Received: 16 July 2021 | Revised: 7 September 2022 | Accepted: 6 October 2022

DOI: 10.1111/mec.16728

ORIGINAL ARTICLE

Diversification linked to larval host plant in the butterfly *Eumedonia eumedon*

Joan C. Hinojosa[1] | Cecilia Montiel-Pantoja[2] | Miguel Sanjurjo-Franch[3] | Isabel Martínez-Pérez[2] | Kyung Min Lee[4,5] | Marko Mutanen[4] | Roger Vila[1]

[1]Institut de Biologia Evolutiva (CSIC-UPF), Barcelona, Spain

[2]C/ Gaspar G. Laviana, Lugones (Siero), Spain

[3]C/ Gozón, Oviedo, Spain

[4]Ecology and Genetics Research Unit, University of Oulu, Oulu, Finland

[5]Zoology Unit, Finnish Museum of Natural History, University of Helsinki, Helsinki, Finland

Correspondence
Roger Vila, Institut de Biologia Evolutiva (CSIC-UPF), Barcelona, Spain.
Email: roger.vila@csic.es

Funding information
European Regional Development Fund, Grant/Award Number: CGL2016-76322-P; Generalitat de Catalunya, Grant/Award Number: 2017-SGR-991; Ministerio de Asuntos Económicos y Transformación Digital, Gobierno de España, Grant/Award Number: BES-2017-080641; Ministerio de Ciencia e Innovación, Grant/Award Number: PID2019-107078GB-I00; Ministerio de Ciencia, Innovación y Universidades

Handling Editor: Jeremy B Yoder

Abstract

It is widely accepted that the relationship between phytophagous insects and their host plants influences insect diversification. However, studies addressed at documenting host-associated genetic differentiation (HAD) and the mechanisms that drive reproductive isolation in host-associated lineages (or host races) are still scarce relative to insect diversity. To uncover further evidence on the HAD processes in Lepidoptera, we investigated the genetic structure of the geranium argus butterfly (*Eumedonia eumedon*) and tested for isolation by ecology (IBE) vs. isolation by distance (IBD). Genomic data revealed an array of host races (three of them in the same mountain range, the Cantabrian Mountains, northern Iberia) at apparently distinct levels of reproductive isolation. We found a pattern of IBE mediated by HAD at both local and European scales, in which genetic differentiation between populations and individuals correlated significantly with the taxonomic relatedness of the host plants. IBD was significant only when considered at the wider European scale. We hypothesize that, locally, HAD between *Geranium*-feeding populations was caused (at least partially) by allochrony, that is via adaptation of adult flight time to the flowering period of each host plant species. Nevertheless, the potential reproductive isolation between populations using *Erodium* and populations using *Geranium* cannot be explained by allochrony or IBD, and other mechanisms are expected to be at play.

KEYWORDS
butterflies, ddRADseq, host races, phylogeography, speciation

1 | INTRODUCTION

Ecological speciation occurs when reproductive isolation evolves between populations as a result of ecologically based divergent selection (Rundle & Nosil, 2005). The environmental elements that cause divergent selection are diverse and comprise a wide range of abiotic and biotic factors, including interactions with other species (Nosil, 2012). In the case of phytophagous insects, divergent selection can occur as a result of a specialization of the insect to its host plant (Tilmon, 2008), which may lead to a variable degree of reproductive isolation and the emergence of genetically distinct host-associated lineages or host races (Drès & Mallet, 2002) in a process commonly referred as host-associated genetic differentiation (HAD). Considering that 66% of all described animal species are insects

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.
© 2022 The Authors. *Molecular Ecology* published by John Wiley & Sons Ltd.

(Zhang, 2011) and that approximately half of them feed on plants (Schoonhoven et al., 2005), the insect–plant interaction is potentially one of the most common relationships among organisms on Earth and, hence, HAD could have a major role in fuelling global diversity.

The influence that host plants exert over phytophagous insects has been intriguing scientists for centuries (Brues, 1924; Walsh, 1864) but, despite all the scientific efforts and the belief that the presence of host races in insects would be probable (Drès & Mallet, 2002), tested examples of HAD have remained rare relative to insect diversity. To understand the role of HAD in diversification, it is essential to retrieve sufficient genetic information that is variable at the population scale and/or between closely related species, which was methodologically difficult and expensive until recently. This hampered not only the detection of HAD itself, but increased the difficulty in distinguishing the impact of the ecology (i.e., isolation by ecology, IBE) from geography (i.e., isolation by distance, IBD) over the observed genetic structure. Nowadays, genomic techniques allow larger amounts of genetic data to be obtained and hence offer more detailed genetic information even at the intraspecific level. In consequence, in recent years, research on HAD is transitioning from a limited number of models—for example, apple maggot flies (*Rhagoletis pomonella*) on apples and hawthorns (Bush, 1969; Feder et al., 1993; McPheron et al., 1988), goldenrod gall fly (*Eurosta solidaginis*) on goldenrod (Craig et al., 1993; Smith et al., 2002), pea aphids (*Acyrthosiphon pisum*) on Fabaceae (Caillaud & Via, 2000; Peccoud et al., 2009), and stick insects (*Timema cristinae*) on redheart and chamise (Nosil et al., 2002; Nosil & Sandoval, 2008)—to a flourishing and more diversified research field.

Among insects, Lepidoptera are particularly suitable to the study of HAD because of their diversity—they are considered the biggest phytophagous order in the world (Schoonhoven et al., 2005; van Nieukerken et al., 2011), although a minority of aphytophagous lineages exist—and their ecology, characterized by larvae typically specialized on a restricted group of plant species (Forister et al., 2015; Schoonhoven et al., 2005). This specialization can trigger adaptations causing isolation between populations. For example, differences in host plant phenologies can result in a barrier to gene flow by adult flight time modification and allochrony (Austin et al., 2008; Bereczki et al., 2020, 2022). Population differentiation can also be driven by the distinct chemistry of the host plants. It has been hypothesized that selection acts on the ability to identify suitable plants and on digestive and physiological traits related to processing chemicals (Matsubayashi et al., 2010; Tilmon, 2008), reducing the performance of specimens that disperse to habitats where their original host plant is absent. Moreover, since lepidopterans can sequester chemicals from their host plants normally during the larval period, the different chemical composition influences the range of available chemicals. Most of the acquired chemicals are presumed to play important roles as defensive agents (Nishida, 2002), but some are incorporated as cuticular hydrocarbons (Otte et al., 2018)—regarded as important for intraspecific semiochemical communication (Howard & Blomquist, 2005)— others are precursors of hormones used for courtship (Landolt & Phillips, 1997; Reddy & Guerrero, 2004) and others can even alter wing colour (e.g., Burghardt et al., 2001). The strength of reproductive isolation that different phytochemical profiles cause is currently poorly understood but, given the correlation between plant taxonomy and phytochemicals (Defossez et al., 2021; Sharma, 2013; Wink et al., 2010), isolating effects due to phytochemicals may be inversely paired with the taxonomic relatedness of the host plants.

The multiple butterfly traits with significance for reproduction that can be modified by the use of alternative host plants (e.g., pheromones, colour, flight time, habitat choice) highlight the impact of host plants on the diversification of Lepidoptera as one of great significance, potentially at the level of factors such as adaptation to climate (e.g., Pitteloud et al., 2017) or reproductive interference (e.g., Hinojosa et al., 2020). Thus, the periodic publication of evidence of HAD in butterflies (Bereczki et al., 2020; Mikheyev et al., 2013; Nice & Shapiro, 2001) and moths (Joyce et al., 2016; Malausa et al., 2007; Silva-Brandao et al., 2018) is not surprising. However, the available knowledge of HAD in Lepidoptera is still highly incomplete since it is unknown how common this phenomenon is across the lepidopteran tree of life, and cases are usually supported by weak genetic data, based on just one or a small set of markers. Furthermore, the sources of selection responsible for reproductive isolation linked to HAD are rarely clarified.

The butterfly geranium argus, *Eumedonia eumedon* (Esper, 1780), is a lycaenid native to the Eurosiberian region of the Palearctic, from the Iberian Peninsula to the Pacific. Its larval host plants constitute several species of the genus *Geranium* (Tolman & Lewington, 2008; Tshikolovets, 2011) although, exceptionally, some populations have been recorded on *Erodium* in specific areas of the southern edges of its range (Eitschberger & Steiniger, 1975; Fuchs, 1989; Gil-T, 2004; Koçak, 1979; Martínez Pérez et al., 2015; Schurian et al., 2014). This plant genus, also a member of the family Geraniaceae and sister to *Geranium*, is not known to be used among other *Eumedonia* species (Tuzov et al., 2000; Zhdanko, 1997) despite being widely dispersed and common. The relationship of *Eu. eumedon* with its larval host plants is tight: it establishes well-defined populations only where they are abundant, its flight time is synchronized with the blooming of the host plants and even their inflorescences are the main source of nectar for the adults (Erhardt, 1995). This extreme dependence on a specific set of host plants and their habitats led us to hypothesize that this species may be prone to experience HAD.

With the aim of improving our knowledge about how host plant preferences contribute to the diversification of lepidopterans, we conducted a genomic study using *Eu. eumedon* as a model to test for HAD. Our prediction was that, if the larval host plant mediates reproductive isolation to some extent, patterns of genetic differentiation should significantly correlate with the host plant of the butterfly populations, taking into account geographical distances (IBD) as a confounding factor. We employed the double digest RADseq (ddRADseq) technique to obtain genomic data from various populations from Europe and Asia that feed on different host plants, including populations associated with *Geranium* and *Erodium*. We first documented intraspecific variation and the relationships among

populations. Since the pattern obtained was congruent with HAD, we then assessed, at the local and European scale, the roles of IBE mediated by HAD and of IBD in shaping the genetic structure of the species.

2 | MATERIAL AND METHODS

2.1 | Sampling

A total of 48 specimens of the butterfly *Eumedonia eumedon* were sequenced using the ddRADseq protocol detailed in the section "ddRADseq library preparation." For the same specimens, the barcode region of the mitochondrial gene cytochrome *c* oxidase I (*COI*) was retrieved from the Barcode of Life Data System (BOLD). Individuals collected from the field were dried as soon as possible, then wings were kept separately as vouchers and bodies were stored in 99% ethanol at −20°C. Information relative to the specimens used for the genetic analyses is detailed in Table S1 and the sampling sites are plotted in Figure 1.

To determine the host plant in each site, direct observations of the presence of immatures on local host plants and the oviposition habits of the females were carried out, except for the Swedish individuals, for which indirect evidence—using bibliographical evidence (Eliasson et al., 2005; Henriksen & Kreutzer, 1982) and records of the host plants present from GBIF—was employed. The locations of the butterfly populations with known host plants are indicated in Figure 1. We cannot assert the diet of the remaining populations, but no other host plant than *Geranium* spp. has been recorded in the ranges where the specimens were collected. Observations of adults were gathered for the populations from the Cantabrian Mountains, including both new and old records published in Martínez Pérez et al. (2015), covering the period from 1999 to 2019 (Table S2).

2.2 | ddRADseq library preparation

For the ddRADseq library preparation, genomic DNA (gDNA) was extracted from half thorax using the DNeasy Blood & Tissue Kit (Qiagen). The amount of gDNA in the extracts was checked using a PicoGreen kit (Molecular Probes) according to the kit instructions. To increase gDNA quantity and quality, whole genome amplification was performed using a REPLI-g Mini Kit (Qiagen). Concentration of the amplified gDNA was estimated with the PicoGreen kit. For every sample, 500 ng of DNA was digested in a reaction consisting of 1 μl *Pst*I (20,000 units ml^{-1}, New England Biolabs), 2 μl *Mse*I (10,000 units ml^{-1}, New England Biolabs), 5 μl of CutSmart Buffer (New England Biolabs) and ultrapure (HPLC quality) water up to 50 μl. It was then incubated for 2 h at 37°C and enzymes were deactivated by freezing. A purification step with

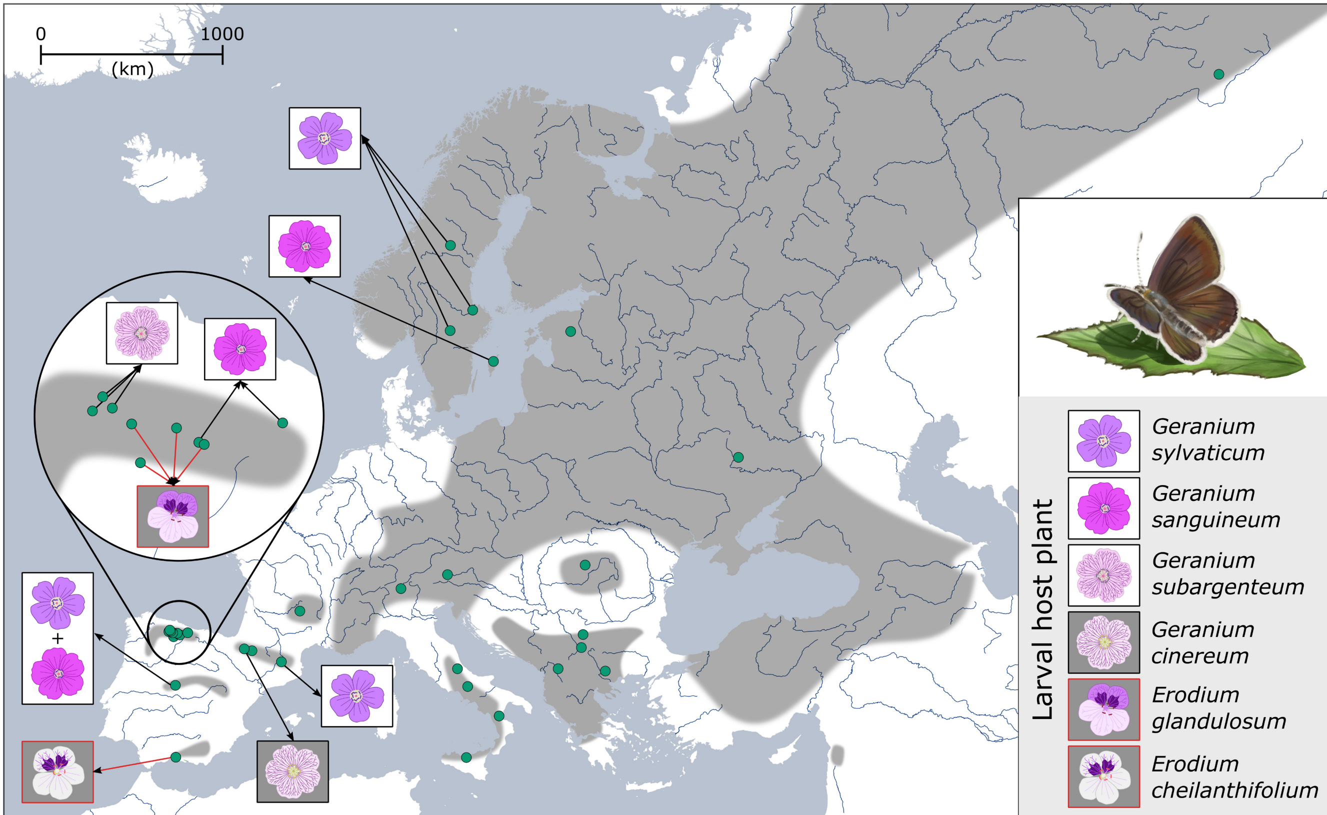


FIGURE 1 Sampling sites of the specimens used for ddRADseq and approximate distribution of *Eumedonia eumedon* (shaded area). For each site, the larval host plant is depicted if known. Butterfly drawing: Nàdia Sentís

AMPure XP x1 magnetic beads (Agencourt) was carried out in a Biomek automated liquid handler (Beckman Coulter) with a final elution in 40 μl. The concentration of purified and digested DNA was measured with PicoGreen; this value was used for the pooling step, after ligation. The listed buffers and enzymes were added in every sample for the ligation step: 5 μl T4 DNA Ligase Buffer (New England Biolabs), 1 μl T4 DNA Ligase (2000,000 units ml^{-1}, New England Biolabs), 0.6 μl rATP (100 mM, Promega), 5 μl P1 adapter (50 nM), 5 μl P2 adapter (50 nM) and 2.4 μl water. Each sample was tagged with unique P1 adapters that differed in a 5-nucleotide barcode sequence; P1 adapters also included a TGCA overhang on the top strand to match the sticky end left by *Pst*I. The P2 adapter included the Illumina sequencing primer sequences that are compatible with paired-end runs, and AT overhangs on the top strand to match the sticky end left by *Mse*I. It also incorporated a "divergent-Y" (Baird et al., 2008) to prevent amplification of fragments with *Mse*I cut sites on both ends. The ligation process was performed for 1 h at 22°C and enzymes were deactivated at 65°C for 20 min. Then, 200 ng from each individual was 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 and eluted in 40 μl. The pools were size-selected at 300 bp with BluePippin (Sage Science) using the cassette type "2% DF Marker V1" and the "tight" option. Finally, PCR (polymerase chain reaction) amplification was performed with primers RAD1.F (5′-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACG-3′) and RAD2.R (5′-CAAGCAGAAGACGGCATACGAGATCGTGATGTGACTGGAGTTCAGACGTGTGC-3′). The ligated 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 denature at 98°C for 30 s, then 98°C for 10 s, 60°C for 30 s and 72°C for 40 s in 16 cycles with the final extension step at 72°C for 5 min. PCR products were purified with AMPure XP magnetic beads and DNA concentration was measured with PicoGreen. The size distribution and concentration of the pools were measured with a Bioanalyzer (Agilent Technologies). Libraries were finally pooled into a single tube in equimolar amounts and sequenced on an Illumina HiSeq 2500 PE 100 at the FIMM Technology Center (Helsinki, Finland). The demultiplexed FASTQ files were archived in the NCBI SRA: SRX10512612–SRX10512659.

2.3 | ddRADseq data set processing

Initial filtering steps, single nucleotide polymorphism (SNP) calling and alignment were carried out using the IPYRAD version 0.6.15 (Eaton & Overcast, 2020) pipeline. After testing distinct parameter combinations, the following parameters were changed from the default settings: datatype was set to *pairddradseq*, restriction overhang to *TGCAG*, *TAA*, maximum low-quality bases to 3, minimum depth (statistical) to 8, clustering threshold to 0.9, minimum trimmed length to 70, maximum Ns to 2, maximum heterozygous bases to 5, minimum number of samples with a given locus to 6, maximum SNPs per locus to 14 and maximum indels per locus to 5.

The software CENTRIFUGE version 1.0.4 (Kim et al., 2016) was used to identify potential contaminant loci, those identified as noninsect. This step was also useful to extract loci identified as *Wolbachia*, a methodology that has been proven efficient to detect individuals infected with this endosymbiont in studies using the same restriction enzymes (Hinojosa et al., 2022). A contamination-free data set was created by keeping only the insect and unidentified loci. Derived from it, five more data sets were built: (i) a data set with a minor allele frequency filtering step, in which alleles with a frequency lower than 5% (*−-maf 0.05*) were excluded with VCFTOOLS version 0.1.13 (Danecek et al., 2011); (ii) a data set with unlinked SNPs obtained by filtering the VCF file using the option *--thin* in VCFTOOLS—this option prevents having two variable sites within the specified distance from one another and, given that the mean locus length was 178 bp, we selected a value of 180; (iii) a data set with very low missing data, obtained using VCFTOOLS and by retaining only the SNPs present in the 95% of the individuals (*−-max-missing 0.95*); (iv) a data set combining unlinked SNPs + low missing data (*−-max-missing 0.50* in VCFTOOLS); and (v) a data set with unlinked SNPs and without missing data (*−-max-missing 1* in VCFTOOLS).

2.4 | Phylogenetic reconstruction (*COI* and ddRAD loci)

COI sequences were aligned in GENEIOUS version 11.0.5 (Kearse et al., 2012). The best fitting model, HKY + G, was found using JMODELTEST version 2.1.7 (Darriba et al., 2012) under the Bayesian information criterion and the phylogeny was reconstructed in BEAST version 2.5.0 (Bouckaert et al., 2014). Base frequencies were estimated, four gamma rate categories were selected and a randomly generated initial tree was used. Estimates of node ages were obtained by applying a strict clock and a normal prior distribution centred on the mean between two substitution rates for invertebrates: 1.5% and 2.3% uncorrected pairwise distance per million years—Quek et al. (2004) and Brower (1994), respectively. The standard deviation was tuned so that the 95% confidence interval of the posterior density coincided with the 1.5% and 2.3% rates. Albeit these substitution rates provide very rough divergence estimates, better calibrations are, as far as we know, unavailable for these taxa due to the absence of fossils or other phylogenetically close calibration points. Parameters were estimated using two independent runs of 20 million generations each, and convergence was checked using TRACER version 1.7.1 (Rambaut et al., 2018). A 10% burn-in was applied and the results from both runs were merged. Genetic distances (d_{XY}) between *COI* sequences were calculated with pairwise deletion using GENEIOUS.

An alignment was built by concatenating all the ddRAD loci except those identified as potential contaminants (i.e., identified as noninsect) by CENTRIFUGE. This alignment was used to construct a phylogeny through maximum-likelihood inference using RAXML version 8.2.4 (Stamatakis, 2014). The GTRGAMMA model

and 1000 bootstrap replicates were selected. The resulting phylogeny was visualized and exported using FIGTREE version 1.4.2 (Rambaut, 2015).

2.5 | Genetic structuring and indicators of genetic diversity

The genetic structure of the species was studied with STRUCTURE version 2.3.4 (Pritchard et al., 2000) using three SNP data sets: the rare-allele filtered, which provided an improvement of the data quality and helped to reduce the computation compared to the contamination-free data set, the unlinked data set, and the unlinked + low missing data set. An admixture model with correlated allele frequencies was chosen. The selected burn-in was 75,000, followed by 250,000 MCMC (Markov chain Monte Carlo) replicates. K (number of populations) values from 1 to 6 were tested, and 10 runs were done for each K, which were afterwards combined in one per group with CLUMPAK version 1.1 (Kopelman et al., 2015). The best K under the Evanno method (Evanno et al., 2005) was calculated using CLUMPAK version 1.1. A plot was constructed with DISTRUCT version 1.1 (Rosenberg, 2004). A principal components analysis (PCA) was performed using the R package *adegenet* version 1.4–1 (Jombart et al., 2010); only data sets with low missing data were employed since this analysis has been suggested to be sensitive to missing data in population genetics studies (Yi & Latch, 2022).

In order to assess differentiation between groups defined by larval host plant and geography, genetic distances (d_{XY}, d_A) and the fixation index (F_{ST}) were calculated. These indexes are affected differently by missing data and sample size and are thus complementary. In particular, F_{ST} and d_A are more informative because they correct genetic distances by intrapopulation variability, but are more affected by low data compared to the less informative d_{XY}. Genetic distances between groups were measured in GENEIOUS using the contamination-free RAD loci alignment and plotted using the R package *complexheatmap* version 1.10.2 (Gu et al., 2016). The groups were built based on the lineages retrieved in the ddRAD phylogeny, which matched with geography, the STRUCTURE clustering, the PCA grouping and, in the Cantabrian Mountains, with larval host plants. All samples were used except the one from central Iberia, since it could not have been placed in any group. F_{ST} was estimated in VCFTOOLS version 0.1.15 using the weighted Weir and Cockerham's estimator. As F_{ST} can be biased by high percentages of missing data we used only SNPs with low missingness, with thresholds of 0% for the analyses focused on the Cantabrian Mountains and of 5% for the rest.

2.6 | Multiple regression of dissimilarity matrix analyses

The weight of IBE and IBD was assessed by multiple regression on distance matrices (MRM) analyses implemented in the *MRM* function of the *ecodist* version 2.0.7 R package (Goslee & Urban, 2007). To do so, two approaches were used to estimate ecological distances: (i) as a binary variable, namely sharing or not the same host plant genus, and (ii) as pairwise genetic distances between larval host plants. Given that distinct *Geranium* and *Erodium* species have chemical differences related to taxonomy (Harborne & Williams, 2002; Munekata et al., 2019), genetic distances between plants can be a good proxy for ecological differences. We downloaded from GenBank all the available sequences of the ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit (*rbcL*)—one of the most popular genes in plant phylogenetics (Vijayan & Tsou, 2010)—of the studied host plants (Table S3); *rbcL* sequences of all host plants were found except for *Er. cheilanthifolium*. A consensus sequence was obtained per plant species and their genetic distances (d_{XY}) were calculated using GENEIOUS. Geographical distances were obtained using the function *shortestPath* of the R package *gdistance* version 1.3–6 (van Etten, 2017).

MRM analyses were conducted at two geographical scales: first focusing on the Cantabrian Mountains and then using samples from all of Europe. At the local scale (Cantabrian Mountains), we defined one population per locality except for the localities associated with *Geranium subargenteum*, which were merged in a single population due to their proximity. F_{ST} among populations was calculated as explained in the section "Genetic structuring and indicators of genetic diversity." At the European scale, we included samples from all the distribution in which host plants were known and sequenced—the individual from central Iberia was excluded because its population feeds indistinctly in two host plant species. In this case, only one sample has been sequenced for some distant localities with tracked host plant, and hence grouping was not possible and genetic distances (d_{XY}) were calculated between individuals. Each MRM analysis was complemented with two linear regressions in which we explored the linear relationship between butterfly genetic distances and the ecological variable, and between butterfly genetic distances and geographical distances.

3 | RESULTS

3.1 | Adult flight period and host plant records

The ecology of populations in the Cantabrian Mountains was studied in greater detail (Table S2). In this region, *Eumedonia eumedon* used *Erodium glandulosum* in 16 sites, where adults were recorded from May to August and observations peaked in June–July. In 21 sites they used only *Geranium subargenteum* as host plant and adults were flying from June to August, with a peak in July. Three populations used only *G. sanguineum* and adults were present from May to July and peaked in May–June. It has to be considered that the methodology did not systematically take into account variation in flight time through the years, but because most sites were visited simultaneously and during a number of years, differences among sites seem to be meaningful.

We observed that butterfly populations linked to distinct host plants inhabit slightly different ecological niches in the Cantabrian Mountains. Populations that use *G. sanguineum* occur in shadowy meadows and torrents at relatively low altitude (950–1350 m) and with some humidity, located in the central and eastern part of the mountain range; populations on *G. subargenteum* live in humid high-altitude (1500–2000 m) grasslands and are exclusive from the western part; populations on *Er. glandulosum* are found in the central part and prefer rocky sunny habitats at mid- and high altitude (1150–2050 m). Occasionally, potential host plants were found in sympatry (*Er. glandulosum* with *G. sylvaticum*, *Er. glandulosum* with *G. subargenteum*, or *G. subargenteum* with *G. sylvaticum*). In these sites, even if the two host plants were equally abundant, only one plant species was chosen for breeding. This behaviour has also been reported by Lafranchis et al. (2015) for French populations, although studies documenting its recurrence are lacking. By contrast, both *G. sanguineum* and *G. sylvaticum* were found in sympatry and used indistinctly in the studied locality of central Iberia.

3.2 | ddRADseq data sets

The IPYRAD pipeline retrieved 21,202 loci and 127,764 SNPs. None of these loci were identified as *Wolbachia* by CENTRIFUGE. After removing potential contaminants, we obtained a data set of 19,610 loci with 106,561 SNPs, used for the maximum-likelihood phylogeny and genetic distance (d_{XY} and d_A) calculations. From this data set, five additional data sets were obtained and the number of SNPs retrieved were: (i) 59,807 SNPs from the data set obtained after the minor allele frequency filtering step, which was used for the STRUCTURE analysis; (ii) 16,561 SNPs from the data set with only unlinked SNPs, also used for STRUCTURE; (iii) 2427 SNPs from the data set with very low missing data, used for the PCA and F_{ST} calculations; (iv) 1898 SNPs from the data set combining unlinked SNPs + a filtering by missing data, used for both STRUCTURE and the PCA; and (v) 249 SNPs from the data set with unlinked SNPs and without missing data, used for the F_{ST} (for analyses only including Cantabrian samples). All these data sets were deposited in figshare (DOI: 10.6084/m9.figshare.14371571).

3.3 | Phylogenetic reconstruction

The Bayesian chronogram based on *COI* sequences (Figure S1) recovered a deeply diverged clade (posterior probability, PP = 1) that appeared ~2.17 (1.32–3.13) million years ago (Ma) and was sister to the rest of Eurasian samples of *Eu. eumedon*. This lineage corresponded to the specimens of populations that feed exclusively on *Er. glandulosum*. The minimum genetic distance of the group with respect to the other individuals was 2.58%. The other Iberian individuals (except those from the Pyrenees) were included in another clade (PP = 1), together with three Italian specimens. The remaining specimens grouped in two more clades: one containing specimens from easternmost Europe and Kazakhstan (PP = 0.77), and another with the rest of the European specimens, including those from the Pyrenees (PP = 0.97).

The maximum-likelihood phylogeny based on ddRADseq loci (Figure 2) recovered two main lineages, one Eurasian and one Iberian. The Eurasian lineage (bootstrap = 100) comprised individuals from the Pyrenees in the west to Kazakhstan in the east. The Iberian lineage (bootstrap = 100) was formed by individuals from the Iberian Peninsula, excluding the Pyrenees. Within this group, we recovered two well-defined clades (bootstrap = 100). One group was formed by the populations from Iberia feeding exclusively on *Erodium*. It included a subgroup with the populations feeding on *Er. glandulosum* (bootstrap = 100) from the Cantabrian Mountains (northwestern Iberia)—the same highly diverged in the *COI* phylogeny (Figure S1)—and another group with the two individuals from the southern Iberia (bootstrap = 100), which breed on *Er. cheilanthifolium*. The second Iberian clade was formed by populations from the Cantabrian Mountains and Central System (central Iberia), all using *Geranium*. Samples from the Cantabrian Mountains were separated into two additional lineages that perfectly correlated with their host plants: either *G. sanguineum* (bootstrap = 95) or *G. subargenteum* (bootstrap = 92).

3.4 | Genetic structuring and genetic distances

The Evanno method highlighted $K = 2$ as best K in the three STRUCTURE analyses; however, it has to be considered that this result is can biased by the "$K = 2$ conundrum" (Janes et al., 2017). At this K, the two clusters are represented by individuals from the Iberian Peninsula (except the Pyrenees and southern Iberia) and by individuals from the rest of Europe and Asia (Figure 3; Figure S2). At $K = 3$ a west European group and an east European + Asia group were differentiated. At $K = 4$, the four clusters corresponded to: (i) an eastern European group, with individuals from eastern Europe + northern Europe + Asia; (ii) a western European group, with individuals from western Europe, except Iberia but including the Pyrenees; (iii) an Iberian group with individuals from this peninsula feeding on *Geranium* (except the Pyrenees); and (iv) another Iberian group with individuals from this peninsula feeding on *Erodium*. At $K = 5$ a southern Iberia group, which previously appeared as a mix of clusters at lower K values, was differentiated. This cluster was also present to some extent in the east European and Asian specimens. At $K = 6$ a cluster present mostly in the Pyrenees and France appeared. Contact zones with gene flow possibly occurred between groups, since individuals with mixed clusters were visible when spatially plotted. The results obtained with the data sets using only unlinked SNPs (Figure S3) and unlinked SNPs + low missing data (Figure S4) were similar, although the separation between the two main Iberian groups appeared less frequently.

The PCA (Figure 4; Figure S5) displayed the *Erodium*-feeding populations from southern Iberia as a very divergent group. In STRUCTURE, these samples displayed as a mix of clusters at several runs, which

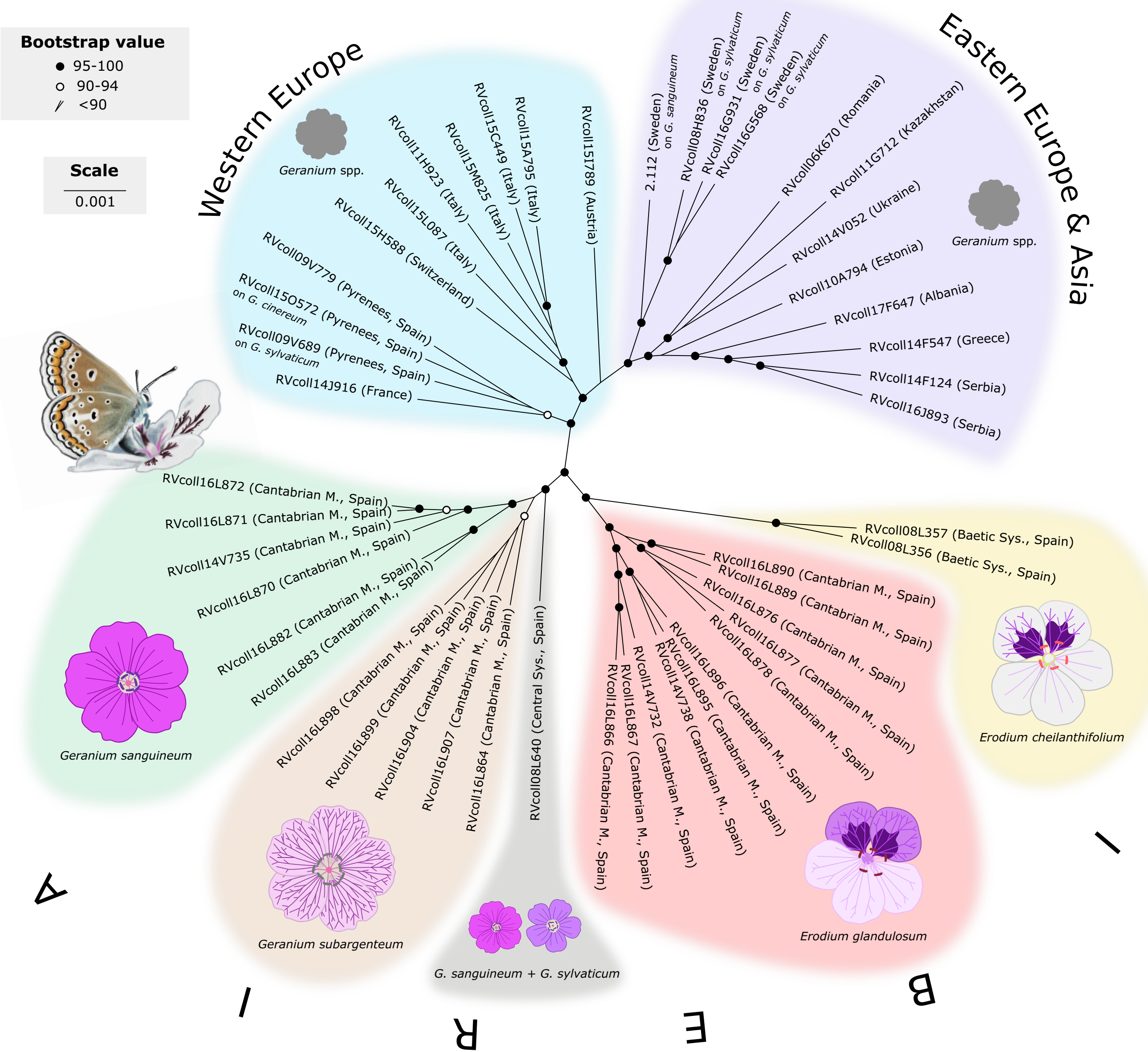


FIGURE 2 Maximum-likelihood inference tree (based on 19,610 loci). Scale units are presented in substitutions per site. Larval host plants are indicated. Butterfly drawing: Nàdia Sentís

is probably explained by the problems of this software when dealing with well-diverged and poorly sampled groups (e.g., Dupuis & Sperling, 2020; Quattrini et al., 2019). The *Geranium*-feeding populations were divided, in PC1, into two groups corresponding to Iberia and Eurasia. Minor differences in the specimens from the Cantabrian Mountains that correlated with the *Geranium* species used as host plant (either *G. sanguineum* or *G. subargenteum*) were displayed in PC2 and PC3. The specimens from the Cantabrian Mountains with *Er. glandulosum* as host plant formed a compact group, distant from the rest of the samples.

Regarding the Cantabrian Mountains (Figure 5), differences between populations feeding on *Er. glandulosum* and populations using *Geranium*, namely *G. sanguineum* (F_{ST} = 0.21; d_A = 0.11%; d_{XY} = 0.29%) and *G. subargenteum* (F_{ST} = 0.16; d_A = 0.09%; d_{XY} = 0.26%), were higher than those between the populations using *G. sanguineum* and *G. subargenteum* (F_{ST} = 0.09; d_A = 0.04%; d_{XY} = 0.21%). The highest d_{XY}, d_A and F_{ST} values were obtained between the population of southern Iberia (on *Er. cheilanthifolium*) and the rest (Figure S6).

3.5 | Multiple regression on distance matrices analyses

The analyses in which host plant genetic distances were set as ecological distances showed that the MRM explained most of the variance when using only Cantabrian populations (R^2 = .821; Figure 6a) and when including individuals from all of Europe (R^2 = .825; Figure 6b). In the Cantabrian Mountains, the butterfly

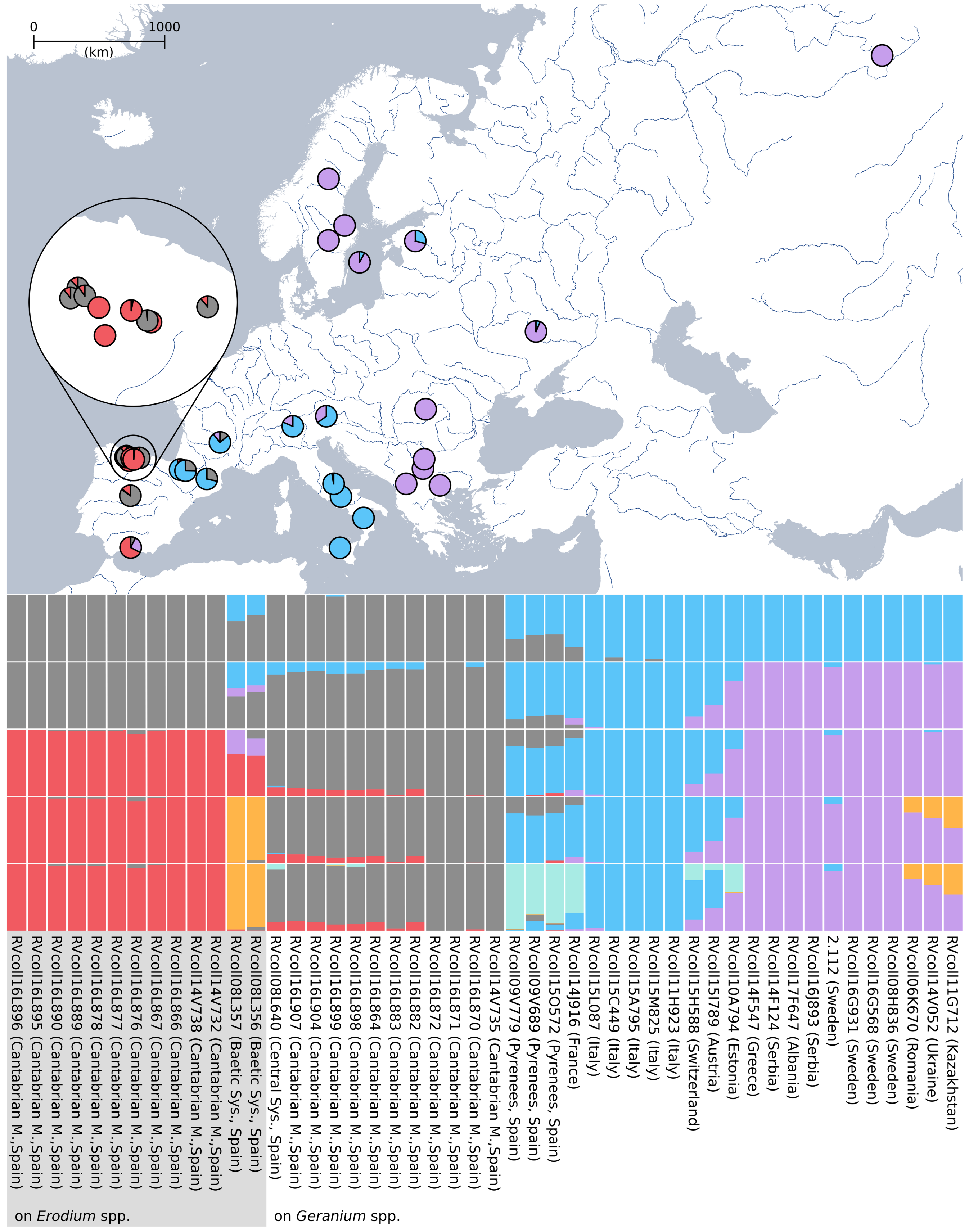


FIGURE 3 STRUCTURE results (based on 59,807 SNPs) represented as bars ($K = 2$–6) and as pie charts ($K = 4$). The area zoomed corresponds to the Cantabrian Mountains

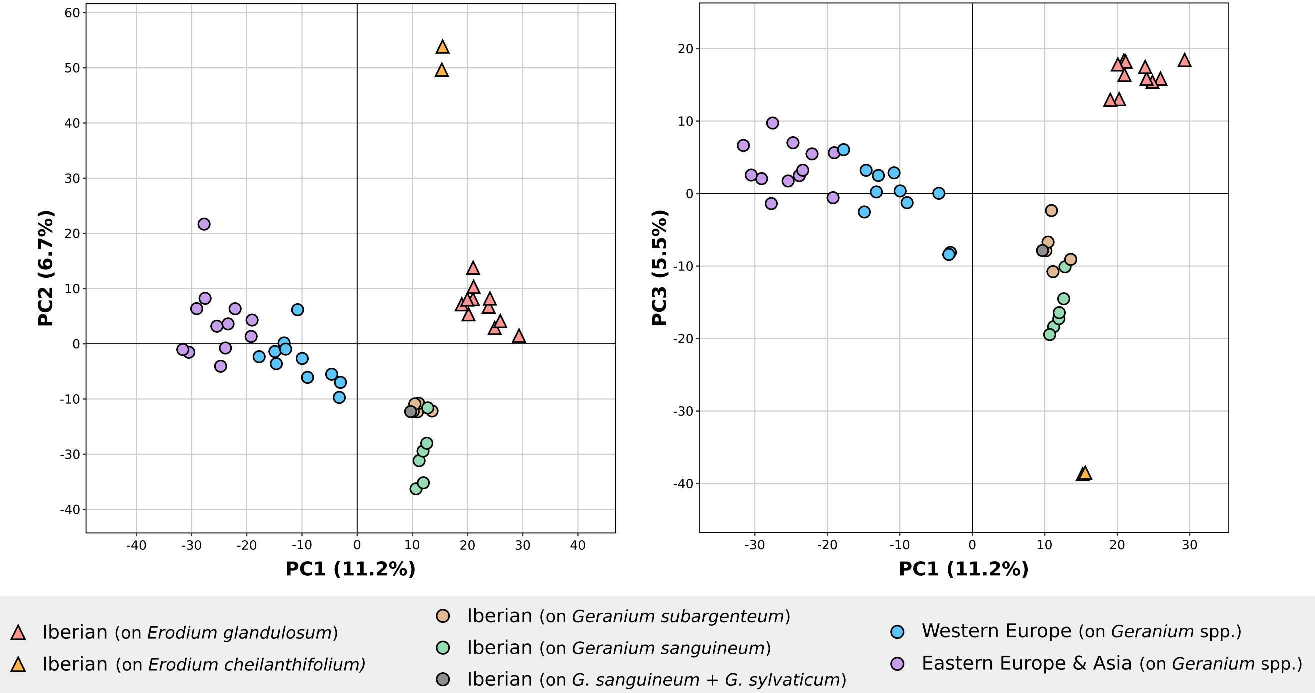


FIGURE 4 Principal component analysis (PCA) results (based on 2427 SNPs)

genetic distances were highly correlated with the host plant genetic distances (regression coefficient = 7.664, p =.005) while they did not correlate with geographical distances (regression coefficient = −0.001, p =.107). Instead, in the analysis including samples from all of Europe (Figure 6b), both host plant genetic distances (regression coefficient = 0.044, p =.001) and geographical distances (regression coefficient = 4.59×10^{-5}, p =.001) correlated significantly with the butterfly genetic distances.

Regarding the analyses using a binary variable (sharing or not host plant genus) as ecological variable, the MRM explained most of the variance (R^2 =.769) at local scale (Figure S7a). The butterfly genetic distances were again significantly correlated with the ecological variable (regression coefficient = 0.148, p =.031), but not with geographical distances (regression coefficient = -1×10^{-4}, p =.774). When samples from all of Europe were included (Figure S7b), the MRM also explained most of the variance (R^2 =.792) and both sharing host plant genus (regression coefficient = 7.12×10^{-4}, p =.001) and geographical distances (regression coefficient = 4.96×10^{-7}, p =.001) correlated significantly with the butterfly genetic distances.

4 | DISCUSSION

4.1 | The influence of geography and host plants on genetic divergence

The overall genetic pattern retrieved by STRUCTURE (Figure 3; Figures S2–S4) fits well with the theory of the three southern European refugia (Habel et al., 2010; Hewitt, 1999). Based on the location of the "pure" individuals, we suggest the current main lineages originated by the action of glacial periods that isolated populations in the Balkan Peninsula (forming the violet cluster at K = 4), Italian Peninsula (blue) and the Iberian Peninsula (grey and red). After the Last Glacial Maximum, Balkan populations seemingly expanded northwards, reaching central Asia and northern Europe, and to central Europe, where they admixed with populations apparently coming from the Italian Peninsula. Italian populations probably dispersed across Europe and colonized the Pyrenees, albeit their contribution fades as we move away from their origin.

Despite the geographical influence pinpointed by STRUCTURE, the complexity of their genetic diversity south to the Pyrenees requires further explanation. In the Iberian Peninsula, the presence of HAD was suggested by additional clusters that separated according to the genus of larval host plant. Further evidence of HAD is visible in the PCA (Figure 4; Figure S5) and especially in the ddRADseq phylogeny (Figure 2), which displayed four lineages tightly linked to the larval host plants, namely *Erodium glandulosum*, *Er. cheilanthifolium*, *Geranium sanguineum* or *G. subargenteum*. Three of them were found in a single mountain range, the Cantabrian Mountains, northwestern Iberia. The effect of IBE was significant according to the MRM analyses (Figure 6; Figure S7) and showed that genetic distances between butterfly populations were significantly correlated with the ecological distances. Butterfly genetic differentiation was explained solely by IBE locally (in the Cantabrian Mountains) but, at European scale, the influence of both ecological and geographical distances was significant. Regarding the chosen ecological variables, the models using host plant genetic distances displayed high R^2 and explained more of the observed variation than models using the binary variable (sharing or not sharing host plant genus). These results may reflect a link between the taxonomy of the host plants and the genetics of

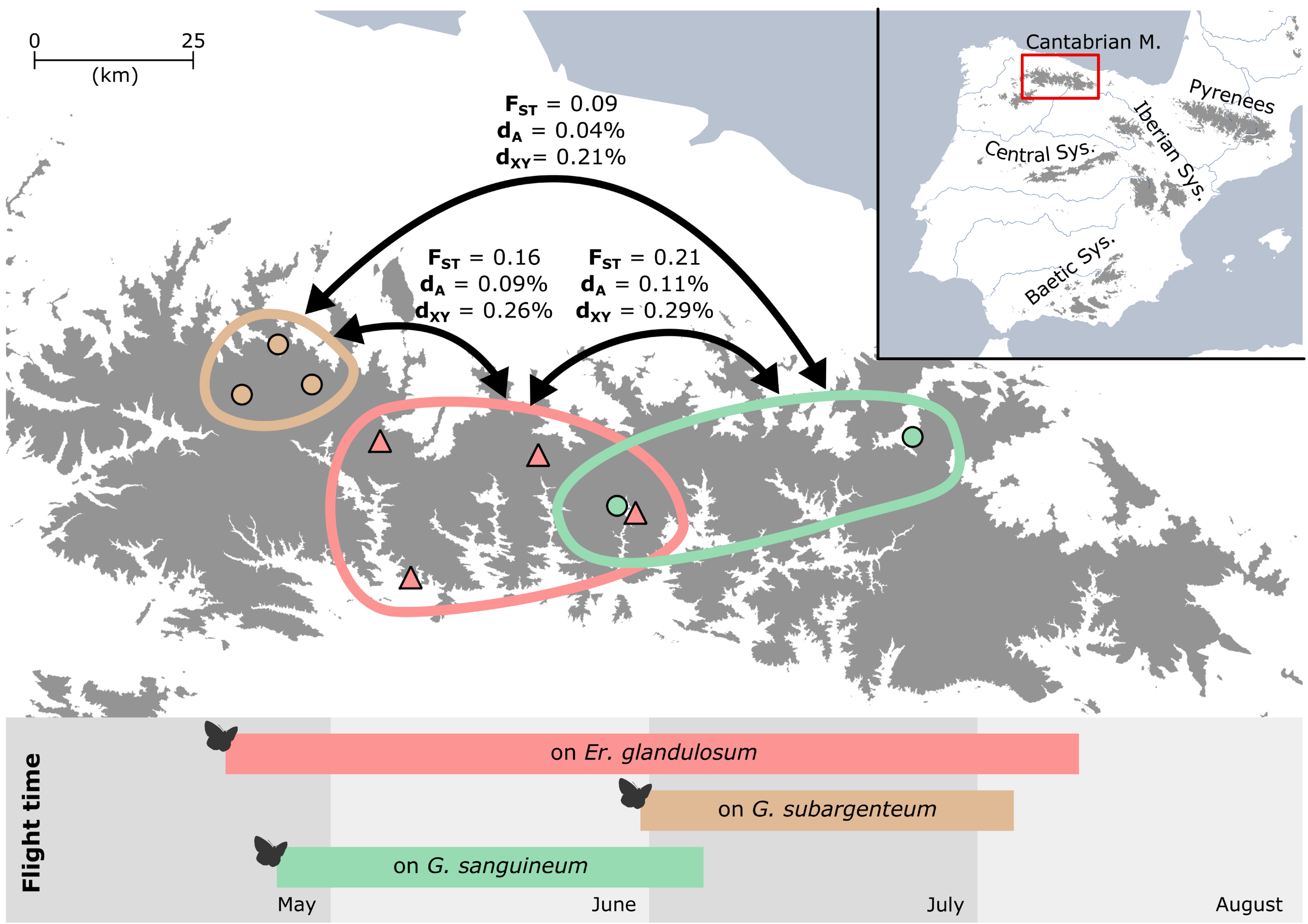


FIGURE 5 Sampling sites of *Eumedonia eumedon* in the Cantabrian Mountains coloured according to larval host plant: *Erodium glandulosum* (red triangles), *Geranium subargenteum* (orange circles) and *Geranium sanguineum* (green circles). Three indicators of population differentiation—F_{ST} (based on 249 SNPs), d_A and d_{XY} (based on 19,610 loci)—Are given; populations were grouped by larval host plant. Flight time period, depicted below, was obtained from the records collected in Table S2

the butterfly, in which a decrease in the taxonomic relatedness of the host plants is paired with an increase in the genetic differences between butterfly populations, and support the hypothesis stating that a more distinct chemical profile among host plants could cause stronger reproductive isolation.

4.2 | Drivers of genetic differentiation in the Cantabrian populations

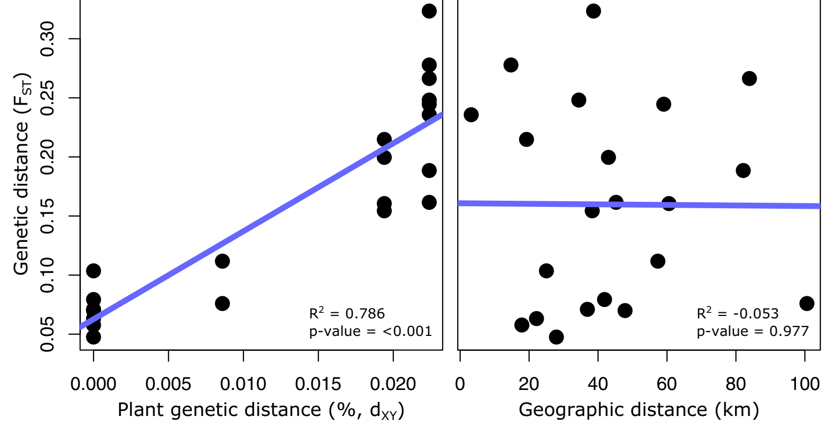
Once IBE is detected, it is equally important to identify the underlying mechanisms that generated it (Wang & Bradburd, 2014), in this case to explore how the host plants could have influenced the origin and/or maintenance of the pattern of HAD. In the Cantabrian Mountains, the situation was paradigmatic, as we retrieved three lineages associated with distinct larval host plants (*Er. glandulosum*, *G. sanguineum* or *G. subargenteum*) and only IBE explained the observed genetic differentiation. One of the processes that has been traditionally thought to be tied to speciation in phytophagous insects is coevolution (Ehrlich & Raven, 1964; Page, 2003; Thompson, 1989) with host plants. The coevolution hypothesis in terms of cospeciation and cocladogenesis is not consistent with the divergence between the butterfly lineages associated with *Geranium* and *Erodium* since both plant genera split 18–34 Ma (Fiz et al., 2008; Fiz-Palacios et al., 2010), far earlier than any division within *Eumedonia*—estimated to be 4.39 Ma at maximum (Figure S1). Cospeciation is also unlikely to explain the origin of the lineages linked to distinct *Geranium* species, as the groups in which *G. sanguineum* (subgenus *Geranium*) and *G. subargenteum* (subgenus *Erodioidea*) are included split ~9 Ma (Marcussen & Meseguer, 2017). Nevertheless, our results reinforce the view that plants generally provide a framework onto which butterflies diversify (Hardy & Otto, 2014; Janz & Nylin, 2008) and, in this sense, coevolution may take place through reciprocal selective pressures.

We observed that butterfly populations using distinct host plants inhabited slightly different habitats in the Cantabrian Mountains, and hence they are not strictly sympatric. On the other hand, it is also important to consider the flowering periods of the host plants since *Eumedonia eumedon* flight time is synchronized with them. This is because adults oviposit in the flowers and the small larva feeds on the

FIGURE 6 Results of the multiple regression on distance matrices (MRM) for the space/ecology model and plots of the linear regression between the butterfly genetic distances vs. plant genetic distances (left) and between the butterfly genetic distances vs. geographical distances (right). In (a) we used samples from the Cantabrian Mountains and F_{ST} (based on 249 SNPs) between populations (grouped by locality) as butterfly genetic distances, while in (b) we used all the European samples with available host plant information and d_{XY} (based on 19,610 loci) between individuals as butterfly genetic distances

(a)

Space/ecology model (Cantabrian Mountains)			
Variable	R. coefficient	p-value	R^2
Plant genetic distance	7.664	0.005	0.821
Geographic distance	-0.001	0.107	



(b)

Space/ecology model (Europe)			
Variable	R. coefficient	p-value	R^2
Plant genetic distance	0.044	0.001	0.825
Geographic distance	4.59×10^{-5}	0.001	

fruit (Tolman & Lewington, 2008), a recurrent strategy in Lycaenidae—for example, *Iolana debilitata* (Rabasa et al., 2005), *Euphilotes ancilla* (Austin et al., 2008). In the Cantabrian Mountains, *G. sanguineum* typically reaches its peak bloom earlier than *G. subargenteum* and indeed *Eu. eumedon* flies earlier in populations associated with the former (Figure 5). Similar to gall-forming wasps over distinct oak species (Zhang et al., 2019) and to the butterfly *Phengaris arion* (Bereczki et al., 2020; Bereczki et al., 2022), differences in breeding time between conspecific individuals (also known as allochrony) could act as a reproductive barrier between the populations found on *G. sanguineum* and *G. subargenteum*. The habitat preferences of the host plants probably strengthened isolation through habitat selection. The slight genetic differentiation between these two *Geranium*-feeding lineages suggests a recent origin, and similar undetected processes of lineage sorting may be taking place locally in other places.

Genetic differentiation between populations linked to *Geranium* vs. *Er. glandulosum* is much higher compared to those feeding on distinct species of *Geranium* (Figure 5; Figure S6), despite being geographically closer (the closest populations found are separated by ~3 km). Thus, they seem to have a stronger barrier to gene flow. Genomic differences are clear in F_{ST} calculations, in which individuals using *Er. glandulosum* vs. individuals using *Geranium* exhibited values (0.16–0.21) that can be considered high for butterflies, although they are still compatible with intraspecific variation (e.g., Arif et al., 2021; Halbritter et al., 2019). The putative reproductive isolation is reflected in the low admixture between these clusters in the STRUCTURE results and in the fact that all the specimens associated with *Er. glandulosum* shared a highly diverged mitochondrial lineage (Figure S1), not present in any other sample. The minimum uncorrected genetic distance of this lineage with respect to the rest of the

specimens analysed was 2.58% for the mitochondrial marker *COI*, which is a high value for butterflies at the intraspecific level (Dincă et al., 2015; Dincă et al., 2021; Huemer et al., 2014); the observed mitochondrial lineage pattern could theoretically be an outcome of *Wolbachia* infection—*Wolbachia* bacteria are maternally inherited and may cause male-killing or cytoplasmic incompatibility (Werren et al., 2008), which can lead to some degree of reproductive isolation (Bordenstein et al., 2001; Shoemaker et al., 1999)—but here all specimens analysed were apparently not infected. Considering that populations found over *Er. glandulosum* and their counterparts that use *Geranium* live nearby and both plants overlap in the blooming period, it is likely that meetings between individuals of these populations occur. In this case, unknown reproductive barriers seem to operate, either preventing dispersal, interbreeding and/or through a reduced fitness of hybrids. Wing patterns and coloration (Fordyce et al., 2002; Knüttel & Fiedler, 2001) and also chemicals (Hernández-Roldán et al., 2014; Oroño et al., 2013; Pinzari et al., 2018) are potential targets for further study, as they are important characters for mate choice that can be influenced by larval diet.

Importantly, it remains unclear whether the lineage sorting between populations feeding on *Erodium* and populations feeding on *Geranium* occurred before or after (as a consequence of) the host plant shift. Even if sorting was preshifting, host plant use could still currently contribute to reproductive isolation, but other possibilities cannot be discarded. To shed new light on this issue, it becomes crucial to study more individuals of the populations feeding on *Er. cheilanthifolium* and *Er. valentinum* from southern Iberia, as well as from other populations that feed on *Erodium* species located in Greece (Fuchs, 1989), Turkey (Koçak, 1979; Schurian et al., 2014) and Lebanon (Eitschberger & Steiniger, 1975). Furthermore, determining if they are related between them or if they have shifted to *Erodium* independently would be also key to reconstructing the evolutionary history of *Eu. eumedon* and reach taxonomic conclusions.

4.3 | Conclusions: Hypotheses over the speciation continuum of *Eu. umedon*

The speciation continuum is defined as the continuous sequence of variation that occurs as lineages diverge from one another on the pathway to full reproductive isolation (Hendry, 2009; Hendry et al., 2009). The Iberian *Eu. eumedon* represents an excellent example of it, since we describe a series of genetically differentiated populations (here also host races) apparently at distinct levels of reproductive isolation. Furthermore, we show that the speciation continuum of *Eu. eumedon* is conditioned by the larval host plants as the genetic divergence between host races correlates with the genetic distance between their host plants: the less related are the host plants taxonomically, the more genetically differentiated are the butterfly populations. As far as we know, this pattern has never been described before. The exact mechanisms that generated this pattern remain largely unknown, although allochrony seems to have contributed to isolate the populations that feed on the most closely related host plants (*G. sanguineum* and *G. subargenteum*). We hypothesize that the more distantly related are the host plants, the adaptations committed to cope with host plant particularities (e.g., its phytochemical profile) could be more divergent and have a greater impact on reproductive isolation and genetic differentiation of the butterfly.

AUTHOR CONTRIBUTIONS

Joan C. Hinojosa, Marko Mutanen and Roger Vila conceived, designed and coordinated the study. Funding was secured by Marko Mutanen and Roger Vila. Laboratory protocols were conducted by Joan C. Hinojosa and Kyung Min Lee. Genetic data were analysed by Joan C. Hinojosa Life history records were collected by Cecilia Montiel-Pantoja, Miguel Sanjurjo-Franch and Isabel Martínez-Pérez. The manuscript was initially written by Joan C. Hinojosa and Roger Vila, along with significant contributions from the remaining authors.

ACKNOWLEDGEMENTS

We acknowledge S. Cuvelier, V. Dincă, R. Vodă, L. Dapporto, F. Gil-T., I. Tot, G. Talavera, Y. Monasterio, N. Ryrholm and E. Karolinskiy for providing samples used in this study. We thank the editor Jeremy Yoder and four anonymous reviewers for providing constructive comments that improved the manuscript. Financial support for this research was provided by projects CGL2016-76322-P funded by AEI/ERDF and EU, PID2019-107078GB-I00 funded by MCIN/AEI/10.13039/501100011033, and 2017-SGR-991 funded by Generalitat de Catalunya to Roger Vila, and by grant BES-2017-080641, funded by MCIN/AEI/10.13039/501100011033 and by "ESF Investing in your future" to Joan C. Hinojosa.

CONFLICT OF INTEREST

The authors have no conflicts of interest to declare.

DATA AVAILABILITY STATEMENT

Raw ddRADseq FASTQ reads: NCBI Bioproject PRJNA719785. Alignments of ddRADseq loci and SNPs: figshare (DOI: 10.6084/m9.figshare.14371571).

ORCID

Joan C. Hinojosa https://orcid.org/0000-0002-6318-4252
Roger Vila https://orcid.org/0000-0002-2447-4388

REFERENCES

Arif, S., Gerth, M., Hone-Millard, W. G., Nunes, M. D. S., Dapporto, L., & Shreeve, T. G. (2021). The dirty north: Evidence for multiple colonisations and *Wolbachia* infections shaping the genetic structure of the widespread butterfly *Polyommatus icarus* in the British Isles. *Molecular Ecology*, *30*(20), 5196–5213. https://doi.org/10.1111/mec.16126

Austin, G. T., Boyd, B. M., & Murphy, D. D. (2008). *Euphilotes ancilla* (Lycaenidae) in the Spring Mountains, Nevada: More than one species. *Journal of the Lepidopterists' Society*, *62*(3), 148–160.

Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., Selker, E. U., Cresko, W. A., & Johnson, E. A. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD

markers. *PLoS One*, *3*(10), e3376. https://doi.org/10.1371/journal.pone.0003376

Bereczki, J., Poliska, S., Váradi, A., & Tóth, J. P. (2020). Incipient sympatric speciation via host race formation in *Phengaris arion* (Lepidoptera: Lycaenidae). *Organisms Diversity & Evolution*, *20*(1), 63–76. https://doi.org/10.1007/s13127-019-00418-y

Bereczki, J., Sielezniew, M., Verovnik, R., Beshkov, S., Kuznetsov, G., Bonelli, S., & Tóth, J. P. (2022). Phylogeography reveals the origin of the two phenological forms of large blue, *Phengaris arion* (Lepidoptera: Lycaenidae). *Biological Journal of the Linnean Society*, *137*, 359–373. https://doi.org/10.1093/biolinnean/blac081

Bordenstein, S. R., O'Hara, F. P., & Werren, J. H. (2001). *Wolbachia* -induced incompatibility precedes other hybrid incompatibilities in *Nasonia*. *Nature*, *409*(6821), 707–710. https://doi.org/10.1038/35055543

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C. H., Xie, D., Suchard, M. A., Rambaut, A., & Drummond, A. J. (2014). BEAST 2: A software platform for Bayesian evolutionary analysis. *PLoS Computational Biology*, *10*(4), e1003537. https://doi.org/10.1371/journal.pcbi.1003537

Brower, A. V. Z. (1994). Rapid morphological radiation and convergence among races of the butterfly *Heliconius erato* inferred from patterns of mitochondrial DNA evolution. *Proceedings of the National Academy of Sciences of the United States of America*, *91*(14), 6491–6495. https://doi.org/10.1073/pnas.91.14.6491

Brues, C. T. (1924). The specificity of food-plants in the evolution of phytophagous insects. *The American Naturalist*, *58*(655), 127–144. https://doi.org/10.1086/279965

Burghardt, F., Proksch, P., & Fiedler, K. (2001). Flavonoid sequestration by the common blue butterfly *Polyommatus icarus*: Quantitative intraspecific variation in relation to larval hostplant, sex and body size. *Biochemical Systematics and Ecology*, *29*(9), 875–889. https://doi.org/10.1016/S0305-1978(01)00036-9

Bush, G. L. (1969). Sympatric host race formation and speciation in frugivorous flies of the genus *Rhagoletis* (Diptera, Tephritidae). *Evolution*, *23*(2), 237–251.

Caillaud, M. C., & Via, S. (2000). Specialized feeding behavior influences both ecological specialization and assortative mating in sympatric host races of pea aphids. *The American Naturalist*, *156*(6), 606–621.

Craig, T. P., Itami, J. K., Abrahamson, W. G., & Horner, J. D. (1993). Behavioral evidence for host-race formation in *Eurosta solidaginis*. *Evolution*, *47*(6), 1696–1710.

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. https://doi.org/10.1093/bioinformatics/btr330

Darriba, D., Taboada, G. L., Doallo, R., & Posada, D. (2012). jModelTest 2: More models, new heuristics and parallel computing. *Nature Methods*, *9*(8), 772. https://doi.org/10.1038/nmeth.2109

Defossez, E., Pitteloud, C., Descombes, P., Glauser, G., Allard, P. M., Walker, T. W., Fernandez-Conradi, P., Wolfender, J. L., Pellissier, L., & Rasmann, S. (2021). Spatial and evolutionary predictability of phytochemical diversity. *Proceedings of the National Academy of Sciences*, *118*(3), e2013344118. https://doi.org/10.1073/pnas.2013344118

Dincă, V., Dapporto, L., Somervuo, P., Vodă, R., Cuvelier, S., Gascoigne-Pees, M., Huemer, P., Mutanen, M., Hebert, P. D., & Vila, R. (2021). High resolution DNA barcode library for European butterflies reveals continental patterns of mitochondrial genetic diversity. *Communications Biology*, *4*(1), 1–11. https://doi.org/10.1038/s42003-021-01834-7

Dincă, V., Montagud, S., Talavera, G., Hernández-Roldán, J., Munguira, M. L., García-Barros, E., Hebert, P. D., & Vila, R. (2015). DNA barcode reference library for Iberian butterflies enables a continental-scale preview of potential cryptic diversity. *Scientific Reports*, *5*(1), 1–12. https://doi.org/10.1038/srep12395

Drès, M., & Mallet, J. (2002). Host races in plant–feeding insects and their importance in sympatric speciation. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *357*(1420), 471–492. https://doi.org/10.1098/rstb.2002.1059

Dupuis, J. R., & Sperling, F. A. (2020). Phylogenomic test of mitochondrial clues to archaic ancestors in a group of hybridizing swallowtail butterflies. *Molecular Phylogenetics and Evolution*, *152*, 106921. https://doi.org/10.1016/j.ympev.2020.106921

Eaton, D. A. R., & Overcast, I. (2020). ipyrad: Interactive assembly and analysis of RADseq datasets. *Bioinformatics*, *36*(8), 2592–2594. https://doi.org/10.1093/bioinformatics/btz966

Ehrlich, P. R., & Raven, P. H. (1964). Butterflies and plants: A study in coevolution. *Evolution*, *18*(4), 586–608. https://doi.org/10.2307/2406212

Eitschberger, U., & Steiniger, H. (1975). Die geographische variation von *Eumedonia eumedon* (Esper, 1780) in der westlichen Paläarktis. *Atalanta*, *6*, 84–125.

Eliasson, C. U., Ryrholm, N., Gärdenfors, U., Holmer, M., & Jilg, K. (2005). *Nationalalnyckeln till Sveriges flora och fauna. Fjärilar: Dagfjärilar. Hesperiidae–Nymphalidae*. SLU ArtDatabanken.

Erhardt, A. (1995). Ecology and conservation of alpine Lepidoptera. In A. S. Pullin (Ed.), *Ecology and conservation of butterflies* (pp. 258–276). Springer Science & Business Media.

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. https://doi.org/10.1111/j.1365-294X.2005.02553.x

Feder, J. L., Hunt, T. A., & Bush, L. (1993). The effects of climate, host plant phenology and host fidelity on the genetics of apple and hawthorn infesting races of *Rhagoletis pomonella*. *Entomologia Experimentalis et Applicata*, *69*(2), 117–135. https://doi.org/10.1111/j.1570-7458.1993.tb01735.x

Fiz, O., Vargas, P., Alarcón, M., Aedo, C., García, J. L., & Aldasoro, J. J. (2008). Phylogeny and historical biogeography of Geraniaceae in relation to climate changes and pollination ecology. *Systematic Botany*, *33*(2), 326–342. https://doi.org/10.1600/036364408784571482

Fiz-Palacios, O., Vargas, P., Vila, R., Papadopulos, A. S., & Aldasoro, J. J. (2010). The uneven phylogeny and biogeography of *Erodium* (Geraniaceae): Radiations in the Mediterranean and recent recurrent intercontinental colonization. *Annals of Botany*, *106*(6), 871–884. https://doi.org/10.1093/aob/mcq184

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. https://doi.org/10.1046/j.1420-9101.2002.00432.x

Forister, M. L., Novotny, V., Panorska, A. K., Baje, L., Basset, Y., Butterill, P. T., Cizek, L., Coley, P. D., Dem, F., Diniz, I. R., Drozd, P., Fox, M., Glassmire, A. E., Hazen, R., Hrcek, J., Jahner, J. P., Kaman, O., Kozubowski, T. J., Kursar, T. A., … Dyer, L. A. (2015). The global distribution of diet breadth in insect herbivores. *Proceedings of the National Academy of Sciences*, *112*(2), 442–447. https://doi.org/10.1073/pnas.1423042112

Fuchs, J. (1989). Ein Vorkommen von *Eumedonia eumedon* Esper (Lycaenidae, Lep.) in den Bergen oberhalb Delfis (Griechenland). *Galathea*, *5*, 153–156.

Gil-T, F. (2004). Una nueva subespecie de *Aricia* (*Eumedonia*) *eumedon* (Esper, 1780) del sur de la Península Ibérica: *axarquia* ssp. nov. (Lepidoptera, Lycaenidae). *Boletín de la SEA*, *34*, 49–54.

Goslee, S. C., & Urban, D. L. (2007). The ecodist package for dissimilarity-based analysis of ecological data. *Journal of Statistical Software*, *22*(7), 1–19. https://doi.org/10.18637/jss.v022.i07

Gu, Z., Eils, R., & Schlesner, M. (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics*, *32*(18), 2847–2849. https://doi.org/10.1093/bioinformatics/btw313

Habel, J. C., Drees, C., Schmitt, T., & Assmann, T. (2010). Review refugial areas and postglacial colonizations in the western palearctic. In J. C. Habel & T. Assmann (Eds.), *Relict species* (pp. 189–197). Springer.

Halbritter, D. A., Storer, C. G., Kawahara, A. Y., & Daniels, J. C. (2019). Phylogeography and population genetics of pine butterflies: Sky islands increase genetic divergence. *Ecology and Evolution*, *9*(23), 13389–13401. https://doi.org/10.1002/ece3.5793

Harborne, J. B., & Williams, C. A. (2002). Phytochemistry of the genus *geranium*. In M. Lis-Balchin (Ed.), *Geranium and pelargonium – The genera geranium and pelargonium* (pp. 32–41). Taylor & Francis.

Hardy, N. B., & Otto, S. P. (2014). Specialization and generalization in the diversification of phytophagous insects: Tests of the musical chairs and oscillation hypotheses. *Proceedings of the Royal Society B: Biological Sciences*, *281*(1795), 20132960. https://doi.org/10.1098/rspb.2013.2960

Hendry, A. P. (2009). Ecological speciation! Or the lack thereof? *Canadian Journal of Fisheries and Aquatic Sciences*, *66*(8), 1383–1398. https://doi.org/10.1139/F09-074

Hendry, A. P., Bolnick, D. I., Berner, D., & Peichel, C. L. (2009). Along the speciation continuum in sticklebacks. *Journal of Fish Biology*, *75*(8), 2000–2036. https://doi.org/10.1111/j.1095-8649.2009.02419.x

Henriksen, H. J., & Kreutzer, I. (1982). *The butterflies of Scandinavia in nature*. Skandinavisk Bogforlag.

Hernández-Roldán, J. L., Bofill, R., Dapporto, L., Munguira, M. L., & Vila, R. (2014). Morphological and chemical analysis of male scent organs in the butterfly genus *Pyrgus* (Lepidoptera: Hesperiidae). *Organisms Diversity & Evolution*, *14*(3), 269–278. https://doi.org/10.1007/s13127-014-0170-x

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

Hinojosa, J. C., Dapporto, L., Pitteloud, C., Koubínová, D., Hernández-Roldán, J., Vicente, J. C., Alvarez, N., & Vila, R. (2022). Hybridization fuelled diversification in *Spialia* butterflies. *Molecular Ecology*, *31*(10), 2951–2967. https://doi.org/10.1111/mec.16426

Hinojosa, J. C., Koubínová, D., Dincă, V., Hernández-Roldán, J., Munguira, M. L., García-Barros, E., Vila, M., Alvarez, N., Mutanen, M., & Vila, R. (2020). Rapid colour shift by reproductive character displacement in *Cupido* butterflies. *Molecular Ecology*, *29*(24), 4942–4955. https://doi.org/10.1111/mec.15682

Howard, R. W., & Blomquist, G. J. (2005). Ecological, behavioral, and biochemical aspects of insect hydrocarbons. *Annual Review of Entomology*, *50*(1), 371–393. https://doi.org/10.1111/j.1095-8312.2010.01511.x

Huemer, P., Mutanen, M., Sefc, K. M., & Hebert, P. D. N. (2014). Testing DNA barcode performance in 1000 species of European Lepidoptera: Large geographic distances have small genetic impacts. *PLoS One*, *9*(12), 1–21. https://doi.org/10.1371/journal.pone.0115774

Janes, J. K., Miller, J. M., Dupuis, J. R., Malenfant, R. M., Gorrell, J. C., Cullingham, C. I., & Andrew, R. L. (2017). The K=2 conundrum. *Molecular Ecology*, *26*(14), 3594–3602. https://doi.org/10.1111/mec.14187

Janz, N., & Nylin, S. (2008). The oscillation hypothesis of host-plant range and speciation. In D. Tilmon (Ed.), *The evolutionary biology of herbivorous insects: Specialization, speciation and radiation* (pp. 203–215). University of California Press.

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), 1–15. https://doi.org/10.1186/1471-2156-11-94

Joyce, A. L., Sermeno Chicas, M., Serrano Cervantes, L., Paniagua, M., Scheffer, S. J., & Solis, M. A. (2016). Host-plant associated genetic divergence of two *Diatraea* spp. (Lepidoptera: Crambidae) stemborers on novel crop plants. *Ecology and Evolution*, *6*(23), 8632–8644. https://doi.org/10.1002/ece3.2541

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., & Drummond, A. (2012). Geneious basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, *28*(12), 1647–1649. https://doi.org/10.1093/bioinformatics/bts199

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. https://doi.org/10.1101/gr.210641.116

Knüttel, H., & Fiedler, K. (2001). Host-plant-derived variation in ultraviolet wing patterns influences mate selection by male butterflies. *Journal of Experimental Biology*, *204*(14), 2447–2459. https://doi.org/10.1242/jeb.204.14.2447

Koçak, A. O. (1979). Studies on the family Lycaenidae II. New taxa and records from Turkey. *Atalanta*, *6*, 309–325.

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. https://doi.org/10.1111/1755-0998.12387

Lafranchis, T., Jutzeler, D., Guillosson, J. Y., Kan, P., & Kan, B. (2015). *La vie des papillons. Ecologie, biologie et comportement des rhopalocères de France*. Diathéo.

Landolt, P. J., & Phillips, T. W. (1997). Host plant influences on sex pheromone behavior of phytophagous insects. *Annual Review of Entomology*, *42*(1), 371–391. https://doi.org/10.1146/annurev.ento.42.1.371

Malausa, T., Dalecky, A., Ponsard, S., Audiot, P., Streiff, R., Chaval, Y., & Bourguet, D. (2007). Genetic structure and gene flow in French populations of two *Ostrinia* taxa: Host races or sibling species? *Molecular Ecology*, *16*(20), 4210–4222. https://doi.org/10.1111/j.1365-294X.2007.03457.x

Marcussen, T., & Meseguer, A. S. (2017). Species-level phylogeny, fruit evolution and diversification history of *Geranium* (Geraniaceae). *Molecular Phylogenetics and Evolution*, *110*, 134–149. https://doi.org/10.1016/j.ympev.2017.03.012

Martínez Pérez, I., Sanjurjo Franch, M. J., & Montiel Pantoja, C. (2015). Plantas nutricias de *Eumedonia eumedon* (Esper, 1780) (Lepidoptera: Lycaenidae) en la provincia de León (noroeste de España). *Arquivos Entomolóxicos*, *13*, 231–250.

Matsubayashi, K. W., Ohshima, I., & Nosil, P. (2010). Ecological speciation in phytophagous insects. *Entomologia Experimentalis et Applicata*, *134*(1), 1–27. https://doi.org/10.1111/j.1570-7458.2009.00916.x

McPheron, B. A., Smith, D. C., & Berlocher, S. H. (1988). Genetic differences between host races of *Rhagoletis pomonella*. *Nature*, *336*(6194), 64–66. https://doi.org/10.1038/336064a0

Mikheyev, A. S., McBride, C. S., Mueller, U. G., Parmesan, C., Smee, M. R., Stefanescu, C., Wee, B., & Singer, M. C. (2013). Host-associated genomic differentiation in congeneric butterflies: Now you see it, now you do not. *Molecular Ecology*, *22*(18), 4753–4766. https://doi.org/10.1111/mec.12423

Munekata, P. E., Alcántara, C., Collado, M. C., Garcia-Perez, J. V., Saraiva, J. A., Lopes, R. P., Barba, F. J., do Prado Silva, L., Sant'Ana, A. S., Fierro, E. M., & Lorenzo, J. M. (2019). Ethnopharmacology, phytochemistry and biological activity of *Erodium* species: A review. *Food Research International*, *126*, 108659. https://doi.org/10.1016/j.foodres.2019.108659

Nice, C. C., & Shapiro, A. M. (2001). Population genetic evidence of restricted gene flow between host races in the butterfly genus *Mitoura* (lepidoptera: Lycaenidae). *Annals of the Entomological*

Society of America, *94*(2), 257–267. https://doi.org/10.1603/0013-8746(2001)094[0257:PGEORG]2.0.CO;2

Nishida, R. (2002). Sequestration of defensive substances from plants by lepidoptera. *Annual Review of Entomology*, *47*, 57–92. https://doi.org/10.1146/annurev.ento.47.091201.145121

Nosil, P. (2012). *Ecological speciation*. Oxford University Press.

Nosil, P., Crespi, B. J., & Sandoval, C. P. (2002). Host-plant adaptation drives the parallel evolution of reproductive isolation. *Nature*, *417*(6887), 440–443. https://doi.org/10.1038/417440a

Nosil, P., & Sandoval, C. P. (2008). Ecological niche dimensionality and the evolutionary diversification of stick insects. *PLoS One*, *3*(4), e1907. https://doi.org/10.1371/journal.pone.0001907

Oroño, L., Paulin, L., Alberti, A. C., Hilal, M., Ovruski, S., Vilardi, J. C., Rull, J., & Aluja, M. (2013). Effect of host plant chemistry on genetic differentiation and reduction of gene flow among *Anastrepha fraterculus* (Diptera: Tephritidae) populations exploiting sympatric, synchronic hosts. *Environmental Entomology*, *42*(4), 790–798. https://doi.org/10.1603/EN13020

Otte, T., Hilker, M., & Geiselhardt, S. (2018). Phenotypic plasticity of cuticular hydrocarbon profiles in insects. *Journal of Chemical Ecology*, *44*(3), 235–247. https://doi.org/10.1007/s10886-018-0934-4

Page, R. D. (2003). *Tangled trees: Phylogeny, cospeciation, and coevolution*. University of Chicago Press.

Peccoud, J., Ollivier, A., Plantegenest, M., & Simon, J. C. (2009). A continuum of genetic divergence from sympatric host races to species in the pea aphid complex. *Proceedings of the National Academy of Sciences*, *106*(18), 7495–7500. https://doi.org/10.1073/pnas.0811117106

Pinzari, M., Santonico, M., Pennazza, G., Martinelli, E., Capuano, R., Paolesse, R., Di Rao, M., D'Amico, A., Cesaroni, D., Sbordoni, V., & Di Natale, C. (2018). Chemically mediated species recognition in two sympatric Grayling butterflies: *Hipparchia fagi* and *Hipparchia hermione* (Lepidoptera: Nymphalidae, Satyrinae). *PLoS One*, *13*(6), e0199997. https://doi.org/10.1371/journal.pone.0199997

Pitteloud, C., Arrigo, N., Suchan, T., Mastretta-Yanes, A., Vila, R., Dincă, V., Hernández-Roldán, J., Brockmann, E., Chittaro, Y., Kleckova, I., Fumagalli, L., Buerki, S., Pellissier, L., & Alvarez, N. (2017). Climatic niche evolution is faster in sympatric than allopatric lineages of the butterfly genus *Pyrgus*. *Proceedings of the Royal Society B: Biological Sciences*, *284*(1852), 20170208. https://doi.org/10.1098/rspb.2017.0208

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

Quattrini, A. M., Wu, T., Soong, K., Jeng, M. S., Benayahu, Y., & McFadden, C. S. (2019). A next generation approach to species delimitation reveals the role of hybridization in a cryptic species complex of corals. *BMC Evolutionary Biology*, *19*(1), 116. https://doi.org/10.1186/s12862-019-1427-y

Quek, S. P., Davies, S. J., Itino, T., & Pierce, N. E. (2004). Codiversification in an ant-plant mutualism: Stem texture and the evolution of host use in *Crematogaster* (Formicidae: Myrmicinae) inhabitants of *Macaranga* (Euphorbiaceae). *Evolution*, *58*(3), 554–570. https://doi.org/10.1111/j.0014-3820.2004.tb01678.x

Rabasa, S. G., Gutierrez, D., & Escudero, A. (2005). Egg laying by a butterfly on a fragmented host plant: A multi-level approach. *Ecography*, *28*(5), 629–639. https://doi.org/10.1111/j.2005.0906-7590.04229.x

Rambaut, A. (2015). FigTree, v1.4.2: tree figure drawing tool. http://tree.bio.ed.ac.uk/software/figtree/

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. https://doi.org/10.1093/sysbio/syy032

Reddy, G. V., & Guerrero, A. (2004). Interactions of insect pheromones and plant semiochemicals. *Trends in Plant Science*, *9*(5), 253–261. https://doi.org/10.1016/j.tplants.2004.03.009

Rosenberg, N. A. (2004). DISTRUCT: A program for the graphical display of population structure. *Molecular Ecology Notes*, *4*(1), 137–138. https://doi.org/10.1046/j.1471-8286.2003.00566.x

Rundle, H. D., & Nosil, P. (2005). Ecological speciation. *Ecology Letters*, *8*(3), 336–352. https://doi.org/10.1111/j.1461-0248.2004.00715.x

Schoonhoven, L. M., van Loon, J. J. A., & Dicke, M. (2005). *Insect-plant biology*. Oxford University Press.

Schurian, K., Castelain, C., & Weiss, J. (2014). *Eumedonia eumedon* (Esper, [1780]) aus der Osttürkei, mit Beschreibung einer neuen Unterart (Lepidoptera: Lycaenidae). *Nachrichten des Entomologischen Vereins Apollo*, *35*, 1–4.

Sharma, G. (2013). Role of phytochemicals and nucleic acids in solving taxonomic problems. *Proceedings of the National Academy of Sciences, India Section B: Biological Sciences*, *83*(2), 141–145. https://doi.org/10.1007/s40011-012-0126-x

Shoemaker, D. D., Katju, V., & Jaenike, J. (1999). *Wolbachia* and the evolution of reproductive isolation between *Drosophila recens* and *Drosophila subquinaria*. *Evolution*, *53*(4), 1157–1164. Portico.

Silva-Brandao, K. L., Peruchi, A., Seraphim, N., Murad, N. F., Carvalho, R. A., Farias, J. R., Omoto, C., Consoli, F. L., Figueira, A., & Brandao, M. M. (2018). Loci under selection and markers associated with host plant and host-related strains shape the genetic structure of Brazilian populations of *Spodoptera frugiperda* (lepidoptera, Noctuidae). *PLoS One*, *13*(5), e0197378. https://doi.org/10.1371/journal.pone.0197378

Smith, P. T., Krager, K., Cronin, J. T., & Kambhampati, S. (2002). Mitochondrial DNA variation among host races of *Eurosta solidaginis* fitch (Diptera: Tephritidae). *Molecular Phylogenetics and Evolution*, *25*(2), 372–376. https://doi.org/10.1016/S1055-7903(02)00261-0

Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*(9), 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

Thompson, J. N. (1989). Concepts of coevolution. *Trends in Ecology & Evolution*, *4*(6), 179–183. https://doi.org/10.1016/0169-5347(89)90125-0

Tilmon, K. J. (2008). *The evolutionary biology of herbivorous insects: Specialization, speciation and radiation*. University of California Press.

Tolman, T., & Lewington, R. (2008). *Collins butterfly guide: The most complete field guide to the butterflies of Britain and Europe*. HarperCollins Publishers.

Tshikolovets, V. V. (2011). *Butterflies of Europe and Mediterranean area*. Tshikolovets Publications.

Tuzov, V. K., Bogdanov, P. V., Churkin, S. V., Devyatkin, A. L., Danchenko, A. V., Murzin, V. S., Samodurov, G. D., & Zhdanko, A. B. (2000). Guide to the butterflies of Russia and adjacent territories (lepidoptera, Rhopalocera). In *Libytheidae*, *Danaidae*, *Nymphalidae*, *Riodinidae*, *Lycaenidae* (Vol. *2*). Pensoft.

van Etten, J. (2017). R package gdistance: Distances and routes on geographical grids. *Journal of Statistical Software*, *76*(13), 21. https://doi.org/10.18637/jss.v076.i13

van Nieukerken, E. J., Kaila, L., Kitching, I. J., Kristensen, N. P., Lees, D. C., Minet, J., Mitter, C., Mutanen, M., Regier, J. C., Simonsen, T. J., Wahlberg, N., Yen, S.-H., Zahiri, R., Adamski, D., Baixeras, J., Bartsch, D., Bengtsson, B. Å., Brown, J. W., Bucheli, S. R., … Zwick, A. (2011). Order Lepidoptera Linnaeus, 1758. In Z.-Q. Zhang (Ed.), *Animal biodiversity: An outline of higher-level classification and survey of taxonomic richness* (Vol. *3148*, pp. 212–221). *Zootaxa*. https://doi.org/10.11646/zootaxa.3148.1.41

Vijayan, K., & Tsou, C. H. (2010). DNA barcoding in plants: Taxonomy in a new perspective. *Current Science*, *99*(11), 1530–1541.

Walsh, B. (1864). On phytophagic varieties and phytophagic species. *Proceedings of the Entomological Society Philadelphia*, *3*, 403–430.

Wang, I. J., & Bradburd, G. S. (2014). Isolation by environment. *Molecular Ecology*, *23*(23), 5649–5662. https://doi.org/10.1111/mec.12938

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

Wink, M., Botschen, F., Gosmann, C., Schäfer, H., & Waterman, P. G. (2010). Chemotaxonomy seen from a phylogenetic perspective and evolution of secondary metabolism. In M. Wink (Ed.), *Annual plant reviews volume 40: Biochemistry of plant secondary metabolism* (pp. 364–433). Blackwell Publishing Ltd.

Yi, X., & Latch, E. K. (2022). Nonrandom missing data can bias PCA inference of population genetic structure. *Molecular Ecology Resources*, *22*(2), 602–611. https://doi.org/10.1111/1755-0998.13498

Zhang, L., Hood, G. R., Ott, J. R., & Egan, S. P. (2019). Temporal isolation between sympatric host plants cascades across multiple trophic levels of host-associated insects. *Biology Letters*, *15*(12), 20190572. https://doi.org/10.1098/rsbl.2019.0572

Zhang, Z. Q. (2011). Animal biodiversity: An introduction to higher-level classification and taxonomic richness. *Zootaxa*, *3148*(1), 7–12.

Zhdanko, B. (1997). Lycaenid foodplants in Kazakhstan and middle Asia. *Atalanta*, *28*, 97–110.

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: Hinojosa, J. C., Montiel-Pantoja, C., Sanjurjo-Franch, M., Martínez-Pérez, I., Lee, K. M., Mutanen, M., & Vila, R. (2023). Diversification linked to larval host plant in the butterfly *Eumedonia eumedon*. *Molecular Ecology*, *32*, 182–197. https://doi.org/10.1111/mec.16728