

ORIGINAL ARTICLE

Systematic
Entomology

Genetic structure, spectral characteristics and evolutionary history of butterflies in the European *Polyommatus eros* species group (Lepidoptera, Lycaenidae)

Nok Ting Lam¹ | Vladislav Ivanov^{1,2} | László Péter Biró³ | Zsolt Bálint^{3,4} |
Krisztián Kertész³ | Gábor Piszter³ | Leonardo Dapporto⁵ | Roger Vila⁶ |
Marko Mutanen¹ | Vlad Dincă^{1,7}

¹Ecology and Genetics Research Unit, University of Oulu, Oulu, Finland

²Institut für Biowissenschaften, University of Rostock, Rostock, Germany

³Institute of Technical Physics and Materials Science, HUN-REN Centre for Energy Research, Budapest, Hungary

⁴Hungarian Natural History Museum, Budapest, Hungary

⁵ZEN lab, Dipartimento di Biologia, University of Florence, Sesto Fiorentino, Italy

⁶Institut de Biologia Evolutiva (CSIC-Universitat Pompeu Fabra), Barcelona, Spain

⁷“Grigore Antipa” National Museum of Natural History, Bucharest, Romania

Correspondence

Vlad Dincă, “Grigore Antipa” National
Museum of Natural History, Șos. Kiseleff
1, RO-011341, Bucharest, Romania.
Email: vlad.dinca@antipa.ro

Nok Ting Lam, Ecology and Genetics Research
Unit, University of Oulu, Oulu, Finland.
Email: nok.lam@oulu.fi

Funding information

Research Council of Finland, Grant/Award
Numbers: 324988, 328895, 352652; National
Research, Development and Innovation Office
of Hungary, Grant/Award Number: 143037;
Ministry of Culture and Innovation of
Hungary, Grant/Award Number:
TKP2021-NKTA-05; Italian Ministero della
Transizione Ecologica; Agencia Estatal de
Investigación, Grant/Award Number:
PID2022-139689NB-I00; Departament de
Recerca i Universitats, Generalitat de
Catalunya, Grant/Award Number: 2021-SGR-
00420; Suomen Kulttuurirahasto,
Grant/Award Number: 00250568; Oulun
Yliopiston Tukisäätiö, Grant/Award Number:
20260116; Ministerio de Ciencia, Innovación y
Universidades, Grant/Award Number:
PID2022-139689NB-I00

Abstract

Species complexes with allopatric populations scattered across a wide geographic area represent some of the most challenging topics for taxonomy, as well as for the prioritisation of limited conservation resources. A prime example is represented by cold-adapted butterflies of the *Polyommatus eros* species group, with a fragmented distribution across Eurasia. In Europe, this group was often regarded as comprising two species (*Polyommatus eros* (Ochsenheimer) and *Polyommatus eroides* (Frivaldszky)), although different taxonomic views recognised from one to four species. Based on limited material and genetic data, a single species (*P. eros*) was recognised by the most recent checklists of European butterflies. Here, we use genome-wide data (double digest RAD sequencing) coupled with information on wing spectral characteristics to investigate evolutionary relationships among European butterflies of the *P. eros* species group. Taken together, results of phylogenetic analyses, genetic structure, gene flow, species delimitation and wing reflectance revealed a complex pattern likely shaped by the Pleistocene glaciations. While differentiation between some populations is relatively low, two lineages (one detected in mountains at the border between Albania and Montenegro, the other, assignable to taxon *menelaos*, endemic to southern Greece) appear to be undergoing accelerated differentiation. The endangered taxon *menelaos* is particularly differentiated in terms of genetics and wing spectral properties, the latter suggesting reinforcement with respect to Balkan *eroides*. Although assessed at a continental scale, a more complete

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2026 The Author(s). *Systematic Entomology* published by John Wiley & Sons Ltd on behalf of Royal Entomological Society.

understanding of the intricate patterns displayed by this species group will only be possible with the inclusion of extra-European populations.

KEYWORDS

admixture, allopatry, double digest RAD-sequencing, spectral properties, taxonomy, *Wolbachia*

INTRODUCTION

As the fundamental unit of taxonomy, the species represents a key element in most biological research fields, including conservation (Sandall et al., 2023). However, classifying organisms into species remains one of the most challenging and debated tasks for biologists (Godfray, 2002). This is particularly true when dealing with taxa occurring in allopatry, when the application of standardised criteria to define species is extremely difficult, even when genomic data is available (Joshi et al., 2024; Lam et al., 2024). Species delimitation is further complicated by the absence of a consensus over the species concept to be applied, because different concepts do not yield the same species boundaries. Therefore, increasing knowledge on taxa with allopatric populations across the tree of life is critical for the search of potentially emerging patterns that may lead to more informed decisions of defining species under such circumstances. Ideally, new knowledge should be generated incorporating large amounts of genetic data, since such datasets provide valuable insights regarding population genetic structure, evolutionary history and levels of gene flow. Additionally, information regarding host-associated microbiota can provide valuable knowledge on potential additional barriers to gene flow, as well as for the conservation and management of threatened populations (Trevelline et al., 2019). In particular, the maternally inherited bacteria *Wolbachia* (Rickettsiales), which are found in numerous arthropods, have the potential to influence the reproductive dynamics of their hosts. They may induce male killing, parthenogenesis, the feminisation of genetic males or cytoplasmic incompatibility (CI) (Kaur et al., 2021; Werren et al., 2008). Some of these phenomena (e.g., CI) can also act as barriers to gene flow, thus promoting speciation (Bruzese et al., 2022; Elias-Costa et al., 2019). Therefore, screening arthropod populations for the presence of *Wolbachia* may be regarded as good practice when addressing speciation and/or conservation of various populations and taxa.

European butterflies include several examples of taxa with geographically isolated populations, many of which are still subject to taxonomic debate. Some of these isolated populations are also threatened by a variety of factors and are protected at national or international level. Hence, taxonomic uncertainties and the lack of a good understanding of genetic structure can also negatively impact conservation measures and the prioritisation of the limited resources (Lam et al., 2024; Vila et al., 2010).

Among butterflies, one of the best examples is provided by subtribe Polyommata (Lycaenidae), a speciose group associated with grasslands including numerous species with debated taxonomy (e.g., Kandul et al., 2004; Lam et al., 2024; Lukhtanov et al., 2015; Vila et al., 2011). Although the higher classification of this group has been

the subject of relatively recent revisions using molecular data (Talavera et al., 2013, 2022), most of the Western Palearctic taxa within this clade have never been assessed using large amounts of genetic data. In Europe, one of the contentious cases involves taxa of the *Polyommatus eros* species group, which include cold- and dry-steppe adapted, allopatric populations occurring from mountains of western and central Europe to steppes of Ukraine and southern Russia, the Caucasus, Türkiye, mountains of northern and central Asia, and to the Far East (Bálint, 1995; Carbonell, 1993, 1994; Coutsis & Tshikolovets, 2025; Tshikolovets, 2011; Tuzov, 2000). Until recently, the European populations of this taxon have usually been regarded as two species: *Polyommatus eros* (Ochsenheimer) occurring in the Pyrenees, Massif Central, Alps, Apennines and locally in the northern Balkans, and *Polyommatus eroides* (Frivaldszky) distributed in the Balkan region (e.g., Higgins, 1975; Higgins & Riley, 1970; Tolman & Lewington, 2008). A third taxon, endemic to the Peloponnese (Taygetos Mountains, southern Greece), has sometimes been regarded as a species, namely *Polyommatus menelaos* Brown (e.g., Coutsis & Tshikolovets, 2025; Lafranchis, 2003; Tolman & Lewington, 2008). On the other hand, some authors (e.g., Haahtela et al., 2011; Tshikolovets, 2011), regarded the entire distribution of the *P. eros* species group as representing a single, almost transpalearctic species (*P. eros*), with various subspecies. Additionally, a few other subspecies have been mentioned from Europe by some authors (*boisduvalii* (Herrich-Schäffer) and *italica* (Oberthür)) (e.g., Leraut, 2016; Tshikolovets, 2011), *boisduvalii* also being occasionally listed as a species (e.g., Akimova, 2009).

Taxonomic decisions for this species group have usually been made based on subtle morphological differences, such as the blue colouration of male wings. Indeed, this character, both in visible and ultraviolet wavelengths, can play an important role as a prezygotic barrier (e.g., Fordyce et al., 2002; Jiggins et al., 2001; Lukhtanov et al., 2005). There is growing evidence that the structural colour of the males plays an important biological role, as the male colours are finely tuned among sympatric and synchronic taxa (Bálint et al., 2012). Additionally, the spectral characteristics can be remarkably well preserved across time and space (Kertész et al., 2019). Nevertheless, to our knowledge, these differences have never been statistically assessed for the *P. eros* species group at the continental scale (cf. Bálint et al., 2007).

Molecular data available for these taxa mostly consists of DNA barcodes (Dapporto et al., 2022; Dincă et al., 2021), which showed fairly low variation across Europe (1.3% maximum p-distance). Very limited nuclear DNA data (a single marker, internal transcribed spacer 2, ITS2) has also been analysed, the results being in agreement with those based on DNA barcodes (Vodolazhsky et al., 2009; Wiemers

et al., 2010). Recent checklists of European butterflies (Dapporto et al., 2022; Wiemers et al., 2018), recognise a single species in Europe (*P. eros*), although this taxonomic arrangement has not yet been corroborated with genomics. Moreover, south-eastern populations of *P. eros* are of high conservation concern in Europe, being included in Annexes II and IV of the Habitats Directive (92/43/EEC), where they are listed as a distinct species (*P. eroides*).

In this study, we used double digest restriction site associated DNA sequencing (ddRADseq) and male wing spectral properties to (1) investigate evolutionary relationships, (2) assess levels of admixture and (3) *Wolbachia* infection, as well as (4) species boundaries in the European representatives of the *P. eros* species group.

METHODS

Specimens

We analysed 52 specimens originating from Europe representing the *Polyommatus eros* species group, as well as two *Polyommatus icarus* (Rottensburg) and one *Polyommatus celina* (Austaut) as outgroup (Table S1). The sampling locations are representative for the European distribution of the target taxa. All specimens have the bodies preserved in absolute ethanol at -20°C , while the wings are stored in glassine envelopes as vouchers. The specimens originate from a public research collection stored at the Institute of Evolutionary Biology (CSIC-UPF), Barcelona, Spain.

DNA extraction and ddRAD library preparation

We extracted genomic DNA from approximately half of the thorax or from legs (if the former was absent) with the Qiagen DNeasy Blood & Tissue Kit (Qiagen, USA) following the manufacturer's protocol. We visually checked the extractions for quality and quantity using 2% agarose gels before proceeding to library preparation. We used the Qiagen REPLI-G Kits (Qiagen, USA) to amplify the genomic DNA for 14 samples with low genomic DNA concentration (gDNA, Table S1). Subsequently, we performed a purification step with AMPure XP magnetic beads (Beckman Coulter Life Sciences, USA) according to the manufacturer's protocol. We prepared the ddRADseq library following the protocol described in Lee et al. (2018) and Ivanov et al. (2024) (Figure S1a,b). We used the PicoGreen kit (Molecular Probes, USA) to measure the amplified gDNA and digested 200 ng of gDNA using the PstI and MseI digestive enzymes. After digestion, we ligated a pair of customised indexes to both ends of the DNA fragment. The PstI end contained both the Illumina primer sequence and our own unique five base pairs individual indexes, while the MseI end only carried the Illumina primer sequence. After ligation, we multiplexed up to 13 samples together and carried out size selection. We used the BluePippin (Sage Science, USA; 2% agarose cartridge), an automated size selection machine, to select fragments with a mean size of 300 bp.

The fragments of the right size underwent PCR enrichment. During the PCR enrichment, we used a primer carrying the Illumina adapter sequence and an additional library index sequence at the MseI end. The DNA fragment of each sample was indexed with a unique index combination at both of its ends (Table S1). Before sequencing, we multiplexed all samples into a single library. The final library was purified and sequenced by MacroGen, Korea, with the HiSeqX platform.

Bioinformatics and datasets

We demultiplexed raw Illumina paired-end reads in a two-step manner according to the unique combination of library and individual indexes (or tags), allowing zero mismatches: First, we demultiplexed the library indexes with Demultiplex developed by Jeroen F.J. Laros and published in GitHub (<https://github.com/jfjlaros/demultiplex#demultiplex-fastafastq-demultiplexer>), then by the individual indexes with lpyrad v0.9.31 (Eaton & Overcast, 2020). Information regarding indexes used is available in Table S1. Raw reads for each sample were first filtered against the built-in potential contaminants reference database of Kraken v2.0.8 (Wood et al., 2019) using default parameters, as done by Lam et al. (2024). Clean reads (reads with no match) were separated from contaminated reads matching the Kraken database. The clean reads were used as input for the lpyrad pipeline.

To decide which assembly approach was more suitable for analyses, we explored both de novo and reference assemblies of the full dataset (i.e., with outgroup included) using lpyrad. The reference assembly was mapped to the reference genome of *Polyommatus icarus* (GCA_937595015.1, Lohse & Vila, 2023), the closest species to *P. eros* in Europe. Furthermore, as previous works demonstrated the effects of different levels of missing data on the results (reviewed in Hemstrom et al., 2024; Shafer et al., 2017), we generated four datasets from each assembly approach by increasing the minimum number of samples per locus (m value) from 15 to 45 with increments of 10 per dataset, and retained the other parameters as default (Figure S1b, Table 1). The full dataset to be used for further analyses was selected based on two criteria: (1) minimal missing data and (2) enough phylogenetic signal indicated by the topology stability and robustness of node supports in the maximum likelihood tree (ML tree). Applying these criteria resulted in the selection of the $m = 35$ reference assembly, named "full dataset" hereafter (assembly ID 1 in Table 1).

Since some downstream analyses cannot include outgroup taxa, we branched two ingroup-only reference datasets (with $m = 27$ and $m = 42$) from the full dataset using the "branch" function in lpyrad. Among these two, we picked the dataset with less missing data ($m = 42$, named "ingroup dataset" hereafter) (assembly ID 2 in Table 1) for analyses which do not handle missing data very well (Figure S1, Table S2). In summary, downstream analyses were based on the full dataset, with the exception of analyses that require the ingroup dataset because including outgroup taxa would violate the assumption of the analyses (Figure S1, Table S2).

TABLE 1 Overview of all the explored datasets and the basic statistics of each dataset. Datasets in bold were selected for downstream analyses.

Assembly ID	Assembly method	Taxa included	Number of samples	Clustering threshold	m value	Reconstructed loci	Sequence length	Sequence missing data (%)	SNPs	SNPs missing data (%)	Unlinked SNPs
1	Reference	All taxa	55	0.85	35	5995	1,290,744	33.71	82,275	28.63	5958
2	Reference	Ingroup only	52	0.85	42	2938	623,009	25.73	30,274	18.82	2916
3	Reference	All taxa	52	0.85	15	13,117	2,652,214	49.80	150,096	42.43	12,904
4	Reference	All taxa	55	0.85	25	9238	1,923,509	42.12	117,235	35.35	9144
5	Reference	All taxa	55	0.85	45	2637	595,490	28.60	39,503	21.21	2628
6	Reference	Ingroup only	52	0.85	27	7982	1,606,272	37.05	74,705	30.94	7887
7	de novo	All taxa	55	0.85	15	10,585	1,793,857	42.36	101,604	45.18	10,517
8	de novo	All taxa	55	0.85	25	6804	1,168,204	34.18	72,740	33.17	6792
9	de novo	All taxa	55	0.85	35	3843	663,674	24.25	43,757	23.89	3834
10	de novo	All taxa	55	0.85	45	1245	215,826	13.77	15,072	13.66	1244

Different formats of lpyrad output

lpyrad can produce several file formats after variant calling and, depending on analysis, we used several file formats, which are summarised in Figure S1c. A metadata file containing sample ID, loci number, variant site coordinates and other metadata was written into a hdf5 formatted file (.hdf5). All the aligned loci were called separately in a PHYLIP file (.loci). The single nucleotide polymorphisms (SNPs) were written out into several formats: variant calling file (.vcf), SNPs alignment PHYLIP file (.snp), diploid STRUCTURE file (.str). We used a function in lpyrad to sample one random SNP per locus to prevent over-representation of SNPs, and the SNPs were assumed unlinked. As a result, some outputs had a subsampled matrix with less SNPs and these outputs were denoted with an “u” before the normal output (e.g., an unlinked structure file is denoted as “.ustr”) (Eaton & Overcast, 2020).

Phylogenetic inference

We built phylogenetic trees to explore evolutionary relationships between our samples with *P. icarus* and *P. celina* as outgroup. We used the SNPs alignment file (.snp) from the full dataset as input for phylogenetic analyses and used IQ-TREE v2.2.2.7 to reconstruct ML trees (Minh et al., 2020). The DNA substitution model applied for each tree was selected by the IQ-TREE ModelFinderPlus (Kalyaanamoorthy et al., 2017) according to the Bayesian information criterion (BIC). We also enabled the -bnni option to correct ultrafast-bootstrap support in case of severe substitution model violation (Hoang et al., 2018). Ultrafast bootstrap and the Shimodaira–Hasegawa approximate likelihood ratio test (SH-aLRT) (5000 replicates in each case) were run for all trees. A branch was considered well supported when there was more than 95% support in ultrafast bootstrap and more than 80% support in SH-aLRT as per the recommendation of the IQ-TREE manual.

Multivariate analyses (PCA, t-SNE and DAPC)

We ran principal component analysis (PCA) with lpyrad v0.9.31 for the ingroup and full datasets to understand global and local variance within our data and to set a foundation for further analyses to investigate biological causes for the patterns observed. The PCA (Patterson et al., 2006) and t-distributed stochastic neighbour embedding (t-SNE) (van der Maaten & Hinton, 2008) summarise variances observed across the unlinked SNPs alignment into manageable planes for visualisation. The latter is capable of identifying non-linear patterns in the data.

We used the metadata file (.hdf5) as input and followed the lpyrad manual to run the PCA. The assignment of individuals to taxa/lineages (Table S1) did not affect the calculations but was necessary for visualisation and data filtering. We only allowed 25% missing data and performed 50 subsampling replications per run to remove the potential linkage effect of our data. For the full dataset, we used its

PCA matrix to perform t-SNE, following the lpyrad manual. We allowed subsampling, set the perplexity to 4.0 and performed 10,000 interactions.

In addition to PCA, which do not need predefined groups, we used discriminant analyses of principal components (DAPC) to quantify the distinctiveness of each predefined group (Jombart et al., 2010). DAPC also facilitated the visualisation of inter-population variance and the proportion of individual alleles contributing to population structure. The DAPC analysis was done with a custom R script (Data S1, and https://github.com/SamuellNT/Peros_manuscript_scripts) using the adegenet package (Jombart, 2008; Jombart & Ahmed, 2011). We ran the analysis with the unlinked SNP alignment from lpyrad to reduce the effect of missing data. Each individual was assigned to groups that were suggested by the find.clusters function according to the lowest BIC. The DAPC results were visualised with the ggplot2 package (Wickham, 2016).

Inbreeding and population differentiation (F_{IT} and F_{ST})

Using the ingroup dataset, we calculated the observed heterozygosity (H_o) and the F_{IT} coefficient of each individual, to obtain a rough estimate of the level of inbreeding in the studied individuals against the whole ingroup dataset (Weir & Cockerham, 1984; Wright, 1951). We obtained the average F_{IT} of each taxon group for an overview of the inbreeding level of each taxon/lineage (as defined in Figure 1 and Table S1). We also calculated F_{ST} between taxa/lineages pairs (Weir & Cockerham, 1984; Wright, 1951), using the full and ingroup datasets, to measure the genetic differentiation of each taxon/lineage. All the calculations were done using the corresponding vcf files with vcftools 0.1.17 (Danecek et al., 2011).

Admixture and gene flow

We used STRUCTURE v2.3.4 (Pritchard et al., 2000) to explore admixture patterns in different predefined K-clusters. We used the unlinked SNPs STRUCTURE file (.ustr) of the ingroup dataset as input. We used StrAuto to automate (Chhatre & Emerson, 2017) and parallelise the software and carried out 1,000,000 Markov Chain Monte Carlo interactions, with 10% burn-in and 10 replicates in each run. To make sure we did not underestimate potential genetic admixture sources, we set the maximum number of K-clusters to eight (taking into account current taxonomy and the topology of the ML tree). We used Structure-Harvester to calculate the ad hoc statistic ΔK and the model with the highest ΔK was inferred as preferred (Earl & vonHoldt, 2012; Evanno et al., 2005). The results of alternative K-clusters were illustrated using the CLUMPAK pipeline (Kopelman et al., 2015) and are presented in Figure S2.

We ran D-statistics (Malinsky et al., 2021) to detect the presence of potential gene flow between triplets of tested taxa, which helps outline the general evolutionary history of the *P. eros* species group. This analysis inferred the deviation of incomplete lineage sorting

against the triplet model as the result of gene flow. We ran the analysis with the full dataset and used *P. icarus* as outgroup (Data S1). The analysis was run assigning samples into groups as *eros*, AL/ME, *eroides*, *menelaos*, *boisduvalii* and *italica*, as shown in Figure 1a. We did not assume any tree topology for the test in order to explore all possible triplet combinations.

We also ran D-statistics using *P. celina* as outgroup to test whether taxon *italica* or the eastern Alps (Austrian) *eros* show notable gene flow with *P. icarus*. For this purpose, we tested the following triplets: (i) Pyrenean *eros* (11 samples), *italica* (3 samples), *P. icarus* (2 samples); (ii) the same dataset as (i), but replacing *italica* with the eastern Alps (Austrian) *eros* samples (2 specimens).

We estimated possible current or historical gene flow between taxa using PhyloNetworks. 0.16.4 (Solís-Lemus et al., 2017). Unlike D-statistics, PhyloNetworks infers gene flow by detecting the hybrid edge through a network based on gene trees of all taxa in the dataset. It can also accommodate models with multiple hybridisation events. We tested two datasets: (a) *P. icarus* as outgroup; (b) only ingroup taxa. The detailed pipeline is available in the Data S1, as well as in GitHub (https://github.com/SamuellNT/Peros_manuscript_scripts). Briefly, the loci alignment PHYLIB file (.loci) was split into separate PHYLIB files for each locus using a slightly modified script written by Bradley T. Martin (https://github.com/btmartin721/file_converters/blob/master/loci2phylib.py). Then, ML trees were computed for each PHYLIB file using RAXML v. 8.2.12 (Stamatakis, 2014) and the resulting best trees were used as input for computing consensus trees with ASTRAL v. 5.7.8 (Mirarab & Warnow, 2015). The best ML trees and the ASTRAL tree were used as input for PhyloNetworks. We tested from zero to five hybrid edges for each dataset. The two networks with the lowest loglikelihood were considered as the best and second-best network reconstructed by PhyloNetworks.

Demographic inference

We estimated changes in effective population size (N_e) for each ingroup taxon/lineage with Stairway Plot 2 v2.1.3 (Liu & Fu, 2020). We removed taxon *boisduvalii* from the analysis because only two samples from the same locality were available and our attempts to run analysis for this group showed a lack of signal for the estimation. Invariant sites were necessary for easySFS (Coffman et al., 2016) to calculate site frequency spectrum (SFS), but the vcf outputs from lpyrad did not contain any invariable sites. We thus used bcftools 1.21 (Danecek et al., 2021; Li, 2011; Li et al., 2009) to generate a new vcf file (all-sites vcf file). We applied a filter only allowing 36% missing data to maintain roughly the same level of missing data across datasets in other analyses (matching the m35 filtering in lpyrad) and retained only the invariant and biallelic sites (more information can be found in the Data S1). As Stairway Plot assumed SNPs to be unlinked (Liu & Fu, 2015) and our loci are 300 bp long on average, we sampled one SNP from every 300 bp window (—thin 300) using vcftools 0.1.17 to reduce the linkage of each SNP (Danecek et al., 2011). While calculating SFS, sites with missing data will be dropped out by

easySFS, therefore, by reducing the number of sample (downsampling), we would utilise more data even with less samples be used to infer SFS (Marth et al., 2004). We used the built-in function in easySFS for downsampling while calculating folded SFS for each taxon using the all-sites vcf file. We chose the number of samples during downsampling based on the highest number of sites available for SFS projection in easySFS and we calculated folded SFS, instead of unfolded SFS, because we could not reliably infer the ancestral state of the alleles. We removed the first SFS bin of the one-dimension SFS from easySFS for Stairway Plot 2 input as instructed by the manual. We assumed all taxa evolved at the same rate and used a mutation rate of 2.9×10^{-9} , which was estimated in *Heliconius melpomene* (L.) (Keightley et al., 2015). Other parameters used in Stairway Plot 2 can be found in the Data S1.

Species delimitation

We used BFD* implemented in SNAPP, BEAST v2.7.1 for species delimitation (Leaché et al., 2014). BFD* is a Bayesian multispecies coalescent species delimitation approach implemented in SNAPP (Bryant et al., 2012; RoyChoudhury et al., 2008). This method recognises the optimal species assignment by ranking multiple species delimitation models according to their corresponding marginal likelihoods (Leaché et al., 2014). In this study, we compared 13 different species assignment models, grouping specimens into one to five species (see Table S3 for assignments). We considered the current taxonomy as the null model, recognising a single species in Europe (*P. eros*) according to the most recent checklists of European butterflies (Dapporto et al., 2022; Wiemers et al., 2018). The marginal likelihood estimation (MLE) of each tested model was calculated using the path sampler option implemented in SNAPP. The alternative models were then ranked against the null model (current taxonomy) according to the Bayes factor (BF) scores following the scale presented in Kass and Raftery (1995). Because better models should return higher (more positive) BF scores, the model with the highest BF score was considered as the best one.

Using the full dataset failed to converge for any two species model within reasonable computational time. Therefore, we ran SNAPP using a subset of specimens. For all taxa except for *eros*, five specimens (if available) were used, selecting the samples with the highest number of loci in the ddRADseq dataset. For *eros*, we retained six specimens, prioritising specimens with the highest number of loci, but also their geographic origin (two representatives from the Pyrenees, Western Alps and Eastern Alps, respectively). Two specimens of *P. icarus* were included as outgroup, resulting in a dataset of 28 specimens (Table S1). We only retained SNPs present in all taxa and removed non-biallelic sites with a customised R script (Data S1 and https://github.com/SamuelLNT/Peros_manuscript_scripts) using the packages bioseq v0.1.4 (Keck, 2020), ape v5.7.1 (Paradis et al., 2004) and phrynomics v2.0 (Leaché et al., 2014). This step allows taxonomic models to be compared with one another because they use the same number of sites for the inference. We used a final alignment matrix

with 2836 sites to prepare XML files with BEAUti, BEAST v2.7.1 for all models with the corresponding species assignment. (Table S3).

The priors of the models were calculated based on the alignment matrix, taking the average uncorrected distance of two different taxa as the expected divergence (θ), estimating the tree height (2θ) as the speciation rate under Yule model (λ), and its hyperprior set as approximately $\lambda/10$. Therefore, all models were run with the following settings: $\theta = 21/250$ (α/β , 0.084), $\lambda = 0.16$ with hyperprior set as $2/200$ (α/β , 0.01) sampled from the input data, mutation rate $U=V = 1$, coalescent rate = 10, 24 path sampler steps, MCMC chain length $\geq 200,000$ until the Effective Sample Size (ESS) reaches at least 300 for each step and burn-in = 10%. The results were inspected with Tracer 1.7.1 (Rambaut et al., 2018).

We also used CLADES (Pei et al., 2018) to explore potential species boundaries among taxa. CLADES is an unsupervised machine learning programme utilising the python package libsvm to classify populations into species by summarising their private position (species specific SNPs) characteristics. In contrast to SNAPP, the general model in CLADES, which was trained by the developer, can handle gene flow and migration between populations prior to speciation. We first used the python script, loci2phyliip.py, written by Bradley T. Martin (https://github.com/btmartin721/file_converters/blob/master/loci2phyliip.py) to split each locus in the loci file into separate PHYLIIP files. Since CLADES is very sensitive to missing data and cannot handle invariant loci, we used bcftools to obtain loci coordinates which contained SNP(s) for at least 50 samples (out of a total of 55). Subsequently, we retained the PHYLIIP files of the selected loci and combined the loci into a single file. This file with all the selected loci was used as the input for CLADES. Then we removed the outgroup (*P. icarus* and *P. celina*) sequences from the alignments to reduce the effect of taxon-based missing data. We assigned population tags to each individual according to taxon/lineage (Table S1) and used the general model ("all") provided by CLADES to check whether the populations could be considered as species. As an output, CLADES provides a probability (from zero to one) for each population pair to determine if they are the same species, then it summarises the probabilities and produces a taxonomic proposal. The protocol and commands used to run CLADES are available in Github (https://github.com/SamuelLNT/Peros_manuscript_scripts).

Wolbachia infection screening

The presence of the maternally inherited bacterium *Wolbachia* was surveyed using specific primers for markers *coxA* and *ftsZ* (Table S4) based on the same DNA extractions used for ddRADseq. These markers, which are part of the multilocus sequence typing system (MLST) for *Wolbachia*, are extensively used to detect infection in a wide array of insects (Baldo et al., 2006). The presence of *Wolbachia* was assessed on agarose gels for all specimens in the dataset and amplicons were sequenced for a subset of 22 specimens representative of all the taxa in the ddRADseq dataset. The resulting amplicons

were blasted against the PubMLST database (Jolley et al., 2018) to infer the corresponding alleles.

As a second approach, the presence of *Wolbachia* was detected using the summary report of the same Kraken run that was used previously to screen for contaminants in the ddRADseq dataset. By comparing this data with the results of the previous screening approach (*coxA* and *ftsZ*), it was possible to determine the suitability of ddRAD-seq to detect *Wolbachia* infection.

Analysis of wing spectral properties

Spectral measurements were performed on the dorsal side of the wings for 45 male specimens with wing vouchers available (Table S1, Figure S3). For the reflectance measurements, an Avantes (Avantes BV, Apeldoorn, The Netherlands) modular spectrophotometer system was used consisting of a UV-visible light source (AvaLight-DH-S-BAL), an integrating sphere for light collection (AvaSphere-30-REFL) and a spectrophotometer (AvaSpec-HSC1024 × 58TEC-EVO) for detection. The components were connected with UV-visible-NIR optical fibres, and a PTFE tile (Avantes WS-2) was used as a reference (Kertész et al., 2021).

RESULTS

Datasets

Out of 10 datasets explored, we selected for further analyses a full (i.e., including outgroup taxa) reference dataset retaining loci shared by at least 35 samples ($m = 35$). This dataset contained 82,275 SNPs and 28.63% missing data. Although it had slightly more missing data compared to the de novo version (Table 1, Table S2), the reference dataset we selected yielded almost double the number of SNPs compared to the de novo version. The ingroup dataset had 30,274 SNPs and 18.82% missing data, and it was extracted from the full dataset by removing outgroup taxa (Table 1). Information of each sample can be found in Table S1.

Phylogeny

Phylogenetic relationships and supports for taxon nodes were congruent in all reference datasets tested (Figure 1, Figure S4). Although not used for further analyses, congruent results were also obtained using the de novo datasets, with the sole exception of the *m45* version (unresolved position of the AL/ME clade) (Figure S4).

Four of the five taxa recognised in Europe at least by some authors were recovered as distinct clades, with very good supports (>99). The exception was the taxon *italica*, which was always recovered as paraphyletic with respect to the clade formed by all other taxa. Within the remaining taxa, *menelaos* was recovered as sister to all other specimens. All *eros* specimens originating from the Pyrenees and the Alps formed a well-supported clade. A second clade

attributable to *eros* consisted of specimens from two neighbouring mountains at the border between Albania and Montenegro (sampling sites about 7 km apart, Table S1). This clade, to which we refer as AL/ME, was sister to the clade formed by *boisduvalii* and *eroides*, thus rendering *eros* paraphyletic.

Multivariate analyses

The first two components of the PCA analysis using the full dataset recovered three main clusters: (1) outgroup, (2) *menelaos* and (3) the remaining taxa (Figure 2a). The first and third components returned two main groups: (1) outgroup and (2) all other taxa (Figure S5a). The second and third components recovered two main clusters: (1) *menelaos* and (2) all other taxa including the outgroup (Figure S5b).

Similarly, the first two components of the PCA analysis using the ingroup dataset also recovered two main clusters: (1) *menelaos* and (2) the remaining taxa (Figure 2b). In the latter cluster, taxon *italica* appeared to be closer to *eros* than to *eroides*, while the AL/ME clade and *boisduvalii* were placed between the *eros* and *eroides* clusters, slightly closer to *eroides*. The first and third components in the ingroup dataset identified three main groups: (1) *menelaos*, (2) AL/ME clade and (3) all other taxa (Figure S5c). The second and third components recovered two main clusters: (1) AL/ME clade and (2) all other taxa, and the AL/ME cluster appeared to be relatively dispersed in other PCA iterations along the second component (Figure S5d).

The t-SNE analysis recovered three main clusters: (1) *menelaos*, (2) AL/ME clade and (3) the rest of the taxa. The third cluster with the remaining taxa was close to the outgroup (Figure 2c).

The DAPC analysis identified five distinct groups representing *eroides* + *boisduvalii*, the AL/ME clade, *eros*, *menelaos* and *italica* (Figure 1a).

Inbreeding and population differentiation (F_{IT} and F_{ST})

The averaged inbreeding coefficients (F_{IT}) in the AL/ME clade and in taxon *menelaos* were notably higher (>0.5) than those in the other taxa (ranging from 0.27 to 0.47, Table S5a) and their H_o values were also the lowest among all the other taxa (AL/ME clade $H_o = 0.046$, *menelaos* $H_o = 0.03$, H_o of the rest ranged from 0.059 to 0.082, Table S5a). The weighted F_{ST} showed the most notable differences ($F_{ST} > 0.25$) in comparisons involving *menelaos* or the AL/ME clade and all the other ingroup taxa, regardless of dataset used (with or without outgroup) (Table S5b,c). The rest of the pairs using the full dataset had F_{ST} ranging from 0.08 to 0.23 and those using the ingroup dataset had F_{ST} ranging from 0.07 to 0.22.

Genetic structure and gene flow

The STRUCTURE analyses recovered admixture from six ancestral lineages ($K = 6$, $\Delta K = 317.3$) (Figure 1a,b, Figure S2). Taxon *menelaos*

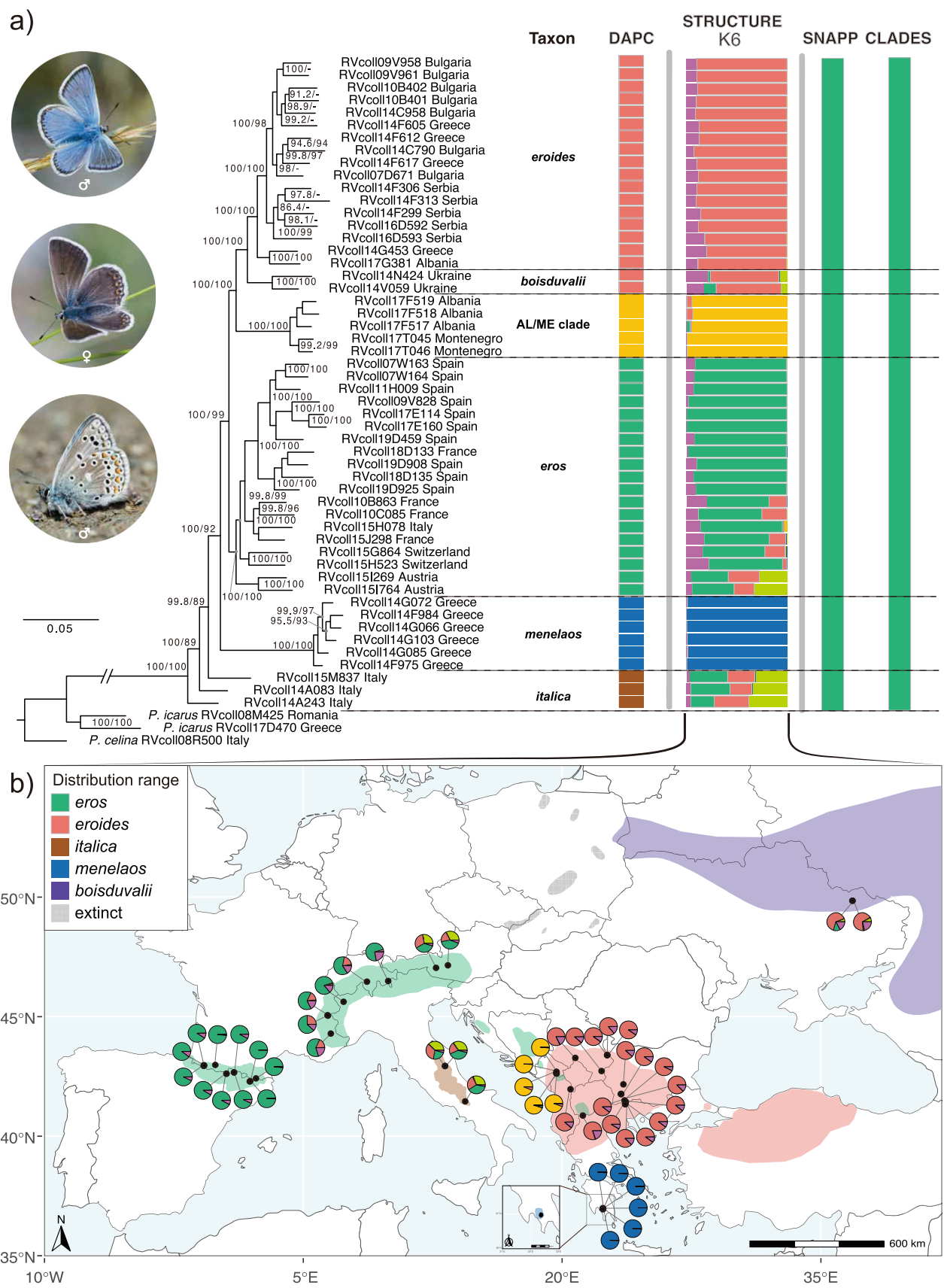
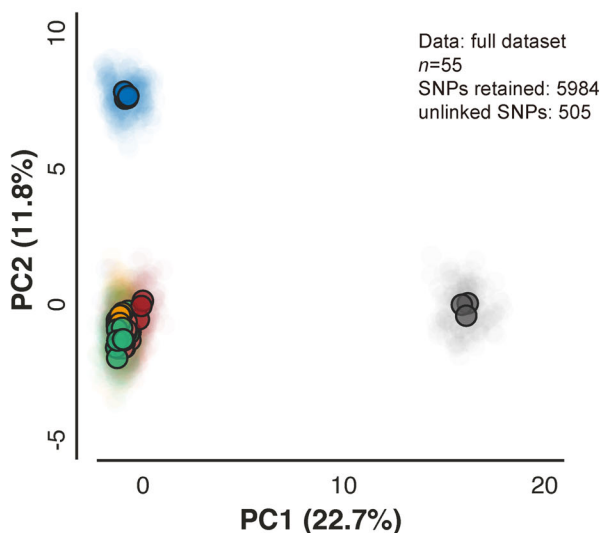
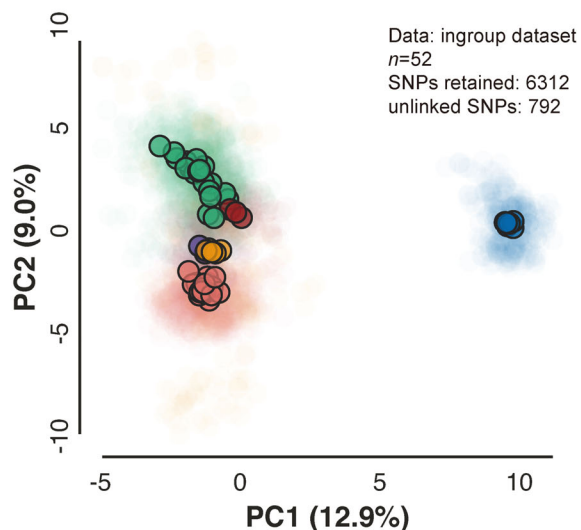


FIGURE 1 Legend on next page.

a) Full dataset PCA: PC1 vs PC2



b) Ingroup dataset PCA: PC1 vs PC2



c) Full dataset t-SNE-PCA

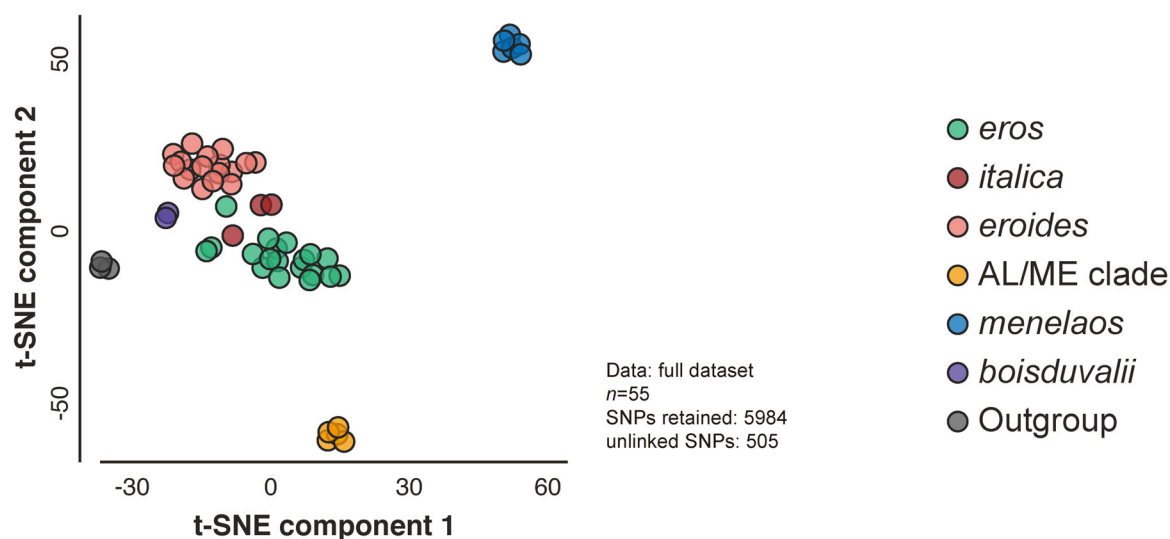


FIGURE 2 (a,b) PCA plots based on the full and ingroup datasets. The shades correspond to the results of different PCA iterations and indicate alternative PCA results. (c) t-SNE plot based on the full dataset. Colours correspond to Figure 1.

showed virtually no admixture. The AL/ME specimens displayed no or very limited admixture (in the latter case from *eroides* or from *eros*). Taxa *eroides* and Pyrenean *eros* had low (to moderate in *eroides*) levels of admixture from the same source (purple colour in Figure 1a,b). The western and central Alps *eros* showed higher levels of admixture from

two main sources, one of them being *eroides*. The other source (purple colour) was the same causing admixture in *eroides* and Pyrenean *eros*. The eastern (Austrian) Alps *eros* exhibited high levels of admixture, mainly from *eroides* and another source (light green). Taxon *italica* displayed an admixture pattern that was very similar to that of *eros* from

FIGURE 1 Phylogenetic relationships, species delimitation and genetic structure for taxa of the *Polyommatus eros* species group in Europe. (a) Maximum likelihood tree based on the ddRADseq reference assembly (55 samples, $m = 35$). Values next to the recovered nodes represent SH-aLRT (left, >70) and ultrafast bootstrap supports (right, >90). Taxa of the *P. eros* species group are indicated. Results of DAPC, STRUCTURE, SNAPP and CLADES correspond to specimens in the ML tree. Images on the left side exemplify (from top to bottom) male, female and ventral view of butterflies representing the *P. eros* species group. Photos: V. Dincă. (b) STRUCTURE results (pies, $K = 6$) mapped to sampling sites (black dots). Approximate European ranges of the taxa are indicated based on mainstream literature, with colours corresponding to the upper left legend. The colours of STRUCTURE pies correspond to those used in a).

the eastern (Austrian) Alps. The taxon *boisduvalii* showed notable levels of admixture from the same sources as *italica* and the eastern Alps *eros*, but the *eroides* ancestry was clearly dominant.

D-statistics detected gene flow in eight triplet combinations. Gene flow was detected between taxa *eros* and *menelaos* ($D = 0.175/0.080$, $P1 = \textit{italica/eroides}$), *eros* and *italica* ($D = 0.161/0.128$, $P1 = \textit{eroides/boisduvalii}$) AL/ME and *menelaos* ($D = 0.176/0.101$, $P1 = \textit{italica/eroides}$), and AL/ME and *italica* ($D = 0.125/0.110$, $P1 = \textit{eroides/boisduvalii}$). We did not detect any significant gene flow between *eroides* or *boisduvalii* and *italica* regardless of what was considered as $P1$ (Table S6a). Similarly, no gene flow was detected between *boisduvalii* and *eros* or *eroides*. Other tested gene flow scenarios were not supported ($Z < 3$).

D-statistics runs screening for hybridisation events between *italica* and *P. icarus* or between the eastern Alps (Austrian) *eros* and *P. icarus* did not detect any significant gene flow (Table S6b,c).

In the PhyloNetworks results using *P. icarus* as outgroup, the best supported network (loglikelihood = 0.207) suggested that an unsampled sister taxon of *eroides* contributed to the hybrid origin of the common ancestor of *menelaos* and *P. icarus* ($\gamma = 0.2520$). Since the hybrid node was located before the divergence of *menelaos* and *P. icarus*, this network cannot be rooted at *P. icarus*. Therefore, according to the manual (Liu & Fu, 2020), we have investigated the second-best supported network (loglikelihood = 0.366), which suggested that an unsampled taxon (sister to all analysed taxa except for *menelaos*) contributed to the hybrid origin of *eroides* ($\gamma = 0.3340$) (Figures S6 and S7).

The ingroup dataset displayed a different pattern, with two hybridisation events. In the best supported network, an unsampled sister taxon to *boisduvalii* had a minor contribution to the hybrid origin of the common ancestor of *italica*, *menelaos* and *eros* ($\gamma = 0.0106$). The gene flow from an unsampled sister taxon of *eros* (or its ancestor) to *italica* ($\gamma = 0.3860$) was the second suggested reticulation. This event was also detected in the second-best supported network ($\gamma = 0.3860$). Additionally, this network also estimated gene flow from an unsampled taxon sister to *boisduvalii*, *eroides* and AL/ME ($\gamma = 0.490$) (Figures S8 and S9).

Demographic inference

The results of Stairway Plot 2 suggested a gradual increase in N_e for all taxa. Since taxa had different sample sizes and used a different number of SNPs (Table S1), the estimated N_e cannot be directly compared with each other, while the variation of N_e is still useful to infer demographic histories. Both *eros* and *eroides* were represented by 28 haploid samples during SFS projection, and their changes in N_e were very similar to each other (Figure S10 and Data S1), showing a fairly steep increase followed by relatively constant levels starting from approximately 400 thousand years ago. The AL/ME clade (10 haploid samples for SFS projections) showed a relatively steep and constant increase in N_e . Taxon *menelaos* (10 haploid samples) displayed an increase in N_e , followed by overall constant levels for the

last 200 thousand years. Taxon *italica* (six haploid samples for SFS projections) had a sudden increase in N_e around 200 thousand years ago, followed by a stable N_e since then (Figure S10).

Species delimitation

All 13 models tested with SNAPP converged, and the best supported model was the one species model (current taxonomy) grouping all taxa as *P. eros*. All the other models had negative BF_s when compared to the current taxonomy (Table S3).

Overall, CLADES considered all taxa as a single species (likelihood = 0.000769). Each taxon combination was also regarded as the same species, except for *eros* – *boisduvalii* where the probability that they represent the same species ($\text{prob}_s = 0.494$) was just below that of them being different ($\text{prob}_d = 0.506$) (Table S7). CLADES performed relatively well ($\text{prob}_s = 0.65\text{--}0.78$) in classifying taxa, especially in comparisons not involving *eros* or *eroides*. When it came to combinations involving *eros* or *eroides*, the prob_s were always below 0.619 (Table S7).

Wolbachia infection

Wolbachia screening based on *coxA* and *ftsZ* markers detected infection in all 52 samples screened (Table S1). A single allele combination was detected (*coxA14* – *ftsZ73*) in 22 specimens, representative for all taxa in the dataset, for which the PCR products have been sequenced.

Using the Kraken database, we detected *Wolbachia* reads in the same 52 samples mentioned above. Fifty-one of the samples had over 0.01% of *Wolbachia* sequences in the raw ddRADseq reads and 34 of them had at least 0.10% (Table S1).

Spectral properties of dorsal wing surfaces

The reflectance spectrum was recorded for each wing of all available male specimens (Table S1). After applying a 20-point fast Fourier transformation smoothing, the peak wavelength values were calculated and presented in the form of a box plot (Figure 3a). Based on the wing reflectance of each individual, averaged spectra were calculated for the taxa of the *P. eros* species group (Figure 3b). Using these two techniques, the structural colour of butterflies with similar photonic nanostructures can be conveniently compared.

The most distinctive sample of the *P. eros* species group is represented by *menelaos* specimens. They displayed an averaged reflectance peak at 470 nm, the highest wavelength among the examined material. Moreover, the shape of their averaged spectra was remarkably asymmetrical showing the lowest reflectivity between 300–400 nm and the highest reflectivity in the range of 600–700 nm, in contrast to the other members of the *P. eros* species group that have an averaged peak reflectance of 440–450 nm and coincident

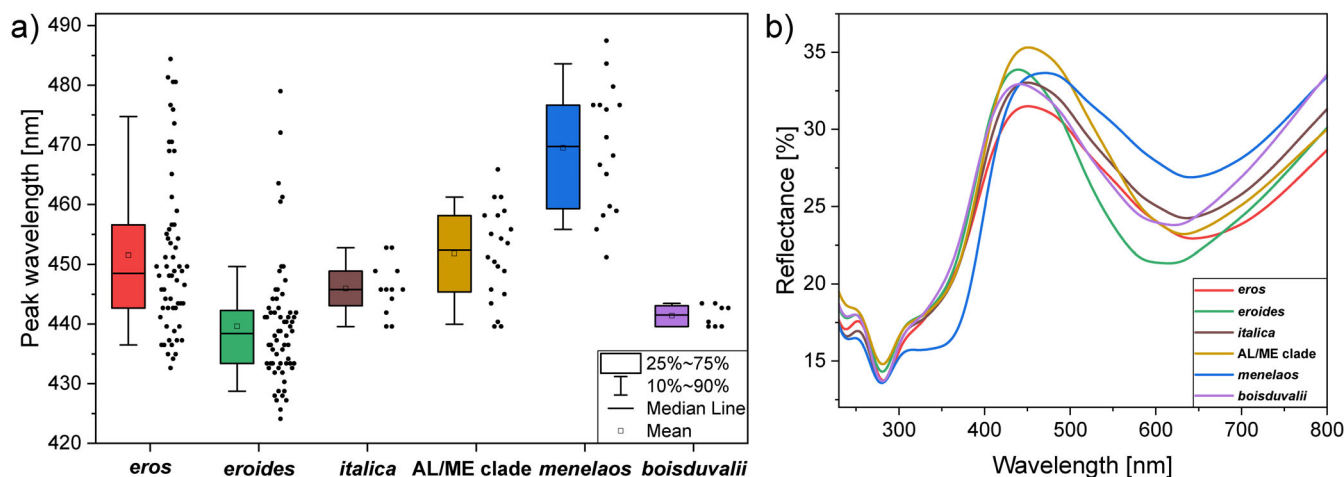


FIGURE 3 (a) Box plots and individual data points of the peak wavelength values of the measured reflectance maxima for taxa of the *P. eros* species group. (b) Averaged reflectance spectra associated with taxa of the *P. eros* species group.

reflectance (higher than that of *menelaos*) in the range of 300–400 nm. Another characteristic group was formed by *eroides* specimens, which, in contrast to *menelaos*, displayed the lowest averaged peak wavelength (slightly below 440 nm) and clearly the lowest reflectivity beyond the peak towards higher wavelengths. The cluster of *boisduvalii* samples was situated close to *eroides* in terms of averaged peak reflectance (slightly over 440 nm), but the shape of its reflectance towards higher wavelengths was similar to that of *eros*, *italica* and the AL/ME clade (i.e., notably higher than *eroides*, but lower than *menelaos*). On the basis of spectral maxima statistical analysis, the samples of *eros* and *italica* clustered very close within a 5 nm range, and the shape of their spectra was almost identical. The spectrum of the AL/ME clade was similar to *eros*, showing congruent reflectance peaks and with an averaged maximum of 450 nm.

DISCUSSION

Evolutionary history of the *P. eros* species group

Overall, populations belonging to the *P. eros* species group appear closely related compared to the outgroup taxa and the patterns of isolation and admixture suggest a complex evolutionary history in Europe possibly shaped by recurrent population expansions and contractions during the Quaternary glaciations. The interpretation of the data is not straightforward due to a series of apparently contrasting results, especially between phylogeny and clustering analyses, which we discuss below.

The monophyly and distinctiveness of the AL/ME clade and especially of *menelaos* are picked as the most striking variation in multivariate analyses, and could be the result of small population sizes (genetic drift) and/or selection processes resulting in accelerated differentiation. This was also supported by F_{IT} , which showed that these two taxa had higher inbreeding indexes than the others. Furthermore, the differentiation of these populations was supported by F_{ST} when they

were compared to other taxa. Although the cause of the accelerated differentiation among these taxa with restricted range remains unclear, based on our Stairway Plot results, differentiation caused by major demographic events such as bottlenecks or rapid expansions is rather improbable. In the case of *menelaos* (endemic to Mt. Taygetos in southern Greece), the hypothesis of accelerated differentiation is also supported by the spectral characteristics of male wings. Males of this taxon show the least UV reflectance (ca 300–400 nm) and the highest reflectance in the orange-red spectrum (ca 600–650 nm) among all examined taxa. Moreover, these differences are most pronounced with respect to taxon *eroides*, which is currently present in central and northern Greece (Pamperis, 2009). This taxon may have had secondary contact with *menelaos* leading to low fitness hybrids and resulting in reproductive character displacement by reinforcement. Indeed, male wing colour plays an important role as a premating signal in Lycaenidae (Bálint et al., 2012; Bernard & Remington, 1991; Piszter et al., 2016). As such, interference in successful breeding due to the presence of other closely related species can act as a strong selective force capable of leading to reproductive character displacement and, in certain cases, to reinforcement (Hinojosa et al., 2020; Lukhtanov et al., 2005). Notably, as shown by the averaged reflectance spectra, the differences between *menelaos* and the other taxa, and in particular *eroides*, cannot be attributed simply to differences in melanin content. A higher melanin content in *menelaos* (potentially explaining the lower reflectance in the UV spectrum) would also generate a lower reflectance in the range of 700–800 nm, which is not the case. This hints at differences in the photonic nanoarchitecture of the wings, which is usually associated with different species or populations experiencing little or no gene flow (Bálint et al., 2012; Kertész et al., 2019).

It is however difficult to state with certainty where *menelaos* and *eroides* started to diverge. One possibility is that differentiation was triggered by the isolation of *menelaos* in the Peloponnese, as the latter is a glacial refugium and a genetic endemism hotspot for butterflies (Dapporto et al., 2024). However, the hypothesis that *menelaos* had

an Anatolian-Iranian connection should not be discarded. This is suggested by the high similarity of taxa *menelaos* and *molleti* Nekrutenko, the latter being an extra-European member of the *P. eros* species group (Talavera et al., 2022). Other butterfly species or species complexes display distributions that illustrate the connection between southern Greece and Anatolia, such as *Polymmatius iphigenia* (Herrich-Schäffer), *Lycaena thetis* Klug, *Kretania eurypilus* (Freyer) or *Turanana taygetica* (Rebel) (Coutsis, 2005; Coutsis & De Prins, 2006; Tshikolovets, 2011). In such a case, the PhyloNetworks result recovering *menelaos* as sister to *P. icarus* could be sensible. In the absence of any other representative from the Anatolian-Iranian region, *menelaos* appeared to be an outgroup in phylogenetic networks, and this is also corroborated by the PCA results. In addition to that, the gene flow detected from an unsampled lineage sister to *eroides* to the common ancestor of *P. icarus* + *menelaos* could have indeed taken place considering that *eroides* is regarded as present in Turkey and it could have contributed to the ancestor of *menelaos* and related populations. However, such a hypothesis must be treated with caution as PhyloNetworks using the ingroup dataset and the analyses based on raw data such as D-statistics and STRUCTURE did not detect any gene flow between *eroides* and *menelaos*.

The reflectance spectrum of the AL/ME clade indicates similarities with *eros*, although genetically it appears to be more related to *eroides* (but note that the second-best STRUCTURE model, $K = 3$, groups AL/ME with *eros*, Figure S2), possibly reflecting its intermediate geographic position between the two taxa. The AL/ME clade is a rather unexpected finding of this study and comprised specimens sampled at over 2000 m a.s.l. across the border between Albania and Montenegro. With a long basal branch and shallow terminals (similarly to *menelaos*), this lineage may also undergo rapid differentiation. It would also be interesting to determine whether these populations have a contact zone with *eroides*, since their location falls close to the western range of the latter, but testing this hypothesis will obviously require additional sampling and analyses.

Another contrasting result between phylogeny and clustering is represented by taxon *italica* which, although recovered as sister (but paraphyletic) to all other taxa of the *P. eros* species group analysed, was not highlighted as well-differentiated in the clustering analyses. The phylogenetic position of *italica* could have been caused by gene flow with closely related taxa, such as *P. icarus*. Indeed, hybridisation among European butterflies appears to be relatively widespread, with approximately 16% of species reported or hypothesised to hybridise in the wild, including mentions involving the *P. eros* species group and *P. icarus* (Descimon & Mallet, 2009). Nevertheless, our D-statistics tests to detect a potentially notable gene flow with *P. icarus* did not reveal such a phenomenon. Thus, the sister-group relationship of *italica* with respect to all other ingroup taxa is either correct and the accelerated differentiation of *menelaos* and the AL/ME clade is truly paraphyletic, or the tree topology is affected by gene flow. The latter is also possible since gene flow analyses suggested complex interactions. For example, D-statistics detected gene flow between *menelaos* and both *eros* and the AL/ME clade, or between *italica* and the AL/ME clade, none of which was captured by STRUCTURE. This discrepancy

could have been caused by a combination of factors such as the long branches (drift and/or selection) of the *menelaos* and AL/ME clades and the possibly more complicated gene flow between tested triplets causing noise in D-statistics (e.g., while there is gene flow between P2 and P3, gene flow also occurred between P1 and the outgroup, creating extra ABBA and returning false positive results, or gene flow between P1 and P3 rendering more ABAB and diluting the signal of ABBA, returning false negative results). Additionally, fairly complex patterns of gene flow from extinct or unsampled taxa were also revealed by PhyloNetworks, which could further interfere with the phylogenetic inference.

The population structure of *P. eros* and *P. eroides* closely reflects the biogeographic history of southern European mountain systems and the effects of Quaternary climatic oscillations on cold-adapted steppe butterflies. These taxa were likely much more widespread during glacial periods, whereas relatively shorter interglacial warming progressively fragmented their distribution, leaving isolated mountain populations within the southern European centres of endemism (Menchetti et al., 2021). Similar scenarios have been proposed for several cold-adapted butterfly species, where extensive population connectivity during glacial periods was followed by relatively recent isolation of mountain populations, resulting in limited genetic divergence despite present-day geographic discontinuity (Hinojosa et al., 2019; Mewis et al., 2025; Schmitt, 2009; Schmitt et al., 2006; Schmitt & Hewitt, 2004). This scenario is well reflected in our data, with the Pyrenean populations representing the genetically purest *eros*, whereas populations from the western and central Alps show increasing admixture with *eroides*. Likewise, the former lowland steppe populations of Central Europe probably represented glacial relicts that disappeared only recently.

The evolutionary history of the Apennine populations appears considerably more complex. Besides showing strong admixture, specimens assigned to *italica* occupy a basal position in the phylogenetic tree while clustering between *eros* and *eroides* in PCA and sharing a distinctive ancestry component with populations from the eastern Alps in the STRUCTURE analysis. The Italian Peninsula is a well-known centre of butterfly endemism harbouring numerous differentiated lineages (Menchetti et al., 2021). At the same time, several butterfly taxa display an amphi-adriatic distribution shared between the Italian and Balkan peninsulas (Correggiari et al., 1996). During glacial periods, lowered sea levels substantially reduced the Adriatic Sea, creating broad land connections between the central Apennines and the Dinaric region. Such geographic continuity may have facilitated secondary contact between lineages related to *eros*, *eroides*, the AL/ME lineage, and an unsampled Italian lineage, generating the complex genomic signature observed today. Subsequent postglacial expansion from the Italian Peninsula towards the eastern Alps, a recurrent pattern documented in European butterflies (Dapporto et al., 2024), could explain why the eastern Alpine populations share a remarkably similar admixture profile with *italica*.

Under this scenario, the intermediate genomic composition of *italica* would naturally explain both its unstable phylogenetic position and its central placement in multivariate analyses. More importantly, it

suggests that *eros* and *eroides* were still exchanging genes during their latest episodes of contact, providing additional support for the current taxonomic treatment that considers all European populations as a single species, *Polyommatus eros*.

The patterns of admixture between the eastern Alps and the Apennines suggest a former connection between those populations, possibly extending further east towards Poland and Ukraine. The latter is suggested (1) by the detection of gene flow between *boisduvalii* and *italica* (see the D-statistics results, Table S6), and (2) by the admixture pattern of *boisduvalii* which, although having dominant *eroides* ancestry, showed admixture from the same sources as *italica* and the eastern Alps *eros*. This general out of Italy route extending towards eastern Europe has been highlighted as one of the main routes of postglacial recolonisation for European butterflies (Dapporto et al., 2024). Furthermore, the differentiation of the eastern Alps falls within a known general pattern in which the Alps relatively often display two main genetic lineages (western and central on one hand, eastern on the other) (Schmitt, 2017). Spectrally, taxon *italica* is almost identical to *eros*, reflecting the possible former connection between the two, while *boisduvalii* and *eroides* display a deeper blue colour (with a peak around 440 nm) compared to *eros*, in accordance with their genetic similarities.

Although allopatry represents one of the most challenging scenarios when making taxonomic decisions (Joshi et al., 2024; Mutanen et al., 2012). The divergence of the *menelaos* and AL/ME clades, coupled with signs of reinforcement shown by the spectral characteristics of the former, suggests that such populations of the *P. eros* species group are rapidly differentiating. However, the detected levels of admixture, estimated gene flow and species delimitation results (SNAPP, CLADES) suggest it is reasonable to regard all analysed populations of the *P. eros* species group as conspecific as proposed in the most recent checklists (Dapporto et al., 2022; Wiemers et al., 2018). Molecular data based on (limited resolution) Sanger sequencing also seem to support the idea of a single species even when some Asian populations are considered (Vodolazhsky et al., 2012). A better understanding of the group and possibly additional explanations for the apparent discordance between some of the analyses done here could be achieved by studying unsampled populations, since the range of the *P. eros* species group extends to Asia, with several other isolated populations (Tshikolovets, 2011). This is for example the case of Anatolian (*P. eroides*, *P. eroides yildizae* Koçak and *P. molleti* Carbonell), Caucasian (*P. eroides erotulus* Nekrutenko, *P. eros tshetverikovi* Nekrutenko) and Altai-Siberian populations (*P. erotides* (Staudinger)) of the group. It is known that these geographic regions have a direct connection and profound influence on the evolution of the Western Palearctic (European) fauna (e.g., Varga, 2022), and genomic analysis of samples from these regions may shed further light especially regarding *menelaos*, *italica* and *boisduvalii*. The fact that the taxonomy of the *P. eros* species group requires further research including extra-European taxa is also reflected by the conclusions of a recent synthesis work that, within the *P. eros* species group, recognises *P. eros* as a superspecies with 13 semispecies distributed from the

Pyrenees to Kamchatka, plus three species (*P. eroides*, *P. menelaos* and *Polyommatus amoratus* (Alphéraky)) (Coutsis & Tshikolovets, 2025).

The lack of variation in *Wolbachia* infection among the analysed taxa provides no additional evidence for reproductive isolation and is consistent with the uncertain species boundaries inferred from the genomic data. These maternally inherited bacteria are known for their capacity to manipulate the reproductive system of the host through CI, male killing, parthenogenesis or the feminisation of genetic males (Kaur et al., 2021; Werren et al., 2008). While the impact of this apparently widespread infection on the populations of the *P. eros* species group remains unclear, the presence of the same allele combination suggests that potential local reintroductions or population reinforcement may not represent a problem when this type of infection is considered. Nevertheless, routine *Wolbachia* screening is beneficial in such cases to mitigate potentially unwanted effects on butterfly populations, such as introducing new *Wolbachia* strains into a population when translocation actions are planned (e.g., Dincă et al., 2018; Lam et al., 2024). Our comparative screening approach (using well-established markers of the MLST system and the ddRAD-seq data itself), supports previous findings indicating that *Wolbachia* detection using ddRADseq data is fairly reliable and can provide a general overview of infection patterns, despite limitations related to setting a clear infection threshold (Lam et al., 2024).

Conservation of taxa within the *P. eros* species group

Across their fragmented range, some populations belonging to the *P. eros* species group are strongly declining or have already become extinct. This is particularly the case for several populations from central and eastern Europe: for example, the butterfly has become extinct in the Czech Republic (Beneš et al., 2002), is very little known (potentially extinct) from Slovakia (Bálint, 1996; Slamka, 2004), regarded as endangered in Poland (Buszko & Maślowski, 2008; Klimczuk, 2005) and as declining and listed in the Red Book of Ukraine (Akimova, 2009; Nekrutenko & Tshikolovets, 2005). Most of these populations occur(ed) in lowland areas, in sites exposed to several century long anthropogenic pressure, such as forest clearings, extending pastures, intensive livestock grazing and breeding, followed more recently by a marked change of land use generated by technical development and innovation.

Taxon *menelaos*, initially described as a distinct species, is endemic to subalpine and alpine areas of the Taygetos Mountains (Peloponnese, southern Greece) and currently highly isolated from all other populations of the *P. eros* species group (Pamperis, 2009; Tshikolovets, 2011). This taxon is also regarded as endangered due to habitat modifications and environmental change (Legakis & Marangou, 2009), with declines over the past decade estimated to 25%–50% (Tzortzakaki et al., 2024).

Members of the *P. eros* species group, treated together under *P. eros*, have recently been assessed as “least concern” (LC) in Europe (van Swaay et al., 2025), a change from the previous status of “near

threatened" (NT) (van Swaay et al., 2010). Taxon *eroides* is regarded as of conservation concern in the European Union, being listed as a species (*P. eroides*) in Annexes II and IV of the Habitats Directive 92/43/EEC. Overall, populations of the *P. eros* species group are probably less threatened in the highest elevations compared to lowlands. In alpine areas, the habitats of the butterfly are less accessible and traditional land use (e.g., extensive livestock breeding) is still maintained. Nevertheless, alpine and subalpine populations may face additional threats given current climatic changes, especially those occurring in areas that, due to elevation limits, do not leave many opportunities for altitudinal shifts (e.g., Apennines, Dinaric Alps, Taygetos etc.). Cases of butterfly populations isolated on mountains and affected by climate warming have already been documented in European butterflies (e.g., Bonifacino et al., 2022; Sistri et al., 2022) and further research is needed to assess the situation of *P. eros* species group populations. Previous modelling attempts suggested that climatic variables explain the distribution of these butterflies only to a limited extent, placing them in the "potential climate change risk" category (Settele et al., 2008). A particularly vulnerable taxon in face of climate warming might be *menelaos*, which, as this study has shown, is genetically unique at least in Europe and highly isolated on Mount Taygetos in southern Greece. This pattern is not unusual for cold-adapted butterfly populations that, as a consequence of climate warming during interglacial periods, remain trapped and isolated in mountain areas where they are often threatened by environmental change (e.g., Hinojosa et al., 2019; Lam et al., 2024; Sistri et al., 2022).

CONCLUSIONS

Our ddRADseq data, complemented by analyses of wing spectral characteristics, has provided the first genomic overview of structure and gene flow in the taxonomically debated European taxa of the *P. eros* species group. Across Europe, these butterflies display a complex pattern of admixture likely shaped by the Quaternary glaciations. Two lineages of the *P. eros* species group, *menelaos* and the AL/ME clade, seem to be undergoing accelerated differentiation on the road to speciation. The AL/ME clade is a rather unexpected discovery that had not been assigned to a particular taxon and will require additional research. Taxon *menelaos* is particularly differentiated from the rest both in terms of genetics and of wing spectral properties, the latter hinting to the occurrence of reinforcement with respect to *eroides*. Populations of *menelaos*, currently highly isolated on Mt. Taygetos (Peloponnese, southern Greece) and regarded as endangered, are thus unique and represent a high-priority conservation unit. Although based on analyses done at the continental scale, the intricate patterns of differentiation and gene flow highlighted by this study may be further explained by expanding research to populations of the *P. eros* species group occurring outside Europe.

AUTHOR CONTRIBUTIONS

Nok Ting Lam: Methodology; data curation; software; visualization; formal analysis; funding acquisition; writing – original draft; writing – review and editing; validation; investigation. **Vladislav**

Ivanov: Methodology; formal analysis; writing – review and editing; validation. **László Péter Biró:** Methodology; formal analysis; writing – review and editing. **Zsolt Bálint:** Formal analysis; writing – review and editing; investigation. **Krisztián Kertész:** Methodology; formal analysis; writing – review and editing. **Gábor Piszter:** Methodology; formal analysis; writing – review and editing. **Leonardo Dapporto:** Resources; investigation; writing – review and editing. **Roger Vila:** Resources; investigation; writing – review and editing. **Marko Mutanen:** Resources; supervision; writing – review and editing. **Vlad Dincă:** Writing – original draft; conceptualization; project administration; funding acquisition; supervision; resources; writing – review and editing; investigation; validation.

ACKNOWLEDGEMENTS

We are grateful to all colleagues who provided samples for this study. We thank Laura Törmälä for support with laboratory work and Shahan Derkarabetian for information regarding the use of CLADES. N.T.L. was supported by the Finnish Cultural Foundation (Suomen Kulttuurirahasto, grant no. 00250568) and University of Oulu scholarship foundation (no. 20260116). V.D. was supported by the Research Council of Finland (Academy Research Fellow, grant nos. 324988, 328895 and 352652). G.P. was supported by the National Research, Development and Innovation Office of Hungary (NKFIH OTKA PD 143037) and by the Ministry of Culture and Innovation of Hungary from the National Research, Development and Innovation Fund, financed under the TKP2021 funding scheme (TKP2021-NKTA-05). L.D. was supported by Direttiva Biodiversità projects 2019–2024 of the Italian Ministero della Transizione Ecologica. R.V. was supported by grant PID2022-139689NB-I00 (MICIU/AEI/<https://doi.org/10.13039/501100011033> and ERDF, EU) and by grant 2021-SGR-00420 (Departament de Recerca i Universitats, Generalitat de Catalunya). Open access publishing facilitated by Oulun yliopisto, as part of the Wiley - FinELib agreement.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The raw ddRADseq data that support the findings of this study are openly available in the NCBI Sequence Read Archive under the project accession PRJNA1231510 (Lam et al., 2026a) and in DS-POLYEROS from the Barcode of Life Data System (*Wolbachia* *coxA* and *ftsZ* sequences). All analytical data and results are openly available in Fairdata services at <https://doi.org/10.23729/fd-a5c032a7-755b-3220-a5f5-ee5cb8968ef> (Lam et al., 2026b).

ORCID

Nok Ting Lam  <https://orcid.org/0000-0003-4774-9142>

Zsolt Bálint  <https://orcid.org/0000-0001-8174-878X>

Krisztián Kertész  <https://orcid.org/0000-0001-8004-9135>

Roger Vila  <https://orcid.org/0000-0002-2447-4388>

Marko Mutanen  <https://orcid.org/0000-0003-4464-6308>

Vlad Dincă  <https://orcid.org/0000-0003-1791-2148>

REFERENCES

- Akimova, I.A. (Ed.). (2009) *Red book of Ukraine. Animal world*. Kyiv: Global Consulting.
- Baldo, L., Hotopp, J.C.D., Jolley, K.A., Bordenstein, S.R., Biber, S.A., Choudhury, R.R. et al. (2006) Multilocus sequence typing system for the endosymbiont *Wolbachia pipientis*. *Applied and Environmental Microbiology*, 72(11), 7098–7110. Available from: <https://doi.org/10.1128/AEM.00731-06>
- Bálint, Z. (1995) Two new *Polyommatus* species from the Himalayan region (Lepidoptera, Lycaenidae, Polyommataini). *Annales Historico-Naturales Musei Nationalis Hungarici*, 87, 93–102.
- Bálint, Z. (1996) *A Kárpát-medence nappali lepkéi. 1. rész. [The Butterflies of the Carpathian Basin. Part I]*. Budapest: Magyar Madártani és Természetvédelmi Egyesület. (in Hungarian).
- Bálint, Z., Horváth, Z.E., Kertész, K., Vértessy, Z. & Biró, L.P. (2007) Observations on scale structures and spectroscopic properties of *Polyommatus* lycaenid butterflies (Lepidoptera: Lycaenidae). *Annales Historico-Naturales Musei Nationalis Hungarici*, 99, 115–127.
- Bálint, Z., Kertész, K., Piszter, G., Vértessy, Z. & Biró, L.P. (2012) The well-tuned blues: the role of structural colours as optical signals in the species recognition of a local butterfly fauna (Lepidoptera: Lycaenidae: Polyommatainae). *Journal of the Royal Society Interface*, 9(73), 1745–1756. Available from: <https://doi.org/10.1098/rsif.2011.0854>
- Beneš, J., Konvička, M., Dvořák, J., Fric, Z., Havelda, Z., Pavlíčko, A. et al. (Eds.). (2002) *Motýli České republiky: Rozšíření a ochrana I, II/butterflies of The Czech Republic: distribution and conservation I, II*. Prague: SOM, p. 857.
- Bernard, G.D. & Remington, C.L. (1991) Color vision in *Lycaena* butterflies: spectral tuning of receptor arrays in relation to behavioral ecology. *Proceedings of the National Academy of Sciences of the United States of America*, 88(7), 2783–2787. Available from: <https://doi.org/10.1073/pnas.88.7.2783>
- Bonifacio, M., Pasquali, L., Sistri, G., Menchetti, M., Santini, L., Corbella, C. et al. (2022) Climate change may cause the extinction of the butterfly *Lasiommata petropolitana* in the Apennines. *Journal of Insect Conservation*, 26, 959–972. Available from: <https://doi.org/10.1007/s10841-022-00441-z>
- Bruzzese, D.J., Schuler, H., Wolfe, T.M., Glover, M.M., Mastroni, J.v., Doellman, M.M. et al. (2022) Testing the potential contribution of *Wolbachia* to speciation when cytoplasmic incompatibility becomes associated with host-related reproductive isolation. *Molecular Ecology*, 31(10), 2935–2950. Available from: <https://doi.org/10.1111/mec.16157>
- Bryant, D., Bouckaert, R., Felsenstein, J., Rosenberg, N.A. & RoyChoudhury, A. (2012) Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. *Molecular Biology and Evolution*, 29(8), 1917–1932. Available from: <https://doi.org/10.1093/molbev/mss086>
- Buszko, J. & Masłowski, J. (2008) *Motyle dzienne Polski*. Nowy Sącz, Poland: Koliber.
- Carbonell, F. (1993) Contribution à la connaissance du genre *Polyommatus* Latreille (1804): le complexe ultraspécifique de *Polyommatus eros-eroides* au Moyen-Orient et en Transcaucasie (Lepidoptera: Lycaenidae). (1re partie: descriptions de nouveaux taxa). *Linneana Belgica*, 14(4), 227–234.
- Carbonell, F. (1994) Contribution à la connaissance du genre *Polyommatus* Latreille (1804): le complexe ultraspécifique de *Polyommatus eros-eroides* au Moyen-Orient et en Transcaucasie (Lepidoptera: Lycaenidae). (2ème partie: diagnose et discussion). *Linneana Belgica*, 14(8), 439–454.
- Chhatre, V.E. & Emerson, K.J. (2017) StrAuto: automation and parallelization of STRUCTURE analysis. *BMC Bioinformatics*, 18(1), 192. Available from: <https://doi.org/10.1186/s12859-017-1593-0>
- Coffman, A.J., Hsieh, P.H., Gravel, S. & Gutenkunst, R.N. (2016) Computationally efficient composite likelihood statistics for demographic inference. *Molecular Biology and Evolution*, 33(2), 591–593. Available from: <https://doi.org/10.1093/molbev/msv255>
- Correggiari, A., Roveri, M. & Trincardi, F. (1996) Late Pleistocene and Holocene evolution of the North Adriatic Sea. *II Quaternario: Italian Journal of Quaternary Sciences*, 9(2), 697–704.
- Coutsis, J.G. (2005) Revision of the *Turanana endymion* species-group (Lycaenidae). *Nota Lepidopterologica*, 27(4), 251–272.
- Coutsis, J.G. & De Prins, J. (2006) The chromosome number and karyotype of *Plebeius* (*Plebejides*) *pylaon brethertoni* from Mt. Helmos, Peloponnisos, Greece, its tentative elevation to species level, and notes about presently existing unsettled taxonomic questions in the *pylaon* species-group complex (Lepidoptera: Lycaenidae). *Phegea*, 34(2), 57–60.
- Coutsis, J.G. & Tshikolovets, V.V. (2025) *Holarctic butterflies of the subtribe Polyommataina (Lepidoptera: Lycaenidae) and their genitalia*. Pardubice: Tshikolovets Publications, p. 368.
- Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A. et al. (2011) The variant call format and VCFtools. *Bioinformatics*, 27(15), 2156–2158. Available from: <https://doi.org/10.1093/bioinformatics/btr330>
- Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O. et al. (2021) Twelve years of SAMtools and BCFtools. *GigaScience*, 10(2), giab008. Available from: <https://doi.org/10.1093/gigascience/giab008>
- Dapporto, L., Menchetti, M., Dincă, V., Talavera, G., Garcia-Berro, A., D'Ercole, J. et al. (2024) The genetic legacy of the quaternary ice ages for west palearctic butterflies. *Science Advances*, 10(38), eadm8596. Available from: <https://doi.org/10.1126/sciadv.adm8596>
- Dapporto, L., Menchetti, M., Vodá, R., Corbella, C., Cuvelier, S., Djemadi, I. et al. (2022) The atlas of mitochondrial genetic diversity for Western Palearctic butterflies. *Global Ecology and Biogeography*, 31(11), 2184–2190. Available from: <https://doi.org/10.1111/geb.13579>
- Descimon, H. & Mallet, J. (2009) Bad species. In: Settele, J., Shreeve, T., Konvička, M. & Dyck, V.H. (Eds.) *Ecology of butterflies in Europe*. Cambridge: Cambridge University Press, pp. 219–249.
- Dincă, V., Bálint, Z., Vodá, R., Dapporto, L., Hebert, P.D.N. & Vila, R. (2018) Use of genetic, climatic, and microbiological data to inform reintroduction of a regionally extinct butterfly. *Conservation Biology*, 32(4), 828–837. Available from: <https://doi.org/10.1111/cobi.13111>
- Dincă, V., Dapporto, L., Somervuo, P., Vodá, R., Cuvelier, S., Gascoigne-Pees, M. et al. (2021) High resolution DNA barcode library for European butterflies reveals continental patterns of mitochondrial genetic diversity. *Communications Biology*, 4, 315. Available from: <https://doi.org/10.1038/s42003-021-01834-7>
- Earl, D.A. & vonHoldt, B.M. (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, 4(2), 359–361. Available from: <https://doi.org/10.1007/s12686-011-9548-7>
- Eaton, D.A.R. & Overcast, I. (2020) Ipyrad: interactive assembly and analysis of RADseq datasets. *Bioinformatics*, 36(8), 2592–2594. Available from: <https://doi.org/10.1093/bioinformatics/btz966>
- Elias-Costa, A.J., Confalonieri, V.A., Lanteri, A.A. & Rodriguez, M.S. (2019) Game of clones: is *Wolbachia* inducing speciation in a weevil with a mixed reproductive mode? *Molecular Phylogenetics and Evolution*, 133, 42–53. Available from: <https://doi.org/10.1016/j.ympev.2018.12.027>
- Evanno, G., Regnaut, S. & Goudet, J. (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology*, 14(8), 2611–2620. Available from: <https://doi.org/10.1111/j.1365-294X.2005.02553.x>
- Fordyce, J.A., Nice, C.C., Forister, M.L. & Shapiro, A.M. (2002) The significance of wing pattern diversity in the Lycaenidae: mate discrimination by two recently diverged species. *Journal of Evolutionary Biology*, 15(5), 871–879. Available from: <https://doi.org/10.1046/j.1420-9101.2002.00432.x>

- Godfray, H.C.J. (2002) Challenges for taxonomy. *Nature*, 417, 17–19. Available from: <https://doi.org/10.1038/417017a>
- Haataela, T., Saarinen, K., Ojaleinen, P., Aarnio, H. & Hanski, I.A. (2011) *Butterflies of Britain and Europe. A photographic guide*. London, UK: A&C Black Publishers Ltd.
- Hemstrom, W., Grummer, J.A., Luikart, G. & Christie, M.R. (2024) Next-generation data filtering in the genomics era. *Nature Reviews Genetics*, 25(11), 750–767. Available from: <https://doi.org/10.1038/s41576-024-00738-6>
- Higgins, L.G. (1975) *The classification of European butterflies*. London: William Collins Sons & Co Ltd., p. 320.
- Higgins, L.G. & Riley, N.D. (1970) *A field guide to the butterflies of Britain and Europe*. London: Collins, p. 380.
- Hinojosa, J.C., Koubínová, D., Dincă, V., Hernández-Roldán, J., Munguira, M.L., García-Barros, E. et al. (2020) Rapid colour shift by reproductive character displacement in *Cupido* butterflies. *Molecular Ecology*, 29(24), 4942–4955. Available from: <https://doi.org/10.1111/mec.15682>
- Hinojosa, J.C., Monasterio, Y., Escobés, R., Dincă, V. & Vila, R. (2019) *Erebia epiphron* and *Erebia orientalis*: sibling butterfly species with contrasting histories. *Biological Journal of the Linnean Society*, 126(2), 338–348. Available from: <https://doi.org/10.1093/biolinnean/bly182>
- Hoang, D.T., Chernomor, O., von Haeseler, A., Minh, B.Q. & Vinh, L.S. (2018) UFBoot2: improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution*, 35(2), 518–522. Available from: <https://doi.org/10.1093/molbev/msx281>
- Ivanov, V., Lee, K.M. & Mutanen, M. (2024) ddRAD sequencing and DNA barcoding. In: DeSalle, R. (Ed.) *DNA barcoding. Methods in molecular biology*, Vol. 2744. New York, NY: Humana. Available from: https://doi.org/10.1007/978-1-0716-3581-0_13
- Jiggins, C.D., Naisbit, R.E., Coe, R.L. & Mallet, J. (2001) Reproductive isolation caused by colour pattern mimicry. *Nature*, 411(6835), 302–305. Available from: <https://doi.org/10.1038/35077075>
- Jolley, K.A., Bray, J.E. & Maiden, M.C.J. (2018) Open-access bacterial population genomics: BIGSdb software, the PubMLST.Org website and their applications. *Wellcome Open Research*, 3, 124. Available from: <https://doi.org/10.12688/wellcomeopenres.14826.1>
- Jombart, T. (2008) Adegnet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, 24(11), 1403–1405. Available from: <https://doi.org/10.1093/bioinformatics/btn129>
- Jombart, T. & Ahmed, I. (2011) Adegnet 1.3–1 : new tools for the analysis of genome-wide SNP data. *Bioinformatics*, 27(21), 3070–3071. Available from: <https://doi.org/10.1093/bioinformatics/btr521>
- Jombart, T., Devillard, S. & Balloux, F. (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics*, 11(1), 94. Available from: <https://doi.org/10.1186/1471-2156-11-94>
- Joshi, M., Espeland, M., Huemer, P., deWaard, J. & Mutanen, M. (2024) Species delimitation under allopatry: genomic insights within and across continents in Lepidoptera. *Insect Systematics and Diversity*, 8(5), 7. Available from: <https://doi.org/10.1093/isd/ixae027>
- Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A. & Jermini, L.S. (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods*, 14(6), 587–589. Available from: <https://doi.org/10.1038/nmeth.4285>
- Kandul, N.P., Lukhtanov, V.A., Dantchenko, A.V., Coleman, J.W.S., Sekercioglu, C.H., Haig, D. et al. (2004) Phylogeny of *Agrodiaetus* Hübner 1822 (Lepidoptera: Lycaenidae) inferred from mtDNA sequences of COI and COII and nuclear sequences of EF1- α karyotype diversification and species radiation. *Systematic Biology*, 53(2), 278–298. Available from: <https://doi.org/10.1080/10635150490423692>
- Kass, R.E. & Raftery, A.E. (1995) Bayes factors. *Journal of the American Statistical Association*, 90(430), 773–795. Available from: <https://doi.org/10.1080/01621459.1995.10476572>
- Kaur, R., Shropshire, J.D., Cross, K.L., Leigh, B., Mansueto, A.J., Stewart, V. et al. (2021) Living in the endosymbiotic world of *Wolbachia*: a centennial review. *Cell Host & Microbe*, 29(6), 879–893. Available from: <https://doi.org/10.1016/j.chom.2021.03.006>
- Keck, F. (2020) Handling biological sequences in R with the bioseq package. *Methods in Ecology and Evolution*, 11(12), 1728–1732. Available from: <https://doi.org/10.1111/2041-210X.13490>
- Keightley, P.D., Pinharanda, A., Ness, R.W., Simpson, F., Dasmahapatra, K.K., Mallet, J. et al. (2015) Estimation of the spontaneous mutation rate in *Heliconius melpomene*. *Molecular Biology and Evolution*, 32(1), 239–243. Available from: <https://doi.org/10.1093/molbev/msu302>
- Kertész, K., Bálint, Z., Piszter, G., Horváth, Z.E. & Biró, L.P. (2021) Multi-instrumental techniques for evaluating butterfly structural colors: a case study on *Polyommatus bellargus* (Rottemburg, 1775) (Lepidoptera: Lycaenidae: Polyommatinae). *Arthropod Structure & Development*, 61, 101010. Available from: <https://doi.org/10.1016/j.asd.2020.101010>
- Kertész, K., Piszter, G., Bálint, Z. & Biró, L.P. (2019) Biogeographical patterns in the structural blue of male *Polyommatus icarus* butterflies. *Scientific Reports*, 9, 2338. Available from: <https://doi.org/10.1038/s41598-019-38827-w>
- Klimczuk, P. (2005) The larval host plant of *Polyommatus eroides* (Frivaldszky, 1835) (Lycaenidae) from Poland with comments on the life history. *Nota Lepidopterologica*, 28(2), 103–111.
- Kopelman, N.M., Mayzel, J., Jakobsson, M., Rosenberg, N.A. & Mayrose, I. (2015) Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources*, 15(5), 1179–1191. Available from: <https://doi.org/10.1111/1755-0998.12387>
- Lafranchis, T. (2003) *Butterflies of Europe*. Paris: Diatheo, p. 351.
- Lam, N.T., Ivanov, V., Biró, L.P., Bálint, Z., Kertész, K., Piszter, G. et al. (2026a) European *Polyommatus eros* species group (Lepidoptera, Lycaenidae) ddRADseq raw data. [Dataset], NCBI. <https://www.ncbi.nlm.nih.gov/sra/PRJNA1231510>
- Lam, N.T., Ivanov, V., Biró, L.P., Bálint, Z., Kertész, K., Piszter, G. et al. (2026b) Dataset used in the manuscript titled “Genetic structure, spectral characteristics and evolutionary history of butterflies in the European *Polyommatus eros* species group (Lepidoptera, Lycaenidae)”. [Dataset], Fairdata. Available from: <https://doi.org/10.23729/fd-a5c032a7-755b-3220-a5f5-ec5cb8968ef>
- Lam, N.T., Ivanov, V., Dapporto, L., Vila, R., Mutanen, M. & Dincă, V. (2024) Conservation implications of a genomic-based taxonomy for threatened allopatric *Agriades* butterflies. *Insect Conservation and Diversity*, 17(6), 1084–1097. Available from: <https://doi.org/10.1111/icad.12771>
- Leaché, A.D., Fujita, M.K., Minin, V.N. & Bouckaert, R.R. (2014) Species delimitation using genome-wide SNP data. *Systematic Biology*, 63(4), 534–542. Available from: <https://doi.org/10.1093/sysbio/syu018>
- Lee, K.M., Kivelä, S.M., Ivanov, V., Hausmann, A., Kaila, L., Wahlberg, N. et al. (2018) Information dropout patterns in restriction site associated DNA Phylogenomics and a comparison with multilocus sanger data in a species-rich moth genus. *Systematic Biology*, 67(6), 925–939. Available from: <https://doi.org/10.1093/sysbio/syy029>
- Legakis, A. & Marangou, P. (Eds.). (2009) *The red book of threatened animals in Greece*. Athens: Hellenic Zoological Society, p. 528.
- Leraut, P. (2016) *Butterflies of Europe and neighbouring regions*. Verrières-le-Buisson: NAP Editions, p. 1100.
- Li, H. (2011) A statistical framework for SNP calling, mutation discovery, association mapping and population genetic parameter estimation from sequencing data. *Bioinformatics*, 27(21), 2987–2993. Available from: <https://doi.org/10.1093/bioinformatics/btr509>
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N. et al. (2009) The sequence alignment/map format and SAMtools. *Bioinformatics*,

- 25(16), 2078–2079. Available from: <https://doi.org/10.1093/bioinformatics/btp352>
- Liu, X. & Fu, Y.-X. (2015) Exploring population size changes using SNP frequency spectra. *Nature Genetics*, 47(5), 555–559. Available from: <https://doi.org/10.1038/ng.3254>
- Liu, X. & Fu, Y.-X. (2020) Stairway plot 2: demographic history inference with folded SNP frequency spectra. *Genome Biology*, 21(1), 280. Available from: <https://doi.org/10.1186/s13059-020-02196-9>
- Lohse, K. & Vila, R. (2023) The genome sequence of the common blue, *Polyommatus icarus* (Rottemburg, 1775). *Wellcome Open Research*, 8, 72. Available from: <https://doi.org/10.12688/wellcomeopenres.18772.1>
- Lukhtanov, V.A., Dantchenko, A.V., Vishnevskaya, M.S. & Saifitdinova, A.F. (2015) Detecting cryptic species in sympatry and allopatry: analysis of hidden diversity in *Polyommatus* (*Agrodiaetus*) butterflies (Lepidoptera: Lycaenidae). *Biological Journal of the Linnean Society*, 116(2), 468–485. Available from: <https://doi.org/10.1111/bij.12596>
- Lukhtanov, V.A., Kandul, N.P., Plotkin, J.B., Dantchenko, A.V., Haig, D. & Pierce, N.E. (2005) Reinforcement of pre-zygotic isolation and karyotype evolution in *Agrodiaetus* butterflies. *Nature*, 436(7049), 385–389. Available from: <https://doi.org/10.1038/nature03704>
- Malinsky, M., Matschiner, M. & Svardal, H. (2021) Dsuite – fast D-statistics and related admixture evidence from VCF files. *Molecular Ecology Resources*, 21(2), 584–595. Available from: <https://doi.org/10.1111/1755-0998.13265>
- Marth, G.T., Czabarka, E., Murvai, J. & Sherry, S.T. (2004) The allele frequency spectrum in genome-wide human variation data reveals signals of differential demographic history in three large world populations. *Genetics*, 166(1), 351–372. Available from: <https://doi.org/10.1534/genetics.166.1.351>
- Menchetti, M., Talavera, G., Cini, A., Salvati, V., Dincă, V., Platania, L. et al. (2021) Two ways to be endemic. Alps and Apennines are different functional refugia during climatic cycles. *Molecular Ecology*, 30(5), 1297–1310. Available from: <https://doi.org/10.1111/mec.15795>
- Mewis, V., Wendt, M. & Schmitt, T. (2025) Phylogeographic analyses reveal recent dispersal and multiple *Wolbachia* infections of the bright-eyed ringlet *Erebia oeme* within the European mountain systems. *Scientific Reports*, 15(1), 1956. Available from: <https://doi.org/10.1038/s41598-024-84551-5>
- Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A. et al. (2020) IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution*, 37(5), 1530–1534. Available from: <https://doi.org/10.1093/molbev/msaa015>
- Mirarab, S. & Warnow, T. (2015) ASTRAL-II: coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. *Bioinformatics*, 31(12), i44–i52. Available from: <https://doi.org/10.1093/bioinformatics/btv234>
- Mutanen, M., Hausmann, A., Hebert, P.D.N., Landry, J.-F., de Waard, J.R. & Huemer, P. (2012) Allopatry as a gordian knot for taxonomists: patterns of DNA barcode divergence in Arctic-Alpine lepidoptera. *PLoS One*, 7(10), e47214. Available from: <https://doi.org/10.1371/journal.pone.0047214>
- Nekrutenko, Y. & Tshikolovets, V. (2005) *The butterflies of Ukraine*. Kyiv: Rayevsky Scientific Publishers, p. 231.
- Pamperis, L.N. (2009) *The butterflies of Greece. Second edition revised and enlarged*. Athens: Editions Pamperis, p. 766.
- Paradis, E., Claude, J. & Strimmer, K. (2004) APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*, 20(2), 289–290. Available from: <https://doi.org/10.1093/bioinformatics/btg412>
- Patterson, N., Price, A.L. & Reich, D. (2006) Population structure and eigenanalysis. *PLoS Genetics*, 2(12), 190. Available from: <https://doi.org/10.1371/journal.pgen.0020190>
- Pei, J., Chu, C., Li, X., Lu, B. & Wu, Y. (2018) CLADES: a classification-based machine learning method for species delimitation from population genetic data. *Molecular Ecology Resources*, 18(5), 1144–1156. Available from: <https://doi.org/10.1111/1755-0998.12887>
- Pisztér, G., Kertész, K., Bálint, Z. & Biró, L.P. (2016) Variability of the structural coloration in two butterfly species with different prezygotic mating strategies. *PLoS One*, 11, e0165857. Available from: <https://doi.org/10.1371/journal.pone.0165857>
- Pritchard, J.K., Stephens, M. & Donnelly, P. (2000) Inference of population structure using multilocus genotype data. *Genetics*, 155(2), 945–959. Available from: <https://doi.org/10.1093/genetics/155.2.945>
- Rambaut, A., Drummond, A.J., Xie, D., Baele, G. & Suchard, M.A. (2018) Posterior summarization in Bayesian phylogenetics using tracer 1.7. *Systematic Biology*, 67(5), 901–904. Available from: <https://doi.org/10.1093/sysbio/syy032>
- RoyChoudhury, A., Felsenstein, J. & Thompson, E.A. (2008) A two-stage pruning algorithm for likelihood computation for a population tree. *Genetics*, 180(2), 1095–1105. Available from: <https://doi.org/10.1534/genetics.107.085753>
- Sandall, E.L., Maureaud, A.A., Guralnick, R., McGeoch, M.A., Sica, Y.v., Rogan, M.S. et al. (2023) A globally integrated structure of taxonomy to support biodiversity science and conservation. *Trends in Ecology & Evolution*, 38(12), 1143–1153. Available from: <https://doi.org/10.1016/j.tree.2023.08.004>
- Schmitt, T. (2009) Biogeographical and evolutionary importance of the European high mountain systems. *Frontiers in Zoology*, 6(1), 9. Available from: <https://doi.org/10.1186/1742-9994-6-9>
- Schmitt, T. (2017) Molecular biogeography of the High Mountain Systems of Europe: an overview. In: *Advances in global change research*, Vol. 62. Cham: Springer. Available from: https://doi.org/10.1007/978-3-319-55982-7_3
- Schmitt, T. & Hewitt, G.M. (2004) Molecular biogeography of the arctic-alpine disjunct burnet moth species *Zygaena exulans* (Zygaenidae, Lepidoptera) in the Pyrenees and Alps. *Journal of Biogeography*, 31(6), 885–893. Available from: <https://doi.org/10.1111/j.1365-2699.2004.01079.x>
- Schmitt, T., Hewitt, G.M. & Müller, P. (2006) Disjunct distributions during glacial and interglacial periods in mountain butterflies: *Erebia epiphron* as an example. *Journal of Evolutionary Biology*, 19(1), 108–113. Available from: <https://doi.org/10.1111/j.1420-9101.2005.00980.x>
- Settele, J., Kudrna, O., Harpke, A., Kühn, I., van Swaay, C., Verovnik, R. et al. (2008) Climatic risk atlas of European butterflies. *BioRisk*, 1-(special issue), 1–712. Available from: <https://doi.org/10.3897/biorisk.1>
- Shafer, A.B.A., Peart, C.R., Tusso, S., Maayan, I., Brelsford, A., Wheat, C.W. et al. (2017) Bioinformatic processing of RAD-seq data dramatically impacts downstream population genetic inference. *Methods in Ecology and Evolution*, 8(8), 907–917. Available from: <https://doi.org/10.1111/2041-210X.12700>
- Sistri, G., Menchetti, M., Santini, L., Pasquali, L., Sapianti, S., Cini, A. et al. (2022) The isolated *Erebia pandrose* Apennine population is genetically unique and endangered by climate change. *Insect Conservation and Diversity*, 15(1), 136–148. Available from: <https://doi.org/10.1111/icad.12538>
- Slamka, F. (2004) *Die Tagfalter Mitteleuropas – östlicher Teil*. Bratislava: František Slamka, p. 288.
- Solís-Lemus, C., Bastide, P. & Ané, C. (2017) PhyloNetworks: a package for phylogenetic networks. *Molecular Biology and Evolution*, 34(12), 3292–3298. Available from: <https://doi.org/10.1093/molbev/msx235>
- Stamatakis, A. (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30(9), 1312–1313. Available from: <https://doi.org/10.1093/bioinformatics/btu033>
- Talavera, G., Lukhtanov, V., Pierce, N.E. & Vila, R. (2022) DNA barcodes combined with multilocus data of representative taxa can generate reliable higher-level phylogenies. *Systematic Biology*, 71(2), 382–395. Available from: <https://doi.org/10.1093/sysbio/syab038>

- Talavera, G., Lukhtanov, V.A., Pierce, N.E. & Vila, R. (2013) Establishing criteria for higher-level classification using molecular data: the systematics of *Polyommatus* blue butterflies (Lepidoptera, Lycaenidae). *Cladistics*, 29(2), 166–192. Available from: <https://doi.org/10.1111/j.1096-0031.2012.00421.x>
- Tolman, T. & Lewington, R. (2008) *Collins butterfly guide*. London: Harper-Collins Publisher, p. 384.
- Trevelline, B.K., Fontaine, S.S., Hartup, B.K. & Kohl, K.D. (2019) Conservation biology needs a microbial renaissance: a call for the consideration of host-associated microbiota in wildlife management practices. *Proceedings of the Royal Society B: Biological Sciences*, 286(1895), 20182448. Available from: <https://doi.org/10.1098/rspb.2018.2448>
- Tshikolovets, V.V. (2011) *Butterflies of Europe & the Mediterranean area*. Pardubice: Tshikolovets Publications, p. 544.
- Tuzov, V.K. (2000) *Guide to the butterflies of Russia and adjacent territories*, Vol. 2. Sofia: Pensoft, 480 p. (vol 1.), 580 p. (vol. 2).
- Tzortzakaki, O., Zografou, K. & Pamperis, L. (2024) *Polyommatus eros*. The Greek Red List of Threatened Species. Available from: redlist.necca.gov.gr.
- van der Maaten, L. & Hinton, G. (2008) Visualizing Data using t-SNE. *Journal of Machine Learning Research*, 9(86), 2579–2605.
- van Swaay, C., Cuttelod, A., Collins, S., Maes, D., López Munguira, M., Šašic, M. et al. (2010) *European red list of butterflies*. Luxembourg: Publications Office of the European Union. Available from: <https://doi.org/10.2779/83897>
- van Swaay, C., Warren, M., Ellis, S., Clay, J., Bellotto, V., Allen, D.J. et al. (2025) *Measuring the pulse of European biodiversity. European red list of butterflies*. Brussels, Belgium: European Commission. Available from: <https://doi.org/10.2779/935927>
- Varga, Z. (2022) Centres of endemism of Noctuidae (Lepidoptera) in the Palaearctic arid mountains: biogeographical and phylogenetic implications. *Contributions to Entomology*, 72(1), 1–35. Available from: <https://doi.org/10.3897/contrib.entomol.72.e87196>
- Vila, R., Bell, C.D., Macniven, R., Goldman-Huertas, B., Ree, R.H., Marshall, C.R. et al. (2011) Phylogeny and palaeoecology of *Polyommatus* blue butterflies show Beringia was a climate-regulated gateway to the New World. *Proceedings of the Royal Society B: Biological Sciences*, 278(1719), 2737–2744. Available from: <https://doi.org/10.1098/rspb.2010.221>
- Vila, R., Lukhtanov, V.A., Talavera, G., Gil-T, F. & Pierce, N.E. (2010) How common are dot-like distributions? Taxonomical oversplitting in western European *Agrodiaetus* (Lepidoptera: Lycaenidae) revealed by chromosomal and molecular markers. *Biological Journal of the Linnean Society*, 101, 130–154. Available from: <https://doi.org/10.1111/j.1095-8312.2010.01481.x>
- Vodolazhsky, D.I., Stradomsky, B.V. & Yakovlev, R.V. (2012) Study of mitochondrial COI and nuclear ITS2 sequences of Mongolian specimens of the *Polyommatus eros*-group (Lepidoptera: Lycaenidae). *Caucasian Entomological Bulletin*, 8(2), 264–266.
- Vodolazhsky, D.I., Wiemers, M. & Stradomsky, B.V. (2009) A comparative analysis of mitochondrial and nuclear DNA sequences in blue butterflies of subgenus *Polyommatus* (s. str.) Latreille, 1804 (Lepidoptera: Lycaenidae: *Polyommatus*). *Caucasian Entomological Bulletin*, 5(1), 115–120. Available from: <https://doi.org/10.23885/1814-3326-2009-5-1-115-120>
- Weir, B.S. & Cockerham, C.C. (1984) Estimating F-statistics for the analysis of population structure. *Evolution*, 38(6), 1358–1370. Available from: <https://doi.org/10.1111/j.1558-5646.1984.tb05657.x>
- Werren, J.H., Baldo, L. & Clark, M.E. (2008) *Wolbachia*: master manipulators of invertebrate biology. *Nature Reviews Microbiology*, 6(10), 741–751. Available from: <https://doi.org/10.1038/nrmicro1969>
- Wickham, H. (2016) *ggplot2: elegant graphics for data analysis*. New York, USA: Springer Cham. Available from: <https://ggplot2.tidyverse.org>
- Wiemers, M., Balletto, E., Dincă, V., Fric, Z.F., Lamas, G., Lukhtanov, V. et al. (2018) An updated checklist of the European butterflies (Lepidoptera, Papilionoidea). *ZooKeys*, 811, 9–45. Available from: <https://doi.org/10.3897/zookeys.811.28712>
- Wiemers, M., Stradomsky, B.V. & Vodolazhsky, D.I. (2010) A molecular phylogeny of *Polyommatus* s. str. and *Plebicula* based on mitochondrial COI and nuclear ITS2 sequences (Lepidoptera: Lycaenidae). *European Journal of Entomology*, 107(3), 325–336. Available from: <https://doi.org/10.14411/eje.2010.041>
- Wood, D.E., Lu, J. & Langmead, B. (2019) Improved metagenomic analysis with kraken 2. *Genome Biology*, 20, 257. Available from: <https://doi.org/10.1186/s13059-019-1891-0>
- Wright, S. (1951) The genetical structure of populations. *Annals of Eugenics*, 15(1), 323–354. Available from: <https://doi.org/10.1111/j.1469-1809.1949.tb02451.x>

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

Figure S1. Library preparation and bioinformatics flowchart.

Figure S2. Structure results that include all the K-clusters along with the summary statistics from Structure Harvester and the graphs generated using CLUMPAK.

Figure S3. Dorsal view of male *P. eros* species group wings used for analyses of spectral properties.

Figure S4. Maximum likelihood trees build based on the ddRAD de novo assembly (above) and reference assembly (below), using different *m* values.

Figure S5. PCA results with different components.

Figures S6–S9. PhyloNetworks results.

Figure S10. Stairway Plot 2 results.

Table S1. Information on specimens used in this study.

Table S2. Datasets explored during the research.

Table S3. Models tested in SNAPP.

Table S4. Primers and PCR protocols used for whole genome amplification and *Wolbachia* screening (coxA and ftsZ).

Table S5. *F*-statistics: F_{IT} and F_{ST} results.

Table S6. Dsuite results for different groupings of European taxa belonging to the *P. eros* species group.

Table S7. Detailed results from CLADES.

Data S1. Supporting Information.

How to cite this article: Lam, N.T., Ivanov, V., Biró, L.P., Bálint, Z., Kertész, K., Piszter, G. et al. (2026) Genetic structure, spectral characteristics and evolutionary history of butterflies in the European *Polyommatus eros* species group (Lepidoptera, Lycaenidae). *Systematic Entomology*, 51(4), e70072. Available from: <https://doi.org/10.1111/syen.70072>