GENOME REPORT

ERGA-BGE genome of *Coenonympha oedippus*: an IUCN endangered European butterfly species occurring in two ecotypes

Tatjana Čelik[1], Tjaša Lokovšek[1], Elena Bužan[2,3], Astrid Böhne[4], Rita Monteiro[4], Rosa Fernández[5], Nuria Escudero[5], Marta Gut[6,7], Laura Aguilera[6,7], Francisco Câmara Ferreira[6,7], Fernando Cruz[6,7], Jèssica Gómez-Garrido[6,7], Tyler S. Alioto[6,7], Chiara Bortoluzzi[8,9*]

[1] ZRC SAZU, Jovan Hadži Institute of Biology, Novi trg 2, SI-1000 Ljubljana, Slovenia

[2] University of Primorska, Faculty of Mathematics, Natural Sciences and Information Technologies, Glagoljaška 8, SI-6000 Koper, Slovenia

[3] Faculty of Environmental Protection, Trg mladosti 7, 3320 Velenje, Slovenia

[4] Leibniz Institute for the Analysis of Biodiversity Change, Museum Koenig Bonn, Adenauerallee 127, 53113 Bonn, Germany

[5] Metazoa Phylogenomics Lab, Institute for Evolutionary Biology (CSIC-UPF). Passeig marítim de la Barceloneta 37-49. 08003 Barcelona, Spain

[6] Centro Nacional de Análisis Genómico (CNAG), Baldiri Reixac 4, 08028 Barcelona, Spain

[7] Universitat de Barcelona (UB), 08028 Barcelona, Spain

[8] University of Florence, Department of Biology, Via Madonna del Piano 6, 50019 Sesto Fiorentino, Firenze, Italy

[9] SIB Swiss Institute of Bioinformatics, Department of Aquatic Ecology, Eawag, Überlandstrasse 133, 8600 Dübendorf, Switzerland

* To whom correspondence should be addressed: chiara.bortoluzzi@unifi.it

Abstract

The reference genome of the false ringlet (*Coenonympha oedippus*) will serve as a valuable resource for uncovering the genetic mechanism underlying the species' adaptability to two ecologically distinct habitats. Through this genome we might be able to determine whether (i) each ecotype is monophyletic, indicating that the ecological divergence represents an early stage of speciation, (ii) the ecotypes have evolved through divergent evolution of habitat preference, or (iii) the differences between ecotypes are solely due to phenotypic plasticity or epigenetic variation. This reference genome is also a prerequisite for the planning, design, and implementation of conservation

measures for this endangered species, taking into account its intraspecific diversity. Furthermore, it holds broader implications for population genomic studies of the species-rich genus *Coenonympha*, which includes some of the most endangered butterfly taxa in Europe. The complete genome sequence was assembled into 30 contiguous chromosomal pseudomolecules (sex chromosomes included). This chromosome-level assembly encompasses 0.39 Gb, composed of 385 contigs and 62 scaffolds, with contig and scaffold N50 values of 2.8 Mb and 14.2 Mb, respectively.

Keywords

Coenonympha oedippus, genome assembly, European Reference Genome Atlas, Biodiversity Genomics Europe, *Satyrinae*, False ringlet

Introduction

Coenonympha oedippus, also known as false ringlet or Barjanski okarček in Slovenian, is a Eurasian butterfly species from the subfamily *Satyrinae*- The species shows remarkable ecotype differences in its European distribution area, which are reflected in the different larval host plants (Čelik et al., 2015), wing morphology (Jugovic et al., 2018), and neutral genetic differentiation (Zupan et al., 2021). Given the phylogeographic distribution of the species in Europe (Després et al., 2019), adaptive genetic differentiation is expected. The recent evolutionary history of *Coenonympha* butterflies has shown that *C. oedippus* (together with the genus *Triphysa*) belongs to an early divergent clade that split from other *Coenonympha* species about 13 million years ago (Greenwood et al., 2024).

Coenonympha oedippus was last assessed for the IUCN Red List of Threatened Species in 2009, during which it was categorised as Endangered under criterion A2c according to Europe's assessment criteria (van Sway et al., 2010). The species is also listed in Annexes II and IV of the EU Habitats Directive (Council Directive 92/43/EEC) and its conservation status is assessed as 'Unfavourable' (U2, U1) in most EU Member States (Article 17 reporting, 2013–2018). The initial conservation translocations (i.e. reintroduction and reinforcements) of the species in Slovenia (2018–2022) highlighted the need to assess genetic diversity within the source and reintroduced populations to evaluate whether genetic variation needs to be increased through dedicated management strategies, such as genetic rescue or assisted gene flow (Čelik et al., 2024).

Coenonympha oedippus is an important indicator species that reflect the health status of oligotrophic wet grassland and fens. The larvae are nutritionally dependent on plant species that are dominant and characteristic of these grasslands and therefore could influence plant population dynamics (Čelik et al., 2021). They are also hosts for parasites (Dierks, 2006) and are an important food source for many predatory invertebrates,

birds, and small mammals in addition to the other three life stages (egg, pupa, butterfly) (Lhonoré, 1998). The complex life cycle of *C. oedippus*, with four developmental stages, each with specific habitat requirements at different scales - from microhabitat to landscape - places the species at the centre of restoration and management planning, which also benefits wider biodiversity (Čelik et al., 2015, 2024).

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 aims 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 chromosomal architecture, Illumina Paired-End (PE) for polishing (i.e. recommended for ONT-only assemblies), and RNA sequencing for transcriptomic profiling, to facilitate genome assembly and annotation.

Sample and Sampling Information

On 15 June 2023, Tatjana Čelik and Tjaša Lokovšek sampled 5 female specimens of *Coenonympha oedippus*. The specimens were identified by Tatjana Čelik based on morphology without an identification key, as the expert knows the species well, having researched it for 20 years. The 5 specimens were collected in Babiči-Zupančiči, Slovenska Istra (Slovenia). Sampling was performed under permission 35601-140/2009 – 6 issued by the Ministry of the Environment and Spatial Planning of the Republic of Slovenia. Specimens were hand-collected and then euthanized using a light mechanical compression of the thorax. At the collection site, each specimen was snap-frozen and subsequently preserved at -80°C until DNA extraction.

Vouchering information

Physical reference materials from a proxy female individual (i.e., an individual different from the one used for sequencing) have been deposited in the Prirodoslovni muzej Slovenije https://www.pms-lj.si/, Ljubljana (Slovenia) under proxy voucher ID PMSL-Invertebrata-22.

Frozen reference tissue material of a whole organism (female) is available in the Prirodoslovni muzej Slovenije https://www.pms-lj.si/, Ljubljana (Slovenia) under proxy tissue voucher ID PMS_TIS_14.

Data Availability

Coenonympha oedippus and the related genomic study were assigned to Tree of Life ID (ToLID) 'ilCoeOedi1' and all sample, sequence,

and assembly information are available under the umbrella BioProject PRJEB77778. The sample information is available at the following BioSample accessions: SAMEA115177279 and SAMEA115177280. The genome assembly is accessible from ENA under accession number GCA_964304575.1 and the annotated genome is available through the Ensembl Beta page (https://beta.ensembl.org/). Sequencing data produced as part of this project are available from ENA at the following accessions: ERX13168332, ERX13168333, and ERX12746377. 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/tree/main/Assembly_Reports/Coenonympha_oedippus/ilCoeOedi1. Further details and data about the project are hosted on the ERGA portal at https://portal.erga-biodiversity.eu/data_portal/554486.

Genetic Information

The estimated genome size, based on ancestral taxa, is 0.48 Gb. This is a diploid genome with a haploid number of 29 chromosomes (2n=58) and unknown sex chromosomes. All information for this species was retrieved from Genomes on a Tree (Challis et al., 2023).

DNA/RNA processing

DNA was extracted from the whole organism using the Blood & Cell Culture DNA Midi 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 Genomic DNA 165 Kb Kit (Agilent) on the Femto Pulse system (Agilent). The DNA was stored at +4oC until used.

RNA was extracted using an RNeasy Mini Kit (Qiagen) according to the manufacturer's instructions. RNA was extracted from three different specimen parts: head, wing and abdomen. RNA quantification was performed using the Qubit RNA BR kit and RNA integrity was assessed using a Bioanalyzer 2100 system (Agilent) RNA 6000 Nano Kit (Agilent). The RNA was pooled equimolarly for the library preparation and stored at-80oC until used.

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 on a PromethION 24 A Series instrument (ONT). A short-read whole genome sequencing library was prepared using the KAPA Hyper Prep Kit (Roche). A Hi-C library was prepared from thorax using the Dovetail Omni-C Kit (Cantata Bio), followed by the KAPA Hyper Prep Kit for Illumina sequencing (Roche). The RNA library was prepared from the pooled sample using

the KAPA mRNA Hyper prep kit (Roche). The short-read libraries were sequenced on a NovaSeq 6000 instrument (Illumina).

In total 226x Oxford Nanopore, 75x Illumina WGS shotgun, and 183x HiC data were sequenced to generate the assembly.

Genome Assembly Methods

The genome was assembled using the CNAG CLAWS pipeline (Jessica Gomez-Garrido, 2024). Briefly, reads were preprocessed for quality and length using Trim Galore v0.6.7 and Filtlong v0.2.1, and initial contigs were assembled using NextDenovo v2.5.0, followed by polishing of the assembled contigs using HyPo v1.0.3, removal of retained haplotigs using purge-dups v1.2.6 and scaffolding with YaHS v1.2a. Finally, assembled scaffolds were curated via manual inspection using Pretext v0.2.5 with the Rapid Curation Toolkit (https://gitlab.com/wtsi-grit/rapid-curation) to remove any false joins and incorporate any sequences not automatically scaffolded into their respective locations in the chromosomal pseudomolecules (or super-scaffolds). Finally, the mitochondrial genome was assembled as a single circular contig of 15,325 bp using the FOAM pipeline (https://github.com/cnag-aat/FOAM) and included in the released assembly (GCA_964304575.1). Summary analysis of the released assembly was performed using the ERGA-BGE Genome Report ASM Galaxy workflow (https://workflowhub.eu/workflows/1103?version=2).

Results

Genome Assembly

The genome assembly has a total length of 397,595,348 bp in 62 scaffolds including the mitogenome (Figures 1 & 2), with a GC content of 37.3%. The assembly has a contig N50 of 2,773,664 bp and L50 of 45 and a scaffold N50 of 14,242,255 bp and L50 of 13. The assembly has a total of 323 gaps, totaling 65.7 kb in cumulative size. The single-copy gene content analysis using the Lepidoptera database with BUSCO (Manni et al. 2021) resulted in 98.2% completeness (97.5% single and 0.7% duplicated). 69.4% of reads k-mers were present in the assembly and the assembly has a base accuracy Quality Value (QV) of 50.4 as calculated by Merqury (Rhie et al., 2020).

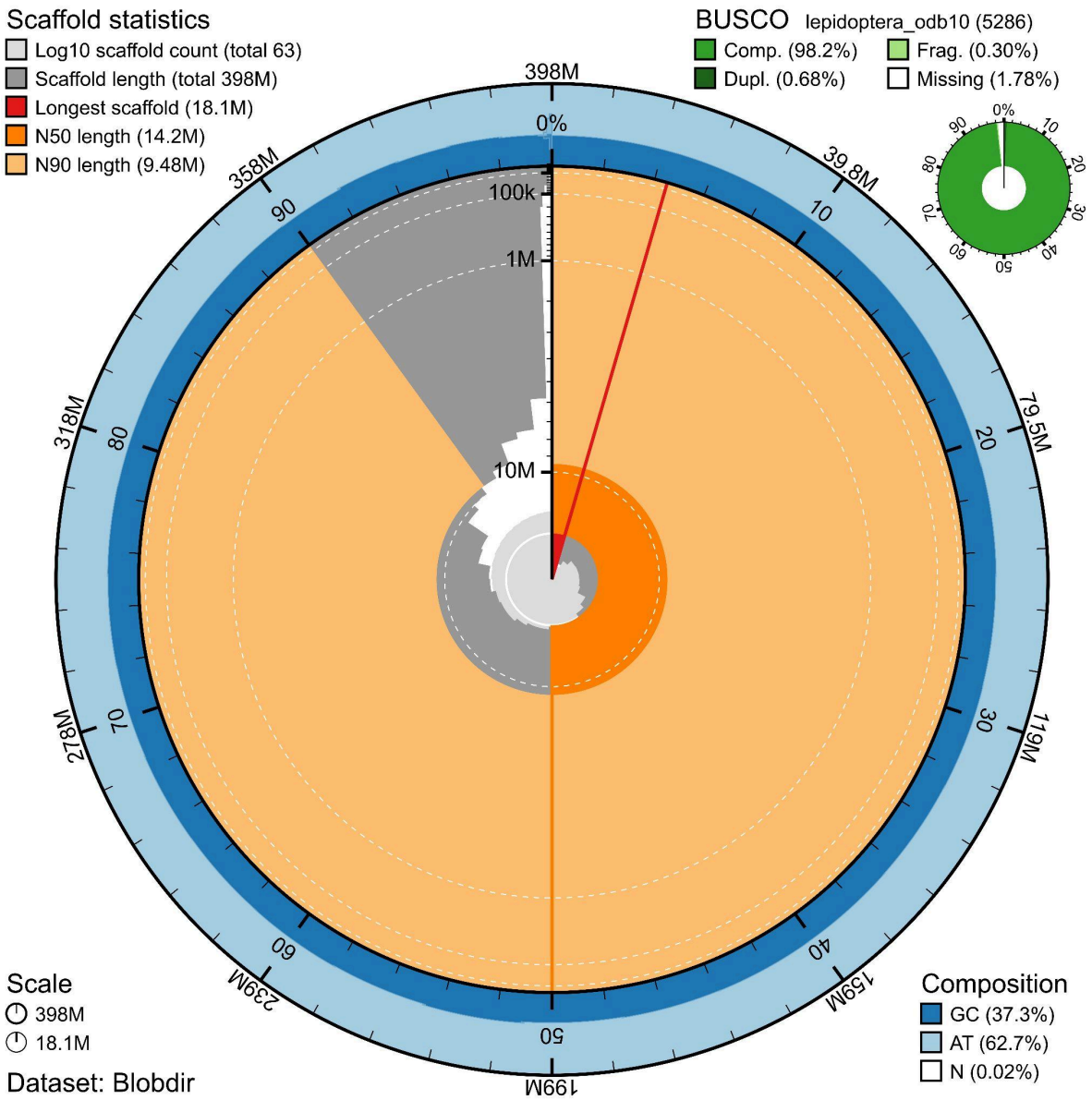

Figure 1. Snail plot summary of assembly statistics. The main plot is divided into 1,000 size-ordered bins around the circumference, with each bin representing 0.1% of the 397,595,348 bp assembly including the mitochondrial genome. The distribution of sequence lengths is shown in dark grey, with the plot radius scaled to the longest sequence present in the assembly (18.1 Mb, shown in red). Orange and pale-orange arcs show the scaffold N50 and N90 sequence lengths (14,242,255 and 9,479,849 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.

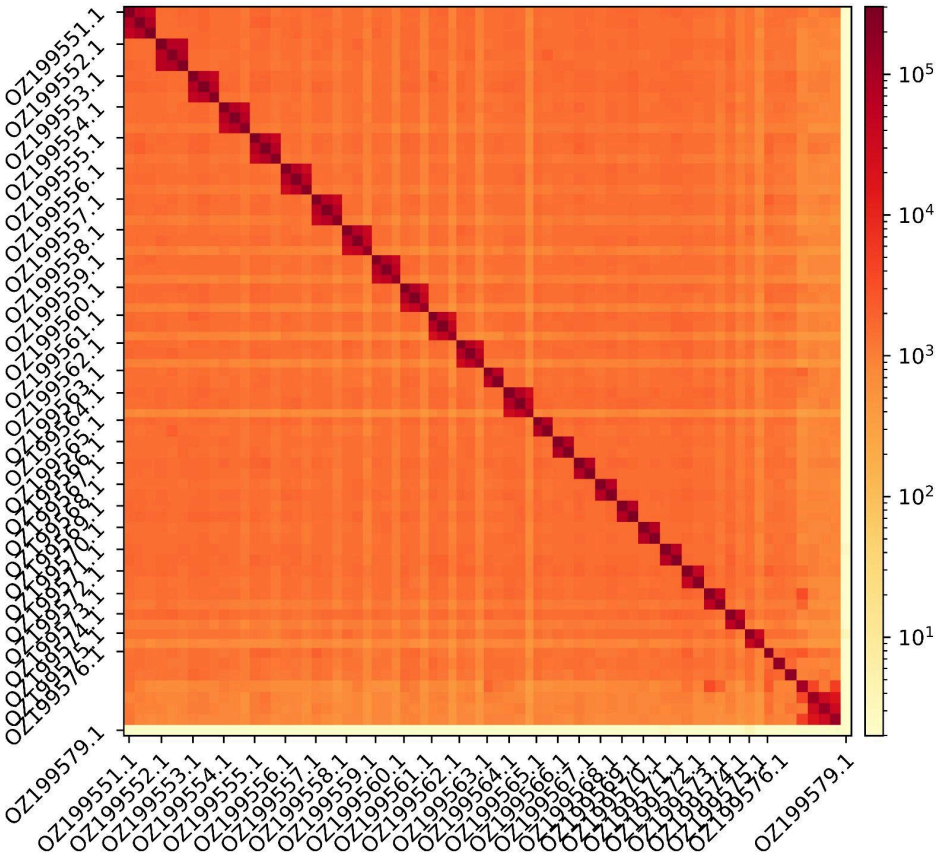

Figure 2. Hi-C contact map showing spatial interactions between regions of the genome. The diagonal corresponds to intra-chromosomal contacts, depicting chromosome boundaries. The frequency of contacts is shown on a logarithmic heatmap scale. Hi-C matrix bins were merged into a 100 kb bin size for plotting. Due to space constraints on the axes, only the GenBank names of the 26th largest chromosomes and the mitochondrial genome (GenBank name: OZ199579.1) are shown.

Acknowledgements

We 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, Emilie Teodori, from Genoscope.

Conflict of Interest

The authors declare no conflict of interest related to this study. The funding sources had no involvement in the study design, collection, analysis, or interpretation of data; in the writing of the manuscript; or in the decision to submit the article for publication. All authors have participated sufficiently in the work to take public responsibility for the content and agree to the submission of this manuscript.

Funder Information

Biodiversity Genomics Europe (Grant no.101059492) 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. We acknowledge the Slovenian Research and Innovation Agency (Research Core Funding No. P1-0236) for funding the sampling and the writing and reviewing of this genome note.

Author Contributions

TČ and EB coordinated the project; TČ collected the species; TČ identified the species; TČ sampled and preserved biological material and provided metadata; RF, NE, and TL provided support in sampling, shipping of biological material, metadata collection, and management; LA and MG extracted DNA, prepared libraries and performed sequencing; FCF, FC and JGG performed genome assembly and curation under the supervision of TA; 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.

Literature Cited

Article 17 Reporting, 2013–2018. Reporting under Article 17 of the Habitats Directive. https://nature-art17.eionet.europa.eu/article17/

Challis R, Kumar S, Sotero-Caio C, Brown M & Blaxter M (2023) 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 Research 8:24. https://doi.org/10.12688/wellcomeopenres.18658.1

Čelik T, Bräu M, Bonelli S, Cerrato C, Vreš B, Balletto E. et al (2015) Winter-green host-plants,litter quantity and vegetation structure are key determinants of habitat quality for *Coenonympha oedippus* in Europe. Journal of Insect Conservation, 19, 359–375. https://doi.org/10.1007/s10841-014-9736-3

Čelik T, Küzmič F, Behrič S, ŠilcU, & Vreš B (2021) Raziskava stanja potencialnih izvornih populacij vrste barjanski okarček (*Coenonympha oedippus*) in stanja njihovega habitata s smernicami za ustrezno upravljanje. Končno poročilo. Biološki inštitut Jovana Hadžija ZRC SAZU, Ljubljana, 69 str.

Čelik T, Kralj-Fišer S, Šilc U, Küzmič F, & Vreš B (2024) Reviving of *Coenonympha oedippus*: A comprehensive approach to the reintroduction of an endangered European butterfly. Insect Conservation and Diversity. https://doi.org/10.1111/icad.12778

Council Directive 92/43/EEC of 21 May 1992 on the conservation of natural habitats and of wild fauna and flora.

Després L, Henniaux C, Rioux D, Capblancq T, Zupan S, Čelik T, Sielezniew M, Bonato L, & Ficetola GF (2019) Inferring the biogeography and demographic history of an endangered butterfly in Europe from multilocus markers. Biological Journal of the Linnean Society, 126, 95–113. https://doi.org/10.1093/biolinnean/bly160

Dierks K (2006) Beobachtungen zur Larvalbiologie von *Coenonympha oedippus* (Fabricius, 1787) im Südwesten Frankreichs (Lepidoptera: Satyridae). Entomologische Zeitschrift, 116, 186–188.

Gomez-Garrido J (2024) CLAWS (CNAG's Long-read Assembly Workflow in Snakemake) [Computer software]. https://doi.org/10.48546/WORKFLOWHUB.WORKFLOW.567.2

Greenwood MP, Capblancq T, Wahlberg N, & Després L (2024) Whole genome data confirm pervasive gene discordance in the evolutionary history of Coenonympha (Nymphalidae) butterflies. Molecular Phylogenetics and Evolution, https://doi.org/10.1016/j.ympev.2024.108222

Jugovic J, Zupan S, Buzan E, & Čelik T (2018) Variation in the morphology of the wings of the

endangered grass-feeding butterfly *Coenonympha oedippus* (Lepidoptera: Nymphalidae) in response to contrasting habitats. European Journal of Entomology, 115, 339–353. https://doi.org/10.14411/eje.2018.034

Lhonoré J (1998) Biologie, écologie et répartition de quatre espèces de Lépidoptères Rhopalocères protégés (Lycaenidae, Satyridae) dans l'ouest de la France. Rapports d'études de l'OPIE vol. 2. OPIE, Guyancourt.

Manni M, Berkeley MR, Seppey M, Simão FA, & Zdobnov EM (2021) BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. Molecular Biology and Evolution, 38:4647–4654. https://doi.org/10.1093/molbev/msab199

Mazzoni C, Ciofi C & Waterhouse R (2023). Biodiversity: An atlas of European reference genomes. Nature 619:252–252. https://doi.org/10.1038/d41586-023-02229-w

Rhie A, Walenz BP, Koren S & Phillippy AM (2020) Merqury: Reference-free quality, completeness, and phasing assessment for genome assemblies. Genome Biology 21:245. https://doi.org/10.1186/s13059-020-02134-9

van Swaay C, Wynhoff I, Verovnik R, Wiemers M, López Munguira M, Maes D, Šašić M, Verstrael T, Warren M, & Settele J (2010) *Coenonympha oedippus* (Europe assessment). The IUCN Red List of Threatened Species 2010: e.T5100A11114633. Accessed on 02 December 2024

Zupan S, Jugovic J, Čelik T, & Bužan E (2021) Population genetic structure of the highly endangered butterfly *Coenonympha oedippus* (Nymphalidae: Satyrinae) at its southern edge of distribution. Genetica, 149, 21–36. https://doi.org/10.1007/s10709-020-00108-0