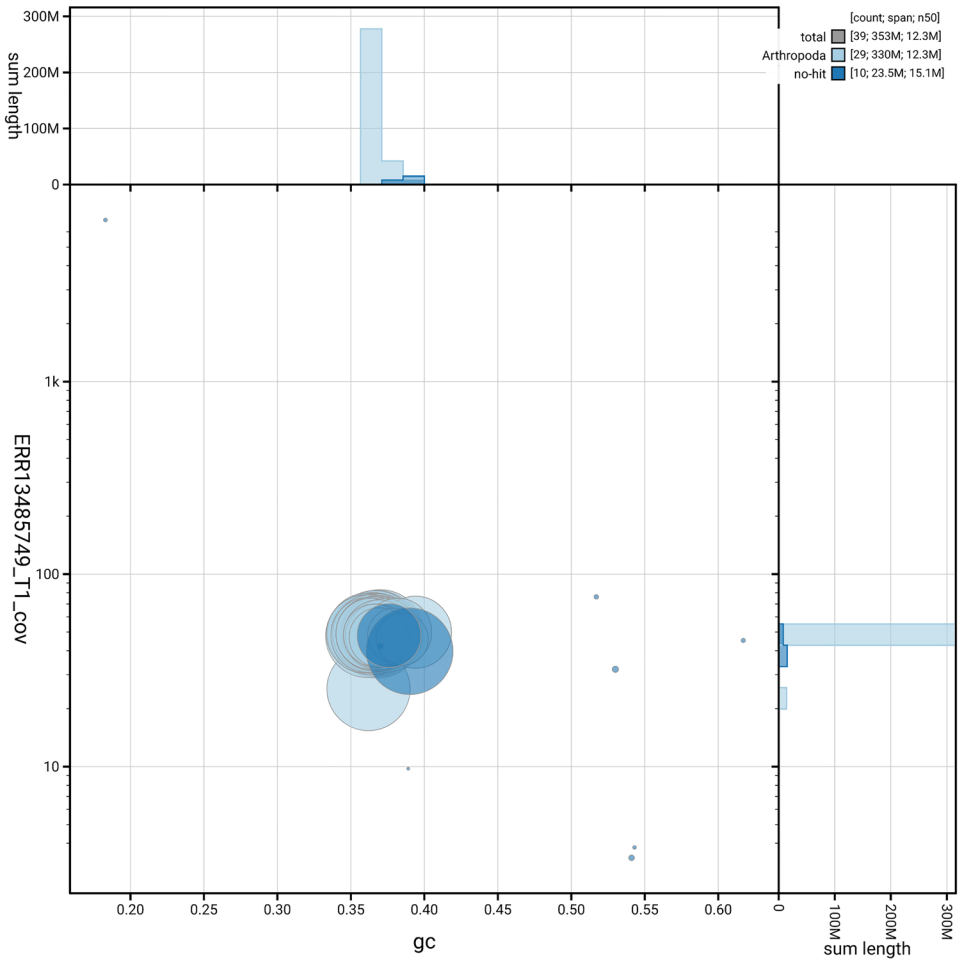

Figure 6. BlobToolKit GC-coverage plot for ilZygCarn1.hap1.1. Blob plot showing sequence coverage (vertical axis) and GC content (horizontal axis). The circles represent scaffolds, with the size proportional to scaffold length and the colour representing phylum membership. The histograms along the axes 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 *Zygaena carniolica* assembly.

Measure (Benchmark)	Value
EBP summary (haplotype 1)	6.C.Q64
Contig N50 length (≥ 1 Mb)	5.74 Mb
Scaffold N50 length (= chromosome N50)	12.29 Mb
Consensus quality (QV) (≥ 40)	Haplotype 1: 64.3; haplotype 2: 64.5; combined: 64.4
k-mer completeness (≥ 95%)	Haplotype 1: 77.49%; Haplotype 2: 73.39%; combined: 99.65%
BUSCO* (S > 90%; D < 5%)	C:98.0%[S:97.6%,D:0.4%],F:0.4%,M:1.6%,n:5286
Percentage of assembly assigned to chromosomes (≥ 90%)	99.94%

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: Zygaena carniolica (crepuscular burnet). Accession number PRJEB78806. The genome sequence is released openly for reuse. The *Zygaena carniolica* 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

- Members of the Wellcome Sanger Institute Tree of Life Management, Samples and Laboratory team
- Members of Wellcome Sanger Institute Scientific Operations – Sequencing Operations
- Members of the Wellcome Sanger Institute Tree of Life Core Informatics team
- Members of the Tree of Life Core Informatics collective
- Members of the 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	1.1.2	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

Software	Version	Source
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

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

Burgeff H, Haupt G: **Chromosomzahlen bei der gattung *zygaena* (lep.).** *Nachrichten der Akademie der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse.* 1967; **9**: 89–94.

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 U, Sobczyk T: **Rote liste der schwärmer und spinnerartigen schmetterlinge**. 2001.

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, Engler JO, Rödder D, *et al.*: **Contrasting genetic and morphologic responses on recent population decline in two burnet moths (*Lepidoptera, zygaenidae*).** *Conserv Genet.* 2012; **13**(5): 1293–1304.
Publisher Full Text

Hofmann A, Tremewan WG: **The Natural History of Burnet Moths. Part III.2.** Munich: Museum Witt, 2020.
Reference Source

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

Larsen TB: **Chromosome numbers in genus *zygaena* f. (*Lepidoptera, zygaenidae*).** *Entomological Gazette.* 1976; **27**: 157–60.

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

Lukhtanov VA, Kuznetsova VC: **Karyotype structure in higher lepidoptera (*Papilionomorpha*).** *Entomol Rev.* 1989; **68**(5): 12–31.
Reference Source

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

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

Rennwald E, Sobczyk T, Hofmann A: **Rote liste und gesamtartenliste der**

spinnerartigen falter (*lepidoptera*: Bombyces, sphinges s.l.) deutschlands. In: M. Binot-Hafke, S. Balzer, N. Becker, et al. (eds), *Rote Liste Gefährdeter Tiere, Pflanzen Und Pilze Deutschlands. Band 3: Wirbellose Tiere (Teil 1).* Bonn: Bundesamt für Naturschutz, 2012; 243–283.
Reference Source

Rhie A, McCarthy SA, Fedrigo O, *et al.*: **Towards complete and error-free genome assemblies of all vertebrate species.** *Nature.* 2021; **592**(7856): 737–746.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Rhie A, Walenz BP, Koren S, *et al.*: **Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies.** *Genome Biol.* 2020; **21**(1): 245.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

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

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

Wachlin V, Kallies A, Hoppe H: **ROTE LISTE der gefährdeten großschmetterlinge mecklenburg-vorpommerns**. 1997.
Reference Source

Wermeille E, Chittaro Y, Gonseth Y: **Rote liste tagfalter und widderchen.** Gefährdete arten der schweiz, stand 2012, 2014.
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 05 January 2026

https://doi.org/10.21956/wellcomeopenres.27126.r139546

✔ **Luana Bataglia**
Universita Cattolica del Sacro Cuore - Campus di Piacenza e Cremona (Ringgold ID: 96985), Piacenza, Emilia-Romagna, Italy

The article reports the sequencing, at the chromosome level, of the genome of the Lepidoptera species *Zygaena carniolica*.
The article is sufficient as a data note, but I suggest a minimum of biological interpretation of the technical data provided. This will proportionate a gain in quality and visibility for the article.
Some points to be considered are:

1) The mitochondrial genome data were poorly reported. The only information is about the length of 15.56 kilobases, reported in the abstract. But this information is missing in the results section.

2) Although it is written that the gene annotation is still to be done, I strongly suggest that the authors complete it as soon as possible and include this information in the article.

3) The authors should include small discussions as:

- Do the length of the nuclear and mitochondrial genomes match the expected size and resemble those of close species?

-The same kind of discussion should be done with the annotation data. Do the number of predicted genes and proteins match the expected number and resemble those of close species?

- Some groups of genes presented a significant increase or decrease in the species?

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

Are the protocols appropriate and is the work technically sound?
Yes

Are sufficient details of methods and materials provided to allow replication by others?
Yes

Are the datasets clearly presented in a useable and accessible format?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Molecular Biology, Genetics, Epigenetics, Entomology, Biotechnology.

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 26 November 2025

https://doi.org/10.21956/wellcomeopenres.27126.r139549

Matthew Merkin
University of Liverpool, Liverpool, UK

The authors present a chromosome-level genome assembly for the Crepuscular Burnet moth (Zygaena carniolica), which was generated for Project Psyche. A female sample was collected in Switzerland and DNA was extracted then sequenced from its thorax and head to obtain HiFi and Hi-C reads respectively. The reads were then assembled, scaffolded and passed through several quality assessment pipelines to obtain the final sequence report. High quality score metrics, such as an EBP score of 6.C.Q64 and BUSCO completeness of 98% suggest that this assembly is suitable for use as a reference genome.

Minor comments:

The introduction to this genome note is excellent at highlighting the interesting features of this species and previous genetic studies. However, there is a typo with "biovoltine" instead of "bivoltine" and the phrase "Flight time is variable" is ambiguous; it is unclear whether this refers to flight season (months of the year in which adults can be found) or time of day- the latter of which seems particularly relevant in this species due to its "crepuscular" title. Additionally, an extra sentence explaining what Project Psyche is and why this species was chosen in particular would be beneficial.

The consistent reference to the primary haplotype containing 31 chromosomes (e.g. abstract and Table 2) is very misleading since it implies that there are 30 autosomes and the sex chromosome. However,there are only 29 autosomes since the Z and W are both included in the primary assembly. As such, a reader skimming the start of the paper may assume that the haploid chromosome number is 31 instead of 30, which is particularly problematic since the introduction

mentions that different populations can have 30 or 31 chromosomes (with this individual belonging to the former). Although this is briefly explained in a sentence below Table 2, I feel that it should be explicitly stated that n=30 in a prominent position e.g. abstract or end of the introduction.

The number of significant figures that the mitochondrial genome is reported to is odd. In the text and Table 2, 15.56kb is used when 15561bp has a similar number of characters. Table 3 reports this as 0.02Mb, which does not seem useful- this is also not a chromosomal pseudomolecule, so the sequence should probably not be in the table.

Although there is a commitment in the data availability section to annotate the genome later on, the lack of discussion about an annotation makes this genome note feel incomplete and its publication premature.

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: Comparative genomics, genome assembly, genetics, entomology, ancient DNA, population 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.

Reviewer Report 18 November 2025

https://doi.org/10.21956/wellcomeopenres.27126.r132211

Annabel Whibley
[1] The University of Auckland, Auckland, New Zealand
[2] Grapevine Improvement, Bragato Research Institute, Lincoln, New Zealand

Espeland and colleagues describe the chromosome-level, haplotype-resolved genome assembly of the Crepuscular Burnet, *Zygaena carniolica,* obtained from a female sample collected in Switzerland. The natural history and population distrubtion/status of this species are well summarized. The standard DToL sequencing, assembly and reporting pipelines and templates have been implemented, with chromosomal painting included to identify Merian elements in this Lepidopteran assembly. The assembly quality is exceedingly good (formally 6.C.Q64), with contiguity, completeness, accuracy and contamination assessments underscoring this. The project metadata is comprehensive and public accessions are active.

I would second the comment of Reviewer 1, that the experimental details of the HiC protocol could be enhanced by including the specific enzyme used for fragmentation.

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.

Reviewer Report 25 September 2025

https://doi.org/10.21956/wellcomeopenres.27126.r132217

Bin Zhang
Qingdao Agricultural University, Shenzhen, China

This manuscript presents a high-quality, haplotype-resolved genome assembly of *Zygaena carniolica* (Crepuscular Burnet), a species of ecological and evolutionary interest in the Lepidoptera: Zygaenidae family. While the manuscript is strong overall, several minor revisions would improve its completeness and scientific impact:
The methods describe using a “restriction enzyme master mix” for Hi-C sample preparation but do

not specify the restriction enzyme(s) (e.g., *DpnII*, *HindIII*) or enzyme recognition site. This information is critical for evaluating Hi-C data quality (e.g., fragment size distribution, valid contact rate) and reproducing the scaffolding workflow.

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.