



DATA NOTE

The genome sequence of the Purple Cloud, *Actinotia polyodon* (Clerck, 1759) (Lepidoptera: Noctuidae)

[version 1; peer review: 3 approved]

oreina, Papillons de France, Michal Rindoš¹, Charlotte J. Wright ²,
Joana I. Meier², Mark L. Blaxter ²,
Wellcome Sanger Institute Tree of Life Management, Samples and Laboratory
team,
Wellcome Sanger Institute Scientific Operations: Sequencing Operations,
Wellcome Sanger Institute Tree of Life Core Informatics team,
Tree of Life Core Informatics collective, Project Psyche Community

¹Biology Centre of the Czech Academy of Sciences, Institute of Entomology, Ceske Budejovice, Czech Republic

²Tree of Life Programme, Wellcome Sanger Institute, Hinxton, England, UK

V1 First published: 16 Jan 2026, 11:40
<https://doi.org/10.12688/wellcomeopenres.25772.1>

Latest published: 16 Jan 2026, 11:40
<https://doi.org/10.12688/wellcomeopenres.25772.1>

Abstract

We present a genome assembly from a female specimen of *Actinotia polyodon* (Purple Cloud; Arthropoda; Insecta; Lepidoptera; Noctuidae). The assembly contains two haplotypes with total lengths of 686.76 megabases and 626.61 megabases. Most of haplotype 1 (99.78%) is scaffolded into 32 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 15.36 kilobases. This work is part of Project Psyche, a collaborative programme generating genomes for European butterflies and moths.

Keywords

Actinotia polyodon; Purple Cloud; genome sequence; chromosomal; Lepidoptera



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

Open Peer Review

Approval Status

	1	2	3
version 1			
16 Jan 2026	view	view	view

1. **Kathy Darragh** , Indiana University
Bloomington, Bloomington, USA
2. **Jonathan Badger** , National Cancer
Institute, Bethesda, USA
3. **Ariel Chipman** , Hebrew University of
Jerusalem School of Pharmacy, Jerusalem,
Israel

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: **Rindoš M:** Writing – Original Draft Preparation, 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 funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*

Copyright: © 2026 oreina, Papillons de France *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: oreina, Papillons de France, Rindoš M, Wright CJ *et al.* **The genome sequence of the Purple Cloud, *Actinotia polyodon* (Clerck, 1759) (Lepidoptera: Noctuidae) [version 1; peer review: 3 approved]** Wellcome Open Research 2026, 11:40 <https://doi.org/10.12688/wellcomeopenres.25772.1>

First published: 16 Jan 2026, 11:40 <https://doi.org/10.12688/wellcomeopenres.25772.1>

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; Noctuoidea; Noctuidae; Xyleninae; *Actinotia*; *Actinotia polyodon* (Clerck, 1759) (NCBI:txid753264)

Background

The Purple Cloud, *Actinotia polyodon* (Clerck, 1759), (Figure 1) is a unique, unmistakable moth that occurs throughout the Palaearctic region (Macek *et al.*, 2009). The forewings are yellowish with purple and brown, the reniform stigma is large and pale with a linear centre and creamy white outline (Fibiger, 1990). The hindwings are pale with a broad dark border and black veins.

The adults are multivoltine and fly in two generations, from May to June and later in August (Fibiger, 1990). They are primarily nocturnal, but in the very warm days they are also active during the day at higher altitudes. The Purple Cloud is a xerothermophilic species that inhabits habitats such as forest edges, sunny slopes, gravel pits and clearings (Macek *et al.*, 2009). The larvae are reddish-brown in colour with white dorsal and spiracular lines. They feed at night on the flowers of various species of the genus *Hypericum* L., and remain hidden in the litter around the host plant during the day (Fibiger, 1990). The Purple Cloud overwinters as a pupa covered by a thin, silky cocoon.

Methods

Sample acquisition

The specimen used for genome sequencing was an adult female *Actinotia polyodon* (specimen ID SAN28000394, ToLID ilActPoly1; Figure 1), collected from Saint-Étienne-De-Baïgorry, Pyrénées-Atlantiques, Pays Basques, France (latitude 43.1744, longitude -1.3301) on 2024-05-18. The specimen was collected and identified by Eric Drouet (Oreina, France).

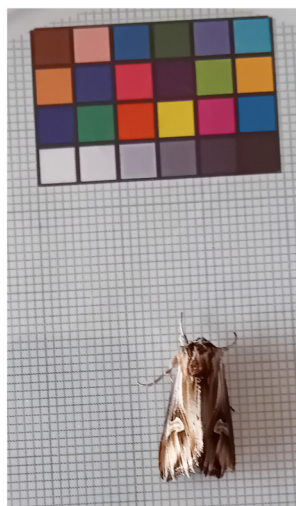


Figure 1. Voucher image of the *Actinotia polyodon* specimen used for genome sequencing (SAN28000394).

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 ilActPoly1 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 74.82 ng/μL and a yield of 3 516.54 ng, and a fragment size of 14.8 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 ilActPoly1 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

Table 1. Specimen and sequencing data for BioProject PRJEB82264.

Platform	PacBio HiFi	Hi-C
ToLID	ilActPoly1	ilActPoly1
Specimen ID	SAN28000394	SAN28000394
BioSample (source individual)	SAMEA115948954	SAMEA115948954
BioSample (tissue)	SAMEA115949434	SAMEA115949431
Tissue	thorax	head
Instrument	Revio	Illumina NovaSeq X
Run accessions	ERR13946493	ERR13947521
Read count total	5.07 million	781.24 million
Base count total	47.18 Gb	117.97 Gb

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 five breaks and 38 joins, which reduced the scaffold count by 26.8%. The curation process is described at <https://gitlab.com/wtsi-grit/rapid-curation>. PretextViewSnapshot 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 database ($k = 31$) 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 plotted along chromosomes drawn to scale.

Genome sequence report

Sequence data

PacBio sequencing of the *Actinotia polyodon* specimen generated 47.18 Gb (gigabases) from 5.07 million reads, which were used to assemble the genome. GenomeScope2.0 analysis estimated the haploid genome size at 654.01 Mb, with a heterozygosity of 0.71% and repeat content of 26.31%

(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 117.97 Gb from 781.24 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 686.76 Mb in 89 scaffolds, with 127 gaps, and a scaffold N50 of 23.16 Mb (Table 2).

Most of the assembly sequence (99.78%) was assigned to 32 chromosomal-level scaffolds, representing 30 autosomes and the W and Z sex chromosomes. 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 and highlights patterns of chromosomal evolution relative to Lepidopteran ancestral linkage groups (Figure 4). During curation, we observed that contigs along chromosome 28 from 4.5 Mb and 5.9 Mb have uncertain order and orientation.

The mitochondrial genome was also assembled (length 15.36 kb, OZ203337.1). 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 65.0, and for haplotype 2, 65.2. When the two haplotypes are combined, the assembly achieves an estimated QV of 65.1. The *k*-mer completeness is

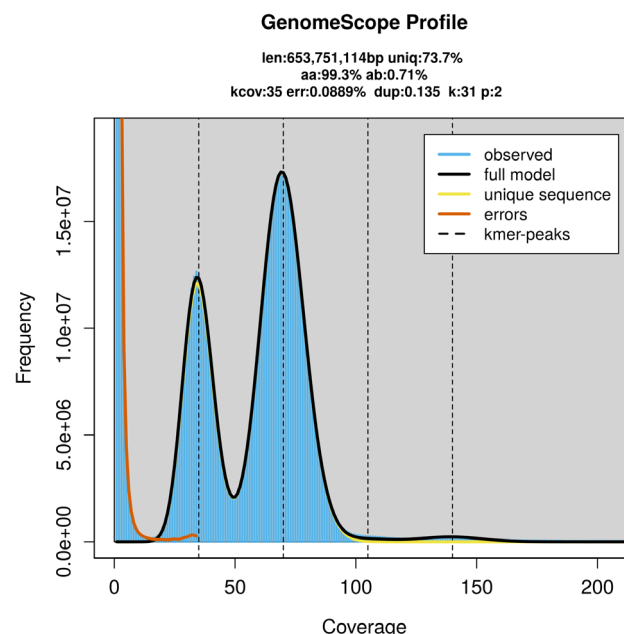


Figure 2. Frequency distribution of *k*-mers generated using GenomeScope2. The plot shows observed and modelled *k*-mer spectra, providing estimates of genome size, heterozygosity, and repeat content based on unassembled sequencing reads.

Table 2. Genome assembly statistics.

Assembly name	ilActPoly1.hap1.1	ilActPoly1.hap2.1
Assembly accession	GCA_964340415.1	GCA_964340405.1
Assembly level	chromosome	scaffold
Span (Mb)	686.76	626.61
Number of chromosomes	32	scaffold-level
Number of contigs	216	161
Contig N50	7.93 Mb	9.21 Mb
Number of scaffolds	89	82
Scaffold N50	23.16 Mb	22.79 Mb
Longest scaffold length (Mb)	34.12	-
Sex chromosomes	W and Z	-
Organelles	Mitochondrion: 15.36 kb	-

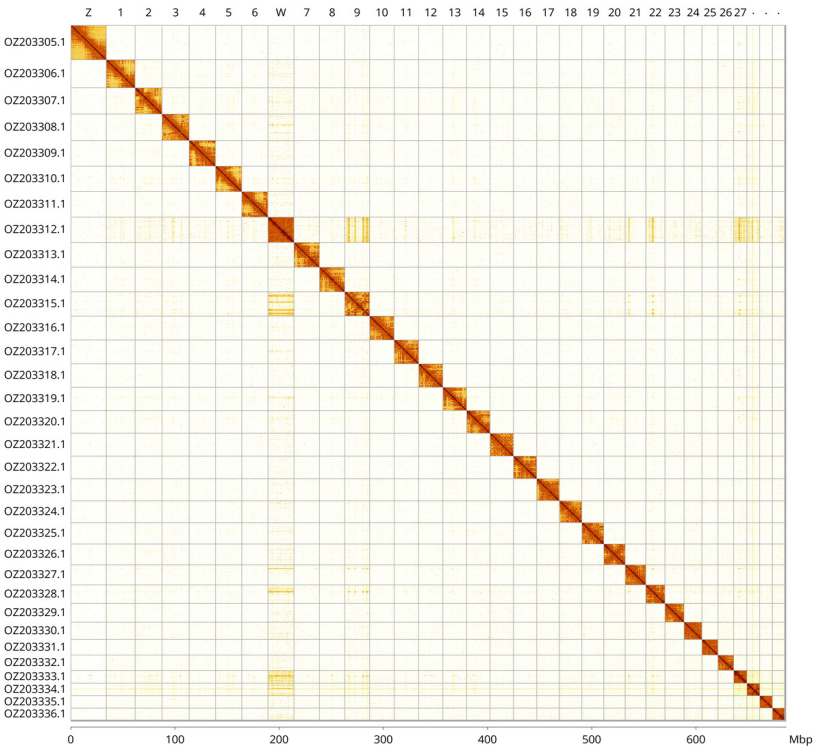


Figure 3. Hi-C contact map of the *Actinotia polyodon* genome assembly. Assembled chromosomes are shown in order of size and labelled along the axes, with a megabase scale shown below. The plot was generated using PretextSnapshot.

86.41% for haplotype 1, 80.75% for haplotype 2, and 99.78% for the combined haplotypes (Figure 5).

BUSCO analysis using the lepidoptera_odb10 reference set ($n = 5\,286$) identified 98.9% of the expected gene set (single = 98.4%, duplicated = 0.5%) in haplotype 1. For haplotype 2, BUSCO analysis identified 94.7% of the expected gene set (single = 94.4%, 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.Q65**, meeting the recommended reference standard.

Table 3. Chromosomal pseudomolecules in the haplotype 1 genome assembly of *Actinotia polyodon* iActPoly1.

INSDC accession	Molecule	Length (Mb)	GC%	Assigned Merian elements
OZ203306.1	1	27.61	37.50	M2
OZ203307.1	2	25.99	37.50	M17;M20
OZ203308.1	3	25.97	37.50	M5
OZ203309.1	4	25.41	37	M9
OZ203310.1	5	25.18	37.50	M1
OZ203311.1	6	25.01	37	M8
OZ203313.1	7	24.56	37	M12
OZ203314.1	8	24.19	37.50	M3
OZ203315.1	9	23.88	37.50	M23
OZ203316.1	10	23.56	37.50	M7
OZ203317.1	11	23.52	37	M16
OZ203318.1	12	23.16	37	M18
OZ203319.1	13	22.89	37.50	M4
OZ203320.1	14	22.53	37.50	M6
OZ203321.1	15	22.46	37	M22

INSDC accession	Molecule	Length (Mb)	GC%	Assigned Merian elements
OZ203322.1	16	22.31	37	M21
OZ203323.1	17	21.78	37.50	M10
OZ203324.1	18	21.58	37.50	M15
OZ203325.1	19	21.01	37.50	M11
OZ203326.1	20	20.58	37.50	M14
OZ203327.1	21	19.67	37	M13
OZ203328.1	22	18.41	38	M19
OZ203329.1	23	18.38	37.50	M24
OZ203330.1	24	17.19	37.50	M26
OZ203331.1	25	15.34	37.50	M28
OZ203332.1	26	15.10	37.50	M27
OZ203333.1	27	12.66	39.50	M30
OZ203334.1	28	12.33	40.50	M31
OZ203335.1	29	12.20	38.50	M25
OZ203336.1	30	11.69	38.50	M29
OZ203312.1	W	24.98	38.50	-
OZ203305.1	Z	34.12	37	MZ

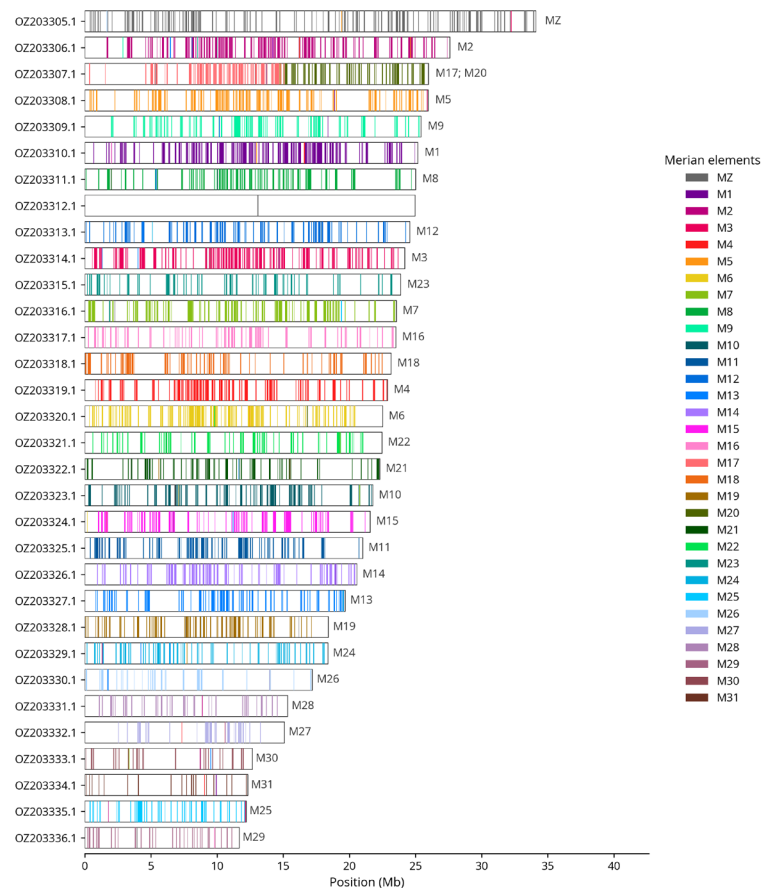


Figure 4. Merian elements painted across chromosomes in the iActPoly1.hap1.1 assembly of *Actinotia polyodon*. Chromosomes are drawn to scale, with the positions of orthologues shown as coloured bars. Each orthologue is coloured by the Merian element that it belongs to. All orthologues which could be assigned to Merian elements are shown.

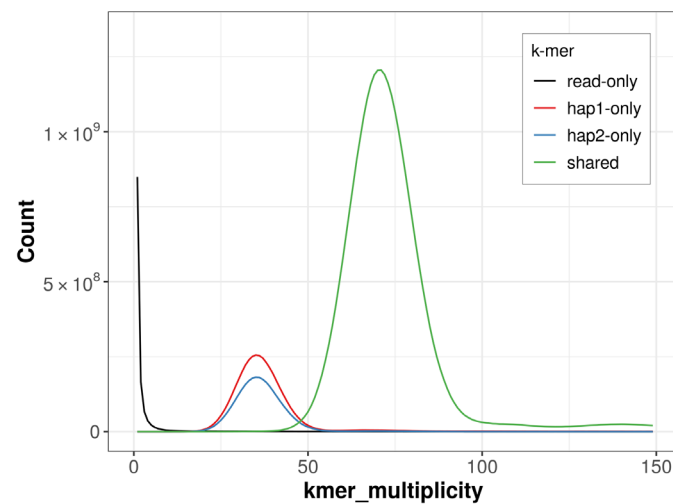


Figure 5. Evaluation of *k*-mer completeness using MerquyFK. This plot illustrates the recovery of *k*-mers from the original read data in the final assemblies. The horizontal axis represents *k*-mer multiplicity, and the vertical axis shows the number of *k*-mers. The black curve represents *k*-mers 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.

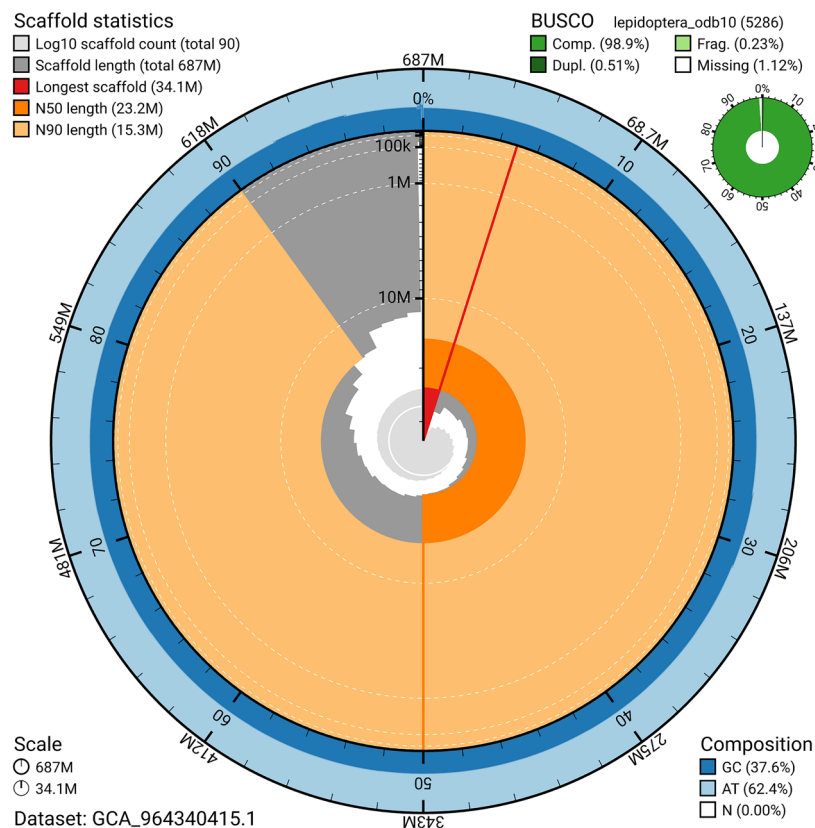


Figure 6. Assembly metrics for ilActPoly1.hap1.1. The BlobToolKit snail plot provides an overview of assembly metrics and BUSCO gene completeness. The circumference represents the length of the whole genome sequence, and the main plot is divided into 1,000 bins around the circumference. The outermost 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](#).

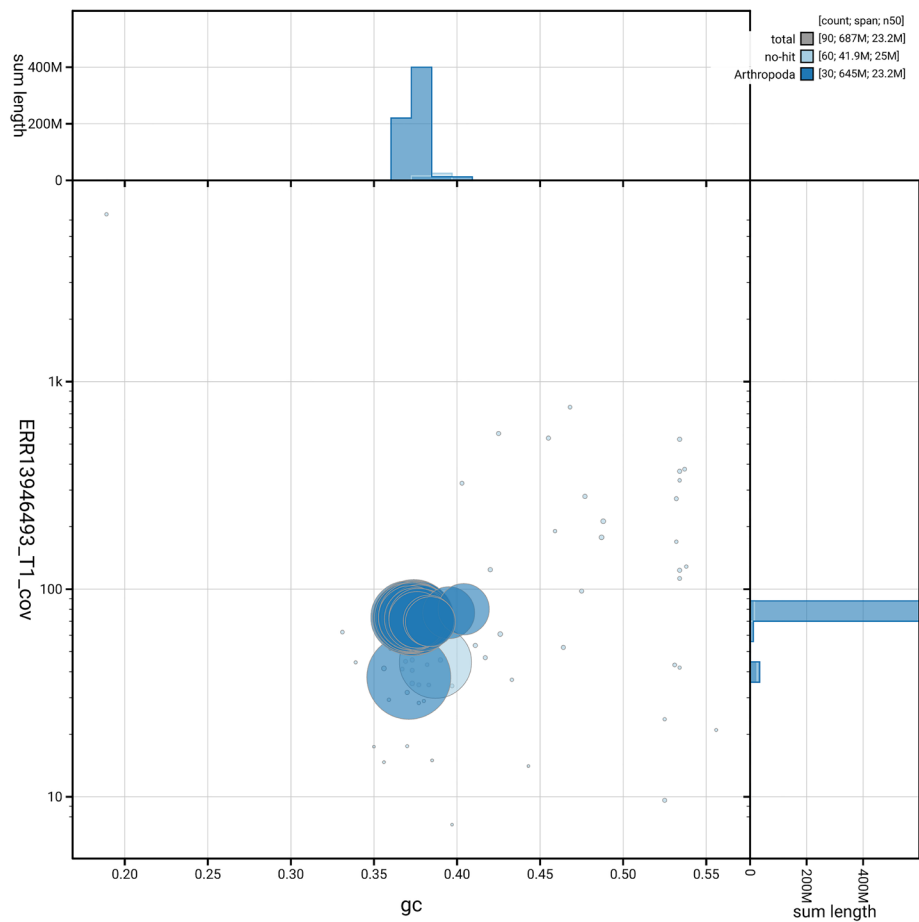


Figure 7. BlobToolKit GC-coverage plot for ilActPoly1.hap1.1. Blob plot showing sequence coverage (vertical axis) and GC content (horizontal axis). The circles represent scaffolds, with the size proportional to scaffold length and the colour representing phylum membership. The histograms along the axes 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 *Actinotia polyodon* assembly.

Measure	Value	Benchmark
EBP summary (haplotype 1)	6.C.Q65	6.C.Q40
Contig N50 length	7.93 Mb	≥ 1 Mb
Scaffold N50 length	23.16 Mb	= chromosome N50
Consensus quality (QV)	Haplotype 1: 65.0; haplotype 2: 65.2; combined: 65.1	≥ 40
k-mer completeness	Haplotype 1: 86.41%; Haplotype 2: 80.75%; combined: 99.78%	≥ 95%
BUSCO	C:98.9% [S:98.4%; D:0.5%]; F:0.2%; M:0.9%; n:5 286	S > 90%; D < 5%
Percentage of assembly assigned to chromosomes	99.78%	≥ 90%
Percentage of assembly assigned to chromosomes	98.16%	≥ 90%

Notes: EBP summary uses log10(Contig N50) . log10(Scaffold N50) . Q (Merquy QV). BUSCO: C=complete; S=single-copy; D=duplicated; F=fragmented; M=missing; n=orthologues

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

Data availability

European Nucleotide Archive: *Actinotia polyodon*. Accession number [PRJEB82264](#). The genome sequence is released openly for reuse. The *Actinotia polyodon* genome sequencing initiative

is part of the Sanger Institute Tree of Life Programme (PRJEB43745) and Project Psyche (PRJEB71705). All raw sequence data and the assembly have been deposited in INSDC databases. The genome will be annotated using available RNA-Seq data and presented through [Ensembl](#) at the European Bioinformatics Institute. Raw data and assembly accession identifiers are reported in [Table 1](#) and [Table 2](#).

Pipelines used for genome assembly at the WSI Tree of Life are available at <https://pipelines.tol.sanger.ac.uk/pipelines>. [Table 5](#) lists software versions used in this study.

Author information

Contributors are listed at the following links:

- Members of [oreina](#) (Papillons de France).
- 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/tolkita/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/chhyllp123/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
Minimap2	2.24-r1122	https://github.com/lh3/minimap2

Software	Version	Source
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
PretextView	0.0.5	https://github.com/sanger-tol/PretextView
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](#)
- Fibiger M: **Noctuinae I, Vol. 1.** Sorø: Entomological Press, 1990.
- 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](#)
- 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](#)
- 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](#)
- Macek J, Dvořák J, Traxler L, *et al.*: **Noční Motýli – můrovití Vol. II.** Praha: Academia, 2009.
- 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.*: **Merquy: 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](#)

Uliano-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](#)

Wright CJ, Stevens L, Mackintosh A, *et al.*: **Comparative genomics reveals the dynamics of chromosome evolution in Lepidoptera.** *Nat Ecol Evol.* 2024; **8**(4): 777–90.

[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 17 February 2026

<https://doi.org/10.21956/wellcomeopenres.28373.r145390>

© 2026 Chipman 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.



Ariel Chipman 

Hebrew University of Jerusalem School of Pharmacy, Jerusalem, Israel

The authors assembled the genome of the noctuid moth *Actinotia polyodon* from a single female collected in France. Sequencing and assembly were done according to well established protocols, by a research group with ample experience in using these protocols. The quality of the assembled genome is very high, with chromosome-length scaffolds corresponding to the expected number of chromosomes, including the W and Z chromosomes, as well as the mitochondria as a separate scaffold. The work adheres to expected standards and is presented as is customary in the field. I have no comments regarding the manuscript.

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 12 February 2026

<https://doi.org/10.21956/wellcomeopenres.28373.r145970>

© 2026 Badger J. 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. The author(s) is/are employees of the US Government and therefore domestic copyright protection in USA does not apply to this work. The work may be protected under the copyright laws of other jurisdictions when used in those jurisdictions.



Jonathan Badger 

National Cancer Institute, Bethesda, Maryland, USA

First of all, please check the author list to make sure it is correct -- currently the first two authors are listed as oreina, Papillons de France. I understand that this butterfly organization contributed to this work, but please recheck the authorlist.

Other than that, I think this is a well written genome note that gives sufficient detail as to background, methods, and results. I particularly liked the inclusion of Table 5, as details of the software used are often missing in these data notes.

You may want to include some discussion of why you had issues with the ordering of contigs on chromosome 28 and not the other ones.

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 am a computational genomicist primarily working in biomedical research, although I have also contributed to insect, fungal, and protist genome projects.

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 06 February 2026

<https://doi.org/10.21956/wellcomeopenres.28373.r145969>

© 2026 Darragh K. 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.



Kathy Darragh 

Indiana University Bloomington, Bloomington, Indiana, USA

This is a standard genome note. All methods are fully described with sufficient background on the species and a specimen photograph. All information is clear and well presented. Data has been made available and this is also included in the note. This genome will contribute to comparative genomics projects. I have no suggestions for the authors.

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: Evolutionary biology, 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.
