DATA NOTE

The genome sequence of the Speedwell Longhorn, *Cauchas fibulella* ([Denis & Schiffermüller], 1775) (Lepidoptera: Adelidae)

[version 1; peer review: 6 approved]

David C. Lees[1], Natural History Museum Genome Acquisition Lab,
Darwin Tree of Life Barcoding Collective,
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, Darwin Tree of Life Consortium

[1]Natural History Museum, London, England, UK

v1 **First published:** 21 Oct 2025, **10**:594
https://doi.org/10.12688/wellcomeopenres.25061.1
Latest published: 21 Oct 2025, **10**:594
https://doi.org/10.12688/wellcomeopenres.25061.1

Abstract

We present a genome assembly from an individual male *Cauchas fibulella* (Speedwell Longhorn; Arthropoda; Insecta; Lepidoptera; Adelidae). The assembly contains two haplotypes with total lengths of 578.63 megabases and 573.48 megabases. The whole sequence for haplotype 1 is scaffolded into 25 chromosomal pseudomolecules, including the Z chromosome. Most of haplotype 2 (97.29%) is scaffolded into 25 chromosomal pseudomolecules, also including a Z chromosome. The mitochondrial genome has also been assembled, with a length of 15.77 kilobases. This assembly was generated as part of the Darwin Tree of Life project, which produces reference genomes for eukaryotic species found in Britain and Ireland.

Keywords

Cauchas fibulella; Speedwell Longhorn; genome sequence; chromosomal; Lepidoptera

This article is included in the Tree of Life gateway.

Open Peer Review

Approval Status ✓ ✓ ✓ ✓ ✓ ✓

	1	2	3	4	5	6
version 1 21 Oct 2025	✓ view	✓ view	✓ view	✓ view	✓ view	✓ view

1. **Jhon Alex Dziechciarz Vidal**, Universidade Federal de SãoCarlos (UFSCar), São Carlos, Brazil
2. **Ariel Chipman**, Hebrew University of Jerusalem School of Pharmacy, Jerusalem, Israel
3. **Ian M Carr**, University of Leeds, Leeds, UK
4. **Peng-Fe Ma**, Chinese Academy of Sciences, Kunming, China
5. **Oliver Rupp**, Justus-Liebig-University, Giessengiessen, Germany

6. **Sara E Miller**, University of Missouri at Saint Louis, St Louis, USA

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

Corresponding author: Darwin Tree of Life Consortium (mark.blaxter@sanger.ac.uk)

Author roles: Lees DC: Investigation, Resources, Writing – Original Draft Preparation, Writing – Review & Editing;

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by Wellcome through core funding to the Wellcome Sanger Institute (220540) and the Darwin Tree of Life Discretionary Award [218328, https://doi.org/10.35802/218328].
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: Lees DC, Natural History Museum Genome Acquisition Lab, Darwin Tree of Life Barcoding Collective *et al.* **The genome sequence of the Speedwell Longhorn, *Cauchas fibulella* ([Denis & Schiffermüller], 1775) (Lepidoptera: Adelidae) [version 1; peer review: 6 approved]** Wellcome Open Research 2025, **10**:594 https://doi.org/10.12688/wellcomeopenres.25061.1

First published: 21 Oct 2025, **10**:594 https://doi.org/10.12688/wellcomeopenres.25061.1

Species taxonomy

Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Protostomia; Ecdysozoa; Panarthropoda; Arthropoda; Mandibulata; Pancrustacea; Hexapoda; Insecta; Dicondylia; Pterygota; Neoptera; Endopterygota; Amphiesmenoptera; Lepidoptera; Glossata; Neolepidoptera; Heteroneura; Incurvarioidea; Adelidae; *Cauchas*; *Cauchas fibulella* ([Denis & Schiffermüller], 1775)(NCBI: txid1100941)

Background

Cauchas fibulella ([Denis & Schiffermüller], 1775), also known as the Speedwell Longhorn, is a longhorn moth in the family Adelidae with an 3.5–5 mm forewing length (Sterling *et al.*, 2023) or 7–9 mm wingspan (Bryner, 2020). The dark brown forewing with a brassy reflection has a more or less diamond shaped white central fascia, with two smaller white spots near the wing base, and the antennae are marginally longer than the forewing. Sexual dimorphism is not pronounced. The Lesser Speedwell Longhorn (*Cauchas leucocerella* (Scopoli, 1763)), which overlaps in distribution with *C. fibulella* in most of Europe, is similar in appearance but has an extra white spot at 4/5 along the costa (Bryner, 2020).

The adult moth is univoltine, flying from the end of April to end of July (Bryner, 2020). It rests in sunshine on the flowers of Germander Speedwell *Veronica chamaedrys* L. (the main larval foodplant) and *V. teucrium* L. in Europe (Bryner, 2020). The oviposition in leaf material, the whitish egg and larva, the larval case and the pupal exuvium are illustrated by Bryner (2020: 314, 316–317).

In the United Kingdom (NBN Atlas Partnership, 2025) the Speedwell Longhorn is widespread almost throughout England, although rare or absent in the southwest, also occurring in southwestern and northeastern Scotland, but is apparently absent from Ireland. The species is otherwise widespread in Europe ranging east to central Asia but absent in northern Scandinavia (GBIF Secretariat, 2025) and is absent from Portugal (Lepiforum, 2025).

The DNA barcode from the mitogenome assembly (OZ245579.1) of the genome presented here represents the COI-5P cluster Barcode Index Number (BIN) BOLD:AAE9284. It is identical to the most common haplotype on BOLD (29/07/2025) sequenced from Europe, and 3.38% *p*-distant from an unidentified cluster (BOLD:AFQ1679) from Samarskaya Oblast, Russia; among identified nearest neighbours. The *p*-distances from *C. rufifrontella* (Treitschke, 1833) (BIN, BOLD:AAW4116) and *C. talyakiella* Korb, 2016 (BIN, BOLD:AFE8349) are around 4.6–4.9%. While also associated with *V. chamaedrys* and *V. teucrium*, *C. leucocerella* (BIN, BOLD:AAF2044) is not closely related by DNA barcode, at about 11% pairwise divergence. Nielsen (1980) conducted a cladistic analysis of Palaearctic Adelidae that elevated *Cauchas* Zeller, 1839 to generic status (*Tinea fibulella* is its type species) and he derived a sister relationship to the genera (*Nemophora* Illiger & Hoffmannsegg, 1798 + *Adela* Latreille, 1796), but did not explore species relationships.

The assembly presented here produced using the Tree of Life pipeline from a specimen collected in Bradenham, England, United Kingdom (Figure 1), as part of the Darwin Tree of Life project. The genome will be useful for more detailed studies on *Cauchas* phylogeny and speciation.

Methods

Sample acquisition and DNA barcoding

The specimen used for genome sequencing was an adult male *Cauchas fibulella* (specimen ID NHMUK013697084, ToLID ilCauFibu1; Figure 1), collected from Bradenham, England, United Kingdom (latitude 51.68, longitude –0.81) on 2022-05-22. The specimen was collected and identified by David Lees (Natural History Museum). A second specimen was used for Hi-C sequencing (specimen ID NHMUK013697052, ToLID ilCauFibu2). It was collected from Ranmore Common, England, United Kingdom (latitude 51.24, longitude –0.37) on 2022-05-07. The specimen was collected by Michael Geiser and identified by David Lees (Natural History Museum). The same specimen was used for RNA sequencing.

The initial identification was verified by an additional DNA barcoding process according to the framework developed by Twyford *et al.* (2024). A small sample was dissected from the specimen and stored in ethanol, while the remaining parts were shipped on dry ice to the Wellcome Sanger Institute (WSI) (see the protocol). The tissue was lysed, the COI marker region was amplified by PCR, and amplicons were sequenced and compared to the BOLD database, confirming the species identification (Crowley *et al.*, 2023). Following whole genome sequence generation, the relevant DNA barcode region was also used alongside the initial barcoding data for sample tracking at the WSI (Twyford *et al.*, 2024). The standard operating procedures for Darwin Tree of Life barcoding are available on protocols.io.

Sample metadata were collected in line with the Darwin Tree of Life project standards described by Lawniczak *et al.* (2022).

Figure 1. Photograph of *Cauchas fibulella* on flower of Germander Speedwell (one of the same population as the specimen used for genome sequencing) taken by David Lees at Bradenham on 2022-05-22.

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 ilCauFibu1 sample was weighed and triaged to determine the appropriate extraction protocol. Tissue from the whole organism 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 8.08 ng/μL and a yield of 379.76 ng, with a fragment size of 15.6 kb. The 260/280 spectrophotometric ratio was 2.59.

RNA was extracted from whole organism tissue of ilCauFibu2 in the Tree of Life Laboratory at the WSI using the RNA Extraction: Automated MagMax™ *mir*Vana protocol. The RNA concentration was assessed using a Nanodrop spectrophotometer and a Qubit Fluorometer using the Qubit RNA Broad-Range Assay kit. Analysis of the integrity of the RNA was done using the Agilent RNA 6000 Pico Kit and Eukaryotic Total RNA assay.

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.

Hi-C

Sample preparation and crosslinking

The Hi-C sample was prepared from 20–50 mg of frozen whole organism tissue of the ilCauFibu2 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 to 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.

RNA library preparation and sequencing

Libraries were prepared using the NEBNext® Ultra™ II Directional RNA Library Prep Kit for Illumina (New England Biolabs), following the manufacturer's instructions. Poly(A) mRNA in the total RNA solution was isolated using oligo(dT) beads, converted to cDNA, and uniquely indexed; 14 PCR cycles were performed. Libraries were size-selected to produce fragments between 100–300 bp. Libraries were quantified, normalised, pooled to a final concentration of 2.8 nM, and diluted to 150 pM for loading. Sequencing was carried out on the Illumina NovaSeq X to generate 150-bp paired-end reads.

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 28 breaks and 79 joins. This reduced the scaffold count by 10.3%. The curation process is described 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

The Merqury.FK tool (Rhie *et al.*, 2020) was run in a Singularity container (Kurtzer *et al.*, 2017) 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 version (Challis *et al.*, 2020). The pipeline aligns PacBio reads using minimap2 (Li, 2018) and SAMtools (Danecek *et al.*, 2021) to generate coverage tracks. It runs BUSCO (Manni *et al.*, 2021) using lineages identified from the NCBI Taxonomy (Schoch *et al.*, 2020). For the three domain-level 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 containerisation through Docker (Merkel, 2014) and Singularity (Kurtzer *et al.*, 2017).

Genome sequence report

Sequence data

PacBio sequencing of the *Cauchas fibulella* specimen generated 16.60 Gb (gigabases) from 1.65 million reads, which were used to assemble the genome. GenomeScope2.0 analysis estimated the haploid genome size at 570.00 Mb, with a heterozygosity of 0.89% and repeat content of 33.73% (Figure 2). These estimates guided expectations for the assembly. Based on the estimated genome size, the sequencing data provided approximately 28× coverage. Hi-C sequencing produced 94.27 Gb from 624.28 million reads, which were used to scaffold the assembly. RNA sequencing data were also generated and are available in public sequence repositories. 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 578.63 Mb in 25 scaffolds, with 85 gaps, and a scaffold N50 of 24.76 Mb (Table 2).

Most of the assembly sequence (100.0%) was assigned to 25 chromosomal-level scaffolds, representing 24 autosomes and the Z sex chromosome. These chromosome-level scaffolds, confirmed by Hi-C data, are named according to size (Figure 3;

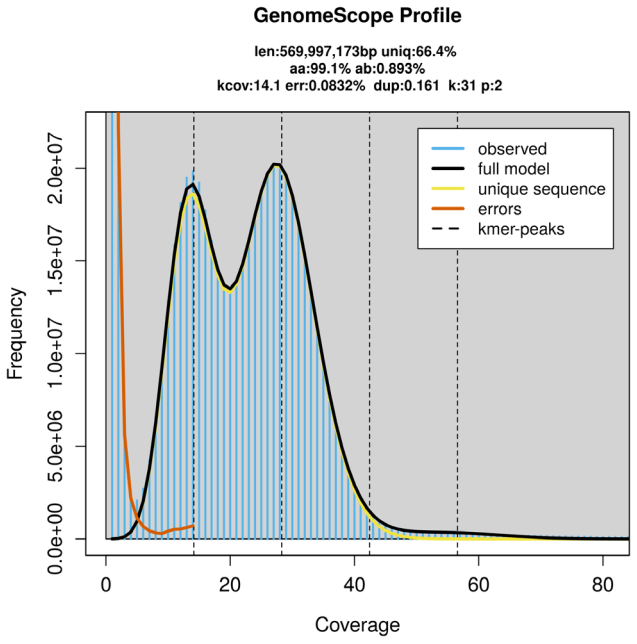

Figure 2. Frequency distribution of *k*-mers generated using GenomeScope2. The plot shows observed and modelled *k*-mer spectra, providing estimates of genome size, heterozygosity, and repeat content based on unassembled sequencing reads.

Table 1. Specimen and sequencing data for BioProject PRJEB84255.

Platform	PacBio HiFi	Hi-C	RNA-seq
ToLID	ilCauFibu1	ilCauFibu2	ilCauFibu2
Specimen ID	NHMUK013697084	NHMUK013697052	NHMUK013697052
BioSample (source individual)	SAMEA114805654	SAMEA114806238	SAMEA114806238
BioSample (tissue)	SAMEA114805773	SAMEA114806374	SAMEA114806374
Tissue	whole organism	whole organism	whole organism
Instrument	Revio	Illumina NovaSeq X	Illumina NovaSeq X
Run accessions	ERR14151095; ERR14151096	ERR14172292	ERR14792859
Read count total	1.65 million	624.28 million	87.95 million
Base count total	16.60 Gb	94.27 Gb	13.28 Gb

Table 2. Genome assembly statistics.

Assembly name	ilCauFibu1.hap1.1	ilCauFibu1.hap2.1
Assembly accession	GCA_965226335.1	GCA_965226765.1
Assembly level	chromosome	chromosome
Span (Mb)	578.63	573.48
Number of chromosomes	25	25
Number of contigs	110	1 468
Contig N50	9.91 Mb	1.05 Mb
Number of scaffolds	25	533
Scaffold N50	24.76 Mb	23.1 Mb
Longest scaffold length (Mb)	40.98	40.63
Sex chromosomes	Z	Z
Organelles	Mitochondrion: 15.77 kb	-

Table **¿tbl:table3?**). This genome has been assembled using PacBio and HiC data and phased. The result is two curated haplotypes. Z chromosome identified based BUSCO gene painting with ancestral Merian elements (Wright *et al.*, 2024). 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.

For haplotype 1, the estimated QV is 66.5, and for haplotype 2, 66.0. When the two haplotypes are combined, the assembly achieves an estimated QV of 66.2. The *k*-mer completeness is 80.72% for haplotype 1, 78.67% for haplotype 2, and 99.49% for the combined haplotypes (Figure 4).

BUSCO analysis using the lepidoptera_odb10 reference set (n = 5 286) identified 92.1% of the expected gene set (single = 91.3%, duplicated = 0.8%) 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 3 lists the assembly metric benchmarks adapted from Rhie *et al.* (2021) and the Earth BioGenome Project Report on Assembly Standards September 2024. The EBP metric, calculated for the haplotype 1, is **6.C.Q66**, 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 Darwin Tree of Life Partner. The submission of materials by a Darwin Tree of Life Partner is

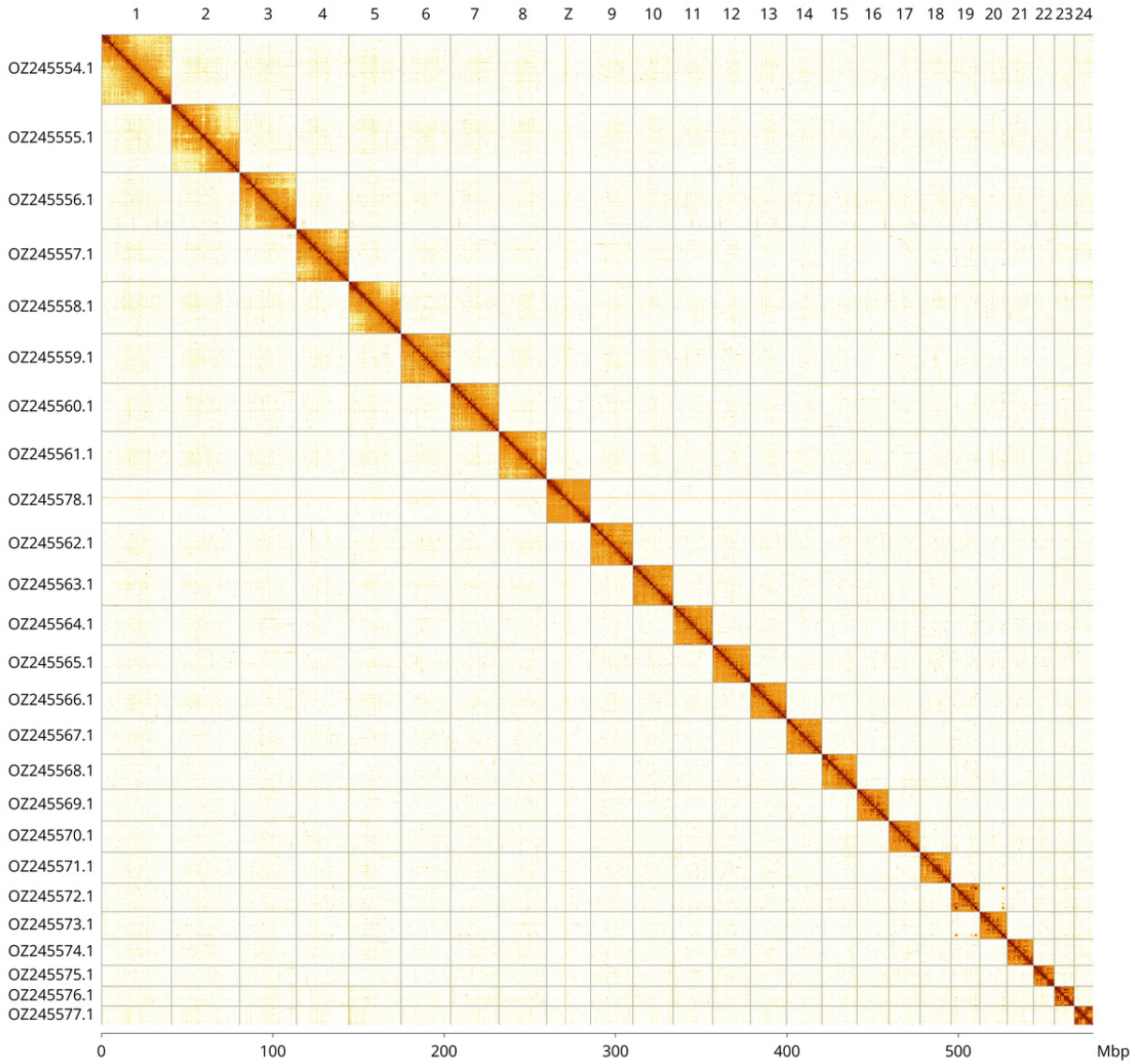

Figure 3. Hi-C contact map of the *Cauchas fibulella* genome assembly. Assembled chromosomes are shown in order of size and labelled along the axes, with a megabase scale shown below. The plot was generated using PretextSnapshot.

Chromosomal pseudomolecules in both haplotypes of the genome assembly of *Cauchas fibulella*, ilCauFibu1{#tbl:table3}.

Haplotype 1				Haplotype 2			
INSDC accession	Name	Length (Mb)	GC%	INSDC accession	Name	Length (Mb)	GC%
OZ245554.1	1	40.98	39.50	OZ245660.1	1	40.63	39.50
OZ245555.1	2	39.75	39	OZ245661.1	2	38.67	39
OZ245556.1	3	33.12	39	OZ245662.1	3	32.78	39
OZ245557.1	4	30.60	39	OZ245663.1	4	30.45	39
OZ245558.1	5	30.39	39.50	OZ245664.1	5	30.15	39.50

Haplotype 1				Haplotype 2			
INSDC accession	Name	Length (Mb)	GC%	INSDC accession	Name	Length (Mb)	GC%
OZ245559.1	6	29.02	39.50	OZ245665.1	6	28.93	39.50
OZ245560.1	7	28.13	39	OZ245666.1	7	27.97	39
OZ245561.1	8	27.84	39.50	OZ245667.1	8	27.16	39.50
OZ245562.1	9	24.76	39	OZ245668.1	9	24.66	39
OZ245563.1	10	23.38	39	OZ245669.1	10	23.10	39
OZ245564.1	11	23.10	38.50	OZ245670.1	11	22.73	38.50
OZ245565.1	12	21.99	38.50	OZ245671.1	12	13.46	38.50
OZ245566.1	13	21.18	39	OZ245672.1	13	20.63	39
OZ245567.1	14	20.57	38.50	OZ245673.1	14	20.49	38.50
OZ245568.1	15	20.56	38.50	OZ245674.1	15	20.19	38.50
OZ245569.1	16	18.44	38.50	OZ245675.1	16	18.44	38.50
OZ245570.1	17	18.21	38.50	OZ245676.1	17	17.88	38.50
OZ245571.1	18	18.17	39	OZ245677.1	18	17.68	39
OZ245572.1	19	16.76	39.50	OZ245678.1	19	16.01	39.50
OZ245573.1	20	15.93	39	OZ245679.1	20	15.85	39
OZ245574.1	21	15.28	39.50	OZ245680.1	21	15.22	39.50
OZ245575.1	22	12.28	39	OZ245681.1	22	12.04	39
OZ245576.1	23	11.65	40	OZ245682.1	23	10.39	40
OZ245577.1	24	10.86	41	OZ245683.1	24	10.48	41
OZ245578.1	Z	25.66	38	OZ245684.1	Z	21.94	38

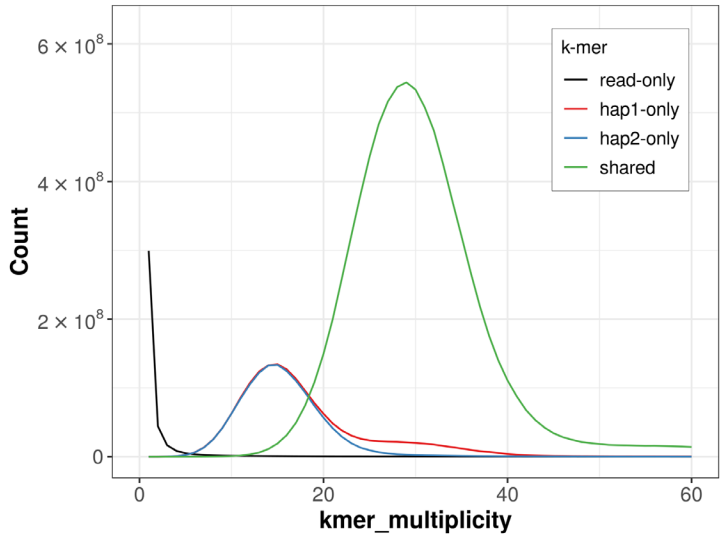

Figure 4. Evaluation of *k*-mer completeness using MerquryFK. This plot illustrates the recovery of *k*-mers from the original read data in the final assemblies. The horizontal axis represents *k*-mer multiplicity, and the vertical axis shows the number of *k*-mers. The black curve represents *k*-mers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

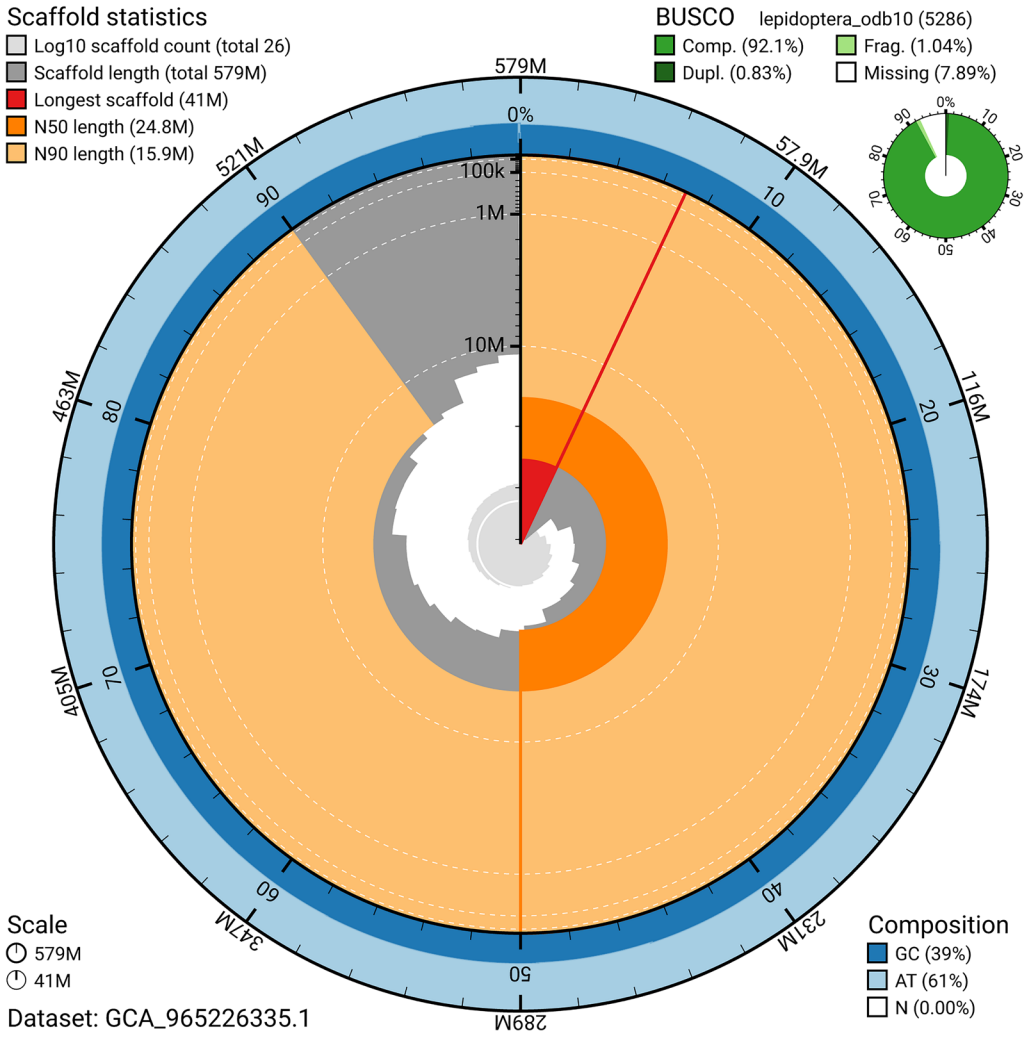

Figure 5. Assembly metrics for ilCauFibu1.hap1.1. The BlobToolKit snail plot provides an overview of assembly metrics and BUSCO gene completeness. The circumference represents the length of the whole genome sequence, and the main plot is divided into 1 000 bins around the circumference. The outermost 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.

subject to the **'Darwin Tree of Life Project Sampling Code of Practice'**, which can be found in full on the Darwin Tree of Life website. By agreeing with and signing up to the Sampling Code of Practice, the Darwin Tree of Life Partner agrees they will meet the legal and ethical requirements and standards set out within this document in respect of all samples acquired for, and supplied to, the Darwin Tree of Life Project. Further, 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 further undertaken according to a Research Collaboration Agreement or Material Transfer Agreement entered into by the Darwin Tree of Life Partner,

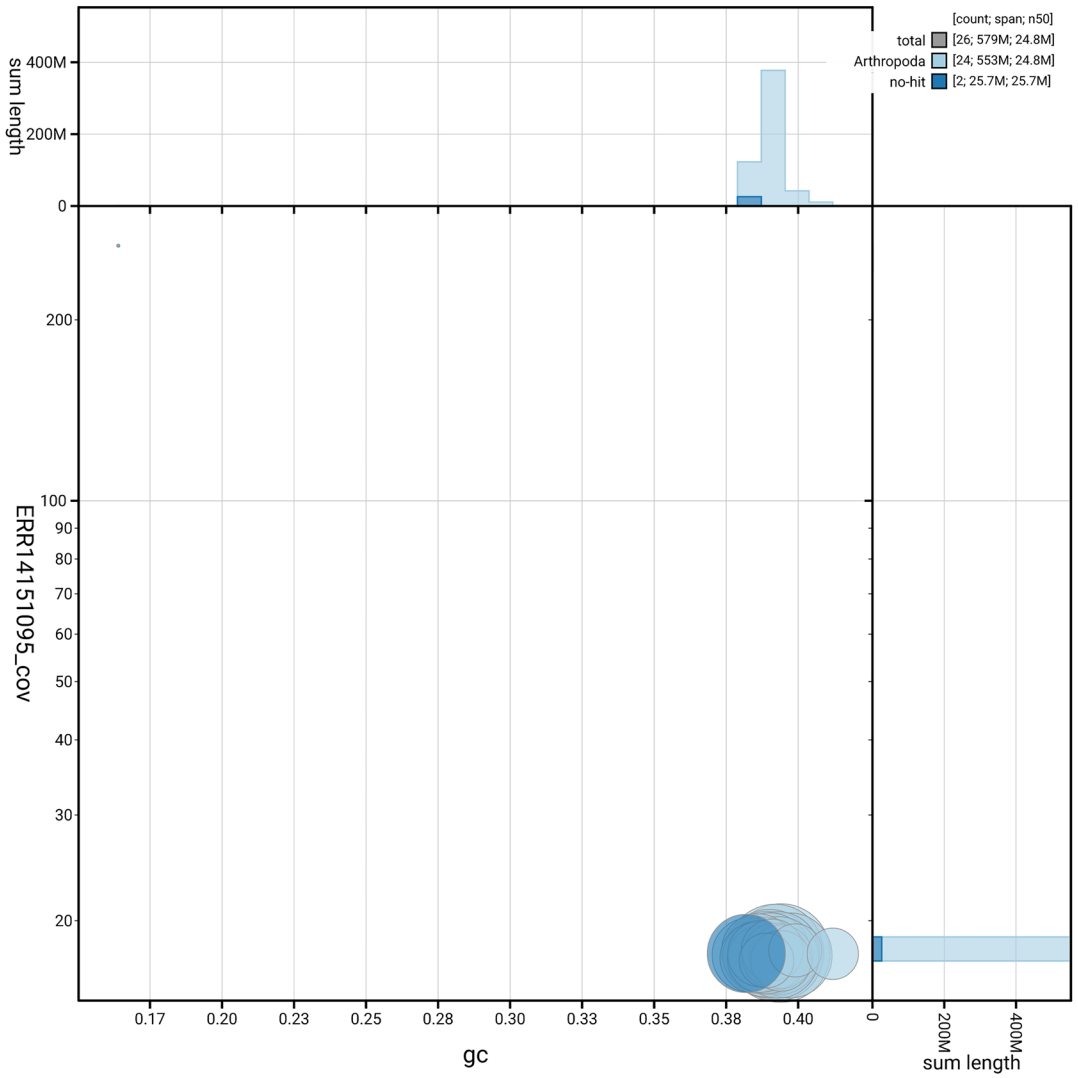

Figure 6. BlobToolKit GC-coverage plot for ilCauFibu1.hap1.1. Blob plot showing sequence coverage (vertical axis) and GC content (horizontal axis). The circles represent scaffolds, with the size proportional to scaffold length and the colour representing phylum membership. The histograms along the axes 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 3. Earth Biogenome Project summary metrics for the *Cauchas fibulella* assembly.

Measure	Value	Benchmark
EBP summary (haplotype 1)	6.C.Q66	6.C.Q40
Contig N50 length	9.91 Mb	≥ 1 Mb
Scaffold N50 length	24.76 Mb	= chromosome N50
Consensus quality (QV)	Haplotype 1: 66.5; haplotype 2: 66.0; combined: 66.2	≥ 40
k-mer completeness	Haplotype 1: 80.72%; Haplotype 2: 78.67%; combined: 99.49%	≥ 95%
BUSCO	C:92.1% [S:91.3%; D:0.8%]; F:1.0%; M:6.8%; n:5 286	S > 90%; D < 5%
Percentage of assembly assigned to chromosomes	100%	≥ 90%

Genome Research Limited (operating as the Wellcome Sanger Institute), and in some circumstances, other Darwin Tree of Life collaborators.

Data availability

European Nucleotide Archive: Cauchas fibulella. Accession number PRJEB84255. The genome sequence is released openly for reuse. The *Cauchas fibulella* genome sequencing initiative is part of the Darwin Tree of Life Project (PRJEB40665), 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 the Ensembl pipeline at the European Bioinformatics Institute. Raw data and assembly accession identifiers are reported in Table 1 and Table 2.

Production code used in genome assembly at the WSI Tree of Life is available at https://github.com/sanger-tol. Table 4 lists software versions used in this study.

Table 4. 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	4.4.5	https://github.com/blobtoolkit/blobtoolkit
BUSCO	5.7.1	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	2.28-r1209	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	24.10.4	https://github.com/nextflow-io/nextflow
PretextSnapshot	-	https://github.com/sanger-tol/PretextSnapshot
PretextView	0.2.5	https://github.com/sanger-tol/PretextView
samtools	1.21	https://github.com/samtools/samtools
sanger-tol/ascc	0.1.0	https://github.com/sanger-tol/ascc
sanger-tol/blobtoolkit	v0.7.1	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	1.2.2	https://github.com/c-zhou/yahs

Author information

Contributors are listed at the following links:

- Members of the Natural History Museum Genome Acquisition Lab
- Members of the Darwin Tree of Life Barcoding collective
- 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 Darwin Tree of Life Consortium

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

Bryner R: **Adelidae (Lepidoptera) – a contribution to the knowledge of biology and identification of European species**. Contributions to Natural History, 2020; **38**: 1–475.
Reference Source

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, 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

Crowley L, Allen H, Barnes I, *et al.*: **A sampling strategy for genome sequencing the British terrestrial arthropod fauna [version 1; peer review: 2 approved].** *Wellcome Open Res.* 2023; **8**: 123.
PubMed Abstract | Publisher Full Text | Free Full Text

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

Ewels P, Magnusson M, Lundin S, *et al.*: **MultiQC: summarize analysis results for multiple tools and samples in a single report.** *Bioinformatics.* 2016; **32**(19): 3047–3048.
PubMed Abstract | Publisher Full Text | Free Full Text

Ewels PA, Peltzer A, Fillinger S, *et al.*: **The nf-core framework for community-curated bioinformatics pipelines.** *Nat Biotechnol.* 2020; **38**(3): 276–278.
PubMed Abstract | Publisher Full Text

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

GBIF Secretariat: ***Cauchas fibulella* ([Denis & Schiffermüller], 1775).** In: GBIF Backbone Taxonomy, 2025.
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

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

Lawniczak MKN, Davey RP, Rajan J, *et al.*: **Specimen and sample metadata standards for biodiversity genomics: a proposal from the Darwin Tree of Life project [version 1; peer review: 2 approved with reservations].** *Wellcome Open Res.* 2022; **7**: 187.
Publisher Full Text

Lepiforum: ***Cauchas fibulella* ([Denis & Schiffermüller], 1775)**. 2025.
Reference Source

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

NBN Atlas Partnership: ***Cauchas fibulella* ([Denis & Schiffermüller], 1775) Speedwell Longhorn**. 2025.
Reference Source

Nielsen ES: **A cladistic analysis of the Holarctic genera of adelid moths (Lepidoptera: Incurvarioidea).** *Entomologica Scandinavica.* 1980; **11**: 161–78.

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

Schoch CL, Ciufo S, Domrachev M, *et al.*: **NCBI Taxonomy: a comprehensive update on curation, resources and tools.** *Database (Oxford).* 2020; **2020**: baaa062.
PubMed Abstract | Publisher Full Text | Free Full Text

Sterling P, Parsons M, Lewington R: **Field guide to the Micro-Moths of Great Britain and Ireland.** London: Bloomsbury Publishing, 2023.
Reference Source

Twyford AD, Beasley J, Barnes I, *et al.*: **A DNA barcoding framework for taxonomic verification in the Darwin Tree of Life project [version 1; peer review: 2 approved].** *Wellcome Open Res.* 2024; **9**: 339.
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

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 30 January 2026

https://doi.org/10.21956/wellcomeopenres.27627.r143608

✔ **Sara E Miller**
University of Missouri at Saint Louis, St Louis, Missouri, USA

The authors present a new high quality nuclear, Z-chromosome, and mitochondrial genome assembly for the Speedwell Longhorn moth, *Cauchas fibulella* (Lepidoptera: Adelidae). The methods are consistent with other genome assemblies from the Darwin Tree of Life project and are well-described. The genome was assembled into two haplotypes. Both haplotype assemblies exceed Earth BioGenome Project Assembly Standards benchmarks. Data and code used to assemble this genome has been archived.

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: Hymenoptera genetics and genomics, evolutionary ecology

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 23 January 2026

https://doi.org/10.21956/wellcomeopenres.27627.r143606

Oliver Rupp
Justus-Liebig-University, Giessengiessen, Germany

This article presents the chromosome level genome assembly of the Speedwell Longhorn moth. The DNA sequencing and assembly followed well established protocols for eukaryotic genomes, including PacBio HiFi and Hi-C sequencing followed by a Hifiasm assembly and Hi-C scaffolding. The assembly includes the mitochondrial genome which was acquired using appropriate tools. The quality of the genome was assessed with different metrics, including k-mer statistics and BUSCO completeness analysis. The assembly is available in standard databases.
Overall this article presents a high quality, chromosome level assembly.

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: genome assembly, genome annotation

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 19 January 2026

https://doi.org/10.21956/wellcomeopenres.27627.r143609

Peng-Fe Ma
Chinese Academy of Sciences, Kunming, China

This study generated a genome assembly for the longhorn moth (*Cauchas fibulella*), which is widespread in the United Kingdom as well as in Europe. The assembled genome is of high quality and contains two haplotypes. And the Z chromosome was also successfully assembled. The related genomic resource will be helpful for understanding the evolution and species identification in the *Cauchas* genus.

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 and evolution.

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 13 January 2026

https://doi.org/10.21956/wellcomeopenres.27627.r143610

Ian M Carr
University of Leeds, Leeds, UK

General comments:
I'm sure the work described here was completed to a high quality, but its write-up lets it down. For instance, the document's internal table referencing system is broken, and acronyms are used before they're described or described twice. A simple proofreading step would have improved the document significantly. I think the figures are taken from analysis software reports and just pasted in the document without any critical thinking about whether they would benefit from slight tweaks.
Specific comments (from the PDF version):
The penultimate paragraph of the *Background* section discusses the genome assembly. This is a result and should be moved to a later section of the paper.
The sentence beginning *"The assembly presented here produced using the Tree of Life pipeline"* is

missing a verb. It should read: *"The assembly presented here was produced using the Tree of Life pipeline."*
The acronym **ToLID** is not explained on first use.
Please define the acronym **COI** in the Methods section.
The acronym **WSI** is introduced twice: once in the Sample Collection section and again in the Nucleic Acid Extraction section. Please remove the second introduction.
Please provide the manufacturers for the **Qubit fluorometer**, **Nanodrop**, and **FemtoPulse** systems when they are first mentioned.
RNA was quantified using the *Agilent RNA 6000 Pico Kit*. Please clarify whether this was run on a **TapeStation** or a **Bioanalyzer**.
The manuscript states: *"Size selection and clean-up were performed using diluted AMPure PB beads (Pacific Biosciences).":* AMPure beads are manufactured by **Beckman Coulter Life Sciences**, not Pacific Biosciences. If the beads were diluted, please specify the dilution factor, as bead concentration determines the size-selection cutoff.
Figure 2 contains text beneath the GenomeScope profile tile that is not clearly explained. Although some values can be inferred from the main text, this information would be better removed from the figure and presented in a clearer format within the figure legend.
In the *Assembly Statistics* section, the sentence *"These chromosome-level scaffolds, confirmed by Hi-C data, are named according to size (Figure 3; Table ¿tbl:table3?)."* contains a broken table reference. Please correct the link.
Following from the last comment: There is an unlabelled table ({#tbl:table3}) between Figure 3 and Figure 4. It appears after Table 2 but before Table 3, which breaks the numbering sequence. The tables should be renumbered and labelled correctly.
The sentence *"Z chromosome identified based BUSCO gene painting with ancestral Merian elements"* should read:
"The Z chromosome was identified based on BUSCO gene painting with ancestral Merian elements."
Figure 6 contains a large amount of empty space due to a contig with a significantly different read depth (>200) and GC content (~0.16). This contig never seems to be mentioned elsewhere.

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: Bioinformatics and NGS data production and analysis

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 30 December 2025

https://doi.org/10.21956/wellcomeopenres.27627.r140700

Ariel Chipman
Hebrew University of Jerusalem School of Pharmacy, Jerusalem, Jerusalem District, Israel

The authors present the genome sequence of the lepidopteran Cauchas fibulella, or speedwell longhorn. The genome was sequenced from two individuals collected in separate location sin England. One individual was used primary sequencing and the second for Hi-C and RNA sequencing. The protocols follow standard Darwin ToL procedures and were carried out as expected. The results are presented in a clear and standardized fashion, and don't include any surprises. The manuscript is standard and straightforward and I see no problems with its Indexing.

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: Arthropod evolution

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 10 November 2025

https://doi.org/10.21956/wellcomeopenres.27627.r137069

Jhon Alex Dziechciarz Vidal
Universidade Federal de SãoCarlos (UFSCar), São Carlos, Brazil

This manuscript presents a high-quality genome assembly for *Cauchas fibulella*, generated using the standard Darwin Tree of Life Consortium assembly pipeline. The work is well structured and contributes valuable genomic information to the growing dataset of Lepidoptera, a group for which many assemblies are already available, yet still essential for evolutionary comparative studies.
I commend the authors for the quality of the genomic data and the clarity of the report. However, I have a few points that require clarification.

Abstract
The abstract states that haplotype 1 was fully scaffolded into 25 chromosomal pseudomolecules, while 97.29% of haplotype 2 was scaffolded.
Please indicate the equivalent percentage for haplotype 1, for consistency and clarity.

Methods
The genome assembly was generated from one individual, whereas the Hi-C data originated from another. Using different individuals for primary sequencing and chromosomal scaffolding is not common and may introduce structural or heterozygosity-related inconsistencies.
Could the authors comment on whether this approach may have affected the assembly, and if any steps were taken to mitigate potential bias?

Figure 3
Although the Z chromosome is described in the text, it does not appear to be labelled or visible as a separate scaffold in Figure 3 (Hi-C map).
Could the authors clarify why the Z chromosome is not displayed, or confirm whether it is included but not annotated in the figure?

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: I have expertise in cytogenomics, repetitive DNA biology, and chromosome-level genome assembly, which qualifies me to evaluate this work.

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.