Check for updates

DATA NOTE

ERGA-BGE genome of *Erebia palarica* Chapman, 1905: a montane butterfly endemic to North-West Spain

[version 1; peer review: 1 approved with reservations]

Marta Vila[1], Laura Torrado-Blanco[1], Nils Ryrholm[2], David Romero-Pedreira[1], María Conejero[3], Astrid Böhne[4], Rita Monteiro[4], Thomas Marcussen[5,6], Rebekah A. Oomen[5-9], Torsten H. Struck[5], Laura Aguilera[10,11], Marta Gut[10,11], Francisco Câmara Ferreira[10,11], Fernando Cruz[10,11], Jèssica Gómez-Garrido[10,11], Tyler S. Alioto[10,11], Leanne Haggerty[12], Fergal Martin[12], Chiara Bortoluzzi[13-15]

[1]Departamento de Bioloxía, Facultade de Ciencias, Zapateira, Universidade da Coruña, Grupo de Investigación en Bioloxía Evolutiva, A Coruña, 15071, Spain
[2]Department of Electronics, Mathematics and Natural Sciences, Faculty of Engineering and Sustainable Development, University of Gävle, Gävle, SE-80176, Sweden
[3]Department of Biodiversity and Evolutionary Biology, c/Jose Gutierrez Abascal 2, Museo Nacional de Ciencias Naturales (CSIC), Madrid, 28006, Spain
[4]Leibniz Institute for the Analysis of Biodiversity Change, Adenauerallee 127, Museum Koenig Bonn, Bonn, 53113, Germany
[5]Natural History Museum, P.O. Box 1172, Blindern, University of Oslo, Oslo, 0318, Norway
[6]Centre for Ecological & Evolutionary Synthesis, Blindernveien 31, University of Oslo, Oslo, 0371, Norway
[7]Department of Biological Sciences, 100 Tucker Park Road, University of New Brunswick Saint John, Saint John, E2K5E2, Canada
[8]Tjärnö Marine Laboratory, Hättebäcksvägen 7, University of Gothenburg, Gothenburg, 45296, Sweden
[9]Centre for Coastal Research, Universitetsveien 25, University of Agder, Kristiansand, 4630, Norway
[10]Centro Nacional de Análisis Genómico (CNAG), Baldiri Reixac 4, Barcelona, 08028, Spain
[11]Universitat de Barcelona (UB), 08028, Barcelona, Spain
[12]European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, Cambridge, CB10 1SD, UK
[13]Department of Biology, Via Madonna del Piano 6, University of Florence, Sesto Fiorentino, Florence,, 50019, Italy
[14]Department of Aquatic Ecology, Überlandstrasse 133, Eawag Swiss Federal Institute of Aquatic Science and Technology, Dübendorf, Zurich, 8600, Switzerland
[15]SIB Swiss Institute of Bioinformatics, Amphipôle, Quartier UNIL-Sorge, Lausanne, 1015, Switzerland

v1 **First published:** 06 Jan 2026, **6**:9
https://doi.org/10.12688/openreseurope.20717.1
Latest published: 06 Jan 2026, **6**:9
https://doi.org/10.12688/openreseurope.20717.1

Abstract

The reference genome of *Erebia palarica* (Chapman's ringlet; Lepidoptera: Nymphalidae, Satyrinae) provides valuable insights into evolutionary and conservation genomics. On one hand, it paves the way to unravel the speciation processes, reproductive barriers, and potential hybridisation between *E. palarica* and its closely related sister species, *Erebia meolans*. On the other hand, this genomic resource will

Open Peer Review

Approval Status ?

1

version 1 ? view
06 Jan 2026

1. **Yu-Feng Huang**, Yuan Ze University, Zhongli District, Taiwan

be key for the genetic monitoring of this narrowly endemic species, restricted to the Cantabrian Mountains (NW Spain), facilitating the use of genomic data to estimate population genetic parameters. The genome of a female individual was assembled into 13 phased contiguous autosomes, two sex chromosomes (W and Z), and one mitochondrial genome. This chromosome-level assembly encompasses 0.51 Gb and has no gaps, making this a telomere-to-telomere (T2T) genome assembly. The assembly is composed of 15 contigs with an N50 value of 36.4 Mb.

Keywords
Erebia palarica, Erebia, genome assembly, European Reference Genome Atlas, Biodiversity Genomics Europe, Earth Biogenome Project, Nymphalidae, Satyrinae, Chapman's ringlet, T2T genome assembly

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

This article is included in the Horizon Europe gateway.

This article is included in the Genome Reports from the Biodiversity Genomics Europe Project collection.

Corresponding author: Chiara Bortoluzzi (chiara.bortoluzzi@sib.swiss)

Author roles: Vila M: Investigation, Resources, Writing – Original Draft Preparation, Writing – Review & Editing; **Torrado-Blanco L**: Investigation, Resources, Writing – Review & Editing; **Ryrholm N**: Investigation, Resources, Writing – Review & Editing; **Romero-Pedreira D**: Investigation, Resources, Writing – Review & Editing; **Conejero M**: Investigation, Resources, Writing – Review & Editing; **Böhne A**: Methodology, Project Administration, Supervision, Writing – Review & Editing; **Monteiro R**: Methodology, Project Administration, Supervision, Writing – Review & Editing; **Marcussen T**: Methodology, Project Administration, Supervision, Writing – Review & Editing; **A. Oomen R**: Methodology, Project Administration, Supervision, Writing – Review & Editing; **H. Struck T**: Methodology, Project Administration, Supervision, Writing – Review & Editing; **Aguilera L**: Investigation, Writing – Original Draft Preparation, Writing – Review & Editing; **Gut M**: Investigation, Writing – Review & Editing; **Câmara Ferreira F**: Data Curation, Formal Analysis, Writing – Review & Editing; **Cruz F**: Data Curation, Formal Analysis, Writing – Review & Editing; **Gómez-Garrido J**: Data Curation, Formal Analysis, Writing – Review & Editing; **S. Alioto T**: Data Curation, Formal Analysis, Supervision, Writing – Review & Editing; **Haggerty L**: Data Curation, Formal Analysis, Writing – Review & Editing; **Martin F**: Data Curation, Formal Analysis, Writing – Review & Editing; **Bortoluzzi C**: Visualization, Writing – Original Draft Preparation

Competing interests: No competing interests were disclosed.

Grant information: This project received funding from Biodiversity Genomics Europe (Grant no.101059492), which is funded by Horizon Europe under the Biodiversity, Circular Economy and Environment call (REA.B.3); co-funded by the Swiss State Secretariat for Education, Research and Innovation (SERI) under contract numbers 22.00173 and 24.00054; and by the UK Research and Innovation (UKRI) under the Department for Business, Energy and Industrial Strategy's Horizon Europe Guarantee Scheme. MV was financially supported by grants from Xunta de Galicia (ED431B 2024/23) and the University of A Coruña. LT-B was funded by Xunta de Galicia (ED481A/2020-300).
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2026 Vila M *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: Vila M, Torrado-Blanco L, Ryrholm N *et al.* **ERGA-BGE genome of *Erebia palarica* Chapman, 1905: a montane butterfly endemic to North-West Spain [version 1; peer review: 1 approved with reservations]** Open Research Europe 2026, **6**:9 https://doi.org/10.12688/openreseurope.20717.1

First published: 06 Jan 2026, **6**:9 https://doi.org/10.12688/openreseurope.20717.1

Introduction

The butterfly *Erebia palarica* Chapman 1905, also known as Chapman's ringlet, is endemic to the Cantabrian Mountains (North-West Spain) and surrounding mountain ranges (i.e., Ancares, Courel, Queixa, Montes de León). This univoltine Satyrinae species is the largest species of its genus in Europe, It usually occurs at altitudes above 1,000 m. *Erebia palarica* is currently classified as "Near Threatened" on the IUCN Red List (Torrado-Blanco *et al.*, 2025; van Swaay *et al.*, 2025).

Overall, adults of *E. palarica* flutter in open scrublands, between late June and early September. However, local flight peaks likely extend for a month. Adults are frequently seen near heath (*Erica* L. species) and heather (*Calluna vulgaris* (L.) Hull.), and use rocky trails for basking. This species plays a relevant role in the montane ecosystem, acting as pollinator (adults use nectar resources provided by different plant taxa) and prey species. The larval cycle of *E. palarica* is far from understood (Clarke, 2024). Caterpillars have been reported to feed on *Festuca* L. grasses, but formal confirmation of the mono- or oligophagy is needed (reviewed by Torrado-Blanco (2024)). Most populations of *E. palarica* are known to host the same endosymbiont *Wolbachia* Hertig, 1936 strain as the surveyed populations of Iberian *Erebia meolans* (Prunner, 1789), its sister species. The chromosome number is the same for both species ($2n = 28$) (De Lesse, 1953; Torrado-Blanc et al. in prep).

The phylogeography of *E. palarica* has only been studied with mitochondrial Sanger sequencing and microsatellite markers (Torrado-Blanco *et al.*, 2024). Therefore, developing a high-quality reference genome for *E. palarica* is essential for advancing our understanding of its unique genetic makeup and adaptive traits.

The generation of this reference resource was coordinated by the European Reference Genome Atlas (ERGA) initiative's Biodiversity Genomics Europe (BGE) project, supporting ERGA's aim of promoting transnational cooperation to promote advances in the application of genomics technologies to protect and restore biodiversity (Mazzoni *et al.*, 2023).

Materials & methods

ERGA's sequencing strategy includes Oxford Nanopore Technology (ONT) and/or Pacific Biosciences (PacBio) for long-read sequencing, along with Hi-C sequencing for chromosomalarchitecture, Illumina Paired-End (PE) for polishing (i.e. recommended for ONT-only assemblies), and RNA sequencing for transcriptome profiling, to facilitate genome assembly and annotation.

Sample and sampling information

On 2 July 2022, four female specimens of *E. palarica* were sampled by Marta Vila and David Romero-Pedreira at Serra do Courel, Lugo, Spain: three from A Cabeza Grande and one from Alto do Couto. On 20 July 2023, a fifth female and a male were sampled by Marta Vila and Laura Torrado-Blanco at A Cabeza Grande. The identification of the species was confirmed based on morphology by Marta Vila, Laura Torrado-Blanco and Nils Ryrholm, and barcoding by María Conejero.

The sampling of all specimens was conducted under permits LU/003/2022 and LU/012/2023, issued by the Xunta de Galicia, Consellería de Medio Ambiente, Territorio e Vivenda, Servizo de Patrimonio Natural (Spain). The biological material collected in Spain, and used to generate digital sequences, was retrieved from wildlife taxa regulated by the Spanish Royal Decree 124/2017 (https://www.boe.es/eli/es/rd/2017/02/24/124). No ABS permits were granted as utilization of the material falls under exemption in Article 3(2) of the Spanish Royal Decree 124/2017 and was confirmed by the national authorities. Sampling was performed using a butterfly net. Specimens were handled in two ways before storage at -80°C. Two females collected at A Cabeza Grande in 2022 were immediately euthanized using dry ice, kept there for 34 hours, and then transferred to -80°C. The remaining three specimens were maintained alive for 24 hours until arrival at the lab and then frozen at -80°C. All samples were stored at -80°C until DNA extraction.

Vouchering information

Physical reference materials for a proxy male specimen have been deposited in the Museo Nacional de Ciencias Naturales (MNCN, CSIC) https://mncn.csic.es/en under the accession number MNCN:Ent:366619.

Frozen reference tissue material for the whole organism (proxy male) is available in the Museo Nacional de Ciencias Naturales (MNCN, CSIC) https://mncn.csic.es/en under tissue voucher ID MNCN:Ent:366618 and proxy tissue voucher ID MNCN:ADN:151.720.

Data availability

Erebia palarica and the related genomic study were assigned to Tree of Life ID (ToLID) 'ilErePala4' and all sample, sequence, and assembly information are available under the umbrella BioProject PRJEB75270. The sample information is available at the following BioSample accession: SAMEA114504880. The genome assembly is accessible from ENA under accession number GCA_964035005.3, while the annotation of a previously released genome assembly (GCA_964035005.1) is available through the Ensembl website (https://projects.ensembl.org/erga-bge/). Sequencing data produced as part of this project are available from ENA at the following accessions: ERX12324510, ERX12324512, ERX13168336, and ERX13168337. Documentation related to the genome assembly and curation can be found in the ERGA Assembly Report (EAR) document available at https://github.com/ERGA-consortium/EARs/blob/main/Assembly_Reports/Erebia_palarica/ilErePala4/. Further details and data about the project are hosted on the ERGA portal at https://portal.erga-biodiversity.eu/data_portal/191414.

Genetic information

The estimated genome size, based on ancestral taxa, is 0.48 Gb. This is a diploid genome with a haploid number of 14 chromosomes (2n=28) and sex chromosomes (W and Z) based on previous studies on closely related species (Lohse *et al.*, 2022), Bioproject PRJEB80714). All information for this species was retrieved from Genomes on a Tree (Challis *et al.*, 2023).

DNA/RNA processing

DNA was extracted from head and thorax using the Blood & Cell Culture DNA Mini Kit (Qiagen) following the manufacturer's instructions. DNA quantification was performed using a Qubit dsDNA BR Assay Kit (Thermo Fisher Scientific) and DNA integrity was assessed using a Femtopulse system (Genomic DNA 165 Kb Kit, Agilent). DNA was stored at 4°C until use.

RNA was extracted using an RNeasyMini Kit (Qiagen) according to the manufacturer's instructions. RNA was extracted from three different tissues: leg, wing and abdomen. RNA quantification was performed using the Qubit RNA BR Kit and RNA integrity was assessed using a Fragment Analyzer system (RNA 15nt Kit, Agilent). RNA was pooled in a 1:1:2 (leg:wing: abdomen) ratio before library preparation and stored at -80°C until use.

Library preparation and sequencing

For long-read whole genome sequencing, a library was prepared using the SQK-LSK114 kit (Oxford Nanopore Technologies, ONT) which was then sequenced in one R10.4.1 flow cell on a PromethION 24 A series instrument (ONT). A short-read WGS library was prepared using the KAPA Hyper Prep Kit (Roche). The RNA library from the pooled sample was prepared using the KAPA mRNA Hyper prep kit (Roche). A Hi-C library was prepared from abdomen tissue using the ARIMA High Coverage Hi-C Kit (ARIMA), followed by the KAPA Hyper Prep Kit for Illumina sequencing (Roche). All the short-read libraries were sequenced on a NovaSeq 6000 instrument (2x151bp Illumina). In total 104X Oxford Nanopore, 64X Illumina WGS shotgun, and 81X HiC data were sequenced to generate the assembly.

Genome assembly methods

The genome was assembled using the CNAG CLAWS pipeline v2.3 (Gomez-Garrido, 2024). Briefly, reads were pre-processed for quality and length using Trim Galore v0.6.7 (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) and Filtlong v0.2.1 (https://github.com/rrwick/Filtlong) with the option -t 50000000000 targeting 50Gb (100x coverage in ONT), and initial contigs were assembled using Hifiasm v0.24.0 (Hu *et al.*, 2024) with the –ont option and Hi-C phasing. The hap2 assembly, containing both the Z and W chromosomes, was scaffolded with YaHS v1.2a (Zhou *et al.*, 2023), however no contigs were joined into scaffolds and the error correction actually broke the chrW contig into multiple sequences. Inspection of the Pretext Hi-C contact map of the original Hifiasm hap2 assembly in Pretext v0.2.5 (Harry, 2020) confirmed the T2T nature of the contigs larger than 20 Mb. One small 327 kb contig was identified as a haplotig and removed. The remaining 1.6 Mb contig was identified as a strain of *Wolbachia* and was also removed. The 15 assembled contigs were then processed with the Rapid Curation Toolkit (https://gitlab.com/wtsi-grit/rapid-curation). Finally, the mitochondrial genome was assembled as a single circular contig of 15,221 bp using the FOAM pipeline (https://github.com/cnag-aat/FOAM) and included in the released assembly (GCA_964035005.3). Summary analysis of the released assembly was performed using the ERGA-BGE Genome Report ASM Galaxy workflow (10.48546/workflowhub.workflow.1103.2).

Genome annotation methods

A gene set was generated using the Ensembl Gene Annotation system on a previously released assembly (GCA_964035005.1). (Aken *et al.*, 2016), primarily by aligning publicly available short-read RNA-seq data from BioSample SAMEA114504880 to the genome. Gaps in the annotation were filled via protein-to-genome alignments of a select set of clade-specific proteins from (UniProt Consortium, 2019), which had experimental evidence at the protein or transcript level. At each locus, data were aggregated and consolidated, prioritising models derived from RNA-seq data, resulting in a final set of gene models and associated non-redundant transcript sets. To distinguish true isoforms from fragments, the likelihood of each open reading frame (ORF) was evaluated against known metazoan proteins. Low-quality transcript models, such as those showing evidence of fragmented ORFs, were removed. In cases where RNA-seq data were fragmented or absent, homology data were prioritised, favouring longer transcripts with strong intron support from short-read data. The resulting gene models were classified into two categories: protein-coding, and long non-coding. Models that did not overlap protein-coding genes, and were constructed from transcriptomic data were considered potential lncRNAs. Potential lncRNAs were further filtered to remove single-exon loci due to their unreliability. Putative miRNAs were predicted by performing a BLAST search of miRbase (Kozomara *et al.*, 2019) against the genome, followed by RNAfold analysis (Gruber *et al.*, 2008). Other small non-coding loci were identified by scanning the genome with Rfam (Kalvari *et al.*, 2018) and passing the results through Infernal (Nawrocki & Eddy, 2013).

Results

Genome assembly

The genome assembly has a total length of 513,180,471 bp in 15 linear chromosomal contigs (W and Z chromosomes included) and the mitogenome (Figure 1 & Figure 2), with a GC content of 36.9%. The assembly has a contig N50 of 36,357,879 bp and L50 of 6. The assembly has no gaps, making this a telomere-to-telomere (T2T) genome assembly. The single-copy gene content analysis using the Lepidoptera database with BUSCO v5.8.0 (Manni *et al.*, 2021) resulted in 99.0% completeness (98.6% single and 0.4% duplicated). 84.0% of reads k-mers were present in the assembly and the assembly has a base accuracy Quality Value (QV) of 68.8 as calculated by Merqury (Rhie *et al.*, 2020).

Genome annotation

The genome annotation consists of 13,711 protein-coding genes with associated 26,969 transcripts, in addition to 3,161 non-coding genes (Table 1). Using the longest isoform per transcript, the single-copy gene content analysis using the Lepidoptera odb10 database with BUSCO resulted in 94.6% completeness. Using the OMAmer Metazoa-v2.0.0.h5 database for OMArk (Nevers *et al.*, 2025) resulted in 93.5% completeness and 85.6% consistency (Table 2).

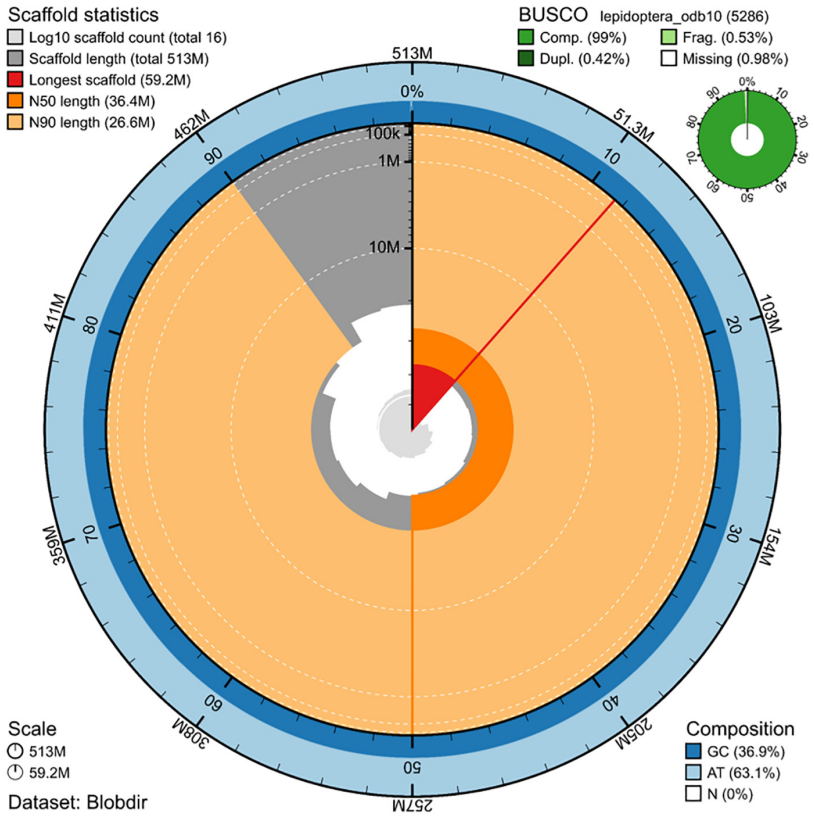


Figure 1. Snail plot summary of assembly statistics. The main plot is divided into 1,000 size-ordered bins around the circumference, with each bin representing 0.1% of the 513,180,471 bp assembly including the mitochondrial genome. The distribution of sequence lengths is shown in dark grey, with the plot radius scaled to the longest sequence present in the assembly (59.2 Mb shown in red). Orange and pale-orange arcs show the scaffold N50 and N90 sequence lengths (36,357,879 and 26,642,400 bp), respectively. The pale grey spiral shows the cumulative sequence count on a log-scale, with white scale lines showing successive orders of magnitude. The blue and pale-blue area around the outside of the plot shows the distribution of GC, AT, and N percentages in the same bins as the inner plot. A summary of complete, fragmented, duplicated, and missing BUSCO genes found in the assembled genome from the Lepidoptera database (odb10) is shown in the top right.

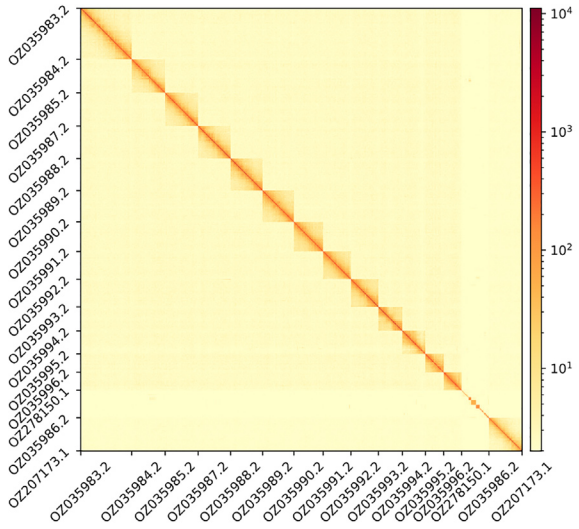


Figure 2. Hi-C contact map showing spatial interactions between regions of the genome. The diagonal corresponds to intra-chromosomal contacts, depicting chromosome boundaries. The frequency of contacts is shown on a logarithmic heatmap scale. Hi-C matrix bins were merged into a 100 kb bin size for plotting. The Hi-C contact map shows the 13 autosomes, the W chromosome (GenBank: OZ278150.1), the Z chromosome (GenBank: OZ035986.2), and the mitochondrial genome (GenBank: OZ207173.1).

Table 1. Statistics from assembled gene models.

	No. genes	No. transcripts	Mean gene length (bp)	No. single-exon genes	Mean exons per transcript
mRNA	13,711	26,969	17,279	859	7.6
pseudogene	0.0	0.0	0.0	0.0	0.0
snoRNA	46	46	114	46	1.0
lncRNA	2,661	3,022	9,739	3	2.3
snRNA	70	70	147	70	1.0
rRNA	47	47	478	47	1.0
tRNA	337	337	75	337	1.0
Other ncRNA	2	2	296 - 296	2	1.0 - 1.0

Table 2. Annotation completeness and consistency scores calculated by BUSCO run in protein mode (lepidoptera_odb10) and OMArk (Metazoa-v2.0.0.h5).

	Complete	Single copy	Duplicated	Fragmented	Missing
BUSCO	4,999 (94.6%)	4,959 (93.8%)	40 (0.8%)	78 (1.5%)	209 (3.9%)
OMArk	7,417 (93.6%)	7,031 (88.7%)	386 (4.9%)	-	511 (6.4%)
	Consistent	**Inconsistent**	**Contaminants**	**Unknown**	
OMArk	11,739 (85.6%)	563 (4.1%)	0 (0.0%)	1,409 (10.3%)	

Author contributions

MV coordinated the project; MV, LT-B, and DR-P collected the species; MV, LT-B, and NR identified the species; MC did the barcoding; MV and LT-B sampled and preserved biological material and provided metadata; AsB, RM, TM, RO, and THS and provided support in sampling, shipping of biological material, metadata collection, and management; LA and MG extracted DNA, prepared libraries, and performed sequencing; FC, FCF and JGG performed genome assembly and curation under the supervision of TSA; LH and FM performed genome annotation; CB generated the analysis and report. All authors contributed to the writing, review, and editing of this genome note and read and approved the final version.

Acknowledgments

We would like to express our gratitude to Dr. Rodolfo Barreiro and Dr. Mercedes París for their assistance with sample storage and vouchering, respectively. We appreciate the suggestions from Nadir Alvarez and ERGA-Spain. We would like to acknowledge the assembly reviewer, Thomas Brown, from the Leibniz Institute for Zoo and Wildlife Research (Germany). The authors acknowledge the support of the Freiburg Galaxy Team: Saim Momin and Björn Grüning, Bioinformatics, University of Freiburg (Germany), funded by the German Federal Ministry of Education and Research BMBF grant 031 A538A de.NBI-RBC and the Ministry of Science, Research and the Arts Baden-Württemberg (MWK) within the framework of LIBIS/de.NBI Freiburg. We would like to acknowledge the assembly reviewer, Tom Brown from the Leibniz Institute of Zoo and Wildlife Research.

References

Aken BL, Ayling S, Barrell D, *et al.*: **The Ensembl gene annotation system.** *Database (Oxford).* 2016; **2016**: baw093.
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

Clarke HE: **A checklist of European butterfly larval foodplants.** *Ecol Evol.* 2024; **14**(1): e10834.
PubMed Abstract | Publisher Full Text | Free Full Text

De Lesse H: **Formules chromosomiques nouvelles du genre *Erebia* (Lepid.**

Rhopal.). *Comptes Rendus Hebdomadaires Des Séances de l'Academie Des Sciences.* 1953; **237**: 758–759.

Gomez-Garrido J: **CLAWS (CNAG's Long-read Assembly Workflow in Snakemake) [Computer software]**. 2024.
Reference Source

Gruber AR, Lorenz R, Bernhart SH, *et al.*: **The Vienna RNA websuite.** *Nucleic Acids Res.* 2008; **36**(suppl_2): W70–W74.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Harry EP: **A desktop application for viewing pretext contact maps**. 2020.

Hu J, Wang Z, Sun Z, *et al.*: **NextDenovo: an efficient error correction and accurate assembly tool for noisy long reads.** *Genome Biol.* 2024; **25**(1): 107.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Kalvari I, Nawrocki EP, Argasinska J, *et al.*: **Non-Coding RNA analysis using the Rfam database.** *Curr Protoc Bioinformatics.* 2018; **62**(1): e51.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Kozomara A, Birgaoanu M, Griffiths-Jones S: **miRBase: from microRNA sequences to function.** *Nucleic Acids Res.* 2019; **47**(D1): D155–D162.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Lohse O, Lohse K, Wellcome Sanger Institute Tree of Life programme, *et al.*: **The genome sequence of the scotch argus butterfly, *Erebia aethiops* (Esper, 1777) [version 1; peer review: 2 approved].** *Wellcome Open Res.* 2022; **7**: 217.
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**

Mazzoni CJ, Ciofi C, Waterhouse RM: **Biodiversity: an atlas of European reference genomes.** *Nature.* 2023; **619**(7969): 252.
PubMed Abstract | **Publisher Full Text**

Nawrocki EP, Eddy SR: **Infernal 1.1: 100-fold faster RNA homology searches.** *Bioinformatics.* 2013; **29**(22): 2933–2935.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Nevers Y, Warwick Vesztrocy A, Rossier V, *et al.*: **Quality assessment of gene repertoire annotations with OMArk.** *Nat Biotechnol.* 2025; **43**(1): 124–133.
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**

Torrado-Blanco L: **Conservation and evolutionary genetics of the butterfly *Erebia palarica* Chapman, 1905.** PhD thesis, 2024.
Reference Source

Torrado-Blanco L, Carballa M, Vila M: **Conservation status of *Erebia palarica* Chapman, 1905 (Lepidoptera, Nymphalidae): local insight and whole distribution perspective.** *Insect Conserv Divers.* 2025; **18**(3): 370–385.
Publisher Full Text

Torrado-Blanco L, Eberle J, Romo H, *et al.*: **Phylogeography of the Iberian endemic butterfly *Erebia palarica* Chapman, 1905 (Lepidoptera: Nymphalidae): an integrative approach.** *Insect Conserv Divers.* 2024; **17**(4): 651–675.
Publisher Full Text

UniProt Consortium: **UniProt: a worldwide hub of protein knowledge.** *Nucleic Acids Res.* 2019; **47**(D1): D506–D515.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

van Swaay C, Ellis S, Warren M: ***Erebia palarica*.** *The IUCN Red List of Threatened Species.* 2025; **2025**: e.T161145A876509.
Publisher 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 02 March 2026

https://doi.org/10.21956/openreseurope.22408.r68418

© **2026 Huang Y.** 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.

?

Yu-Feng Huang
Computer Science and Engineering, Yuan Ze University, Zhongli District, Taoyuan City, Taiwan

The manuscript reports a chromosome-level, gapless genome assembly generated from a female *Erebia palarica* individual, comprising 13 autosomes plus W and Z, along with a mitochondrial genome. The released assembly spans 513,180,471 bp, organized into 15 linear chromosomal contigs (including W and Z) plus the mitogenome, with strong assembly metrics (contig N50 = 36.36 Mb, BUSCO = 99.0%, Merqury QV = 68.8). Overall, this dataset is potentially valuable for conservation and speciation studies; however, several technical and reporting issues should be clarified to support the manuscript's strongest claims (particularly the "T2T" framing) and to improve transparency and reproducibility.

Major Comments

1. Clarify and justify the "telomere-to-telomere (T2T)" claim and resolve internal inconsistencies

The Results state that the assembly "has no gaps, making this a telomere-to-telomere (T2T) genome assembly." However, the Methods indicate that scaffolding/error correction "broke the chrW contig into multiple sequences," which appears inconsistent with the claim of complete telomere-to-telomere continuity. Please reconcile these statements and clarify what evidence supports the "T2T" designation (e.g., detection/confirmation of telomeric repeats at both ends of each chromosome, and whether each chromosome - including W - remains a single contiguous sequence in the final released assembly).

2. Define precisely what "phased" means here and specify what assembly product is actually released

The Abstract refers to "13 phased contiguous autosomes" plus W and Z, while the Methods describe a Hi-C–phased Hifiasm run and then focus on the "hap2 assembly, containing both the Z and W chromosomes," which appears to form the basis of downstream processing and release. Please explicitly state:

- whether both haplotypes are publicly available (and where), or whether only a single

selected haplotype/pseudohaplotype is released as GCA_964035005.3;
- how “phasing” is represented in the final deliverable (e.g., true diploid haplotype-resolved assemblies vs. one haplotype chosen as the primary release);
- why hap2 was selected (e.g., improved representation of W/Z, higher contiguity, curation considerations).

3. Gene annotation appears to correspond to a different assembly version than the reported final release

The manuscript states that genome annotation was generated “on a previously released assembly (GCA_964035005.1).” Data availability indicates the assembly release is GCA_964035005.3, while the annotation is available for GCA_964035005.1 via Ensembl. Yet the Results present detailed annotation outputs and completeness metrics (e.g., 13,711 protein-coding genes; BUSCO 94.6%; OMArk 93.5%). Please clarify whether these annotation statistics correspond strictly to GCA_964035005.1, whether any liftover/remapping was performed to GCA_964035005.3, and how users should interpret the reported annotation quality relative to the claimed T2T “.3” assembly.

4. Citation mismatch for Hifiasm should be corrected

The Methods cite “Hifiasm v0.24.0 (Hu et al., 2024),” but Hu et al. (2024) does not appear to be the canonical reference for Hifiasm and is commonly associated with other assembly tools. Please correct the citation to the appropriate Hifiasm publication(s) to ensure methodological accuracy and proper attribution.

5. Explain genome size discrepancy and interpret the reported k-mer presence

The manuscript reports an estimated genome size of ~0.48 Gb, whereas the assembly span is ~0.513 Gb. Additionally, the Results state that 84.0% of read k-mers are present in the assembly. Given the strong claim of a gapless chromosome-level/T2T assembly, it would be helpful to briefly explain: (i) why the assembly is larger than the genome size estimate, and (ii) what the 84% k-mer presence reflects (e.g., dataset used, how phasing and filtering affect k-mer accounting, and whether readers should *not* interpret this value as evidence of missing sequence).

6. Mitochondrial genome: include a minimal validation/completeness statement

The mitochondrial genome is reported as a single circular contig of 15,221 bp generated using the FOAM pipeline. A brief validation statement would improve confidence in its correctness and completeness, including (i) explicit confirmation of circularization and (ii) confirmation that the expected canonical Lepidoptera gene set is present (13 protein-coding genes, 22 tRNAs, and 2 rRNAs). If available, a short note on screening for NUMT-related artifacts would further strengthen this section.

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

Yes

Are the protocols appropriate and is the work technically sound?

Partly

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

Partly

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

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Bioinformatics, Mitogenomics, Cancer Bioinformatics, Genome assembly and gene annotation

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.