



DATA NOTE

The genome sequence of the Purple Treble-bar, *Aplocera praeformata* (Hübner, 1826) (Lepidoptera: Geometridae)

[version 1; peer review: 2 approved]

Camille Cornet ¹, Kay Lucek ¹, Charlotte J. Wright ², Joana I. Meier², Mark L. Blaxter²,
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

¹University of Neuchâtel, Neuchâtel, Switzerland²Tree of Life, Wellcome Sanger Institute, Hinxton, England, UK

V1 First published: 16 Jan 2026, 11:31
<https://doi.org/10.12688/wellcomeopenres.25721.1>
 Latest published: 16 Jan 2026, 11:31
<https://doi.org/10.12688/wellcomeopenres.25721.1>

Abstract

We present a genome assembly from a male specimen of *Aplocera praeformata* (Purple Treble-bar; Arthropoda; Insecta; Lepidoptera; Geometridae). The assembly contains two haplotypes with total lengths of 380.66 megabases and 380.72 megabases. Most of haplotype 1 (99.91%) is scaffolded into 31 chromosomal pseudomolecules, including the Z sex chromosome. Haplotype 2 was assembled to scaffold level. The mitochondrial genome has also been assembled, with a length of 17.22 kilobases.

Keywords

Aplocera praeformata; Purple Treble-bar; genome sequence; chromosomal; Lepidoptera

Open Peer Review

Approval Status

	1	2
version 1		
16 Jan 2026	view	view

1. **Aqeel Alyousuf** , University of Basrah, Basrah, Iraq
2. **Gulab D Khedkar** , Dr Babasaheb Ambedkar Marathwada University, Aurangabad, India

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



This article is included in the [Tree of Life](#) gateway.

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

Author roles: **Cornet C:** Investigation, Resources; **Lucek K:** Supervision, Writing – Review & Editing; **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 and CC.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2026 Cornet C *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Cornet C, Lucek K, Wright CJ *et al.* **The genome sequence of the Purple Treble-bar, *Aplocera praeformata* (Hübner, 1826) (Lepidoptera: Geometridae) [version 1; peer review: 2 approved]** Wellcome Open Research 2026, 11:31 <https://doi.org/10.12688/wellcomeopenres.25721.1>

First published: 16 Jan 2026, 11:31 <https://doi.org/10.12688/wellcomeopenres.25721.1>

Species taxonomy

Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Protostomia; Ecdysozoa; Panarthropoda; Arthropoda; Mandibulata; Pancrustacea; Hexapoda; Insecta; Dicondylia; Pterygota; Neoptera; Endopterygota; Amphiesmenoptera; Lepidoptera; Glossata; Neolepidoptera; Heteroneura; Ditrysia; Obtectomera; Geometroidea; Geometridae; Larentiinae; *Aplocera*; *Aplocera praeformata* (Hübner, 1826) (NCBI:txid934918)

Background

The Purple Treble-bar, *Aplocera praeformata* (Hübner, 1826), is a geometrid moth that flies in a single generation from June to early September. Larvae feed on St John's-wort (*Hypericum*) and are often recorded on *H. perforatum*. The species overwinters as a caterpillar. It is slightly larger and darker than the other large Treble-bar species and is usually straightforward to recognise (Wagner, 2025).

In Britain *A. praeformata* is a suspected rare immigrant, with two accepted records – Chinnor on the Oxfordshire–Buckinghamshire border (1919) and Lawrenny Park, Pembrokeshire (1946) (Kimber, 2025; Waring & Townsend, 2017). Across Europe *A. praeformata* is linked to montane and subalpine habitats. Typical sites include grasslands, pastures, forest clearings and edges. It is widespread in mountain regions and is particularly frequent in the Alps, with gaps in parts of western Europe and towards the far south and north. Occurrence data show concentrations in central European ranges and records extending into northern and eastern Europe (GBIF Secretariat, 2023; Wagner, 2025).

We report a chromosome-level genome sequence for *A. praeformata*, sequenced as part of Project Psyche. The sequence data were derived from a male specimen (Figure 1) collected from Bern, Switzerland.



Figure 1. Voucher photograph of the *Aplocera praeformata* (ilAplPrae1) specimen used for genome sequencing.

Methods

Sample acquisition

The specimen used for genome sequencing was an adult male *Aplocera praeformata* (specimen ID SAN28000096, ToLID ilAplPrae1; Figure 1), collected from Steingletscher, Innertkirchen, Bern, Switzerland (latitude 46.731, longitude 8.426; elevation 214 m) on 2023-07-14. The specimen was collected and identified by Camille Cornet (University of Neuchâtel).

Nucleic acid extraction

Protocols for high molecular weight (HMW) DNA extraction developed at the Wellcome Sanger Institute (WSI) Tree of Life Core Laboratory are available on protocols.io (Howard et al., 2025). The ilAplPrae1 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 28.34 ng/μL and a yield of 1 331.98 ng, with a fragment size of 14.6 kb.

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), according to 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 tissue from the head of the ilAplPrae1 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 to create equimolar and/or weighted 2.8 nM pools. Pool concentrations were checked using

the Agilent 4200 TapeStation (Agilent) with High Sensitivity D500 reagents before sequencing. Sequencing was performed using paired-end 150 bp reads on the Illumina NovaSeq X.

Specimen details, sequencing platforms, and data yields are summarised in [Table 1](#).

Genome assembly

Prior to assembly of the PacBio HiFi reads, a database of k -mer counts ($k = 31$) was generated from the filtered reads using [FastK](#). GenomeScope2 ([Ranallo-Benavidez et al., 2020](#)) was used to analyse the k -mer frequency distributions, providing estimates of genome size, heterozygosity, and repeat content.

The HiFi reads were assembled using Hifiasm in Hi-C phasing mode ([Cheng et al., 2021](#); [Cheng et al., 2022](#)), producing two haplotypes. Hi-C reads ([Rao et al., 2014](#)) were mapped to the primary contigs using bwa-mem2 ([Vasimuddin et al., 2019](#)). Contigs were further scaffolded with Hi-C data in YaHS ([Zhou et al., 2023](#)), using the --break option for handling potential misassemblies. The scaffolded assemblies were evaluated using Gfastats ([Formenti et al., 2022](#)), BUSCO ([Manni et al., 2021](#)) and MERQURY.FK ([Rhie et al., 2020](#)).

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

Assembly curation

The assembly was decontaminated using the Assembly Screen for Cobionts and Contaminants (ASCC) pipeline. [TreeVal](#) was used to generate the flat files and maps for use in curation. Manual curation was conducted primarily in PretextView and HiGlass ([Kerpedjiev et al., 2018](#)). Scaffolds were visually inspected and corrected as described by [Howe et al. \(2021\)](#). Manual corrections included six breaks and 16 joins. This

Table 1. Specimen and sequencing data for BioProject PRJEB78786.

Platform	PacBio HiFi	Hi-C
ToLID	ilAplPrae1	ilAplPrae1
Specimen ID	SAN28000096	SAN28000096
BioSample (source individual)	SAMEA115109874	SAMEA115109874
BioSample (tissue)	SAMEA115109906	SAMEA115109908
Tissue	thorax	head
Instrument	Revio	Illumina NovaSeq X
Run accessions	ERR13485739	ERR13493996
Read count total	2.42 million	502.86 million
Base count total	27.46 Gb	75.93 Gb

reduced the scaffold count by 8.5% and increased scaffold N50 by 0.9%. 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), 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 (Di Tommaso *et al.*, 2017) 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).

We used lep_buscoPainter to paint Merian elements along chromosomes (Wright *et al.*, 2024). Merian elements represent the 32 ancestral linkage groups in Lepidoptera. The painting process utilised BUSCO gene locations from the lepidoptera_odb10 set (Kriventseva *et al.*, 2019) and chromosome lengths from NCBI Datasets. Each complete BUSCO gene (both single-copy and duplicated) was assigned to a Merian element based on a reference database, then coloured and plotted along chromosomes drawn to scale.

Genome sequence report

Sequence data

PacBio sequencing of the *Aplocera praeformata* specimen generated 27.46 Gb (gigabases) from 2.42 million reads, which were used to assemble the genome. GenomeScope2.0 analysis estimated the haploid genome size at 377.89 Mb, with a heterozygosity of 1.01% and repeat content of 15.73% (Figure 2). These estimates guided expectations for the assembly. Based on the estimated genome size, the sequencing data provided approximately 70× coverage. Hi-C sequencing produced 75.93 Gb from 502.86 million reads, which were used to scaffold the assembly.

Table 1 summarises the specimen and sequencing details.

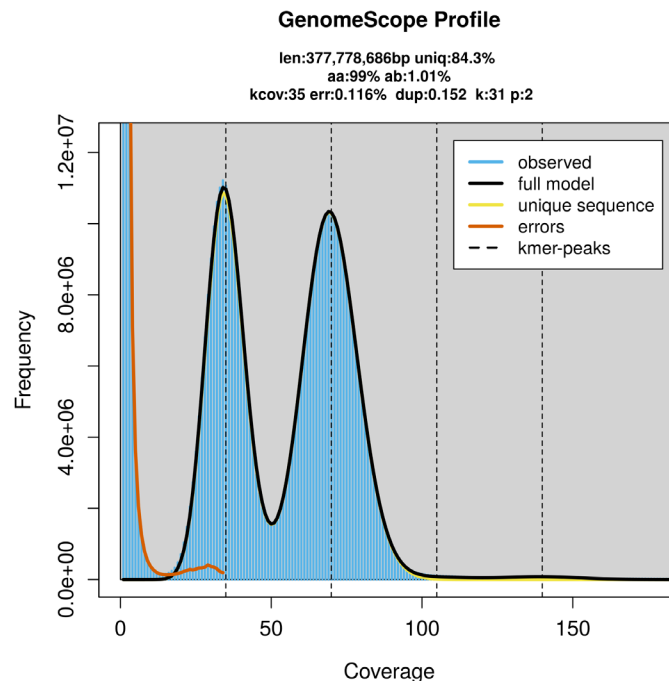


Figure 2. Frequency distribution of *k*-mers generated using GenomeScope2. The plot shows observed and modelled *k*-mer spectra, providing estimates of genome size, heterozygosity, and repeat content based on unassembled sequencing reads.

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 380.66 Mb in 42 scaffolds, with 49 gaps, and a scaffold N50 of 13.79 Mb (Table 2).

Most of the assembly sequence (99.91%) was assigned to 31 chromosomal-level scaffolds, representing 30 autosomes and the Z sex chromosome. These chromosome-level scaffolds, confirmed by Hi-C data, are named according to size (Figure 3; Table 3). Chromosome painting with Merian elements illustrates the distribution of orthologues along chromosomes

Table 2. Genome assembly statistics.

Assembly name	ilAplPrae1.hap1.1	ilAplPrae1.hap2.1
Assembly accession	GCA_964260645.1	GCA_964260625.1
Assembly level	chromosome	scaffold
Span (Mb)	380.66	380.72
Number of chromosomes	31	scaffold-level
Number of contigs	91	90
Contig N50	9.29 Mb	9.43 Mb
Number of scaffolds	42	44
Scaffold N50	13.79 Mb	13.76 Mb
Longest scaffold length (Mb)	22.78	-
Sex chromosomes	Z	-
Organelles	Mitochondrion: 17.22 kb	-

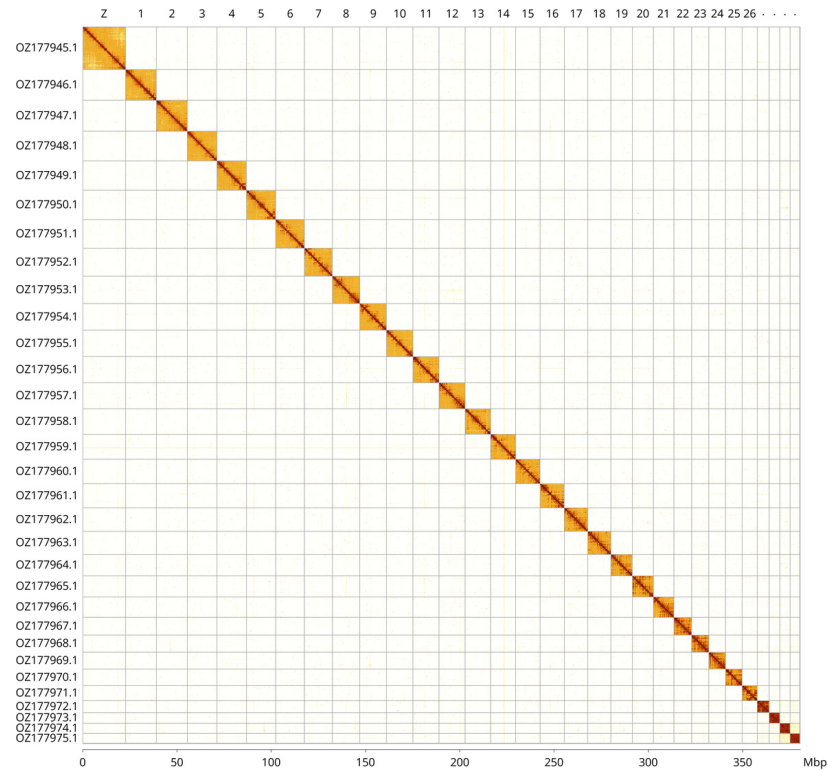


Figure 3. Hi-C contact map of the *Aplocera praeformata* genome assembly. Assembled chromosomes are shown in order of size and labelled along the axes. The plot was generated using PretextSnapshot.

Table 3. Chromosomal pseudomolecules in the haplotype 1 genome assembly of *Aplocera praeformata* ilAplPrae1.

INSDC accession	Molecule	Length (Mb)	GC%	Assigned Merian elements
OZ177946.1	1	16.37	37.50	M1
OZ177947.1	2	16.35	37	M2
OZ177948.1	3	15.79	37	M8
OZ177949.1	4	15.62	37.50	M17;M20
OZ177950.1	5	15.50	37.50	M3
OZ177951.1	6	15.22	37	M9
OZ177952.1	7	14.85	37	M12
OZ177953.1	8	14.55	37	M5
OZ177954.1	9	14.14	37	M18
OZ177955.1	10	13.96	37	M16
OZ177956.1	11	13.93	37.50	M6
OZ177957.1	12	13.79	37.50	M7
OZ177958.1	13	13.51	37.50	M4
OZ177959.1	14	13.31	37	M21
OZ177960.1	15	12.91	37	M22
OZ177961.1	16	12.87	37.50	M15
OZ177962.1	17	12.45	37.50	M10
OZ177963.1	18	12.23	37.50	M11
OZ177964.1	19	11.36	37.50	M13
OZ177965.1	20	11.18	37.50	M23
OZ177966.1	21	10.92	38	M14
OZ177967.1	22	9.35	37.50	M24
OZ177968.1	23	9.12	38	M19
OZ177969.1	24	8.95	37.50	M26
OZ177970.1	25	8.82	38	M28
OZ177971.1	26	7.95	37.50	M27
OZ177972.1	27	6.34	38	M25
OZ177973.1	28	5.59	38	M29
OZ177974.1	29	5.49	40	M30
OZ177975.1	30	5.12	38.50	M31
OZ177945.1	Z	22.78	37	MZ

and highlights patterns of chromosomal evolution relative to Lepidopteran ancestral linkage groups (Figure 4). Chromosome Z was assigned based on synteny to the genome of *Aplocera efformata* (GCA_921293045.1).

The mitochondrial genome was also assembled (length 17.22 kb, OZ177976.1). 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 68.8, and for haplotype 2, 67.4. When the two haplotypes are combined, the assembly achieves an estimated QV of 68.0. The *k*-mer completeness is 78.78% for haplotype 1, 78.76% for haplotype 2, and 99.82% for the combined haplotypes (Figure 5).

BUSCO analysis using the lepidoptera_odb10 reference set ($n = 5\,286$) identified 98.3% of the expected gene set (single = 98.0%, duplicated = 0.3%) in haplotype 1. For haplotype 2, BUSCO analysis identified 98.4% of the expected gene set (single = 98.1%, duplicated = 0.3%).

The snail plot in Figure 6 summarises the scaffold length distribution and other assembly statistics for haplotype 1. The blob plot in Figure 7 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) and the Earth BioGenome Project Report on Assembly Standards September 2024. The EBP metric, calculated for the haplotype 1, is **6.C.Q68**, meeting the recommended reference standard.

Wellcome Sanger Institute – Legal and Governance

The materials that have contributed to this genome note have been supplied by a Tree of Life collaborator. The Wellcome Sanger Institute employs a process whereby due diligence is carried out proportionate to the nature of the materials themselves, and the circumstances under which they have been/are to be collected and provided for use. The purpose of this is to address and mitigate any potential legal and/or ethical implications of receipt and use of the materials as part of the research project, and to ensure that in doing so, we align with best practice wherever possible. The overarching areas of consideration are:

- Ethical review of provenance and sourcing of the material
- Legality of collection, transfer and use (national and international).

Each transfer of samples is undertaken according to a Research Collaboration Agreement or Material Transfer Agreement entered into by the Tree of Life collaborator, Genome Research

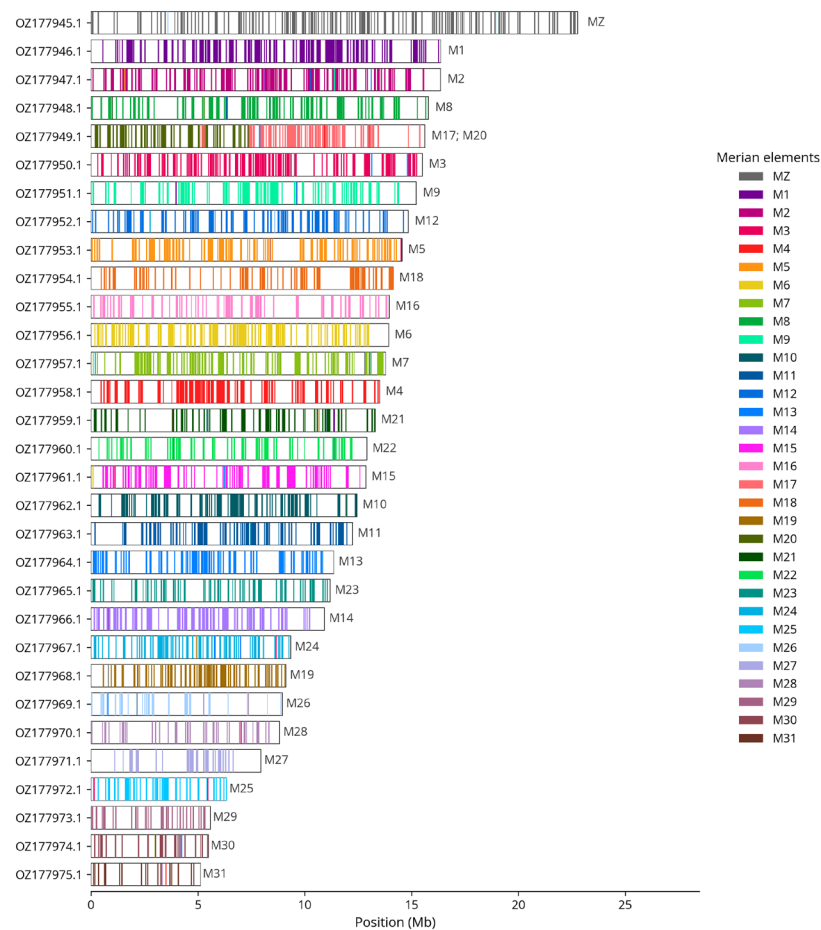


Figure 4. Merian elements painted across chromosomes in the iAplPrae1.hap1.1 assembly of *Aplocera praeformata*. Chromosomes are drawn to scale, with the positions of orthologues shown as coloured bars. Each orthologue is coloured by the Merian element that it belongs to. All orthologues which could be assigned to Merian elements are shown.

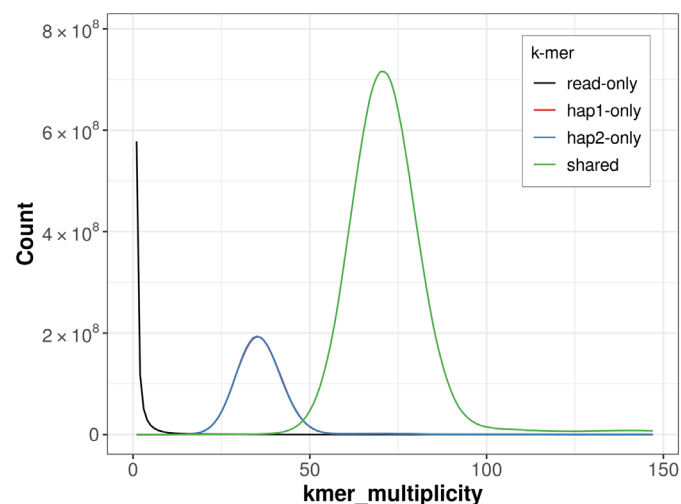


Figure 5. Evaluation of *k*-mer completeness using MerquryFK. This plot illustrates the recovery of *k*-mers from the original read data in the final assemblies. The horizontal axis represents *k*-mer multiplicity, and the vertical axis shows the number of *k*-mers. The black curve represents *k*-mers that appear in the reads but are not assembled. The green curve (the homozygous peak) corresponds to *k*-mers shared by both haplotypes and the red and blue curves (the heterozygous peaks) show *k*-mers found only in one of the haplotypes.

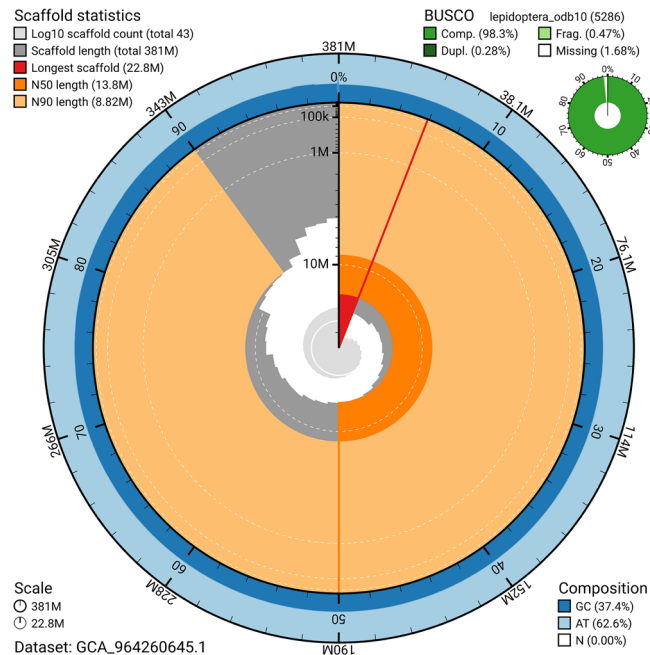


Figure 6. Assembly metrics for ilAplPrae1.hap1.1. The BlobToolKit snail plot provides an overview of assembly metrics and BUSCO gene completeness. The circumference represents the length of the whole genome sequence, and the main plot is divided into 1,000 bins around the circumference. The outermost 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](#).

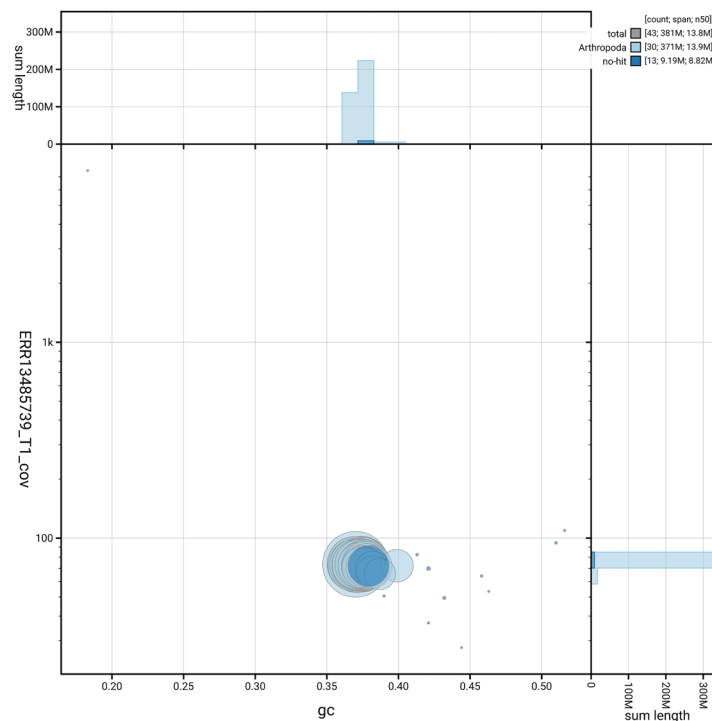


Figure 7. BlobToolKit GC-coverage plot for ilAplPrae1.hap1.1. Blob plot showing sequence coverage (vertical axis) and GC content (horizontal axis). The circles represent scaffolds, with the size proportional to scaffold length and the colour representing phylum membership. The histograms along the axes 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 *Aplocera praeformata* assembly.

Measure	Value	Benchmark
EBP summary (haplotype 1)	6.C.Q68	6.C.Q40
Contig N50 length	9.29 Mb	≥ 1 Mb
Scaffold N50 length	13.79 Mb	= chromosome N50
Consensus quality (QV)	Haplotype 1: 68.8; haplotype 2: 67.4; combined: 68.0	≥ 40
k-mer completeness	Haplotype 1: 78.78%; Haplotype 2: 78.76%; combined: 99.82%	≥ 95%
BUSCO	C:98.3% [S:98.0%; D:0.3%]; F:0.5%; M:1.2%; n:5 286	S > 90%; D < 5%
Percentage of assembly assigned to chromosomes	99.91%	≥ 90%

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

Data availability

European Nucleotide Archive: *Aplocera praeformata* (purple treble-bar). Accession number [PRJEB78786](https://www.ebi.ac.uk/ena/record/PRJEB78786). The genome sequence is released openly for reuse. The *Aplocera praeformata* 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](https://www.ebi.ac.uk/ena/record/PRJEB78786) 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.

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
Hifiasm	0.19.8-r603	https://github.com/chhy123/hifiasm
HiGlass	1.13.4	https://github.com/higlass/higlass
lep_buscoPainter	1.0.0	https://github.com/charlottewright/lep_buscoPainter
MercuryFK	1.1.2	https://github.com/thegenemyers/MERQURY.FK

Software	Version	Source
Minimap2	2.24-r1122	https://github.com/lh3/minimap2
MitoHiFi	3	https://github.com/marcelauliano/MitoHiFi
MultiQC	1.14; 1.17 and 1.18	https://github.com/MultiQC/MultiQC
Nextflow	23.10.0	https://github.com/nextflow-io/nextflow
PretextSnapshot	0.0.5	https://github.com/sanger-tol/PretextSnapshot
PretextView	1.0.3	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
sanger-tol/curationpretext	1.4.2	https://github.com/sanger-tol/curationpretext
Seqtk	1.3	https://github.com/lh3/seqtk
Singularity	3.9.0	https://github.com/sylabs/singularity
TreeVal	1.4.0	https://github.com/sanger-tol/treeval
YaHS	1.2a.2	https://github.com/c-zhou/yahs

References

- Allio R, Schomaker-Bastos A, Romiguier J, *et al.*: **MitoFinder: efficient automated large-scale extraction of mitogenomic data in target enrichment phylogenomics.** *Mol Ecol Resour.* 2020; **20**(4): 892–905.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Altschul SF, Gish W, Miller W, *et al.*: **Basic Local Alignment Search Tool.** *J Mol Biol.* 1990; **215**(3): 403–410.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Bateman A, Martin MJ, Orchard S, *et al.*: **UniProt: the Universal Protein Knowledgebase in 2023.** *Nucleic Acids Res.* 2023; **51**(D1): D523–D531.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Buchfink B, Reuter K, Drost HG: **Sensitive protein alignments at Tree-of-Life scale using DIAMOND.** *Nat Methods.* 2021; **18**(4): 366–368.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Challis R, 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](#)
- Di Tommaso P, Chatzou M, Floden EW, *et al.*: **Nextflow enables reproducible computational workflows.** *Nat Biotechnol.* 2017; **35**(4): 316–319.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Ewels P, Magnusson M, Lundin S, *et al.*: **MultiQC: summarize analysis results for multiple tools and samples in a single report.** *Bioinformatics.* 2016; **32**(19): 3047–3048.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Ewels PA, Peltzer A, Fillinger S, *et al.*: **The nf-core framework for community-curated bioinformatics pipelines.** *Nat Biotechnol.* 2020; **38**(3): 276–278.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Formenti G, Abueg L, Brajuka A, *et al.*: **Gfastats: conversion, evaluation and manipulation of genome sequences using assembly graphs.** *Bioinformatics.* 2022; **38**(17): 4214–4216.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- GBIF Secretariat: ***Aplocera praeformata* (Hübner, 1826).** *GBIF Backbone Taxonomy.* GBIF, 2023.
[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): g1aa153.
[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](#)
- Kimber I: **Purple treble-bar, *Aplocera praeformata* (Hübner, 1826).** *UKMoths,* 2025.
[Reference Source](#)
- Kriventseva EV, Kuznetsov D, Tegenfeldt F, *et al.*: **OrthoDB v10: sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs.** *Nucleic Acids Res.* 2019; **47**(D1): D807–D811.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Kurtzer GM, Sochat V, Bauer MW: **Singularity: scientific containers for mobility of compute.** *PLoS One.* 2017; **12**(5): e0177459.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Li H: **Minimap2: pairwise alignment for nucleotide sequences.** *Bioinformatics.* 2018; **34**(18): 3094–3100.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Manni M, Berkeley MR, Seppely M, *et al.*: **BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes.** *Mol Biol Evol.* 2021; **38**(10): 4647–4654.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Merkel D: **Docker: lightweight Linux containers for consistent development and deployment.** *Linux J.* 2014; **2014**(239): 2.
[Reference Source](#)

Ranallo-Benavidez TR, Jaron KS, Schatz MC: **GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes.** *Nat Commun.* 2020; **11**(1): 1432.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Rao SSP, Huntley MH, Durand NC, *et al.*: **A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping.** *Cell.* 2014; **159**(7): 1665–1680.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

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

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

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

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

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](#)

Ulliano-Silva M, Ferreira JGRN, Krashennikova 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](#)

Wagner W: ***Aplocera praeformata (Hübner, 1826).*** European Lepidoptera; their ecology, 2025.

[Reference Source](#)

Waring P, Townsend M: **Field guide to the moths of Great Britain and Ireland.** London: Bloomsbury Publishing, 2017.

[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 23 June 2026

<https://doi.org/10.21956/wellcomeopenres.28319.r146876>

© 2026 Khedkar G. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Gulab D Khedkar 

Dr Babasaheb Ambedkar Marathwada University, Aurangabad, Maharashtra, India

This manuscript presents a high-quality chromosome-level genome assembly of *Aplocera praeformata*, generated using PacBio HiFi and Hi-C sequencing technologies. The study is technically robust, and the assembly meets current reference genome standards with excellent contiguity, completeness, and consensus quality. The methodological workflow is comprehensive, transparent, and highly reproducible, making this genome resource valuable for future studies on Lepidopteran evolution, comparative genomics, and chromosome biology.

The manuscript is generally well organized and clearly written. The sequencing strategy, assembly pipeline, quality assessment, and data availability are described in sufficient detail, and the authors provide extensive supporting metrics demonstrating the reliability of the assembly. The inclusion of Merian element analysis further enhances the resource's biological relevance.

Optional minor points could further improve the manuscript's utility:

The manuscript would also benefit from a short statement outlining future directions, particularly regarding functional genome annotation and comparative genomic analyses.

Recommendation: Accept for indexing in its present form. The manuscript reports a valuable, high-quality genomic resource generated using state-of-the-art sequencing and assembly approaches, and the data will be of considerable interest to the genomics and Lepidoptera research communities.

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, population genetics, DNA Barcoding

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 02 February 2026

<https://doi.org/10.21956/wellcomeopenres.28319.r145119>

© 2026 Alyousuf A. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Aqeel Alyousuf 

University of Basrah, Basrah, Iraq

The Methods section is thorough and clearly documented, covering all key stages of the workflow, including sample collection, DNA extraction, library preparation, sequencing, assembly, curation, and quality assessment. The use of established protocols alongside PacBio HiFi and Arima Hi-C technologies demonstrates a high technical standard. The inclusion of detailed information on fragment size, DNA concentration, and purity is particularly helpful for reproducibility.

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

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.