Received: 7 June 2023 | Revised: 13 February 2024 | Accepted: 15 February 2024

DOI: 10.1111/mec.17304

ORIGINAL ARTICLE

MOLECULAR ECOLOGY WILEY

The opposed forces of differentiation and admixture across glacial cycles in the butterfly *Aglais urticae*

Valéria Marques[1] | Joan Carles Hinojosa[1] | Leonardo Dapporto[2] | Gerard Talavera[3] | Constantí Stefanescu[4,5] | David Gutiérrez[6,7] | Roger Vila[1]

[1]Institut de Biologia Evolutiva (CSIC-Universitat Pompeu Fabra), Barcelona, Spain

[2]Dipartimento di Biologia, Università degli Studi di Firenze, Sesto Fiorentino, Italy

[3]Institut Botànic de Barcelona (IBB), CSIC-CMCNB, Barcelona, Spain

[4]Natural Sciences Museum of Granollers, Granollers, Spain

[5]CREAF, Cerdanyola del Vallès, Spain

[6]Instituto de Investigación en Cambio Global (IICG), Universidad Rey Juan Carlos, Madrid, Spain

[7]Área de Biodiversidad y Conservación, Universidad Rey Juan Carlos, Madrid, Spain

Correspondence
Roger Vila and Valéria Marques, Institut de Biologia Evolutiva (CSIC-UPF) Passeig Marítim de la Barceloneta 37-49, Barcelona 08003, Spain.
Email: roger.vila@csic.es and valeria.marques@ibe.upf-csic.es

Funding information
Ministerio de Ciencia e Innovación (MCIN)/Agencia Estatal de Investigación (AEI), Grant/Award Number: PID2019-107078GB-I00, PID2022-139689NB-I00, CGL2014-57784-P, PRE2020-094870 and BES-2017-080641; Departament de Recerca i Universitats de la Generalitat de Catalunya, Grant/Award Number: 2021-SGR-00420; National Biodiversity Future Center NBFC (Italian Ministry of University and Research, PNRR), Grant/Award Number: ProjectCN00000033

Handling Editor: Paula Arribas

Abstract

Glacial cycles lead to periodic population interbreeding and isolation in warm-adapted species, which impact genetic structure and evolution. However, the effects of these processes on highly mobile and more cold-tolerant species are not well understood. This study aims to shed light on the phylogeographic history of *Aglais urticae*, a butterfly species with considerable dispersal ability, and a wide Palearctic distribution reaching the Arctic. Through the analysis of genomic data, four main genetic lineages are identified: European, Sierra Nevada, Sicily/Calabria/Peloponnese, and Eastern. The results indicate that the Sardo-Corsican endemic taxon *ichnusa* is a distinct species. The split between the relict lineages in southern Europe and the main European lineage is estimated to have happened 400–450 thousand years ago, with admixture observed during the Quaternary glacial cycles, and still ongoing, albeit to a much smaller extent. These results suggest that these lineages may be better treated as subspecific parapatric taxa. Ecological niche modelling supported the existence of both Mediterranean and extra-Mediterranean refugia during the glacial periods, with the main one located on the Atlantic coast. Nevertheless, gene flow between populations was possible, indicating that both differentiation and admixture have acted continuously across glacial cycles in this cold-tolerant butterfly, generally balancing each other but producing differentiated lineages in the southern peninsulas. We conclude that the population dynamics and the processes shaping the population genetic structure of cold-adapted species during the Quaternary ice ages may be different than those classically accepted for warm-adapted species.

KEYWORDS
Aglais urticae, gene flow, glacial cycles, isolation, relict lineages

1 | INTRODUCTION

Cyclical processes are essential drivers of biological diversity and play significant roles in shaping the distribution and genetic structure of species across geologic, evolutionary, and ecological time scales (Linck & Battey, 2019). Among these processes, glacial cycles (Roy et al., 1996) periodically influence species' ranges, causing them to expand and contract in response to changes in temperatures and ice cover (Hewitt, 1999; Petit et al., 2003; Stewart et al., 2010). These shifts generally involve alternating phases of population interbreeding and isolation, which might lead to differentiation and, even, speciation with periodic gene flow. This alternative to the classical view of divergence in either complete isolation or sympatry is not uncommon, as shown by recent theoretical models and empirical evidence (Butlin et al., 2008; Feder et al., 2012).

Numerous studies have demonstrated the impact of glacial cycles on genetic diversity and population structure in various species (Alcala & Vuilleumier, 2014; Duckworth & Semenov, 2017; Otto & Whitlock, 1997). In Europe, the primary genetic lineages in thermophilous species often align with the presumed glacial refugia in the southern peninsulas. This pattern has been repeatedly documented using mitochondrial DNA (mtDNA) in butterfly species such as *Polyommatus daphnis*, *Melitaea trivia*, *Lasiommata maera*, and *Pyrgus serratulae* (Dapporto et al., 2022). However, the influence of glacial cycles on species' range and, therefore, genetic diversity and structure may vary depending on adaptations and habitat preferences, as illustrated by differences between cold-adapted (range contractions during interglacial periods) and thermophilous (range contractions during glacial periods) species (Menchetti et al., 2021; Mutanen et al., 2012).

The small tortoiseshell, *Aglais urticae*, is a highly dispersive nymphalid butterfly with a wide distribution throughout the Palearctic region. It prefers open habitats with abundance of *Urtica dioica*, its main larval host plant. In the north, it reaches high latitudes well above the Arctic Polar Circle, while, in the southern edge of its European distribution, it is only present in the mountain ranges in Sierra Nevada, Sicily and Calabria, and the Peloponnese—we therefore consider it temperate and cold-tolerant. The taxon *ichnusa* Bonelli, 1826, occurring on the islands of Corsica and Sardinia, displays differences in wing pattern. It was traditionally treated as a subspecies of *A. urticae* (Higgins & Riley, 1978; Strobino, 1971; Vandewoestijne et al., 2004), but it is now generally considered a separate species (e.g. Wiemers et al., 2018), even if a genetic analysis supporting this view is lacking.

A first study on population structure in *A. urticae* using allozyme electrophoresis revealed an apparent homogeneity in populations of Belgium, the Netherlands, and northern France, with most genetic diversity occurring within rather than between populations (Vandewoestijne et al., 1999). The first analysis of COI over the entire Palearctic reported high gene flow and no geographical differentiation, with no subspecies identified as separate evolutionary units (Vandewoestijne et al., 2004). Surprisingly, the most recent documentation of COI pattern for this butterfly (Dapporto et al., 2022) indicated that a single lineage is present in southern mountain ranges of three Mediterranean peninsulas (Iberian, Italian, and Balkan), which suggests a relict distribution (i.e. currently present in restricted and fragmented areas, being more widespread and connected sometime in the past). Two other lineages were found in the rest of Europe, being substantially divergent (ca. 1% uncorrected pairwise distance): a common and widespread one and a much less frequent one restricted to the Alps. The apparent existence of genetic structure in such a dispersive taxon suggests that *A. urticae* may be an adequate model to study differentiation with gene flow.

In this study, we use genome-wide data obtained through double-digest RAD sequencing (ddRADseq (Peterson et al., 2012)) to investigate the phylogeography of this Palearctic butterfly, focusing on the western part of its distribution. In particular, we aim to: (1) assess the specific differentiation of the Sardo-Corsican taxon *ichnusa*; (2) examine the presence of diverged lineages in the three peninsulas of southern Europe (Iberia, Italy, and the Balkans) based on previous mtDNA evidence; (3) assess the degree of genetic admixture, informing on reproductive isolation among these lineages, shedding light on potential cryptic species; and (4) integrate genomic results and ecological niche modelling to propose a hypothesis for the evolutionary history of this cold-tolerant and highly mobile butterfly, with a focus on the dynamic interplay of allopatric differentiation and homogenization through gene flow during glacial cycles.

2 | MATERIALS AND METHODS

2.1 | Sampling

Aglais urticae has a wide distribution throughout the Palearctic, inhabiting places where *U. dioica* is common, from as high as northern Scandinavia to the southern mountain ranges in Sierra Nevada, Sicily and Calabria, and the Peloponnese. In this study, we tried to cover all main populations and mountain ranges in Europe, analysing a total of 79 samples. These included 79 samples, including 73 specimens of *A. urticae* (covering the European part of the distribution, plus two samples from Kazakhstan, one from Nepal and one from China), 5 *A. ichnusa* (here considered a distinct species; covering the entire distribution, Corsica and Sardinia), and 1 sample from Canada of *A. milberti*, sister to the rest of the taxa, used as outgroup (Table S1). Butterfly bodies were stored in 99% ethanol at −20°C, and wings were kept separately as vouchers.

2.2 | ddRADseq library preparation

A ddRAD protocol was used for genome-wide representation. We extracted genomic DNA (gDNA) from half of the thorax using the DNeasy Blood & Tissue Kit (Qiagen). We checked the quantity of gDNA extracts using the PicoGreen kit (Molecular Probes). In order to increase gDNA quantity, we performed whole-genome

amplification using the REPLI-g Mini Kit (Qiagen). We estimated again the concentration of the amplified gDNA with PicoGreen kit (Molecular Probes). For every sample, 500 ng of DNA were digested with 1 μL PstI, 2 μL MseI, and 5 μL of CutSmart Buffer (New England Biolabs), and water was added until a total volume of 50 μL. Samples were incubated for 2 h at 37°C and then frozen for enzyme deactivation. We did a purification step with AMPure XP magnetic beads (Agencourt) in a Biomek automated liquid handler (Beckman Coulter) with a final elution in 40 μL. DNA concentration was measured with PicoGreen, and this value was used for the pooling step. For ligation of the adapters used to later identify individuals, in every sample we added 5 μL T4 DNA Ligase Buffer (NewEngland Biolabs), 1 μL T4 DNA Ligase (New England Biolabs), 0.6 μL rATP (Promega), 5 μL P1 adapter (50 nM), 5 μL P2 adapter (50 nM), and 2.4 μL water. The P1 adapter included 45 unique Illumina primer sequences, 5 bp barcodes, and a TGCA overhang on the top strand to match the sticky end left by PstI. The P2 adapter included the Illumina primer sequences and AT overhangs on the top strand to match the sticky end left by MseI. It also incorporated a 'divergent-Y' to prevent amplification of fragments with MseI cutsites on both ends. Ligation process was extended for 1 h at 22°C, and enzymes were deactivated at 65°C for 20 min. 200 ng of each individual were pooled in tubes making three pools in three different tubes with a final volume of ~450 μL each. Every pool was purified with AMPure XP magnetic beads. Size selection was done at 300 bp with BluePippin (Sage Science). Finally, we performed PCR amplification, with primers RAD1.F (5′-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACG-3′) and RAD2.R (5′-CAAGCAGAAGACGGCATACGAGATCGTGATGTGACTGGAGTTCAGACGTGTGC-3′). DNA was amplified in 60 μL volume reactions: 9 μL water, 30 μL Phusion High-Fidelity PCR Master Mix (Finnzymes), 3 μL of each primer (10 mM), and 15 μL of DNA. Reaction conditions comprised a first denaturation at 98°C for 30 s, then 16 cycles of denaturation at 98°C for 10 s, annealing at 60°C for 30 s, and extension at 72°C for 40 s, with a final extension step at 72°C for 5 min. The PCR products were purified with AMPure XP magnetic beads, and DNA concentration was measured with PicoGreen. We measured size distribution and concentration of the pools with a Bioanalyzer (Agilent Technologies). Finally, we pooled libraries in equimolar amounts and sequenced them on an Illumina HiSeq 2500 PE 100 at FIMM Technology Center (Helsinki, Finland). The demultiplexed fastq data are archived in the NCBI SRA SSR24894924-SRR24895002 and BioProject PRJNA980705.

2.3 | Data processing

Initial filtering steps, single-nucleotide polymorphism (SNP) calling, and alignment were carried out using IPYRAD v0.7.23 (Eaton & Overcast, 2020). We used as reference the genome of *A. urticae* from the Wellcome Sanger Institute (Bishop et al., 2021): GCA_905147175.1 ilAglUrti1.1 (PRJEB41896, INSDC: CAJHUP000000000.1). The following parameters were changed from the default settings: assembly was set to reference; datatype to pairddrad; restriction overhang to TGCAG, TAA; minimum trimmed length to 70; minimum number of samples per locus to five; maximum number of SNPs per locus to 10; and maximum number of Indels per locus to five. This dataset obtained from IPYRAD had 15,832 loci, with a total of 3,812,432 base pairs including 180,539 SNPs.

2.4 | Population genetic structure

Patterns of population genetic structure were explored into a Principal Component Analysis (PCA) and a model-based Bayesian clustering approach using STRUCTURE v2.3.4 (Pritchard et al., 2000). For the PCA, the *A. milberti* and *A. ichnusa* samples were filtered out of the initial SNPs dataset, followed by removing indels and SNPs with >50% of missing data, and keeping a minimum allele frequency of 0.014, using VCFtools v0.1.17 (Danecek et al., 2011). We then filtered for only unlinked SNPs, with the variant pruning option, –indep-pairwise (window size of 40 kb, step size of 10 bp, and r^2 threshold of 0.1) of PLINK v1.90b6.24 (Chang et al., 2015). This resulted in a new filtered dataset with 6408 SNPs. We performed the PCA using PLINK v1.07 (Purcell et al., 2007) and visualized it in the R (R Core Team, 2021) environment, with the package tidyverse (Wickham et al., 2019). For the STRUCTURE analysis, the *A. milberti* and *A. ichnusa* samples were filtered out of the initial allele genotype dataset using VCFtools. We tested values of K from one to six. The selected burnin was 75,000, followed by 200,000 MCMC replicates. For each K, ten runs were done and afterwards averaged with CLUMPAK v1.1 (Kopelman et al., 2015; Pritchard et al., 2000). We calculated the best K under the Evanno method using STRUCTURE HARVESTER v0.6.94 (Earl & vonHoldt, 2012).

Based on STRUCTURE results, filtered datasets with 'pure' individuals from each recovered lineage (with pp. = 100) were used to generate six hybrid classes of the lineage pairs (Europe–Sierra Nevada and Europe–Sicily/Calabria) in HYBRIDLAB (Nielsen et al., 2006), using the recom-sim tool, found in https://github.com/salanova-elliott/recom-sim. This simulated dataset included 500 individuals of each: first-generation hybrids (F1), second-generation hybrids (F2), second-generation backcrosses with Europe (B2Europe), second-generation backcrosses with Sierra Nevada or Sicily/Calabria (B2SierraNevada; B2Sicily), third-generation backcrosses with Europe (B3Europe), and third-generation backcrosses with Europe–Sierra Nevada or Sicily/Calabria (B3SierraNevada; B3Sicily). The simulated data were then analysed in ADMIXTURE (Alexander et al., 2009) for $K = 2$, applying the default 200 bootstrap replicates. In order to remove 'noise' from other lineages in the attribution of putative hybrid specimens to each cluster, a new ADMIXTURE analysis using the filtered dataset with the same 'pure' individuals from each lineage and putative hybrid specimens was performed, using the same parameters as for the simulated data. The resulting clustering coefficients of each simulated hybrid group were used as empirically derived assignment criteria for the putative hybrid individuals to the six categories. 'Pure' and putative hybrids are listed in Table S2.

2.5 | Phylogenetic inference and gene flow analyses

A phylogeny was used to understand the relationships between the 79 individuals included in the study. We obtained a maximum likelihood inference with IQTREE v2.1.2 (Nguyen et al., 2015) for the initial loci dataset using the TVM+F+R6 model as automatically selected by ModelFinder (Kalyaanamoorthy et al., 2017) and 5000 bootstrap replicates. We visualized the resulting phylogeny and assessed bootstrap support using FIGTREE v1.4.2 (Rambaut & Drummond, 2012).

We tested for gene flow between lineages with Treemix v1.13 (Pickrell & Pritchard, 2012), considering the outcome of the STRUCTURE analysis, while also separating individuals from the three European peninsulas. For this analysis, the filtered SNPs dataset was used; this dataset was separated into five different groups/populations (Sierra Nevada, Sicily/Calabria, Peloponnese, European, Eastern), according to geography and the clustering obtained in the STRUCTURE analysis for $K=4$, as shown in Table S1. We tested windows of 9 to 20 SNPs. The R package OptM v0.1.6 (Fitak, 2021) was used to estimate the optimal number of migration edges from the Treemix analysis. Subsequently, we also used the Julia v1.7.2 (Bezanson et al., 2017) package PhyloNetworks v0.14 (Solís-Lemus et al., 2017), with the initial loci dataset. As input, we previously inferred gene trees for each locus using RAxML v8 (Stamatakis, 2014), the GTRGAMMA model and 100 bootstrap replicates per locus, and a species-tree using the coalescent summary method in ASTRAL v5.7.8 (Zhang et al., 2018) as a starting topology. We tested values of reticulations from 0 to 4. We performed 20 independent runs per reticulation value to ensure convergence on a global optimum.

2.6 | Genetic diversity analyses

In order to estimate the origin of the European lineage of *A. urticae*, we used the software TASSEL v5.2.80 (Bradbury et al., 2007) to study its genetic diversity variation across longitude. We considered the software-computed measures of total allele frequency, total number of segregating sites, and average pairwise divergence (π). For this analysis, from the initial SNPs dataset, we filtered out three sub-datasets with VCFtools, representing western Europe, central Europe, and eastern Europe, in order to have a continuous representation of this part of the range, each with three specimens, chosen considering geographical longitude and their highly pure representation of the European lineage according to the STRUCTURE analysis (i.e. those showing very little to no admixture) (Table S3).

2.7 | Demographic history

To explore demographic history for the population pairs Europe–Sierra Nevada and Europe–Sicily/Calabria, we used dadi (Gutenkunst et al., 2009) to analyse joint site frequency spectra (JSFS). The observed JSFS were built using easySFS (Gutenkunst et al., 2009) from the filtered SNP dataset (no indels, only unlinked SNPs with <50% of missing data—13,008 SNPs). We excluded putative hybrid individuals to avoid the confounding effect of contemporary hybridization on the inference of long-term gene flow. We fit 11 demographic models—strict isolation (no_mig); continuous symmetric (sym_mig), and asymmetric (asym_mig) migration; ancient symmetric (anc_sym_mig) and asymmetric (anc_asym_mig) migration; secondary contact symmetric (sec_contact_sym_mig) and asymmetric (sec_contact_asym_mig) migration; continuous symmetric (sym_mig_twoepoch) and asymmetric (asym_mig_twoepoch) migration that varies across two epochs; secondary contact symmetric (sec_contact_sym_mig_three_epoch) and asymmetric (sec_contact_asym_mig_three_epoch) migration followed by isolation—using the dadi_pipeline v3.1.5 (Portik et al., 2017). For all models, we performed four rounds of optimizations. For each round, we ran multiple replicates and used parameter estimates from the best scoring replicate (highest log-likelihood) to seed searches in the following round. We used the default settings in the dadi_pipeline for each round (replicates = 10, 20, 30, 40; maxiter = 3, 5, 10, 15; fold = 3, 2, 2, 1) and optimized parameters using the Nelder–Mead method (optimize_log_fmin). The model providing the best fit to the observed JSFS was determined using the Akaike information criterion (AIC). The run providing the lowest AIC was kept for each model to make comparisons among them. For the conversion of the model parameter values (nu, m, and t) into population sizes in number of individuals (N), migration rates (the fraction of the total population that are new immigrants in each generation, M, later translated into number of individuals per generation), and time in years (T), we used the value of theta (θ) estimated by the software for each model to calculate the ancestral population size (Nref) using the formula $N\text{ref} = \theta/4L\mu$ (Rougeux et al., 2017). L, the proportion of sites included in the analysis, is calculated as $L = zy/x$ (x being the number of SNPs [180,539] originally called from y base pairs in the IPYRAD alignment [3,812,432], and z being the number of SNPs included in the analysis [13,008]). We considered the mutation rate, μ, from the estimation for *Heliconius melpomene* (Keightley et al., 2015), commonly used for butterflies, at 2.9×10^{-9} per base per generation. We assumed a generation time, g, of 0.5 years (average of two generations/year). We used the following formulas to convert the parameters: population size—$N = \text{nu} \times N\text{ref}$; migration rate—$M = \text{m}/2N\text{ref}$; time—$T = 2t \times N\text{ref} \times \text{g}$.

2.8 | Ecological niche modelling (ENM)

Occurrence data of *A. urticae* from 1979 to 2013 (611,218 observations) were downloaded from GBIF (GBIF.org, 2022 [30 May 2023]) GBIF Occurrence Download https://doi.org/10.15468/dl.2xcya5) using the function occ_download from the R package rgbif v3.7.2 (Chamberlain & Boettiger, 2017). A polygon was set (latitude: 29.26758, 72.158; longitude: −11.64551, 48.20801) representing Europe and part of north Africa (included in this analysis to cover

areas of absence). The R package spThin v0.2.0 (Aiello-Lammens et al., 2015) was used to spatially balance species occurrence data, setting up a minimum distance of 10 km between records. Data for the standard 19 bioclimatic variables at 5 arc-min (~10 km) resolution were downloaded from the Paleoclim database (http://www.paleoclim.org/) (Brown et al., 2018). A current time period (1979–2013) was used for the modelling exercise (Karger et al., 2017) and the Last Interglacial (~130,000 YBP) (Otto-Bliesner et al., 2006) and Last Glacial Maximum (~21,000 YBP [Karger et al., 2021]) for projections to the past. Autocorrelation between variables was tested through Pearson correlation coefficients, and variables with correlations with absolute values above .7 were discarded. The five variables retained for use in the model were mean diurnal air temperature range, mean minimum temperature of the coldest month, mean temperatures of the wettest quarter, mean precipitation of the warmest quarter, and mean precipitation of the coldest quarter. The R package biomod2 v3.5.1 (Thuiller et al., 2023) was used to predict environmental suitability areas in Europe for *A. urticae* in the different time periods studied. Occurrence data were supplemented with pseudo-absence data: 5 sets of 1000 points chosen randomly to represent the available environment in the studied area. An ensemble of forecasts of species distribution models was obtained, including projections from six statistical models, namely Generalized Linear Model (GLM), Generalized Additive Model (GAM), Multiple Adaptive Regression Splines (MARS), Classification Tree Analysis (CTA), Flexible Discriminant Analysis (FDA), and Random Forests (RF). Models were calibrated using 80% randomly selected occurrence data, using the True Skill Statistic (TSS). The analysis was replicated 10 times, thus providing a 10-fold internal cross-validation of the models. For present and past projections, we obtained one occurrence probability raster for each statistical model by calculating the mean of all the projections of models with a TSS>0.7. To obtain a single projection for each scenario, we used the option total.consensus, which provides one output of the mean of all projections. Subsequently, occurrence data of the main host plant of *A. urticae*—*U. dioica*—from 1979 to 2013 (885,610 observations) were downloaded from GBIF (GBIF.org [01 September 2023] GBIF Occurrence Download https://doi.org/10.15468/dl.ejdb63), and all steps to obtain a single projection for each time period were repeated. The maps from the final rasters were produced with QGIS v3.16.16 (http://www.qgis.org).

2.9 | Wolbachia infection analysis

Infections by *Wolbachia* bacteria are maternally inherited and may cause male-killing or cytoplasmic incompatibility (Jiggins, 2003; Werren et al., 2008). These infections may trigger selective sweeps, where a particular mitochondrial genome is associated with the expansion of a *Wolbachia* strain. Hence, testing the presence of *Wolbachia* informs on the potential role of this endosymbiont on genetic differentiation. *Wolbachia* sequences were identified using CENTRIFUGE v1.0.4 (Kim et al., 2016), as it has been demonstrated to be a quick and efficient method to detect these bacteria (Hinojosa, Koubínová, et al., 2019; Hinojosa, Monasterio, et al., 2019). A new, de novo assembly had to be produced with IPYRAD v0.7.23 (Eaton & Overcast, 2020), in order to keep all the sequenced genetic information, including from organisms that do not map to the reference genome of *A. urticae*. The loci file was used as input for CENTRIFUGE, which mapped and classified them according to a complete bacterial and viral genome index of the NCBI BLAST's nucleotide database.

3 | RESULTS

3.1 | Principal component analysis

PC1 explained 10.4% of the observed variance and clearly separates the samples from Sierra Nevada from all the rest, with an individual from Cáceres (SW Iberia) appearing approximately midway;

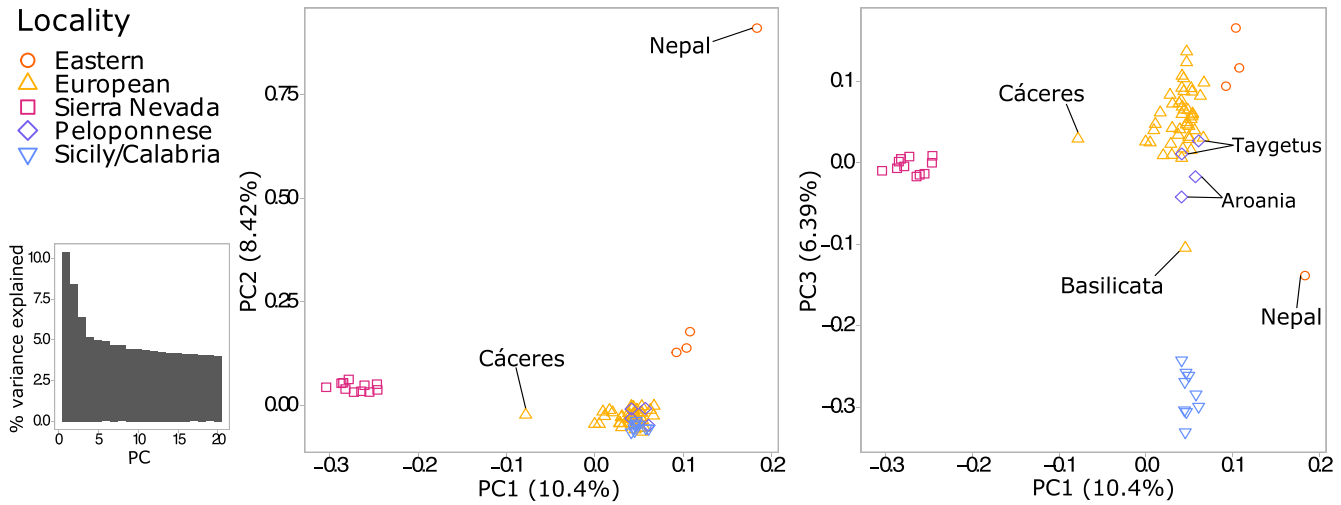

FIGURE 1 Principal component analysis (PCA) for *Aglais urticae*, considering PC1, 2, and 3, according to the highest percentage of variance explained (below legend). Individual samples are represented with a symbol that indicates genetic lineage/locality.

it also separated the eastern samples, especially the individual representative from Nepal; samples from the Peloponnese and Sicily/Calabria were clustered with other European samples (Figure 1). PC2 explained 8.42% of the variance and chiefly separated the sample from Nepal, and some degree of variance is also visible for samples from between Kazakhstan/China, Sierra Nevada, and Europe. PC3 explained 6.39% of the variance and showed separation between the samples from Sicily/Calabria and those from Europe, being the individual from Basilicata intermediate. It also splits individuals from southern Peloponnese (Aroania) from the rest of European; the representative from Nepal appeared again distant from the rest of the eastern samples.

3.2 | Genetic structuring and hybridization between lineages

The best K under the Evanno method was $K=2$ ($\Delta K=240.8$). For $K=2$, a distinct genetic cluster dominated in the individuals from Sierra Nevada, also present, to a lesser degree, in other Iberian samples (Figure 2). For $K=3$, a new cluster with individuals from Sicily and southern Italy appeared, being also well represented in the individuals from the Peloponnese. Some other individuals from Europe showed a small fraction of this component. $K=4$ is the most biologically informative number of clusters, highlighting a new cluster that we refer to as 'Eastern', which includes all samples

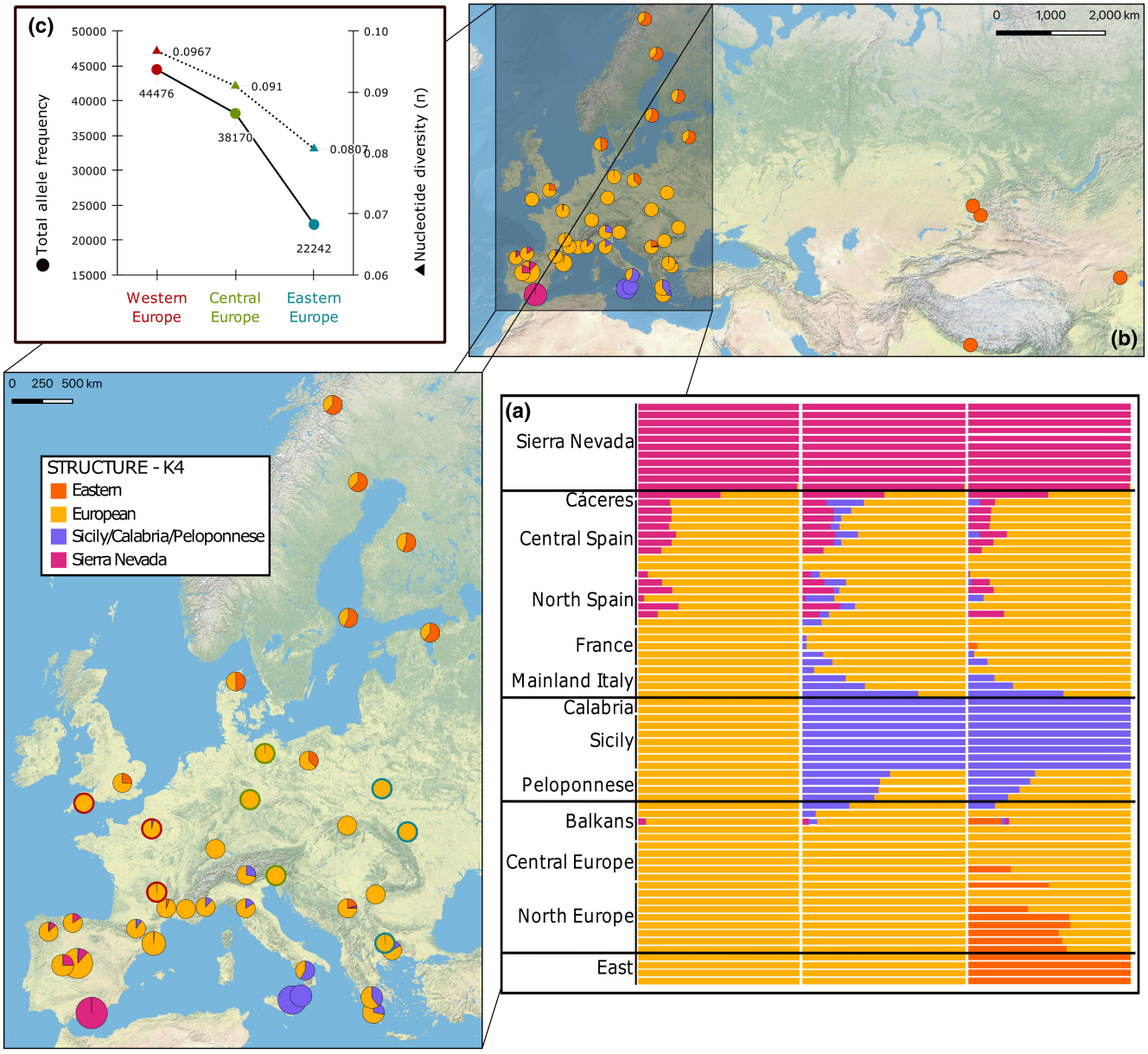

FIGURE 2 STRUCTURE results for *Aglais urticae*, represented as (a) bar plots ($K=2$–4) and as (b) pie charts on a map of the Palearctic and zoomed in Europe ($K=4$) (samples from the same or similar geographic location are pooled in the same pie charts with increased radius); (c) genetic diversity analysis using TASSEL for subsets of the European lineage of *A. urticae* (individuals marked in the map in red, green, and blue), showing west-to-east decline of genetic diversity.

from Asia and is well-represented among samples from eastern and northern Europe. The 'Sierra Nevada' and 'Sicily/Calabria/Peloponnese' clusters maintained their composition as in lower Ks. The remaining samples were mostly clustered in a fourth group, which we refer to as 'European'. This 'European' cluster was present in most of Europe, with the exception of Sierra Nevada and Sicily/Calabria. One individual from Cáceres (Central Spain) and one from Basilicata (South Italy) had ca. 50% of the European cluster and ca. 50% of the clusters typical of Sierra Nevada and Sicily/Calabria/Peloponnese, respectively. The distributions of clustering coefficients (*q*-values) of the simulated hybrid groups between lineages obtained in the second STRUCTURE analysis (Figure 3) were used as assignment criteria for the real genotypes to the four hybrid categories. Here, we use 'hybridization' as a designation for crosses between intraspecific lineages. It is also worth noting that HYBRIDLAB was designed to analyse a small number of nuclear genetic markers such as microsatellites, allozymes, or SNPs and might not work as well with large amounts of data such as those of ddRADseq. It was impossible to distinguish between F1 and F2 due to overlapping distributions for both lineage pairs. For the analysis Europe–Sierra Nevada, the individual from Cáceres was placed between F1/F2 and B2Europe; two other individuals from central Iberian Peninsula (Madrid) were recovered close to the category of B3Europe. The rest of the tested putative hybrids overlapped with the pure Europe category and were thus better interpreted as non-hybrid specimens. As for Europe–Sicily/Calabria, the individual from Basilicata was placed close to the F1/F2 hybrid category; the individual from Tuscany seemed to represent a second-generation backcross with Europe (B2Europe). The rest of the tested putative hybrids were placed with the pure Europe category and are thus unlikely to be hybrid specimens. Considering that the simulated data only go as far as the third-generation hybridization, it is likely that individuals recovered between categories represent admixed individuals from later generations.

3.3 | Phylogenetic inference

A maximum likelihood phylogeny obtained with IQTREE recovered *A. ichnusa* as a well-supported clade (bootstrap support [bs] = 100) that is sister to *A. urticae* (Figure 4). Three main clades were recovered, with good supports (bs = 81–92). A first clade contained samples from the Balkan and Italian peninsulas; within it, a well-supported group (bs = 100) contained all samples from Sicily and southern Italy (Calabria). A second clade was mainly composed of samples from Iberia, except one sample from the United Kingdom. The samples from Sierra Nevada (southern Iberia) formed a well-supported clade (bs = 100), nested within it. A third clade broadly included European samples; within it, a monophyletic clade contained the four samples of the East.

3.4 | Gene flow analyses

OptM estimated that the optimal number of migration edges (that explains 99.8% of the observed variation) was one (Δm = 8.2178, Figure 5). For one migration edge, Treemix suggested the occurrence of admixture from Sierra Nevada to the European lineage (mean weight = 0.292, mean *p*-value = .005), for all the different windows of SNPs. The best phylogenetic network inferred by PhyloNetworks identified one introgression event ($h_{max} = 1$, −loglik = 0.032) from the European lineage to Sierra Nevada.

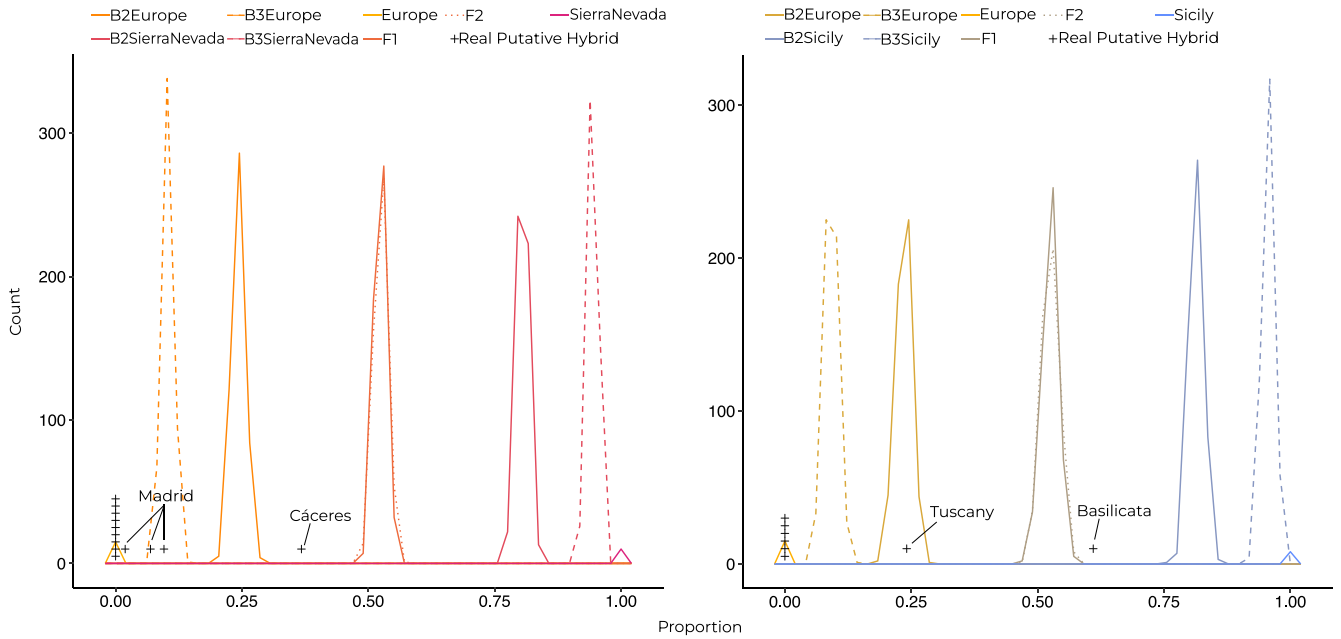

FIGURE 3 Frequency plots of clustering coefficients (*q*-values) for the simulated hybrid groups between the European and Sierra Nevada lineages (left) and the European and Sicily/Calabria lineages (right), and for putative hybrid specimens obtained in nature (crosses).

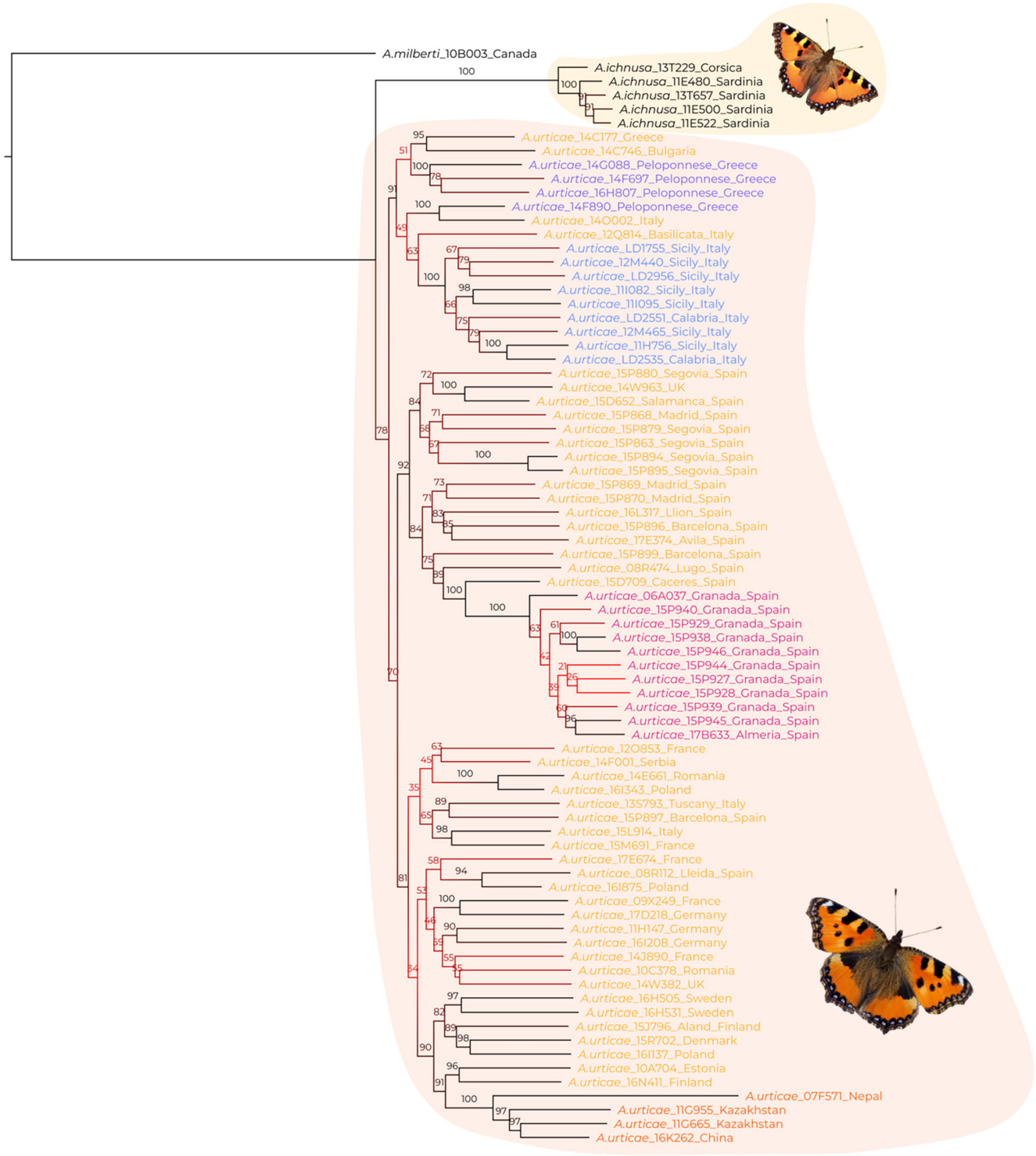

FIGURE 4 Maximum likelihood phylogeny recovered using IQTREE for *Aglais urticae*, *A. ichnusa*, and *A. milberti* (root). Scale represents 0.004 substitutions per site. Branch labels indicate support values.

3.5 | Genetic diversity analyses

The TASSEL analyses on the purest specimens of the European lineage (i.e. those showing very little to no admixture according to the results of structure) revealed a gradual decrease in diversity from west to east in all the indexes: total allele frequency, number of segregating sites, and nucleotide diversity (π, Figure 2, Table S3).

3.6 | Demographic history

The best-supported model, with the lowest AIC values, for both Europe–Sierra Nevada and Europe–Sicily/Calabria was that of ancient asymmetric migration (anc_asym_mig), followed by a general asymmetric migration (asym_mig) model (Table S4). The strict isolation (no_mig) and secondary contact migration followed by

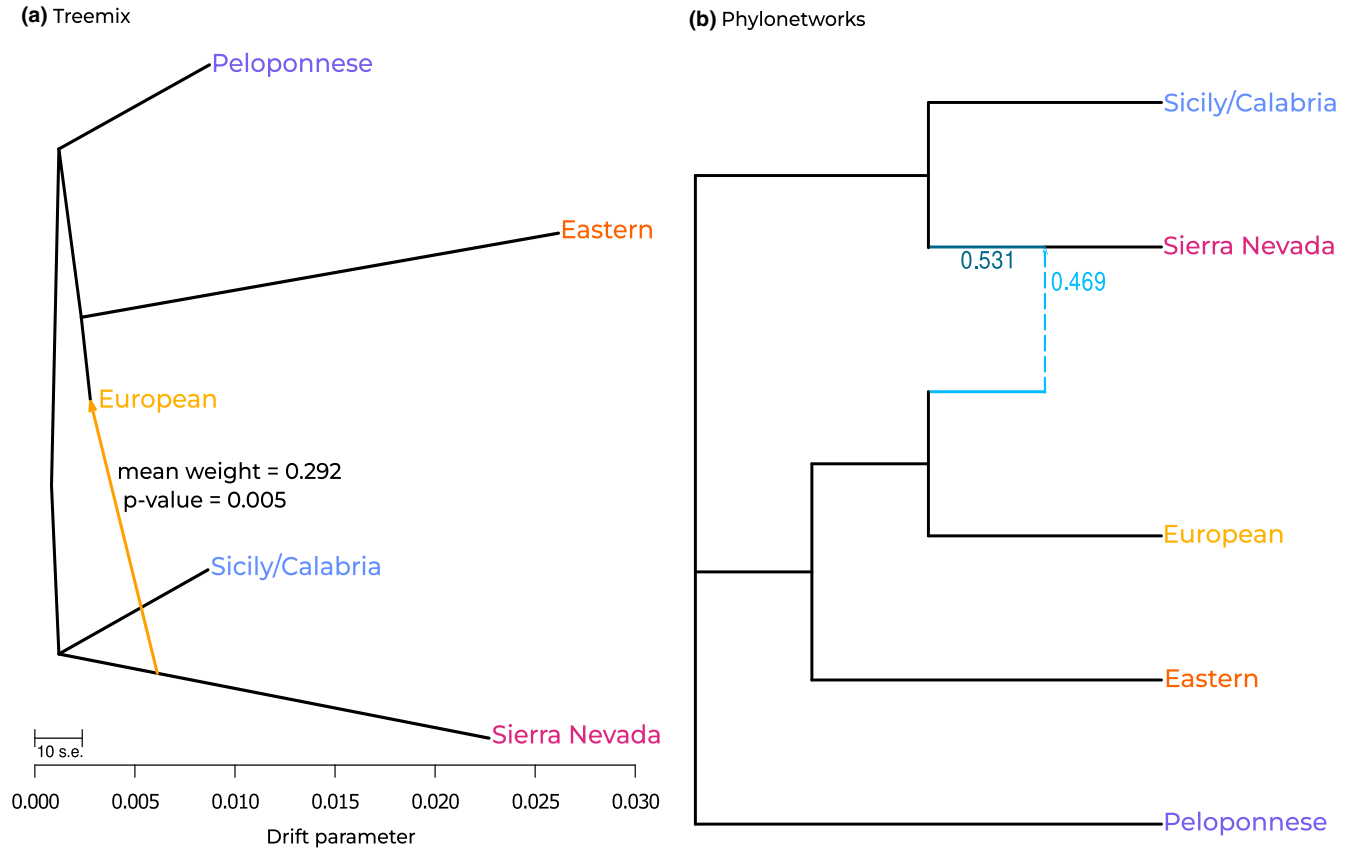

FIGURE 5 (a) Gene flow analyses performed with Treemix (Δm = 8.2178), with the occurrence of admixture between Sierra Nevada and the European lineage, as identified by OptM. The drift parameter is a relative temporal measure, and the scale bar indicates 10 times the average standard error of the relatedness among populations based on the variance–covariance matrix of allele frequencies. (b) Gene flow analysis performed with PhyloNetworks, with one introgression event identified (hmax = 1, −loglik = 0.032) from Europe to Sierra Nevada: this network has the Sierra Nevada lineage as 53.1% sister to Sicily/Calabria and 46.9% sister to the European lineage.

isolation models (sec_contact_sym_mig_three_epoch and sec_contact_asym_mig_three_epoch) were the least supported, presenting high AIC values for both population pairs. For Europe–Sierra Nevada, the best-supported model suggests a split about 400 thousand years ago, with approximately 0.41 individuals migrating from Sierra Nevada to Europe and 1.02 in the opposite direction, per generation, until the start of an isolation period, about 12 thousand years ago (Figure 6). Similarly, for Europe–Sicily/Calabria, the model suggests a split about 450 thousand years ago, and lower values of migration, with approximately 3.00 individuals migrating from Sicily/Calabria to Europe and 3.32 in the opposite direction, per generation, until the start of an isolation period, about 8 thousand years ago (Figure 6). Exact values of model parameters and converted units are presented in Table S5.

3.7 | Ecological niche modelling

Aglais urticae and *U. dioica* seem to respond similarly to all tested bioclimatic variables (Figure S1). Furthermore, both species appear to prefer locations with more humid summers and drier winters. The niche modelling analyses for the current time period agreed with the known distribution of both *A. urticae* and *U. dioica* (Figure 7). They also made visible the relative isolation of the Sierra Nevada population (no corridors in the form of suitable climatic habitat exist), more so than the other southern peninsulas. Model projections for *A. urticae* for the Last Glacial Maximum showed that the southern populations were quite well connected with the rest of Europe and that the lower sea levels also allowed a greater connection between the southern Italian peninsula and the Balkans. A similar result is obtained in the LGM model projection for *U. dioica*, which narrows the location of the extra-Mediterranean refugium in Western Europe to the Atlantic coast. The distributions estimated for the Last Interglacial were relatively similar to the one for the present, although more restricted in the east, especially in northern latitudes.

3.8 | Wolbachia infection analysis

Centrifuge did not identify any locus as belonging to *Wolbachia*, which suggests the absence of infection, in line with previous results by Ilinsky and Kosterin (2017).

4 | DISCUSSION

4.1 | Genomic data clarify previous mitochondrial results

Our study of the genomics of *Aglais urticae* in the western Palearctic recovers four lineages: one in Sierra Nevada, a second one in Sicily and southern Italy, a third one in Scandinavia and northern Europe, and a fourth main lineage throughout Europe. The nuclear DNA (nuDNA) results disagree with the conclusion of virtually no genetic differentiation in mtDNA by Vandewoestijne et al. (2004). On the contrary, they approximate the latest mtDNA patterns published (Dapporto et al., 2022). These concordant results indicate the existence of a mitochondrial DNA lineage in the European peninsulas, separated by regions occupied by a widespread European lineage. However, nuDNA recovers two different southern lineages: one of them in Sierra Nevada, and the other in Sicily/Calabria and in the Peloponnese. The results of population structure analysis and the phylogenetic inference show that populations in Scandinavia, as well as specimens in the United Kingdom, Eastern Europe, and the Baltics, appear to be genetically admixed with an Eastern lineage, with pure specimens from Asia. Interestingly, specimens in north Fennoscandia have a slightly differentiated mtDNA (Dapporto et al., 2022).

Our genomic analyses did not capture the lineage previously detected in the Alps with mtDNA (Dapporto et al., 2022), possibly because our sampling was rather limited for this region. This mtDNA lineage may indicate an extra-Mediterranean glacial refuge close to the Alps that has not expanded postglacially and that may be strongly admixed with the European lineage in the nuclear genome. The genomic results also support the status of species of *A. ichnusa*, as shown by the clear independent clustering of samples from Corsica and Sardinia.

Due to high mutation rate, lower effective population size, haploidy, and no recombination, mitochondrial markers tend to show a higher number of and more contrasted lineages than nuclear markers (Hinojosa, Koubínová, et al., 2019; Hinojosa, Monasterio, et al., 2019; Ivanov et al., 2018; Marshall et al., 2021; Shults et al., 2022). Nevertheless, in the case of *A. urticae*, a similar pattern is visible with both types of markers, albeit it is more gradual in the nuclear genome. In addition, using genomic data, we can capture and study not only differentiation by isolation but also the phenomenon of admixture through gene flow.

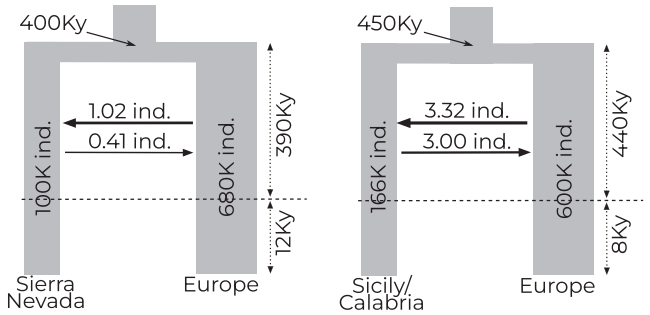

FIGURE 6 Representation of the best demographic models for population pairs Europe–Sierra Nevada (left) and Europe–Sicily/Calabria (right): ancient migration and current isolation. Parameters are approximations and are represented in population sizes—number of individuals; migration rates—number of individuals migrating each generation; time—years.

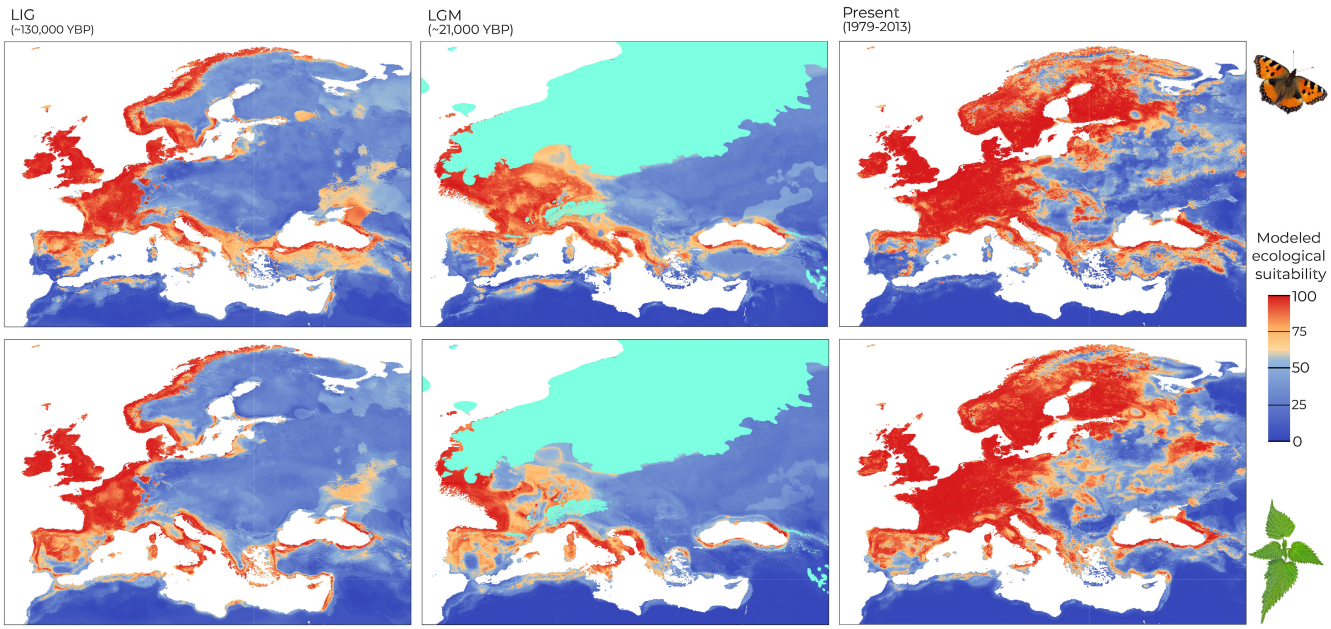

FIGURE 7 Ecological niche modelling for *Aglais urticae* (above) and its main host plant *Urtica dioica* (below) for the Last Interglacial (LIG), Last Glacial Maximum (LGM), and present. LGM ice sheets are represented in aquamarine.

4.2 | Evidence of admixture

There are signs of admixture between the lineages in the three southern European peninsulas and the European one: (1) Treemix and PhyloNetworks analyses both demonstrate the existence of gene flow at least in the Iberian Peninsula, between the Sierra Nevada and the European lineages. The topologies recovered with these programs show differences from that obtained with IQTREE. In fact, a maximum likelihood phylogenetic inference and the models implemented are not designed to incorporate gene flow, and this could explain incongruences and low support values (Leaché et al., 2014). (2) All the models recovered as likely in the demographic history modelling included migration and the model of strict isolation showed the worse fit to the data. (3) Structure shows specimens with a combination of the two components in all the peninsulas. The hybrid simulations recovered the individuals from Cáceres (southwestern Iberia), Basilicata (south Italian peninsula), and Tuscany (Central Italian peninsula) as admixed between the European and relict southern lineage. (4) The PCA shows the specimen from Cáceres, from Basilicata, and from the Peloponnese, especially those from Aroania, in between the clusters of their respective relict lineages and the main European lineage. Moreover, all the specimens studied from Scandinavia show substantial admixture between the European and the Eastern lineage, and this region possibly represents another secondary contact zone. Scandinavia and north-eastern Europe were previously found to harbour contact zones between eastern and southern recolonization routes after Glacial Maximum deglaciation for multiple organisms (Maresova et al., 2021). Furthermore, cytoplasmic incompatibility caused by *Wolbachia* should not limit gene flow in any part of the distribution, as no individual seems to be infected. Thus, we conclude that admixture is a general phenomenon in *A. urticae* that takes place in every area studied where lineages become in contact.

4.3 | Genetic diversity

While we document genetic diversity in the form of four main genetic lineages (European, Sierra Nevada, Sicily/Calabria/Peloponnese, and Eastern), the substantial degree of admixture that we infer between them suggests that they may be better treated as infraspecific parapatric taxa. However, gene flow at the contact zones makes it difficult to define subspecies with precise geographic boundaries. The European lineage corresponds to *A. u. urticae* (Linnaeus, 1758), ranging from the central latitudes in the Mediterranean peninsulas to Scandinavia, but admixture in the north with the Eastern lineage blurs the picture.

The population genetic analysis suggests that Italian and Greek samples of the Sicily/Calabria/Peloponnese relict lineage are very closely related, likely due to these locations allowing admixture during glacial periods, when the sea levels dropped and the connectivity between the two peninsulas increased (Figure 7; Mazza et al., 2013; Velasquez et al., 2021). This relict lineage, found in Sicily, Calabria, and the Peloponnese, seemingly corresponds to the subspecies *A. u. turcica* (Staudinger, 1871) originally described from southern Balkans and Asia Minor. Molecular studies on specimens from Asia Minor would be necessary in order to assess if this Sicily/Calabria/Peloponnese lineage is present in this region. Note that the populations in the Peloponnese seem to be highly admixed with the main European lineage at present, which again makes translation to taxonomy difficult.

The Sierra Nevada lineage would correspond to the taxon *fidelis* (Gomez Bustillo & Fernandez-Rubio, 1974), but this is a *nomen nudum* (García-Barros et al., 2013). This taxon has apparently developed at least one adaptation: the shortening of the breeding season to a bivoltine cycle, which results in the avoidance, through loss of synchrony, of a late-season parasitoid fly, *Sturmia bella* (Stefanescu et al., 2022).

Given the enormous potential distribution of the Eastern lineage and the limitations of our sampling, there is not a strong basis to attribute a subspecific name to this taxon. It could correspond to the subspecies *A. u. polaris* (Staudinger, 1871), *A. u. eximia* (Sheljuzhko, 1919), or *A. u. baicalensis* (Kleinschmidt, 1929), but knowing the detailed distribution limits of the lineage is needed to accurately assign it to one of these categories. The specimen from Nepal was recovered as highly differentiated in the PCA, and it apparently corresponds to the morphologically differentiated taxon *caschmirensis* (Kollar, 1844). This taxon has been poorly studied and, although generally considered a different species, its status is debated (Cotton et al., 2015; Nylin, 2001). Unfortunately, with a single specimen in the analysis, we cannot reach strong conclusions.

4.4 | The opposed forces of differentiation by isolation and admixture by gene flow

Species' isolation because of repeated range contractions during glacial cycles is considered an important source of diversity (Hewitt, 1999). The outcomes of secondary contacts are mostly dictated by the amount and type of genetic differences accumulated during the time in isolation of the populations in question, and how populations are locally adapted to their environmental conditions, including both abiotic and biotic factors. In this study, we document one case of the diversity present in the southern European peninsulas and a snapshot of the intraspecific processes of differentiation and homogenization working in opposite directions across the glacial and interglacial periods of the Quaternary. We show the presence of not only relict lineages in the southern European peninsulas but also another lineage present in mainland Europe. According to genetic diversity results (Figure 2), the main European lineage could have originated in the west, possibly from extra-Mediterranean refugia in western Europe (Stewart et al., 2010), which agrees well with our ENM results (Figure 7) and with previous studies finding glacial refugia in central Europe and western Asia (Bilton et al., 1998; Kühne et al., 2017; Schmitt & Varga, 2012). These results might also

be compatible with the continental/oceanic longitudinal gradient concept of Stewart et al. (2010), where Atlantic (oceanic) species expand to the east during interglacial periods, thanks to the decrease of continentality, and retreat to the west during glacial periods. It is hard to completely discard the possibility of an origin from an unsampled Asian region and a recent colonization westward, although our results suggest otherwise. Moreover, this lineage likely predates the last glaciation and differentiated across multiple glacial cycles, as suggested by the divergence of the mtDNA (Dapporto et al., 2022). Indeed, the demographic analyses estimate that the split between the European and peninsular lineages occurred between 400 and 450 thousand years ago, during the middle Pleistocene, at the beginning of the Mindel-Riss interglacial period according to Penck and Brückner's scheme of glacial/interglacial periods in the Alpines (Milankovitch, 1941; Mustafa, 2021), and coincided with the marine isotopic stage 11 (MIS 11, Raynaud et al., 2005).

We document ongoing admixture between the European and peninsular lineages and suggest that it could have been higher in the past. Indeed, the demographic analyses estimate the main decrease in gene flow approximately 122 to 8 thousand years ago, after the end of the Last Glacial Maximum and the onset of the Holocene, matching the time period of the Younger Dryas (YD) stadial, the last extreme cold event detected in the northern hemisphere during the last deglaciation (Broecker et al., 2010). This agrees with the fact that *A. urticae* is a cold-tolerant species that exhibits a notable dispersal capacity. The outcome of sustained gene flow may be either the formation of a stable hybrid zone or complete admixture (lineage fusion or dilution, depending on relative population sizes and symmetry of gene flow). Gene flow with the widely distributed European lineage, allied to the recent changes in climate which make survival more difficult in the southern range of the species (Christidis & Stott, 2021; Stagge et al., 2017), can hypothetically lead to the disappearance of the southern relict lineages. We argue that the probability of extinction is remarkably different for the different populations studied. On one extreme, prospects for the relict lineage present in the Peloponnese are gloomy: it is seemingly undergoing massive admixture, and we did not record any pure specimen. Nevertheless, this lineage is also present in Sicily/Calabria, where only pure specimens were found. In Sierra Nevada, no signs of admixture from the European lineage were detected, although occasional dispersal seems to take place, given a case of ca. 50% admixture in Central Spain. This is possibly due to the relatively high population numbers in this mountain range compared to the dramatically dwindling populations in central Iberia (Caro-Miralles & Gutiérrez, 2023; Stefanescu et al., 2022). Therefore, we note here that population demography may be a key factor in the outcome of gene flow between lineages.

4.5 | A complex scenario of differentiation with gene flow

The fact that the species is relatively rare at the contact zones between southern isolates and the European lineage (García-Barros et al., 2004; Pamperis, 2009) may illustrate a cyclical process that has been repeated at each glacial cycle, resulting in the maintenance of the southern lineages, perhaps better adapted to warm climates. Given the presented genetic and distribution modelling data, we propose the following scenario: (1) In the glacial periods, the European lineage mostly gathered in a wide extra-Mediterranean refugium in Western Europe. However, potential connections with the southern lineages at the Mediterranean peninsulas are visible in the models, and gene flow was probably maintained to a certain degree, although not sufficient as to completely homogenize the main gene pools. We can only speculate on whether any barrier to gene flow existed in addition to geography, in the form of hybrid depression or natural selection for local adaptation. Gene flow between the Italian and Balkan peninsulas was particularly facilitated, and admixture was intense. (2) In the interglacial periods, the contact zones between the southern and European lineages seem to be rendered less suitable, as is the case at present. Nevertheless, long-range dispersing individuals may still maintain occasional gene flow. Rather surprisingly, the southern lineages seem to not be so strongly affected by climate warming (e.g. in Sierra Nevada) compared to the central parts of the peninsulas (Olivares et al., 2011). A potential explanation could be a higher tolerance of the southern lineages to climate warming. In summary, we infer dynamic north–south parapatric systems that have been maintained across multiple glacial cycles, where a certain degree of gene flow existed (higher in the glacial, lower in the interglacial periods) that was low enough to maintain lineage differentiation.

This represents the first genome-wide phylogeographic study of a cold-tolerant butterfly species with considerable dispersal ability and a wide distribution, from Scandinavia to the Mediterranean. The complex result (differentiation with gene flow, existence of both Mediterranean and extra-Mediterranean refugia) contrasts with those obtained for more localized and less dispersive taxa. In previously studied warm-adapted taxa (Dapporto, 2009; Schmitt, Habel, et al., 2006; Schmitt, Hewitt, & Müller, 2006) and cold-adapted taxa like *Erebia* (Hinojosa, Koubínová, et al., 2019; Hinojosa, Monasterio, et al., 2019; Schmitt, Habel, et al., 2006; Schmitt, Hewitt, & Müller, 2006; Sistri et al., 2022), glacial or interglacial periods produce complete isolation and strong demographic bottlenecks, respectively. Thus, the comparison with results from studies on other taxa shows the importance of species traits in determining population dynamics and responses to climatic changes.

AUTHOR CONTRIBUTIONS

C.S., D.G., and R.V. conceived the original idea; J.C.H., L.D., G.T., C.S., D.G., and R.V. collected specimens; V.M. analysed the data; V.M. wrote the original manuscript with support from J.C.H. and R.V.; all the authors reviewed and edited the article.

ACKNOWLEDGEMENTS

We thank Mutanen, M. for the support in the laboratory work required for obtaining the data used in this study. We thank the colleagues

who provided samples: Dincă, V.; Vodă, R.; Requejo, R.; Viader, S.; Lukhtanov, V.; Hernández-Roldán, J.; Pitteloud, C.; Barea-Azcón, J.M.; Miralles, A.; Ubach, A.; Mølgaard, M.S.; Gascoigne-Pees, M.; Escuer, P.; Blasco, J.; Cuvelier, S.; Bordoni, A.; Gaunet, A. This research was supported by grants PID2019-107078GB-I00 and PID2022-139689NB-I00 (funded by Ministerio de Ciencia e Innovación (MCIN)/Agencia Estatal de Investigación (AEI)/10.13039/501100011033 and ERDF A way of making Europe) and by grant 2021-SGR-00420 from Departament de Recerca i Universitats de la Generalitat de Catalunya to RV, by grant CGL2014-57784-P funded by MCIN/AEI/10.13039/501100011033 to DG, by PRE2020-094870 funded by MCIN/AEI/10.13039/501100011033 and 'European Social Fund (ESF) Investing in your future' to VM, and by BES-2017-080641 funded by MCIN/AEI/10.13039/501100011033 and by 'ESF Investing in your future' to JCH. L.D. acknowledges the support of NBFC to University of Florence, Department of Biology, funded by the Italian Ministry of University and Research, PNRR, Missione 4 Componente 2, 'Dalla ricerca all'impresa', Investimento 1.4, Project CN00000033 and by several projects with the Direttiva Biodiversità of the Italian Environmental Ministry. G.T. was supported by grant PID2020-117739GA-100 funded by Ministerio de Ciencia e Innovación (MCIN)/Agencia Estatal de Investigación (AEI)/10.13039/501100011033 and by grant 2021-SGR-01334 from Departament de Recerca i Universitats de la Generalitat de Catalunya. We acknowledge support for the publication fee by the CSIC Open Access Publication Support Initiative thorugh its Unit of Information Resources for Research (URICI). Access and research permits specific to this study were provided by Comunidad de Madrid, Parque Nacional de la Sierra de Guadarrama, Parque Regional de la Cuenca Alta del Manzanares, and Parque Regional del Curso Medio del Río Guadarrama and Generalitat de Catalunya, among other authorities that provided general collecting permits.

CONFLICT OF INTEREST STATEMENT

The authors have no conflicts of interest to declare.

DATA AVAILABILITY STATEMENT

Raw ddRADseq FASTQ reads: NCBI Bioproject PRJNA980705. Alignments of ddRADseq loci and SNPs: figshare (DOI: https://doi.org/10.6084/m9.figshare.23303741.v3).

ORCID

Leonardo Dapporto https://orcid.org/0000-0001-7129-4526
Gerard Talavera https://orcid.org/0000-0003-1112-1345
Roger Vila https://orcid.org/0000-0002-2447-4388

REFERENCES

Aiello-Lammens, M. E., Boria, R. A., Radosavljevic, A., Vilela, B., & Anderson, R. P. (2015). spThin: An R package for spatial thinning of species occurrence records for use in ecological niche models. *Ecography*, *38*(5), 541–545.

Alcala, N., & Vuilleumier, S. (2014). Turnover and accumulation of genetic diversity across large time-scale cycles of isolation and connection of populations. *Proceedings. Biological Sciences/The Royal Society*, *281*(1794), 20141369.

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, *19*(9), 1655–1664. https://doi.org/10.1101/gr.094052.109

Bezanson, J., Edelman, A., Karpinski, S., & Shah, V. B. (2017). Julia: A fresh approach to numerical computing. *SIAM Review. Society for Industrial and Applied Mathematics*, *59*(1), 65–98.

Bilton, D. T., Mirol, P. M., Mascherett, S., Fredg, K., Zima, J., & Searle, J. B. (1998). Mediterranean Europe as an area of endemism for small mammals rather than a source for northwards postglacial colonization. *Proceedings of the Royal Society of London B*, *265*, 1219–1226. https://doi.org/10.1098/rspb.1998.0423

Bishop, G., Ebdon, S., Lohse, K., Vila, R., & Darwin Tree of Life Barcoding collective, Wellcome Sanger Institute Tree of Life programme, Wellcome Sanger Institute Scientific Operations: DNA Pipelines collective, Tree of Life Core Informatics collective, & Darwin Tree of Life Consortium. (2021). The genome sequence of the small tortoiseshell butterfly, *Aglais urticae* (Linnaeus, 1758). *Wellcome Open Research*, *6*, 233.

Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., & Buckler, E. S. (2007). TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics*, *23*(19), 2633–2635. https://doi.org/10.1093/bioinformatics/btm308

Broecker, W. S., Denton, G. H., Edwards, R. L., Cheng, H., Alley, R. B., & Putnam, A. E. (2010). Putting the Younger Dryas cold event into context. *Quaternary Science Reviews*, *29*(9–10), 1078–1081. https://doi.org/10.1016/j.quascirev.2010.02.019

Brown, J. L., Hill, D. J., Dolan, A. M., Carnaval, A. C., & Haywood, A. M. (2018). PaleoClim, high spatial resolution paleoclimate surfaces for global land areas. *Scientific Data*, *5*(1), 180254. https://doi.org/10.1038/sdata.2018.254

Butlin, R. K., Galindo, J., & Grahame, J. W. (2008). Review. Sympatric, parapatric or allopatric: The most important way to classify speciation? *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *363*(1506), 2997–3007.

Caro-Miralles, E., & Gutiérrez, D. (2023). Butterfly population trends track vegetation encroachment but not climate warming in a Mediterranean mountain. *Biodiversity and Conservation*, *32*, 2017–2035. https://doi.org/10.1007/s10531-023-02589-9

Chamberlain, S. A., & Boettiger, C. (2017). R Python, and Ruby clients for GBIF species occurrence data. *PeerJ*. https://doi.org/10.7287/peerj.preprints.3304v1

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. *GigaScience*, *4*(1). https://doi.org/10.1186/s13742-015-0047-8

Christidis, N., & Stott, P. A. (2021). The influence of anthropogenic climate change on wet and dry summers in Europe. *Science Bulletin*, *66*, 813–823. https://doi.org/10.1016/j.scib.2021.01.020

Cotton, A. M., Fric, Z. F., Gupta, I. J., Van Gasse, P., Inayoshi, Y., Khoon, K. S., Kirton, L. G., Larsen, T. B., Markhasiov, M., Majumdar, M., Pequin, O., Rieger, J., Saito, M., Smith, C. P., Varshney, R. K., & Smetacek, P. (2015). *A synoptic catalogue of the butterflies of India*. Butterfly Research Centre, Bhimtal & Indinov Publishing.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R., & 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158.

Dapporto, L. (2009). Speciation in Mediterranean refugia and post-glacial expansion of *Zerynthia polyxena* (Lepidoptera, Papilionidae). *Journal of Zoological Systematics and Evolutionary Research*, *48*(3), 229–237. https://doi.org/10.1111/j.1439-0469.2009.00550.x

Dapporto, L., Menchetti, M., Vodă, R., Corbella, C., Cuvelier, S., Djemadi, I., Gascoigne-Pees, M., Hinojosa, J. C., Lam, N. T., Serracanta, M., Talavera, G., Dincă, V., & Vila, R. (2022). The atlas of mitochondrial

genetic diversity for Western Palaearctic butterflies. *Global Ecology and Biogeography: A Journal of Macroecology*, *31*, 2184–2190. https://doi.org/10.1111/geb.13579

Duckworth, R. A., & Semenov, G. A. (2017). Hybridization associated with cycles of ecological succession in a passerine bird. *The American Naturalist*, *190*, 94–105.

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. 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.

Feder, J. L., Egan, S. P., & Nosil, P. (2012). The genomics of speciation-with-gene-flow. *Trends in Genetics: TIG*, *28*(7), 342–350.

Fitak, R. R. (2021). OptM: Estimating the optimal number of migration edges on population trees using Treemix. *Biology Methods & Protocols*, *6*(1), bpab017.

García-Barros, E., Munguira, M. L., Martín, J., Romo, H., García-Pereira, P., & Maravalhas, E. S. (2004). Atlas de las Mariposas Diurnas de la Península Ibérica y Baleares (Lepidoptera: Papilionoidea & Hesperioidea). *Monografías* SEA (11), Zaragoza, 228 pp.

García-Barros, E., Munguira, M. L., Stefanescu, C., & Vives Moreno, A. (2013). *Fauna Ibérica Volumen 37: Lepidoptera: Papilionoidea*. Consejo Superior de Investigaciones Científicas.

GBIF.org. (25 May 2022). *GBIF Occurrence Download*. https://doi.org/10.15468/dl.nbyd85

Gomez Bustillo, M. R., & Fernandez-Rubio, F. (1974). *Mariposas de la peninsula iberica*. *Ropaloceros* (Vol. *2*, pp. 198–258). ICONA.

Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H., & Bustamante, C. D. (2009). Inferring the joint demographic history of multiple populations from multidimensional SNP data. *PLoS Genetics*, *5*, e1000695. https://doi.org/10.1371/journal.pgen.1000695

Hewitt, G. M. (1999). Post-glacial re-colonization of European biota. *Biological Journal of the Linnean Society*, *68*(1–2), 87–112. https://doi.org/10.1111/j.1095-8312.1999.tb01160.x

Higgins, L. G., & Riley, N. D. (1978). *Die Tagfalter Europas und Nordwestafrikas*. Übersetzt und bearbeitet von Dr. Walter Forster. (2. Auflage ed.). Paul Parey, 377 pp.

Hinojosa, J. C., Koubínová, D., Szenteczki, M. A., Pitteloud, C., Dincă, V., Alvarez, N., & Vila, R. (2019). A mirage of cryptic species: Genomics uncover striking mitonuclear discordance in the butterfly *Thymelicus sylvestris*. *Molecular Ecology*, *28*(17), 3857–3868.

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. https://doi.org/10.1093/biolinnean/bly182

Ilinsky, Y., & Kosterin, O. E. (2017). Molecular diversity of *Wolbachia* in Lepidoptera: Prevalent allelic content and high recombination of MLST genes. *Molecular Phylogenetics and Evolution*, *109*, 164–179. https://doi.org/10.1016/j.ympev.2016.12.034

Ivanov, V., Lee, K. M., & Mutanen, M. (2018). Mitonuclear discordance in wolf spiders: Genomic evidence for species integrity and introgression. *Molecular Ecology*, *27*(7), 1681–1695.

Jiggins, F. M. (2003). Male-killing Wolbachia and mitochondrial DNA: Selective sweeps, hybrid introgression and parasite population dynamics. *Genetics*, *164*(1), 5–12. https://doi.org/10.1093/genetics/164.1.5

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., & Jermiin, L. S. (2017). ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods*, *14*(6), 587–589.

Karger, D. N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R. W., Zimmermann, N. E., Linder, H. P., & Kessler, M. (2017). Climatologies at high resolution for the earth's land surface areas. *Scientific Data*, *4*, 170122.

Karger, D. N., Nobis, M. P., Normand, S., Graham, C. H., & Zimmermann, N. E. (2021). *CHELSA-TraCE21k v1.0. Downscaled transient temperature and precipitation data since the last glacial maximum*. https://doi.org/10.5194/cp-2021-30

Keightley, P. D., Pinharanda, A., Ness, R. W., Simpson, F., Dasmahapatra, K. K., Mallet, J., Davey, J. W., & Jiggins, C. D. (2015). Estimation of the spontaneous mutation rate in *Heliconius melpomene*. *Molecular Biology and Evolution*, *32*(1), 239–243. https://doi.org/10.1093/molbev/msu302

Kim, D., Song, L., Breitwieser, F. P., & Salzberg, S. L. (2016). Centrifuge: Rapid and sensitive classification of metagenomic sequences. *Genome Research*, *26*(12), 1721–1729.

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.

Kühne, G., Kosuch, J., Hochkirch, A., & Schmitt, T. (2017). Extra-Mediterranean glacial refugia in a Mediterranean faunal element: The phylogeography of the chalk-hill blue *Polyommatus coridon* (Lepidoptera, Lycaenidae). *Scientific Reports*, *7*, 43533. https://doi.org/10.1038/srep43533

Leaché, A. D., Harris, R. B., Rannala, B., & Yang, Z. (2014). The influence of gene flow on species tree estimation: A simulation study. *Systematic Biology*, *63*(1), 17–30.

Linck, E., & Battey, C. J. (2019). On the relative ease of speciation with periodic gene flow. *bioRxiv*. https://doi.org/10.1101/758664

Maresova, J., Bartonova, A. S., Konvicka, M., Høye, T. T., Gilg, O., Kresse, J. C., Shapoval, N. A., Yakovlev, R. V., & Fric, Z. F. (2021). The story of endurance: Biogeography and the evolutionary history of four Holarctic butterflies with different habitat requirements. *Journal of Biogeography*, *48*, 590–602. https://doi.org/10.1111/jbi.14022

Marshall, T. L., Chambers, E. A., Matz, M. V., & Hillis, D. M. (2021). How mitonuclear discordance and geographic variation have confounded species boundaries in a widely studied snake. *Molecular Phylogenetics and Evolution*, *162*, 107194.

Mazza, P., Lovari, S., Masini, F., Marco, M., & Rustioni, M. (2013). A multidisciplinary approach to the analysis of multifactorial land mammal colonization of islands. *Bioscience*, *63*, 939–951. https://doi.org/10.1525/bio.2013.63.12.7

Menchetti, M., Talavera, G., Cini, A., Salvati, V., Dincă, V., Platania, L., Bonelli, S., Balletto, E., Vila, R., & Dapporto, L. (2021). Two ways to be endemic. Alps and Apennines are different functional refugia during climatic cycles. *Molecular Ecology*, *30*(5), 1297–1310.

Milankovitch, M. (1941). *Kanon der Erdbestrahlungen und seine Anwendung auf das Eiszeitenproblem*. Royal Serbian Academy, Section of Mathematical and Natural Sciences, 33, Belgrade.

Mustafa, A. (2021). *Pioneers of the ice age models: A brief history from Agassiz to Milankovitch*. https://doi.org/10.5194/hgss-2021-17

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.

Nielsen, E. E., Bach, L. A., & Kotlicki, P. (2006). Hybridlab (version 1.0): A program for generating simulated hybrids from population samples. *Molecular Ecology Notes*, *6*(4), 971–973. https://doi.org/10.1111/j.1471-8286.2006.01433.x

Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, *32*(1), 268–274.

Nylin, S. (2001). Phylogeny of Polygonia, Nymphalis and related butterflies (Lepidoptera: Nymphalidae): A total-evidence analysis. *Zoological Journal of the Linnean Society*, *132*(4), 441–468.

Olivares, F. J., Barea-Azcón, J. M., Pérez-López, F. J., Tinaut, A., & Henares, I. (2011). *Las Mariposas Diurnas de Sierra Nevada*. Consejería de Medio Ambiente, 512 pp.

Otto, S. P., & Whitlock, M. C. (1997). The probability of fixation in populations of changing size. *Genetics*, *146*(2), 723–733.

Otto-Bliesner, B. L., Marshall, S. J., Overpeck, J. T., Miller, G. H., & Hu, A. (2006). Simulating Arctic climate warmth and icefield retreat in the last interglaciation. *Science*, *311*(5768), 1751–1753.

Pamperis, L. N. (2009). *The butterflies of Greece* (2nd revised ed.). Editions Pamperis, 766 pp.

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS One*, *7*(5), e37135.

Petit, R. J., Aguinagalde, I., de Beaulieu, J.-L., Bittkau, C., Brewer, S., Cheddadi, R., Ennos, R., Fineschi, S., Grivet, D., Lascoux, M., Mohanty, A., Müller-Starck, G., Demesure-Musch, B., Palmé, A., Martín, J. P., Rendell, S., & Vendramin, G. G. (2003). Glacial refugia: Hotspots but not melting pots of genetic diversity. *Science*, *300*(5625), 1563–1565.

Pickrell, J. K., & Pritchard, J. K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genetics*, *8*(11), e1002967.

Portik, D. M., Leaché, A. D., Rivera, D., Barej, M. F., Burger, M., Hirschfeld, M., Rödel, M., Blackburn, D. C., & Fujita, M. K. (2017). Evaluating mechanisms of diversification in a Guineo-congolian tropical forest frog using demographic model selection. *Molecular Ecology*, *26*(19), 5245–5263. https://doi.org/10.1111/mec.14266

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*(2), 945–959. https://doi.org/10.1093/genetics/155.2.945

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., de Bakker, P. I. W., Daly, M. J., & Sham, P. C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, *81*(3), 559–575. https://doi.org/10.1086/519795

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rambaut, A., & Drummond, A. J. (2012). *FigTree version 1.4.0*.

Raynaud, D., Barnola, J. M., Souchez, R., Lorrain, R., Petit, J. R., Duval, P., & Lipenkov, V. Y. (2005). The record for marine isotopic stage 11. *Nature*, *436*, 39–40. https://doi.org/10.1038/43639b

Rougeux, C., Bernatchez, L., & Gagnaire, P. A. (2017). Modeling the multiple facets of speciation-with-gene-flow toward inferring the divergence history of lake whitefish species pairs (*Coregonus clupeaformis*). *Genome Biology and Evolution*, *9*, 2057–2074. https://doi.org/10.1093/gbe/evx150

Roy, K., Valentine, J. W., Jablonski, D., & Kidwell, S. M. (1996). Scales of climatic variability and time averaging in Pleistocene biotas: Implications for ecology and evolution. *Trends in Ecology & Evolution*, *11*(11), 458–463.

Schmitt, T., Habel, J. C., Zimmermann, M., & Müller, P. (2006). Genetic differentiation of the marbled white butterfly, Melanargia galathea, accounts for glacial distribution patterns and postglacial range expansion in southeastern Europe. *Molecular Ecology*, *15*(7), 1889–1901.

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.

Schmitt, T., & Varga, Z. (2012). Extra-Mediterranean refugia: The rule and not the exception? *Frontiers in Zoology*, *9*, 1–12. https://doi.org/10.1186/1742-9994-9-22

Shults, P., Hopken, M., Eyer, P.-A., Blumenfeld, A., Mateos, M., Cohnstaedt, L. W., & Vargo, E. L. (2022). Species delimitation and mitonuclear discordance within a species complex of biting midges. *Scientific Reports*, *12*(1), 1730.

Sistri, G., Menchetti, M., Santini, L., Pasquali, L., Sapienti, S., Cini, A., Platania, L., Balletto, E., Barbero, F., Bonelli, S., Casacci, L. P., Dincă, V., Vila, R., Mantoni, C., Fattorini, S., & Dapporto, L. (2022). The isolated *Erebia pandrose* Apennine population is genetically unique and endangered by climate change. *Insect Conservation and Diversity*, *15*(1), 136–148. https://doi.org/10.1111/icad.12538

Solís-Lemus, C., Bastide, P., & Ané, C. (2017). PhyloNetworks: A package for phylogenetic networks. *Molecular Biology and Evolution*, *34*(12), 3292–3298.

Stagge, J. H., Kingston, D. G., Tallaksen, L. M., & Hannah, D. M. (2017). Observed drought indices show increasing divergence across Europe. *Scientific Reports*, *7*(1), 14045.

Stamatakis, A. (2014). RaxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*(9), 1312–1313.

Stefanescu, C., Colom, P., Barea-Azcón, J. M., Horsfield, D., Komac, B., Miralles, A., Shaw, M. R., Ubach, A., & Gutiérrez, D. (2022). Larval parasitism in a specialist herbivore is explained by phenological synchrony and host plant availability. *The Journal of Animal Ecology*, *91*(5), 1010–1023.

Stewart, J. R., Lister, A. M., Barnes, I., & Dalén, L. (2010). Refugia revisited: Individualistic responses of species in space and time. *Proceedings. Biological Sciences/The Royal Society*, *277*(1682), 661–671.

Strobino, R. (1971). Observations concernant *Aglais urticae ichnusa* Hb. En Corse. *Entomops*, *3*, 267–272.

Thuiller, W., Georges, D., Gueguen, M., Engler, R., Breiner, F., Lafourcade, B., & Patin, R. (2023). *Biomod2: Ensemble platform for species distribution modeling*. R package version 4.2-3.

Vandewoestijne, S., Baguette, M., Brakefield, P. M., & Saccheri, I. J. (2004). Phylogeography of *Aglais urticae* (Lepidoptera) based on DNA sequences of the mitochondrial COI gene and control region. *Molecular Phylogenetics and Evolution*, *31*(2), 630–646.

Vandewoestijne, S., Neve, G., & Baguette, M. (1999). Spatial and temporal population genetic structure of the butterfly *Aglais urticae* L. (Lepidoptera, Nymphalidae). *Molecular Ecology*, *8*(9), 1539–1543. https://doi.org/10.1046/j.1365-294x.1999.00725.x

Velasquez, P., Kaplan, J. O., Messmer, M., Ludwig, P., & Raible, C. C. (2021). The role of land cover in the climate of glacial Europe. *Climate of the Past*, *17*, 1161–1180. https://doi.org/10.5194/cp-17-1161-2021

Werren, J. H., Baldo, L., & Clark, M. E. (2008). Wolbachia: Master manipulators of invertebrate biology. In. *Nature Reviews Microbiology*, *6*(10), 741–751. https://doi.org/10.1038/nrmicro1969

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., … Yutani, H. (2019). Welcome to the Tidyverse. *In Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Wiemers, M., Balletto, E., Dincă, V., Faltynek Fric, Z., Lamas, G., Lukhtanov, V., Munguira, M. L., van Swaay, C. A. M., Vila, R., Vliegenthart, A., Wahlberg, N., & Verovnik, R. (2018). An updated checklist of the European Butterflies (Lepidoptera, Papilionoidea). *ZooKeys*, *811*, 9–45.

Zhang, C., Rabiee, M., Sayyari, E., & Mirarab, S. (2018). ASTRAL-III: Polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics*, *19*(Suppl 6), 153.

SUPPORTING INFORMATION

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

How to cite this article: Marques, V., Hinojosa, J. C., Dapporto, L., Talavera, G., Stefanescu, C., Gutiérrez, D., & Vila, R. (2024). The opposed forces of differentiation and admixture across glacial cycles in the butterfly *Aglais urticae*. *Molecular Ecology*, *33*, e17304. https://doi.org/10.1111/mec.17304