DATA NOTE

# The genome sequence of *Glacies alpinata* (Scopoli, 1763) (Lepidoptera: Geometridae)

[version 1; peer review: 2 approved]

Paula Escuer[1], Kay Lucek[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]University of Neuchâtel, Neuchâtel, Switzerland
[2]Tree of Life, Wellcome Sanger Institute, Hinxton, England, UK

**v1** **First published:** 30 Jul 2025, **10**:392
https://doi.org/10.12688/wellcomeopenres.24664.1
**Latest published:** 30 Jul 2025, **10**:392
https://doi.org/10.12688/wellcomeopenres.24664.1

## Abstract

We present a genome assembly from a male specimen of *Glacies alpinata* (Arthropoda; Insecta; Lepidoptera; Geometridae). The assembly contains two haplotypes with total lengths of 575.96 megabases and 573.82 megabases. Most of haplotype 1 (99.77%) 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 17.2 kilobases.

## Keywords

Glacies alpinata, genome sequence, chromosomal, Lepidoptera

This article is included in the Tree of Life gateway.

**Open Peer Review**

**Approval Status** ✓ ✓

| | 1 | 2 |
|---|---|---|
| **version 1** 30 Jul 2025 | ✓ view | ✓ view |

1. **Annabel Whibley**, Bragato Research Institute, Lincoln, New Zealand

2. **Arun Arumugaperumal**, Rajalakshmi Engineering College, Thandalam ,Chennai, India

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: Escuer P**: Investigation, Resources, Writing – Original Draft Preparation; **Lucek K**: 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). PE and KL were supported by a grant (PCEFP3_202869) from the Swiss National Science Foundation.
*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:** Escuer P, Lucek K, Wright CJ *et al.* **The genome sequence of *Glacies alpinata* (Scopoli, 1763) (Lepidoptera: Geometridae) [version 1; peer review: 2 approved]** Wellcome Open Research 2025, **10**:392
https://doi.org/10.12688/wellcomeopenres.24664.1

**First published:** 30 Jul 2025, **10**:392 https://doi.org/10.12688/wellcomeopenres.24664.1

## 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; Ennominae; *Glacies*; *Glacies alpinata* (Scopoli, 1763) (NCBI:txid206174)

## Background

*Glacies alpinata* is a moth of the Geometridae family. The species can be found in the Alps and across Central-Eastern Europe in countries such as Czech Republic, Poland, Romania and Slovakia in altitudes up to 3 000m (Forster & Wohlfahrt, 1981). Their habitats are alpine flower-rich meadows with rocky edges (Scholley-Pfab, 2025).

They have a wingspan of 18–24 mm and exhibit dark brown colouration with some greyish tones in the wings, with subtle transverse lines. The hindwings have a barely discernible pattern with black stripes along the outer edge of the wings. They are active during the day and fly close to the ground and fly from June to September in one generation. The caterpillars feed on leaves of low plants, where they overwinter (Müller *et al.*, 2019).

This species is not included in the Red List of Endangered Species. In Germany, it is classified as Least Concern for its high abundance in the Alpine regions (Federal Agency for Nature Conservation, 1998).

The genome of *Glacies alpinata* presented here will be an important resource for deep exploration into the genomic architecture of this understudied species, shedding light on evolution of the Geometridae family. The sequence data were derived from a male specimen (Figure 1) collected from Grindelwald, Switzerland.

## Methods

### Sample acquisition

The specimen used for genome sequencing was an adult male *Glacies alpinata* (specimen ID SAN28000073, ToLID ilGlaAlpi1; Figure 1), collected from Grindelwald, Switzerland (latitude 46.6576, longitude 8.0999; elevation 1 975 m) on 07/07/2023. The specimen was collected and identified by Paula Escuer (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 ilGlaAlpi1 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


**Figure 1. Voucher photograph of the *Glacies alpinata* (ilGlaAlpi1) specimen used for genome sequencing.**

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 27.9 ng/μL and a yield of 1 311.30 ng, with a fragment size of 16.2 kb. The 260/280 spectrophotometric ratio was 1.96, and the 260/230 ratio was 2.64.

### 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 protocol (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 ilGlaAlpi1 sample using the Arima-HiC v2 kit (Arima Genomics). Following the manufacturer's instructions, tissue was fixed and DNA crosslinked using TC buffer to a final formaldehyde concentration of 2%. The tissue was homogenised using the Diagnocine Power Masher-II. Crosslinked DNA was digested with a restriction enzyme master mix, biotinylated, and ligated. Clean-up was performed with SPRISelect beads before library preparation. DNA concentration was measured with the Qubit Fluorometer (Thermo Fisher Scientific) and Qubit HS Assay Kit. The biotinylation percentage was estimated using the Arima-HiC v2 QC beads.

***Hi-C library preparation and sequencing***

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

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

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

**Table 1. Specimen and sequencing data for BioProject PRJEB79180.**

| Platform | PacBio HiFi | Hi-C |
|---|---|---|
| **ToLID** | ilGlaAlpi1 | ilGlaAlpi1 |
| **Specimen ID** | SAN28000073 | SAN28000073 |
| **BioSample (source individual)** | SAMEA115109805 | SAMEA115109805 |
| **BioSample (tissue)** | SAMEA115109834 | SAMEA115109833 |
| **Tissue** | thorax | head |
| **Sequencing platform and model** | Revio | Illumina NovaSeq X |
| **Run accessions** | ERR13605503 | ERR13602166 |
| **Read count total** | 1.51 million | 718.24 million |
| **Base count total** | 19.21 Gb | 108.45 Gb |

### Genome assembly

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

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

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

### Assembly curation

The assembly was decontaminated using the Assembly Screen for Cobionts and Contaminants (ASCC) pipeline. TreeVal was used to generate the flat files and maps for use in curation. Manual curation was conducted primarily in PretextView and HiGlass (Kerpedjiev *et al.*, 2018). Scaffolds were visually inspected and corrected as described by Howe *et al.* (2021). Manual corrections included 12 breaks and 24 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 *Glacies alpinata* was sequenced using Pacific Biosciences single-molecule HiFi long reads, generating 19.21 Gb (gigabases) from 1.51 million reads, which were used to assemble the genome. GenomeScope2.0 analysis estimated the haploid genome size at 562.92 Mb, with a heterozygosity of 1.00% and repeat content of 33.28%. These estimates guided expectations for the assembly. Based on the estimated genome size, the sequencing data provided approximately 33× coverage. Hi-C sequencing produced 108.45 Gb from 718.24 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 575.96 Mb in 63 scaffolds, with 103 gaps, and a scaffold N50 of 20.79 Mb (Table 2).

Most of the assembly sequence (99.77%) was assigned to 31 chromosomal-level scaffolds, representing 30 autosomes and the Z sex chromosome. Chromosome Z was assigned based on synteny to the genome of *Idaea aversata* (GCA_907269075.1) (Boyes *et al.*, 2023). 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 66.2, and for haplotype 2, 66.5. When the two haplotypes are combined, the assembly achieves an estimated QV of 66.3. The *k*-mer completeness is 78.93% for haplotype 1, 78.87% for haplotype 2, and 99.19% for the combined haplotypes (Figure 4). BUSCO analysis using the lepidoptera_odb10 reference set ($n$ = 5 286) (Kriventseva *et al.*, 2019) identified 98.6% of the expected gene set (single = 97.9%, duplicated = 0.6%) 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.Q66**, meeting the recommended reference standard.

**Table 2. Genome assembly statistics.**

| | | |
|---|---|---|
| **Assembly name** | ilGlaAlpi1.hap1.1 | ilGlaAlpi1.hap2.1 |
| **Assembly accession** | GCA_964264065.1 | GCA_964263945.1 |
| **Assembly level** | chromosome | scaffold |
| **Span (Mb)** | 575.96 | 573.82 |
| **Number of chromosomes** | 31 | N/A |
| **Number of contigs** | 166 | 158 |
| **Contig N50** | 9.01 Mb | 9.45 Mb |
| **Number of scaffolds** | 63 | 72 |
| **Scaffold N50** | 20.79 Mb | 20.72 Mb |
| **Longest scaffold length (Mb)** | 29.41 | N/A |
| **Sex chromosomes** | Z | N/A |
| **Organelles** | Mitochondrial genome: 17.2 kb | N/A |

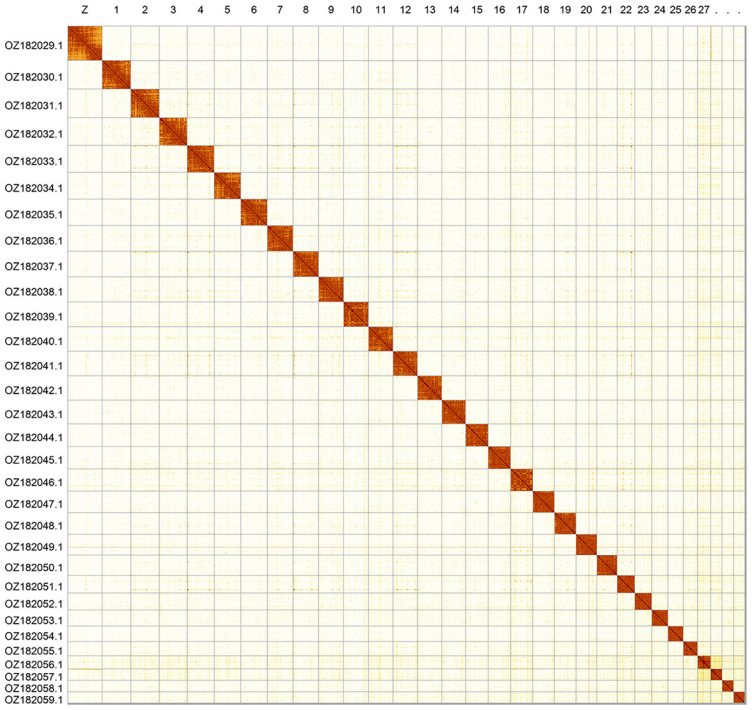


**Figure 2. Hi-C contact map of the *Glacies alpinata* genome assembly.** Assembled chromosomes are shown in order of size and labelled along the axes. The plot was generated using PretextSnapshot.

**Table 3. Chromosomal pseudomolecules in the haplotype 1 genome assembly of *Glacies alpinata* ilGlaAlpi1.**

| INSDC accession | Molecule | Length (Mb) | GC% | Assigned Merian elements |
|---|---|---|---|---|
| OZ182030.1 | 1 | 24.21 | 37 | M1 |
| OZ182031.1 | 2 | 24.19 | 37 | M2 |
| OZ182032.1 | 3 | 23.78 | 37 | M17;M20 |
| OZ182033.1 | 4 | 22.80 | 37 | M8 |
| OZ182034.1 | 5 | 22.56 | 37 | M3 |
| OZ182035.1 | 6 | 22.49 | 36.50 | M9 |
| OZ182036.1 | 7 | 21.83 | 37 | M5 |
| OZ182037.1 | 8 | 21.70 | 37 | M12 |
| OZ182038.1 | 9 | 21.28 | 37 | M7 |
| OZ182039.1 | 10 | 21.08 | 36.50 | M18 |
| OZ182040.1 | 11 | 20.80 | 37 | M4 |
| OZ182041.1 | 12 | 20.79 | 37 | M6 |
| OZ182042.1 | 13 | 20.60 | 37 | M16 |
| OZ182043.1 | 14 | 20.17 | 37 | M21 |
| OZ182044.1 | 15 | 19.21 | 37 | M22 |
| OZ182045.1 | 16 | 19.05 | 37 | M10 |
| OZ182046.1 | 17 | 18.73 | 37.50 | M23 |
| OZ182047.1 | 18 | 18.32 | 37 | M15 |
| OZ182048.1 | 19 | 18.32 | 37.50 | M11 |
| OZ182049.1 | 20 | 17.74 | 37 | M13 |
| OZ182050.1 | 21 | 17.20 | 37 | M14 |
| OZ182051.1 | 22 | 14.90 | 37.50 | M19 |
| OZ182052.1 | 23 | 14.44 | 37 | M24 |
| OZ182053.1 | 24 | 13.72 | 37 | M26 |
| OZ182054.1 | 25 | 13.01 | 37 | M28 |
| OZ182055.1 | 26 | 12.12 | 37 | M27 |
| OZ182056.1 | 27 | 11.01 | 39 | M30 |
| OZ182057.1 | 28 | 9.80 | 38.50 | M31 |
| OZ182058.1 | 29 | 9.70 | 37.50 | M25 |
| OZ182059.1 | 30 | 9.65 | 37.50 | M29 |
| OZ182029.1 | Z | 29.41 | 36.50 | MZ |
| OZ182060.1 | MT | 0.02 | 18 | N/A |

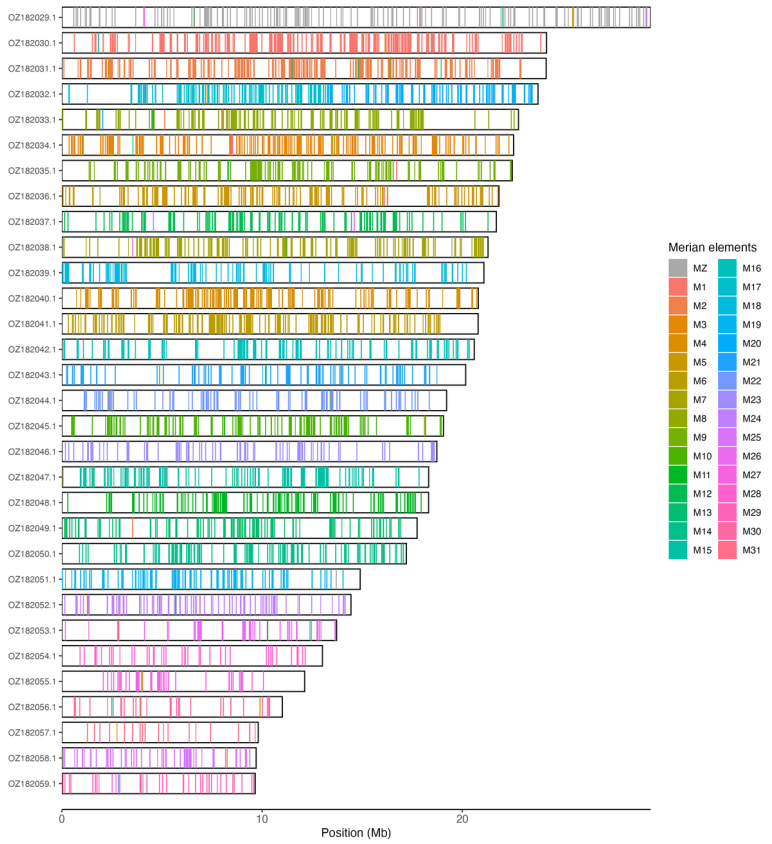


**Figure 3. Merian elements painted across chromosomes in the ilGlaAlpi1.hap1.1 assembly of *Glacies alpinata*.** Chromosomes are drawn to scale, with the positions of orthologues shown as coloured bars. Each orthologue is coloured by the Merian element that it belongs to. All orthologues which could be assigned to Merian elements are shown.

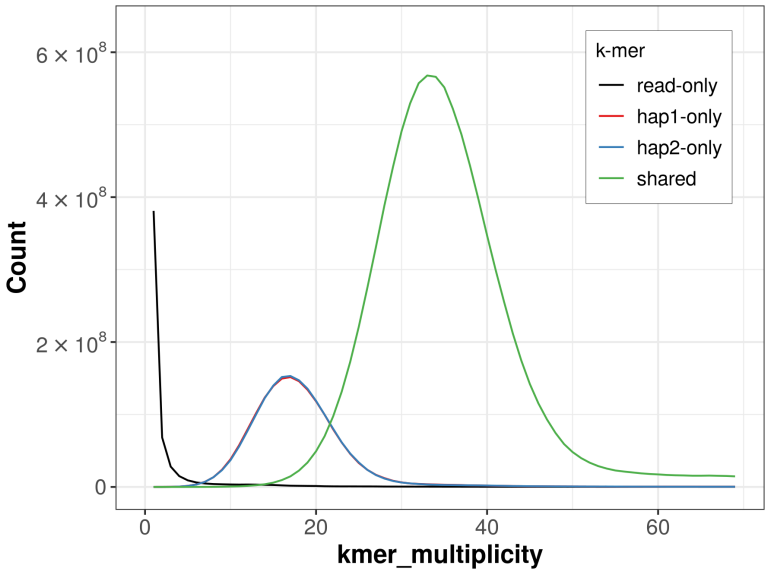


**Figure 4. Evaluation of *k*-mer completeness using MerquryFK.** This plot illustrates the recovery of *k*-mers from the original read data in the final assemblies. The horizontal axis represents *k*-mer multiplicity, and the vertical axis shows the number of *k*-mers. The black curve represents *k*-mers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

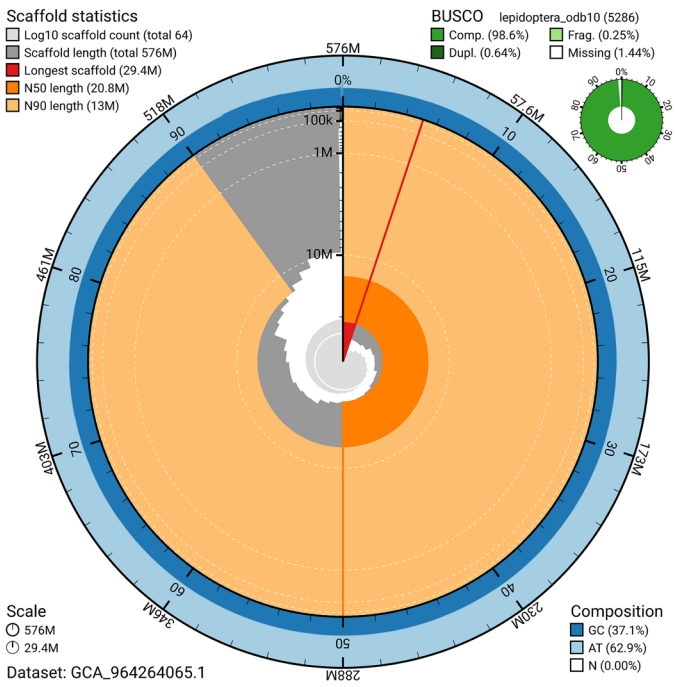


**Figure 5. Assembly metrics for ilGlaAlpi1.hap1.1.** The BlobToolKit snail plot provides an overview of assembly metrics and BUSCO gene completeness. The circumference represents the length of the whole genome sequence, and the main plot is divided into 1,000 bins around the circumference. The outermost 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.

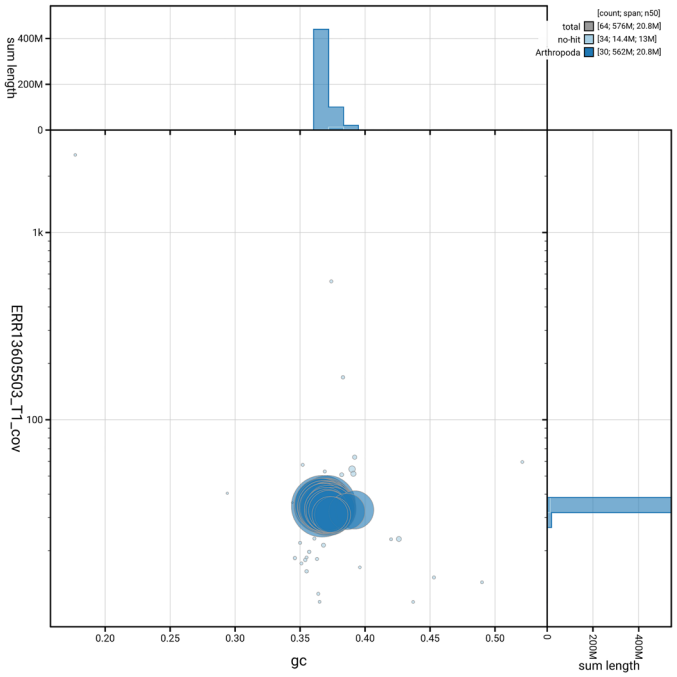


**Figure 6. BlobToolKit GC-coverage plot for ilGlaAlpi1.hap1.1.** Blob plot showing sequence coverage (vertical axis) and GC content (horizontal axis). The circles represent scaffolds, with the size proportional to scaffold length and the colour representing phylum membership. The histograms along the 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 *Glacies alpinata* assembly.**

| Measure | Value | Benchmark |
|---|---|---|
| EBP summary (haplotype 1) | 6.C.Q66 | 6.C.Q40 |
| Contig N50 length | 9.01 Mb | ≥ 1 Mb |
| Scaffold N50 length | 20.79 Mb | = chromosome N50 |
| Consensus quality (QV) | Haplotype 1: 66.2; haplotype 2: 66.5; combined: 66.3 | ≥ 40 |
| *k*-mer completeness | Haplotype 1: 78.93%; Haplotype 2: 78.87%; combined: 99.19% | ≥ 95% |
| BUSCO | C:98.6%[S:97.9%,D:0.6%], F:0.2%,M:1.2%,n:5 286 | S > 90%; D < 5% |
| Percentage of assembly assigned to chromosomes | 99.77% | ≥ 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: Glacies alpinata. Accession number PRJEB79180. The genome sequence is released openly for reuse. The *Glacies alpinata* 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 |

| Software | Version | Source |
|---|---|---|
| bwa-mem2 | 2.2.1 | https://github.com/bwa-mem2/bwa-mem2 |
| Cooler | 0.8.11 | https://github.com/open2c/cooler |
| DIAMOND | 2.1.8 | https://github.com/bbuchfink/diamond |
| fasta_windows | 0.2.4 | https://github.com/tolkit/fasta_windows |
| FastK | 1.1 | https://github.com/thegenemyers/FASTK |
| GenomeScope2.0 | 2.0.1 | https://github.com/tbenavi1/genomescope2.0 |
| Gfastats | 1.3.6 | https://github.com/vgl-hub/gfastats |
| GoaT CLI | 0.2.5 | https://github.com/genomehubs/goat-cli |
| Hifiasm | 0.19.8-r603 | https://github.com/chhylp123/hifiasm |
| HiGlass | 1.13.4 | https://github.com/higlass/higlass |
| lep_busco_painter | 1.0.0 | https://github.com/charlottewright/lep_busco_painter |
| MerquryFK | 1.1.1 | 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

Boyes D; University of Oxford and Wytham Woods Genome Acquisition Lab, Darwin Tree of Life Barcoding collective, *et al.*: **The genome sequence of the Riband Wave, *Idaea aversata* (Linnaeus, 1758) [version 1; peer review: 2 approved].** *Wellcome Open Res.* 2023; **8**: 45.
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**

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

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

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

Federal Agency for Nature Conservation: **Red List of Endangered Animals in Germany**. Landwirtschaftsverlag, Münster, 1998.

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**

Forster W, Wohlfahrt TA: ***Spanner (Geometridae)***. Franckh'sche Verlagshandlung, 1981.

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**

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

Müller B, Erlacher S, Hausmann A, *et al.*: **The Geometrid Moths of Europe. Volume 6. Subfamily Ennominae II.** Leiden; Boston: Brill; 2019.

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**

von Scholley-Pfab A: ***Psodos alpinata* (v9).** 2025.
**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**

Vasimuddin M, Misra S, Li H, *et al.*: **Efficient architecture-aware acceleration of BWA-MEM for multicore systems.** In: *2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS).* IEEE, 2019; 314–324.
**Publisher Full Text**

Wright CJ, Stevens L, Mackintosh A, *et al.*: **Comparative genomics reveals the dynamics of chromosome evolution in Lepidoptera.** *Nat Ecol Evol.* 2024; **8**(4): 777–790.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

Zhou C, McCarthy SA, Durbin R: **YaHS: Yet another Hi-C Scaffolding tool.** *Bioinformatics.* 2023; **39**(1): btac808.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

# Open Peer Review

## Current Peer Review Status: ✔ ✔

**Version 1**

Reviewer Report 08 September 2025

https://doi.org/10.21956/wellcomeopenres.27179.r128449



✔ **Arun Arumugaperumal**
Department of Biotechnology, Rajalakshmi Engineering College, Thandalam ,Chennai, Tamil Nadu, India

The authors have presented the genome sequence of *Glacies alpinata*, a moth species. This is the first report of the high-quality genome sequence of the species. The assembly reported is of size 575.96 Mb. The genome sequence was distributed among 31 chromosomes. They have also assembled the mitogenome which is of size 17.2 kb. The links to the databases are working correctly. The authors have used high-end sequencing facility to sequence the genome which is of very high quality. The BUSCO completeness score of 98.6% is also an indicator of near-completeness of the genome assembly. The article may be indexed. However, the following typing errors need to be corrected.

1. Change the typo in background section "in altitudes up to 3 000m"
2. In methods section, under 'Nucleic acid extraction' correct the typo "and a yield of 1 311.30 ng"

**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 02 September 2025

https://doi.org/10.21956/wellcomeopenres.27179.r128450



**Annabel Whibley**
Grapevine Improvement, Bragato Research Institute, Lincoln, Lincoln, New Zealand

In this Data Note, Escuer and colleagues present a reference genome assembly of the Geometrid moth *Glacies alpinata.* Combining Revio HiFi data and Arima Hi-C, the obtained chromosomal reference haplotype assembly is of very high quality (6.C.Q66) with excellent completeness metrics. The templated report includes extensive QC including assembly statistics, a contact map, chromosome painting with Merian elements, k-mer profiling and contamination scanning with Blobtoolkit. The report follows the standard template- methods are current and appropriate, the protocols and sample metadata are comprehensively documented in a well-structured report and public links to data accessions are functional.

I will register here a query I expressed in an earlier review of a similar Data Note: the sample processing methods section contains hyperlinks to protocols which have been given DOIs. I assume that a new DOI would be minted when the protocol version is updated, but I wonder whether it would be helpful to also note the protocol versions in the text itself.

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

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

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

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

***Competing Interests:*** No competing interests were disclosed.

***Reviewer Expertise:*** Genomics, Bioinformatics

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