DATA NOTE

The genome sequence of the Titania's Fritillary, *Boloria titania* (Esper, 1793) (Lepidoptera: Nymphalidae)

[version 1; peer review: 3 approved]

Irena Klečková[1], Pavel Matos-Maraví[1], Charlotte J. Wright[2], Joana I. Meier[2], 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

[1]Biology Centre CAS, Institute of Entomology, České Budějovice, Czech Republic
[2]Tree of Life, Wellcome Sanger Institute, Hinxton, England, UK

v1 **First published:** 30 Jul 2025, **10**:388
https://doi.org/10.12688/wellcomeopenres.24643.1
Latest published: 30 Jul 2025, **10**:388
https://doi.org/10.12688/wellcomeopenres.24643.1

Abstract

We present a genome assembly from a female specimen of *Boloria titania* (Titania's Fritillary; Arthropoda; Insecta; Lepidoptera; Nymphalidae). The assembly contains two haplotypes with total lengths of 496.64 megabases and 454.08 megabases. Most of haplotype 1 (97.13%) 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 15.17 kilobases.

Keywords

Boloria titania, Titania's Fritillary, genome sequence, chromosomal, Lepidoptera

This article is included in the Tree of Life gateway.

Open Peer Review

Approval Status ✓ ✓ ✓

	1	2	3
version 1 30 Jul 2025	✓ view	✓ view	✓ view

1. **Aqeel Alyousuf**, University of Basrah, Basrah, Iraq
2. **Arun Arumugaperumal**, Rajalakshmi Engineering College, Thandalam 602105, Chennai, India
3. **Annabel Whibley**, Bragato Research Institute, Lincoln, New Zealand

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: Klečková I: Investigation, Resources, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing; **Matos-Maraví P**: Resources, Supervision; **Wright CJ**: Funding Acquisition, Project Administration, Supervision; **Meier JI**: Funding Acquisition, Project Administration, Supervision; **Blaxter ML**: Funding Acquisition, Project Administration, Supervision;

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by Wellcome through core funding to the Wellcome Sanger Institute (220540).
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

How to cite this article: Klečková I, Matos-Maraví P, Wright CJ *et al.* **The genome sequence of the Titania's Fritillary, *Boloria titania* (Esper, 1793) (Lepidoptera: Nymphalidae) [version 1; peer review: 3 approved]** Wellcome Open Research 2025, **10**:388 https://doi.org/10.12688/wellcomeopenres.24643.1

First published: 30 Jul 2025, **10**:388 https://doi.org/10.12688/wellcomeopenres.24643.1

Species taxonomy

Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Protostomia; Ecdysozoa; Panarthropoda; Arthropoda; Mandibulata; Pancrustacea; Hexapoda; Insecta; Dicondylia; Pterygota; Neoptera; Endopterygota; Amphiesmenoptera; Lepidoptera; Glossata; Neolepidoptera; Heteroneura; Ditrysia; Obtectomera; Papilionoidea; Nymphalidae; Heliconiinae; Argynnini; *Boloria*; *Clossiana*; *Boloria titania* (Esper, 1793) (NCBI:txid596536)

Background

Boloria titania is a mountain species with a Holarctic distribution (Tolman & Lewington, 2009). In Europe, it occurs in the Massif Central, the Alps, southern Finland, Latvia, Romania, and the Balkans. Beyond Europe, it is found in the Urals, Siberia, the mountains of Asia, and North America (Cuvelier & Dincă, 2007; Tolman & Lewington, 2009). However, the taxonomic status of the North American population remains unclear (McFarland, 2003; Simonsen *et al.*, 2010). The species is likely extinct in Poland (Glowacinski *et al.*, 2002).

It belongs to the family Nymphalidae. Similar to other fritillaries, the upper side of the wings is orange with black markings. In *B. titania*, triangular submarginal markings on the upper side of the hindwings are often connected to postdiscal spots (Tolman & Lewington, 2009). The underside of the hindwings presents a mosaic of soft pinks, reds, and purples. Additionally, it is larger compared to related species (Simonsen *et al.*, 2010).

The species is univoltine, with a flight period from mid-June to August. Larvae overwinter in the L1 stage (Wagner, 2005) and feed on common bistort (*Polygonum bistorta*) and several *Viola* species (Cozzi *et al.*, 2008). *Boloria titania* inhabits wet meadows and clearings within coniferous forests (Cuvelier & Dincă, 2007; Tolman & Lewington, 2009). Females lay their eggs near forests, particularly close to spruce trees (Ebert, 2005).

The species is classified as Near Threatened on the IUCN Red List for Europe (Van Swaay *et al.*, 2010). *Boloria titania* is threatened by the drainage of wet meadows (Cuvelier & Dincă, 2007) and changes in landscape structure due to its specialisation in tree-rich wetlands (Cozzi *et al.*, 2008). Additionally, habitat alterations and ongoing climate change pose risks, particularly due to their impact on the larval food plant *P. bistorta* (Schweiger *et al.*, 2008; Singer *et al.*, 2018).

We present a chromosome-level, haplotype-resolved genome sequence of the Titania's Fritillary, *Boloria titania*, sequenced as part of Project Psyche.

Methods

Sample acquisition

The specimen used for genome sequencing was an adult female *Boloria titania* (specimen ID SAN28000080, ToLID ilCloTita1; Figure 1), collected from Brochwald, Meiringen, Switzerland (latitude 46.6743, longitude 8.1334; elevation 1 470 m) on 09/07/2023. The specimen was collected and identified by Irena Klečková (Czech Institute of Entomology).

Figure 1. Voucher photograph of the Boloria titania (ilCloTita1) specimen used for genome sequencing.

Nucleic acid extraction

Protocols for high molecular weight (HMW) DNA extraction developed at the Wellcome Sanger Institute (WSI) Tree of Life Core Laboratory are available on protocols.io (Howard *et al.*, 2025). The ilCloTita1 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 55.8 ng/μL and a yield of 2 622.60 ng, with a fragment size of 13.4 kb. The 260/280 spectrophotometric ratio was 1.92, and the 260/230 ratio was 2.51.

PacBio HiFi library preparation and sequencing

Library preparation and sequencing were performed at the WSI Scientific Operations core. Libraries were prepared using the SMRTbell Prep Kit 3.0 (Pacific Biosciences, California, USA), following the manufacturer's instructions. The kit includes reagents for end repair/A-tailing, adapter ligation, post-ligation SMRTbell bead clean-up, and nuclease treatment. Size selection and clean-up were performed using diluted AMPure PB beads (Pacific Biosciences). DNA concentration was quantified using a Qubit Fluorometer v4.0 (ThermoFisher Scientific) and the Qubit 1X dsDNA HS assay kit. Final library fragment size was assessed with the Agilent Femto Pulse

Automated Pulsed Field CE Instrument (Agilent Technologies) using the gDNA 55 kb BAC analysis kit.

The sample was sequenced on a Revio instrument (Pacific Biosciences). The prepared library was normalised to 2 nM, and 15 µL was used for making complexes. Primers were annealed and polymerases bound to generate circularised complexes, following the manufacturer's instructions. Complexes were purified using 1.2X SMRTbell beads, then diluted to the Revio loading concentration (200–300 pM) and spiked with a Revio sequencing internal control. The sample was sequenced on a Revio 25M SMRT cell. The SMRT Link software (Pacific Biosciences), a web-based workflow manager, was used to configure and monitor the run and to carry out primary and secondary data analysis.

Specimen details, sequencing platforms, and data yields are summarised in Table 1.

Hi-C

Sample preparation and crosslinking

The Hi-C sample was prepared from 20–50 mg of frozen head tissue of the ilCloTita1 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.

Table 1. Specimen and sequencing data for BioProject PRJEB78776.

Platform	PacBio HiFi	Hi-C
ToLID	ilCloTita1	ilCloTita1
Specimen ID	SAN28000080	SAN28000080
BioSample (source individual)	SAMEA115109804	SAMEA115109804
BioSample (tissue)	SAMEA115109832	SAMEA115109830
Tissue	thorax	head
Sequencing platform and model	Revio	Illumina NovaSeq X
Run accessions	ERR13485734	ERR13493991
Read count total	1.76 million	717.90 million
Base count total	18.19 Gb	108.40 Gb

Hi-C library preparation and sequencing

Biotinylated DNA constructs were fragmented using a Covaris E220 sonicator and size selected to 400–600 bp using SPRISelect beads. DNA was enriched with Arima-HiC v2 kit Enrichment beads. End repair, A-tailing, and adapter ligation were carried out with the NEBNext Ultra II DNA Library Prep Kit (New England Biolabs), following a modified protocol where library preparation occurs while DNA remains bound to the Enrichment beads. Library amplification was performed using KAPA HiFi HotStart mix and a custom Unique Dual Index (UDI) barcode set (Integrated DNA Technologies). Depending on sample concentration and biotinylation percentage determined at the crosslinking stage, libraries were amplified with 10–16 PCR cycles. Post-PCR clean-up was performed with SPRISelect beads. Libraries were quantified using the AccuClear Ultra High Sensitivity dsDNA Standards Assay Kit (Biotium) and a FLUOstar Omega plate reader (BMG Labtech).

Prior to sequencing, libraries were normalised to 10 ng/µL. Normalised libraries were quantified again and equimolar and/or weighted 2.8 nM pools. Pool concentrations were checked using the Agilent 4200 TapeStation (Agilent) with High Sensitivity D500 reagents before sequencing. Sequencing was performed using paired-end 150 bp reads on the Illumina NovaSeq X.

Specimen details, sequencing platforms, and data yields are summarised in Table 1.

Genome assembly

Prior to assembly of the PacBio HiFi reads, a database of *k*-mer counts (k = 31) was generated from the filtered reads using FastK. GenomeScope2 (Ranallo-Benavidez *et al.*, 2020) was 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 5 breaks and 27 joins. The curation process is documented at https://gitlab.com/wtsi-grit/rapid-curation. PretextSnapshot was used to generate a Hi-C contact map of the final assembly.

Assembly quality assessment

Chromosomal painting was performed using lep_busco_painter using Merian elements, which represent the 32 ancestral linkage groups in Lepidoptera (Wright *et al.*, 2024). Painting was based on gene locations from the lepidoptera_odb10 BUSCO analysis and chromosome lengths from the genome index produced using SAMtools faidx (Danecek *et al.*, 2021). Each complete BUSCO (including both single-copy and duplicated BUSCOs) was assigned to a Merian element using a reference database, and coloured positions were plotted along chromosomes drawn to scale.

The Merqury.FK tool (Rhie *et al.*, 2020), run in a Singularity container (Kurtzer *et al.*, 2017), was used to evaluate *k*-mer completeness and assembly quality for both haplotypes using the *k*-mer databases (k = 31) computed prior to genome assembly. The analysis outputs included assembly QV scores and completeness statistics.

The genome was analysed using the BlobToolKit pipeline, a Nextflow implementation of the earlier Snakemake BlobToolKit pipeline (Challis *et al.*, 2020). The pipeline aligns PacBio reads using minimap2 (Li, 2018) and SAMtools (Danecek *et al.*, 2021) to generate coverage tracks. Simultaneously, it queries the GoaT database (Challis *et al.*, 2023) to identify relevant BUSCO lineages and runs BUSCO (Manni *et al.*, 2021). For the three domain-level BUSCO lineages, BUSCO genes are aligned to the UniProt Reference Proteomes database (Bateman *et al.*, 2023) using DIAMOND blastp (Buchfink *et al.*, 2021). The genome is divided into chunks based on the density of BUSCO genes from the closest taxonomic lineage, and each chunk is aligned to the UniProt Reference Proteomes database with DIAMOND blastx. Sequences without hits are chunked using seqtk and aligned to the NT database with blastn (Altschul *et al.*, 1990). The BlobToolKit suite consolidates all outputs into a blobdir for visualisation. The BlobToolKit pipeline was developed using nf-core tooling (Ewels *et al.*, 2020) and MultiQC (Ewels *et al.*, 2016), with package management via Conda and Bioconda (Grüning *et al.*, 2018), and containerisation through Docker (Merkel, 2014) and Singularity (Kurtzer *et al.*, 2017).

Genome sequence report

Sequence data

The genome of a specimen of *Boloria titania* was sequenced using Pacific Biosciences single-molecule HiFi long reads, generating 18.19 Gb (gigabases) from 1.76 million reads, which were used to assemble the genome. GenomeScope2.0 analysis estimated the haploid genome size at 475.66 Mb, with a heterozygosity of 1.25% and repeat content of 37.99%. These estimates guided expectations for the assembly. Based on the estimated genome size, the sequencing data provided approximately 37× coverage. Hi-C sequencing produced 108.40 Gb from 717.90 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 496.64 Mb in 87 scaffolds, with 89 gaps, and a scaffold N50 of 16.29 Mb (Table 2).

Most of the assembly sequence (97.13%) was assigned to 32 chromosomal-level scaffolds, representing 30 autosomes and the W and Z sex chromosomes. Chromosomes W and

Table 2. Genome assembly statistics.

Assembly name	ilCloTita1.hap1.1	ilCloTita1.hap2.1
Assembly accession	GCA_964270595.1	GCA_964270805.1
Assembly level	chromosome	scaffold
Span (Mb)	496.64	454.08
Number of chromosomes	32	N/A
Number of contigs	176	142
Contig N50	7.97 Mb	6.93 Mb
Number of scaffolds	87	70
Scaffold N50	16.29 Mb	16.34 Mb
Longest scaffold length (Mb)	23.89	N/A
Sex chromosomes	W and Z	N/A
Organelles	Mitochondrial genome: 15.17 kb	N/A

Z were assigned based on Hi-C signal. These chromosome-level scaffolds, confirmed by Hi-C data, are named according to size (Figure 2; Table 3). Chromosome painting with Merian elements illustrates the distribution of orthologues along chromosomes and highlights patterns of chromosomal evolution relative to Lepidopteran ancestral linkage groups (Figure 3).

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

Assembly quality metrics

For haplotype 1, the estimated QV is 61.9, and for haplotype 2, 62.1. When the two haplotypes are combined, the assembly achieves an estimated QV of 62.0. The *k*-mer completeness is 77.36% for haplotype 1, 73.39% for haplotype 2, and 99.46% for the combined haplotypes (Figure 4). BUSCO analysis using the lepidoptera_odb10 reference set (n = 5 286) (Kriventseva *et al.*, 2019) identified 98.7% of the expected gene set (single = 98.2%, duplicated = 0.5%) for haplotype 1. The snail plot in Figure 5 summarises the scaffold length distribution and other assembly statistics for haplotype 1. The blob plot in Figure 6 shows the distribution of scaffolds by GC proportion and coverage for haplotype 1.

Table 4 lists the assembly metric benchmarks adapted from Rhie *et al.* (2021) the Earth BioGenome Project Report on Assembly Standards September 2024. The EBP metric, calculated for the haplotype 1, is **6.C.Q61**, meeting the recommended reference standard.

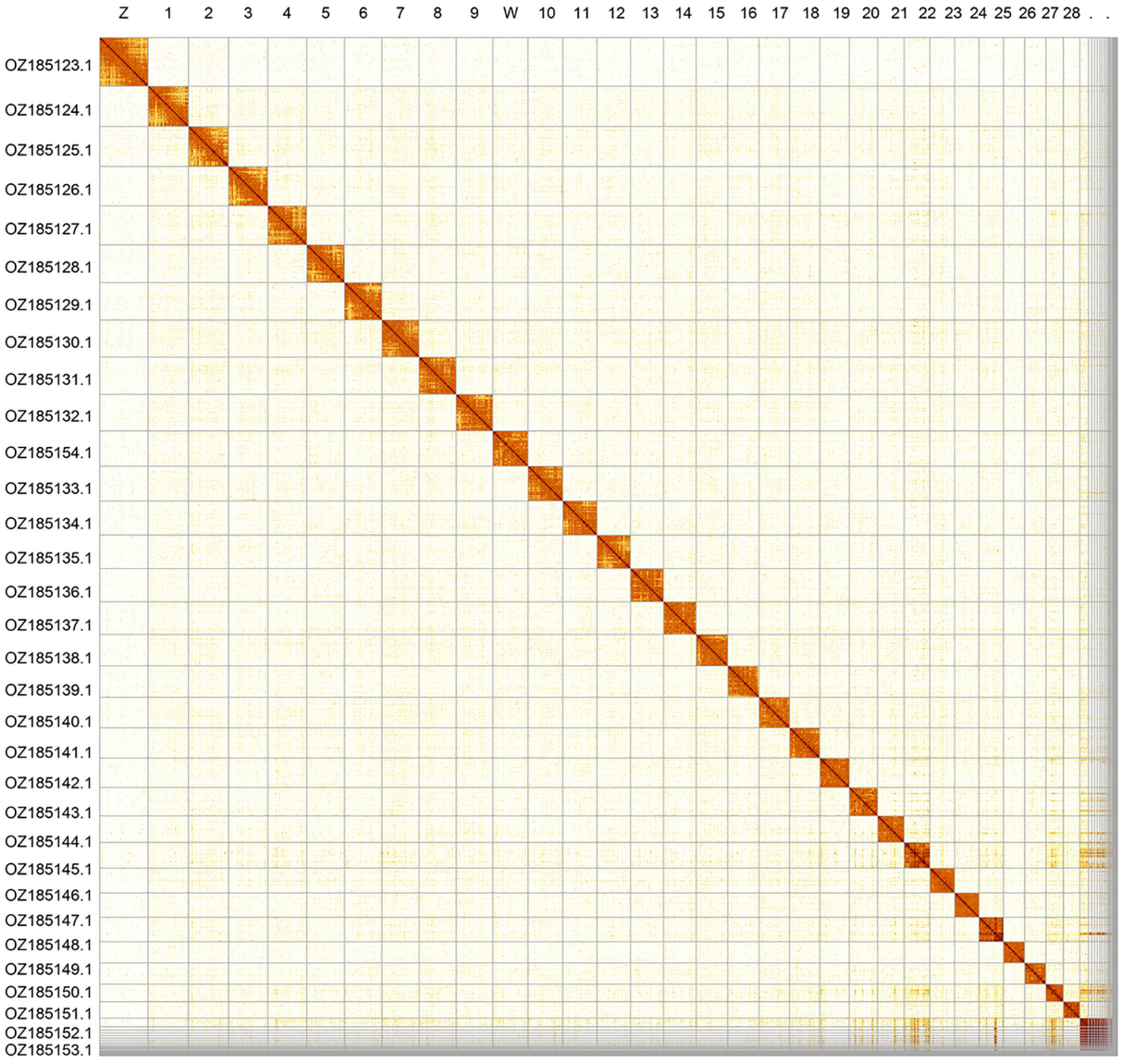

Figure 2. Hi-C contact map of the *Boloria titania* genome assembly. Assembled chromosomes are shown in order of size.

Table 3. Chromosomal pseudomolecules in the haplotype 1 genome assembly of *Boloria titania* ilCloTita1.

INSDC accession	Molecule	Length (Mb)	GC%	Assigned Merian elements
OZ185124.1	1	19.63	33	M2
OZ185125.1	2	19.53	33.50	M17;M20
OZ185126.1	3	19.08	33.50	M1
OZ185127.1	4	19.01	33.50	M3
OZ185128.1	5	18.45	33	M8
OZ185129.1	6	18.27	33	M9
OZ185130.1	7	18.11	33.50	M7
OZ185131.1	8	18.06	33.50	M5
OZ185132.1	9	17.77	33	M12
OZ185133.1	10	17.31	33	M18
OZ185134.1	11	16.84	33.50	M6
OZ185135.1	12	16.78	33	M16
OZ185136.1	13	16.29	33	M4
OZ185137.1	14	16.10	33	M21
OZ185138.1	15	15.92	33.50	M22
OZ185139.1	16	15.35	33.50	M10
OZ185140.1	17	15.32	33	M15
OZ185141.1	18	14.90	33.50	M11
OZ185142.1	19	14.68	33.50	M14
OZ185143.1	20	14.31	33.50	M13
OZ185144.1	21	13.87	33.50	M23
OZ185145.1	22	12.99	34	M19
OZ185146.1	23	12.47	36	M30
OZ185147.1	24	12.04	33.50	M24
OZ185148.1	25	11.93	33	M26
OZ185149.1	26	11.85	36	M25
OZ185150.1	27	10.40	33.50	M28
OZ185151.1	28	10.39	33	M27
OZ185152.1	29	8.55	35	M31
OZ185153.1	30	7.97	34.50	M29
OZ185154.1	W	17.76	36.50	N/A
OZ185123.1	Z	23.89	33	MZ

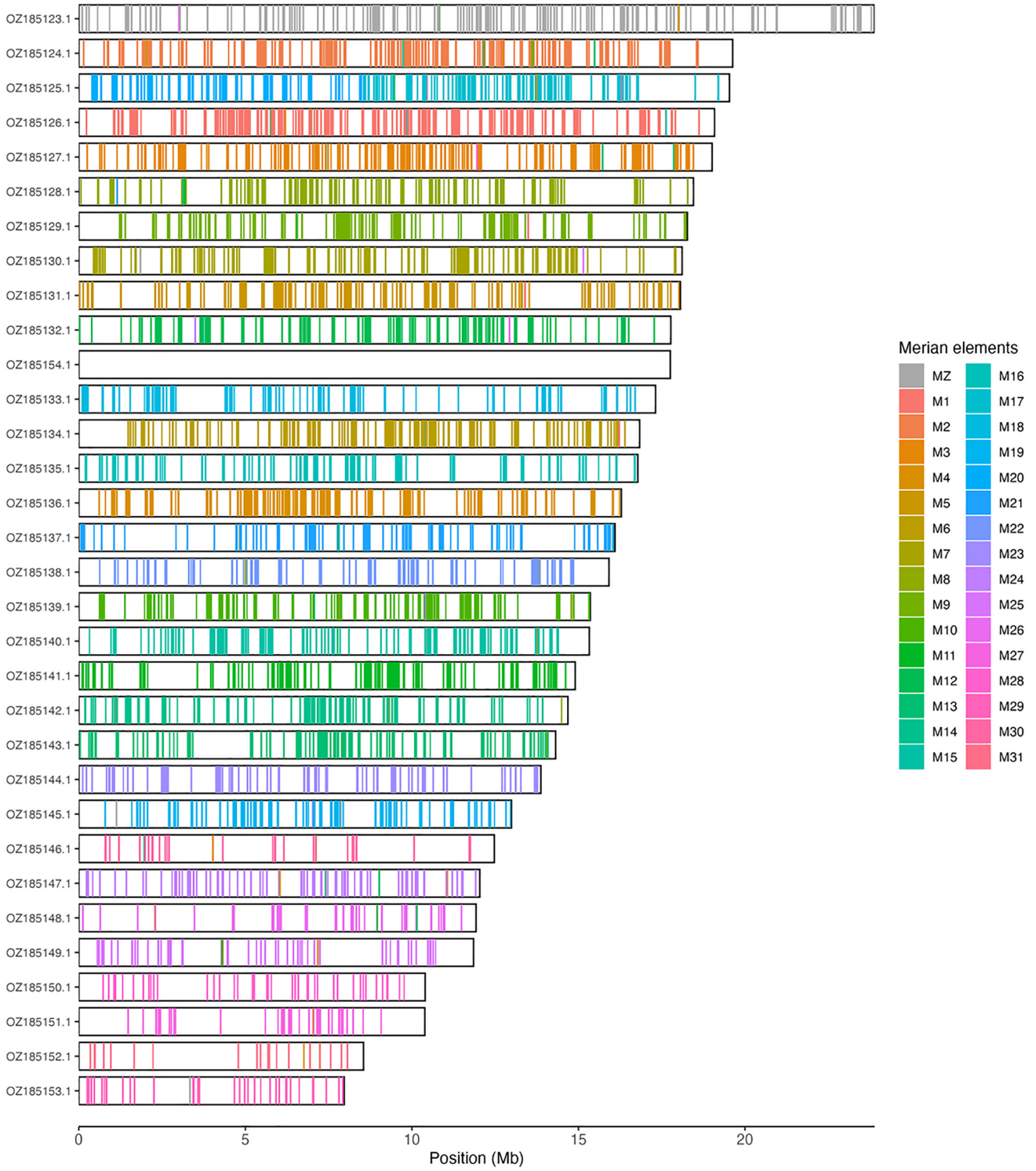

Figure 3. Merian elements painted across chromosomes in the ilCloTita1.hap1.1 assembly of *Boloria titania*. Chromosomes are drawn to scale, with the positions of orthologues shown as coloured bars. Each orthologue is coloured by the Merian element that it belongs to. All orthologues which could be assigned to Merian elements are shown.

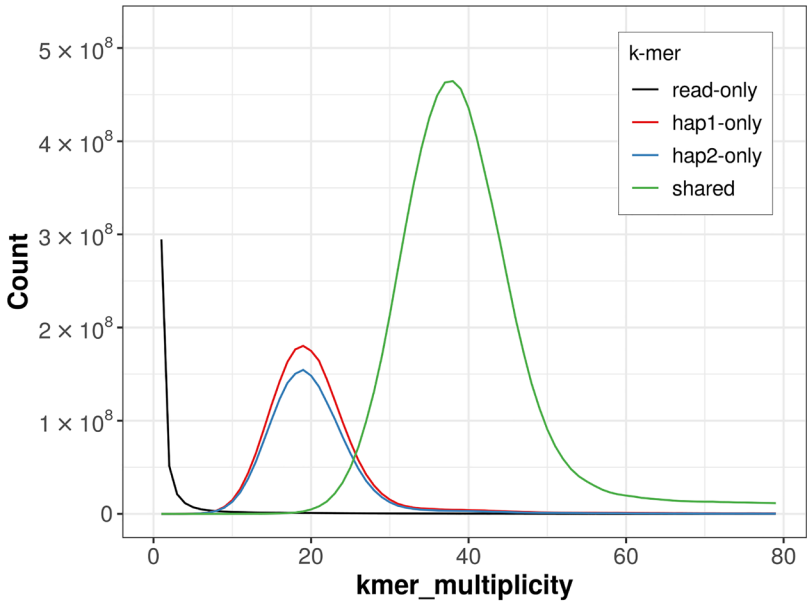

Figure 4. Evaluation of *k*-mer completeness using MerquryFK. This plot illustrates the recovery of *k*-mers from the original read data in the final assemblies. The horizontal axis represents *k*-mer multiplicity, and the vertical axis shows the number of *k*-mers. The black curve represents *k*-mers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

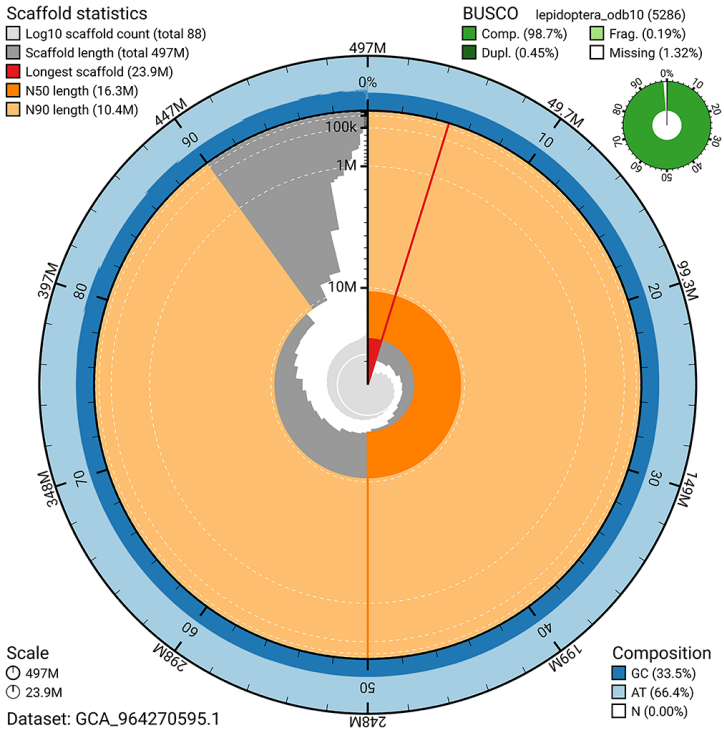

Figure 5. Assembly metrics for ilCloTita1.hap1.1. The BlobToolKit snail plot provides an overview of assembly metrics and BUSCO gene completeness. The circumference represents the length of the whole genome sequence, and the main plot is divided into 1,000 bins around the circumference. The outermost blue tracks display the distribution of GC, AT, and N percentages across the bins. Scaffolds are arranged clockwise from longest to shortest and are depicted in dark grey. The longest scaffold is indicated by the red arc, and the deeper orange and pale orange arcs represent the N50 and N90 lengths. A light grey spiral at the centre shows the cumulative scaffold count on a logarithmic scale. A summary of complete, fragmented, duplicated, and missing BUSCO genes in the set is presented at the top right. An interactive version of this figure can be accessed on the BlobToolKit viewer.

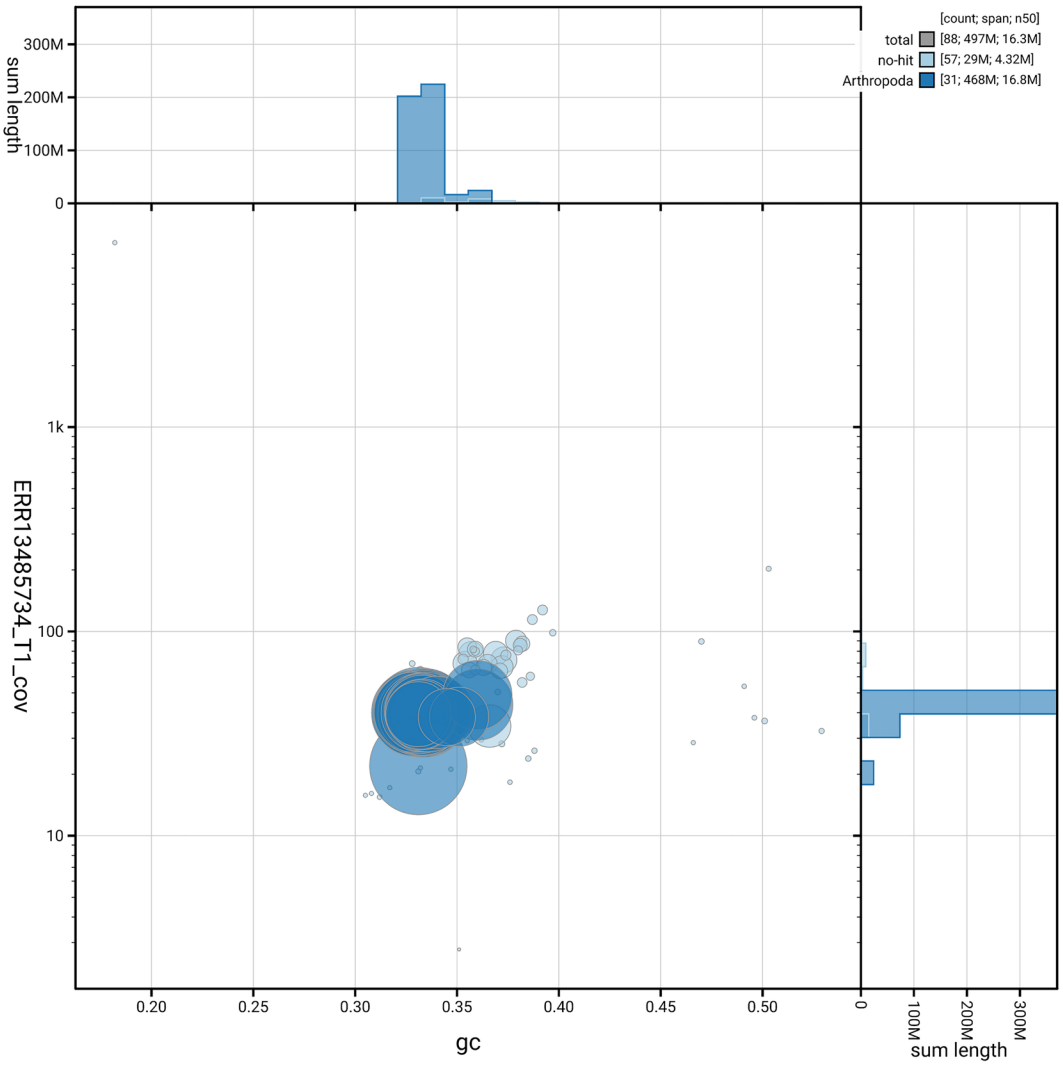

Figure 6. BlobToolKit GC-coverage plot for ilCloTita1.hap1.1. Blob plot showing sequence coverage (vertical axis) and GC content (horizontal axis). The circles represent scaffolds, with the size proportional to scaffold length and the colour representing phylum membership. The histograms along the axes 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 *Boloria titania* assembly.

Measure	Value	Benchmark
EBP summary (haplotype 1)	6.7.Q61	6.C.Q40
Contig N50 length	7.97 Mb	≥ 1 Mb
Scaffold N50 length	16.29 Mb	= chromosome N50
Consensus quality (QV)	Haplotype 1: 61.9; haplotype 2: 62.1; combined: 62.0	≥ 40
k-mer completeness	Haplotype 1: 77.36%; Haplotype 2: 73.39%; combined: 99.46%	≥ 95%
BUSCO	C:98.7%[S:98.2%,D:0.5%], F:0.2%,M:1.1%,n:5 286	S > 90%; D < 5%
Percentage of assembly assigned to chromosomes	97.13%	≥ 90%

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: Clossiana titania (Titania's fritillary). Accession number PRJEB78776. The genome sequence is released openly for reuse. The *Boloria titania* genome sequencing initiative is part of the Sanger Institute Tree of Life Programme (PRJEB43745) and Project Psyche (PRJEB71705). All raw sequence data and the assembly have been deposited in INSDC databases. The genome will be annotated using available RNA-Seq data and presented through Ensembl at the European Bioinformatics Institute. Raw data and assembly accession identifiers are reported in Table 1 and Table 2.

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

Author information

Contributors are listed at the following links:

- Wellcome Sanger Institute Tree of Life Management, Samples and Laboratory team
- Wellcome Sanger Institute Scientific Operations – Sequencing Operations
- Wellcome Sanger Institute Tree of Life Core Informatics team
- Tree of Life Core Informatics collective
- Project Psyche Community.

Table 5. Software versions and sources.

Software	Version	Source
BEDTools	2.30.0	https://github.com/arq5x/bedtools2
BLAST	2.14.0	ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/
BlobToolKit	4.3.9	https://github.com/blobtoolkit/blobtoolkit
BUSCO	5.5.0	https://gitlab.com/ezlab/busco
bwa-mem2	2.2.1	https://github.com/bwa-mem2/bwa-mem2
Cooler	0.8.11	https://github.com/open2c/cooler
DIAMOND	2.1.8	https://github.com/bbuchfink/diamond
fasta_windows	0.2.4	https://github.com/tolkit/fasta_windows
FastK	1.1	https://github.com/thegenemyers/FASTK
GenomeScope2.0	2.0.1	https://github.com/tbenavi1/genomescope2.0
Gfastats	1.3.6	https://github.com/vgl-hub/gfastats
GoaT CLI	0.2.5	https://github.com/genomehubs/goat-cli
Hifiasm	0.19.8-r603	https://github.com/chhylp123/hifiasm
HiGlass	1.13.4	https://github.com/higlass/higlass
lep_busco_painter	1.0.0	https://github.com/charlottewright/lep_busco_painter

Software	Version	Source
MerquryFK	1.1.1	https://github.com/thegenemyers/MERQURY.FK
Minimap2	2.24-r1122	https://github.com/lh3/minimap2
MitoHiFi	3	https://github.com/marcelauliano/MitoHiFi
MultiQC	1.14, 1.17, and 1.18	https://github.com/MultiQC/MultiQC
Nextflow	23.10.0	https://github.com/nextflow-io/nextflow
PretextSnapshot	N/A	https://github.com/sanger-tol/PretextSnapshot
PretextView	0.2.5	https://github.com/sanger-tol/PretextView
samtools	1.19.2	https://github.com/samtools/samtools
sanger-tol/ascc	0.1.0	https://github.com/sanger-tol/ascc
sanger-tol/blobtoolkit	0.6.0	https://github.com/sanger-tol/blobtoolkit
Seqtk	1.3	https://github.com/lh3/seqtk
Singularity	3.9.0	https://github.com/sylabs/singularity
TreeVal	1.2.0	https://github.com/sanger-tol/treeval
YaHS	1.2a.2	https://github.com/c-zhou/yahs

References

Allio R, Schomaker-Bastos A, Romiguier J, *et al.*: **MitoFinder: efficient automated large-scale extraction of mitogenomic data in target enrichment phylogenomics.** *Mol Ecol Resour.* 2020; **20**(4): 892–905.
PubMed Abstract | Publisher Full Text | Free Full Text

Altschul SF, Gish W, Miller W, *et al.*: **Basic Local Alignment Search Tool.** *J Mol Biol.* 1990; **215**(3): 403–410.
PubMed Abstract | Publisher Full Text

Bateman A, Martin MJ, Orchard S, *et al.*: **UniProt: the Universal Protein Knowledgebase in 2023.** *Nucleic Acids Res.* 2023; **51**(D1): D523–D531.
PubMed Abstract | Publisher Full Text | Free Full Text

Buchfink B, Reuter K, Drost HG: **Sensitive protein alignments at Tree-of-Life scale using DIAMOND.** *Nat Methods.* 2021; **18**(4): 366–368.
PubMed Abstract | Publisher Full Text | Free Full Text

Challis R, Kumar S, Sotero-Caio C, *et al.*: **Genomes on a Tree (GoaT): a versatile, scalable search engine for genomic and sequencing project metadata across the eukaryotic Tree of Life [version 1; peer review: 2 approved].** *Wellcome Open Res.* 2023; **8**: 24.
PubMed Abstract | Publisher Full Text | Free Full Text

Challis R, Richards E, Rajan J, *et al.*: **BlobToolKit – interactive quality assessment of genome assemblies.** *G3 (Bethesda).* 2020; **10**(4): 1361–1374.
PubMed Abstract | Publisher Full Text | Free Full Text

Cheng H, Concepcion GT, Feng X, *et al.*: **Haplotype-resolved *de novo* assembly using phased assembly graphs with hifiasm.** *Nat Methods.* 2021; **18**(2): 170–175.
PubMed Abstract | Publisher Full Text | Free Full Text

Cheng H, Jarvis ED, Fedrigo O, *et al.*: **Haplotype-resolved assembly of diploid genomes without parental data.** *Nat Biotechnol.* 2022; **40**(9): 1332–1335.
PubMed Abstract | Publisher Full Text | Free Full Text

Cozzi G, Müller CB, Krauss J: **How do local habitat management and landscape structure at different spatial scales affect fritillary butterfly distribution on fragmented wetlands?** *Landscape Ecology.* 2008; **23**: 269–83.
Publisher Full Text

Cuvelier S, Dincă V: **New data regarding the butterflies (Lepidoptera: Rhopalocera) of Romania, with additional comments (general distribution in Romania, habitat preferences, threats and protection) for ten localized Romanian species.** *Phegea.* 2007; **35**(3): 93–115.
Reference Source

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

Ebert G: **Die Schmetterlinge Baden-Württembergs Band 10 Ergänzungsband.** Stuttgart, Germany: Eugen Ulmer Verlag, 2005.
Reference Source

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

Glowacinski Z, Makomaska-Juchiewicz M, Polczynska-Konior G: **Red List of Threatened Animals in Poland.** Kraków: Wydawnictwo Instytutu Ochrony Przyrody PAN, 2002.
Reference Source

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

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

Howe K, Chow W, Collins J, *et al.*: **Significantly improving the quality of genome assemblies through curation.** *GigaScience.* 2021; **10**(1): giaa153.
PubMed Abstract | Publisher Full Text | Free Full Text

Kerpedjiev P, Abdennur N, Lekschas F, *et al.*: **HiGlass: web-based visual exploration and analysis of genome interaction maps.** *Genome Biol.* 2018; **19**(1): 125.
PubMed Abstract | Publisher Full Text | Free Full Text

Kriventseva EV, Kuznetsov D, Tegenfeldt F, *et al.*: **OrthoDB v10: sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for**

evolutionary and functional annotations of orthologs. *Nucleic Acids Res.* 2019; **47**(D1): D807–D811.
PubMed Abstract | Publisher Full Text | Free Full Text

Kurtzer GM, Sochat V, Bauer MW: **Singularity: scientific containers for mobility of compute.** *PLoS One.* 2017; **12**(5): e0177459.
PubMed Abstract | Publisher Full Text | Free Full Text

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

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

McFarland KP: **Conservation assessment of two endemic butterflies (White Mountain Arctic, *Oeneis melissa semidea*, and White Mountain Fritillary, *Boloria titania montinus*) in the Presidential Range alpine zone White Mountains, New Hampshire.** 2003.
Publisher Full Text

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

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

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

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

Schweiger O, Settele J, Kudrna O, *et al.*: **Climate change can cause spatial mismatch of trophically interacting species.** *Ecology.* 2008; **89**(12): 3472–79.
PubMed Abstract | Publisher Full Text

Simonsen TJ, Wahlberg N, Warren AD, *et al.*: **The evolutionary history of *Boloria* (Lepidoptera: Nymphalidae): Phylogeny, zoogeography and larval-foodplant relationships.** *Syst Biodivers.* 2010; **8**(4): 513–29.
Publisher Full Text

Singer A, Schweiger O, Kühn I, *et al.*: **Constructing a hybrid species distribution model from standard large-scale distribution data.** *Ecol Modell.* 2018; **373**: 39–52.
Publisher Full Text

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

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

Van Swaay C, Cuttelod A, Collins S, *et al.*: **European Red List of Butterflies.** Luxembourg: Publications Office of the European Union, 2010.
Reference Source

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: **Boloria titania.** 2005.

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 28 February 2026

https://doi.org/10.21956/wellcomeopenres.27157.r135193

✔ **Annabel Whibley**
Bragato Research Institute, Lincoln, New Zealand

Kleckova and colleagues report the genome assembly of Titania's Fritillary (*Boloria titania*). Two haplotypes are reported, with haplotype 1 scaffolded into chromosomal pseudomolecules. Sequenced under the auspices of Project Psyche, the report follows standard genome note templating and presents the background natural history and taxonomy of the species, extensive sample metadata, comprehensive methods, clearly described and publicly available genomic resources and extensive evaluation of the quality of these resources. The reported assembly metrics are excellent in all respects (completeness, contiguity, accuracy, contamination scans, chromosome painting), formally reaching 6.C.61 for the scaffolded haplotype 1.

I note that specimen identification was not confirmed by barcoding and that a genome annotation has not been released at this time.

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, Genomics

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

Reviewer Report 25 October 2025

https://doi.org/10.21956/wellcomeopenres.27157.r135190

Arun Arumugaperumal
Department of Biotechnology, Rajalakshmi Engineering College, Thandalam 602105, Chennai, Tamilnadu, India

The genome note describes the whole genome sequence of *Boloria titania*, also known as the Titania's Fritillary. The authors have used standard Darwin's tree of life protocols to sequence the genome. The assembly reported here is of size 496.64 Mb spread among 32 chromosomes. A female specimen has been used and the Z and W sex chromosomes have been captured. The mitochondrial genome of size 15.17 kb has also been sequenced. But the annotation results are not available as Ensemble release has not happened yet. The assembly has confirmed its quality aligned with the earth biogenome project quality standards. BUSCO completeness of 98.7% indiccates that the genome is almost complete with reference to the sequences available in the databases.

The EBP summary metric has been mentioned as 6.C.Q61 in text. But in Table 4 it has been mentioned as 6.7.Q61. Kindly correct this mistake. Then the article can be indexed.

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; Genomics

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

Reviewer Report 25 October 2025

https://doi.org/10.21956/wellcomeopenres.27157.r135187

Aqeel Alyousuf
University of Basrah, Basrah, Iraq

The **Methods** section is technically comprehensive and well-documented, covering all major stages of the workflow, including sample acquisition, DNA extraction, library preparation, sequencing, assembly, curation, and quality assessment. The application of the Wellcome Sanger Institute (WSI) protocols, together with PacBio HiFi and Arima Hi-C technologies, reflects a high technical standard. The inclusion of detailed information on fragment size, concentration, and purity is also commendable.

Minor Issues:

- In the **Species taxonomy** paragraph, *Clossiana* is generally treated as a subgenus or synonym within *Boloria* in modern classifications. Listing "*Boloria; Clossiana; Boloria titania*" without clarification is ambiguous and could be interpreted as a taxonomic error. Please revise to a consistent format, such as *Boloria (Clossiana)* or *Boloria* (syn. *Clossiana*).
- Please correct the scientific name (*Boloria titania*) in **Figure 1** and in the **Keywords** section.
- The **reference** "Wagner W: *Boloria titania*. 2005." is incomplete and does not conform to standard citation guidelines. It should include the full source details, such as database name, URL, and access date.

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.