DR. LEONARDO DAPPORTO (Orcid ID : 0000-0001-7129-4526)

PROF. MARTIN WIEMERS (Orcid ID : 0000-0001-5272-3903)

DR. ROGER VILA (Orcid ID : 0000-0002-2447-4388)

Article type : Resource Article

Integrating three comprehensive datasets shows that mitochondrial DNA variation is linked to species traits and paleogeographic events in European butterflies.

Leonardo Dapporto[1]*, Alessandro Cini[1, 2], Raluca Vodă[3], Vlad Dincă[4], Martin Wiemers[5,6], Mattia Menchetti[1], Giulia Magini[1], Gerard Talavera[7], Tim Shreeve[8], Simona Bonelli[3], Luca Pietro Casacci[9,3], Emilio Balletto[3], Stefano Scalercio[10], Roger Vila[7]

[1] Dipartimento di Biologia dell'Università di Firenze, Via Madonna del Piano 6 50019 Sesto Fiorentino, Firenze, Italy

[2] Centre for Biodiversity and Environment Research, University College London, Gower Street, London WC1E 6BT, UK.

[3] Department of Life Sciences and Systems Biology, University of Turin, Via Accademia Albertina 13, 10123 Turin, Italy

[4] Department of Ecology and Genetics, PO Box 3000, University of Oulu, 90014 Oulu, Finland

[5] Senckenberg Deutsches Entomologisches Institut, Eberswalder Str. 90, 15374 Müncheberg, Germany

[6] UFZ – Helmholtz Centre for Environmental Research, Department of Community Ecology, Theodor-Lieser-Str. 4, 06120 Halle, Germany

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/1755-0998.13059

[7] Institut de Biologia Evolutiva (CSIC-Universitat Pompeu Fabra), Passeig Marítim de la Barceloneta 37, 08003, Barcelona, Spain

[8] Centre for Ecology, Environment and Conservation, Faculty of Health and Life Sciences, Oxford Brookes University, Headington, Oxford, OX3 0BP, UK

[9] Museum and Institute of Zoology, Polish Academy of Sciences, Wilcza 64, 00-679 Warsaw, Poland.

[10] Consiglio per la ricerca in agricoltura e l'analisi dell'economia agraria, Centro di Ricerca Foreste e Legno, Contrada Li Rocchi, 87036 Rende, Italy.

* Corresponding author, leonardo.dapporto@unifi.it

Abstract

Understanding the dynamics of biodiversity, including the spatial distribution of genetic diversity, is critical for predicting responses to environmental changes, as well as for effective conservation measures. This task requires tracking changes in biodiversity at large spatial scales and correlating with species functional traits. We provide three comprehensive resources to understand the determinants for mitochondrial DNA differentiation represented by i) 15,609 COI sequences and ii) 14 traits belonging to 307 butterfly species occurring in Western-Central Europe and iii) the first multi-locus phylogenetic tree of all European butterfly species. By applying phylogenetic regressions we show that mitochondrial DNA spatial differentiation (as measured with Gst, G'st, D and Dst) is negatively correlated with species traits determining dispersal capability and colonization ability. Thanks to the high spatial resolution of the COI data, we also provide the first zoogeographic regionalization maps based on intraspecific genetic variation. The overall pattern obtained by averaging the spatial differentiation of all Western-Central European butterflies shows that the paradigm of long-term glacial isolation followed by rapid pulses of post-glacial expansion has been a pervasive phenomenon in European butterflies. The results and the extensive datasets we provide here constitute the basis for genetically-informed conservation plans for a charismatic group in a continent where flying insects are under alarming decline.

Introduction

Genetic diversity within populations and its spatial differentiation among populations are central concepts in biology. Within population diversity provides opportunities for populations to respond to shifting ecological pressures and inter-population differentiation triggers the processes of allopatric speciation (Coyne & Orr, 2004; Hughes, Inouye, Johnson, Underwood, & Vellend, 2008). Understanding the emergence and maintenance of genetic differentiation exposes fundamental evolutionary processes over a range of spatial and temporal scales. The resolution of such studies has advanced greatly since the onset of DNA sequencing (Allio, Donega, Galtier, & Nabholz, 2017; Bazin, Glémin, & Galtier, 2006; Lewontin, 1974; Nabholz, Mauffrey, Bazin, Galtier, & Glemin, 2008).

A substantial effort has been devoted to verifying the prediction that neutral genetic diversity should equate to the product of mutation rate and effective population size. Despite this clear theoretical statement, DNA polymorphism appeared to be weakly correlated to population size and, when correlations have been found, the genetic diversity revealed is orders of magnitude smaller than expected based on differences in population size (Bazin et al., 2006; Leffler et al., 2012; Nabholz et al., 2008; Romiguier et al., 2014). Moreover, the results greatly varied among studies comparing genetic diversity for different taxa as well as when using different genetic markers (such as allozymes, nuclear or mitochondrial markers) (Allio et al., 2017; Bazin et al., 2006; Fujisawa, Vogler, & Barraclough, 2015; Leffler et al., 2012; Nabholz, Glémin, & Galtier, 2009; Nabholz et al., 2008; Romiguier et al., 2014).

As a major example, differentiation in nuclear (nDNA) and mitochondrial DNA (mtDNA) is expected to show different determinants even in the same model organisms. First of all, mtDNA has a faster mutation rate compared to nDNA and can show signatures of recent differentiation (e.g. intraspecific) as well as relatively old (Avise, 2009; Hebert, Cywinska,

Ball, & deWaard, 2003). Secondly, because mtDNA is haploid, maternally inherited and recombination is limited to rare cases of heteroplasmy, its effective population size is four times smaller and coalescence times shorter than in nuclear DNA (Allio et al., 2017; Nabholz et al., 2009). mtDNA is involved in respiration processes and has been found to be under strong selection (Galtier, Nabholz, GléMin, & Hurst, 2009; Nabholz et al., 2009; Pentinsaari, Salmela, Mutanen, & Roslin, 2016). Finally, mtDNA differentiation can be influenced by infections of microorganisms like *Wolbachia* (Galtier et al., 2009; Smith et al., 2012; Werren, Baldo, & Clark, 2008). Selection for variants determining different respiration performance and improved fitness in association with microorganisms, associated with high potential for genetic hitchhiking in the non-recombinant mitochondrial genome, make it difficult to disentangle neutral from adaptive mutations (Gillespie, 2000, 2001). Finally, population size may rapidly vary in geological time following environmental perturbations. It is thus expected that current effective population size and other species traits are poor predictors for the assumed consistent mutation rates and the resulting mtDNA polymorphism as expected by the neutral theory (Nabholz et al., 2009; Romiguier et al., 2014).

While several studies searching for fingerprints of effective population size and other species traits on DNA polymorphism have been carried out through inter-specific comparisons (Allio et al., 2017; Bazin et al., 2006; Fujisawa et al., 2015; Leffler et al., 2012; Nabholz et al., 2009; Romiguier et al., 2014), there are very few comparative phylogeographic studies which adopted a spatially explicit framework (Burney & Brumfield, 2009; Dapporto et al., 2017; Moritz et al., 2009), facilitating understanding inter-population patterns of genetic diversity and its determinants. The primary challenge for phylogeography is to adequately map current genetic diversity to allow the testing hypotheses that explain such variation. Usually patterns are explained by relatively recent events, such as Quaternary climatic oscillations (Avise, 2009; Hewitt, 2004). The increase in phylogeographic studies of multiple taxa opens the door

to comparative work, which adds a layer of complexity in searching for shared sources of interspecific patterns, particularly in relating intraspecific genetic variation to environmental features and species traits (Bowen et al., 2016; Papadopoulou & Knowles, 2016). The final goal of comparative phylogeography is to disentangle deterministic historical/contemporary and biotic/abiotic processes that determine the detected diversity (Dawson, 2014; Papadopoulou & Knowles, 2016).

Because of the relatively faster rates of divergence and coalescence compared to nDNA, mtDNA is a primary marker to study the distribution of diversity at the intraspecific level, with an almost ubiquitous use in phylogeography (Avise, 2009; Avise et al., 1987). This is particularly true for the cytochrome c oxidase subunit I (COI), a section of which has become the standard DNA barcode for animals (Hebert et al., 2003). Currently, public DNA barcode libraries contain millions of sequences (Kress, García-Robledo, Uriarte, & Erickson, 2015; Ratnasingham & Hebert, 2007) and now allow unknown samples to be identified, often to species level. The accumulation of DNA barcode data for an increasing number of groups, and in particular European butterflies, in public repositories (GenBank, BOLD) generated by wide scale research surveys (Dapporto et al., 2017; Dincă et al., 2015; Hausmann et al., 2011; Huemer, Mutanen, Sefc, & Hebert, 2014) is now extensive.

Here, we provide a novel assessment of which species traits correlate with different layers of intra-specific mtDNA differentiation, providing an overview of comparative phylogeography in western European butterflies. We provide the first zoogeographic regionalization map based on intra-specific genetic variation at the subcontinental scale of an entire superfamily (Papilionoidea). Our analyses are based on three novel resources now available for future studies: i) a DNA barcode dataset for the 307 butterfly species occurring in western Europe (15,609 COI sequences of which 5,380 sequences were new for this study) (Fig.1); b) a database of 14 species features including feeding, morphological, natural history and

ecophysiological traits for each of the 307 barcoded species and c) a phylogenetic tree for all 496 European butterflies based on the mitochondrial gene COI and 13 nuclear markers. By integrating these datasets, we test three main predictions about mtDNA genetic diversification and its spatial structure.

First, since high selection, absence of recombination, erratic mutation rate and stochastic variation in population size make overall mtDNA divergence highly unpredictable (Allio et al., 2017; Galtier et al., 2009; Nabholz et al., 2009; Romiguier et al., 2014), we expect to find no correlations between mtDNA diversity (haplotype diversity) and species traits related to population size, dispersal capability, number of generations and climatic tolerance (prediction 1).

Second, a plethora of studies demonstrated that mtDNA shows strong differentiation among populations particularly in poorly dispersive species. Consequently, we predict that when spatial information and genetic variation are assessed together, as typically done with widely used indices of population differentiation (Whitlock, 2011), a relationship with species traits should emerge (Burney & Brumfield, 2009; Dapporto et al., 2017) (prediction 2).

Finally, the Quaternary history of Europe has been dominated by climatic pulses which rendered most of Northern and Central Europe unsuitable for many ectothermic species, which became restricted to the southern peninsulas (Iberia, Italy, Balkans) and Mediterranean islands (Hewitt, 2004; Schmitt, 2007) during cold periods. These refugia were separated from each other by conspicuous physical barriers such as sea channels and mountain chains (mostly represented by the Alps and Pyrenees) (Fig. 2) (Hewitt, 2004; Schmitt, 2007). We predict that a zoogeographic regionalization based on the intraspecific COI variation in our dataset will produce diversity patterns coherent with those expected on the basis of

theoretical, geomorphological and paleoclimatic expectations, as well as with those obtained by comparing communities based on faunistic data (prediction 3).

Materials and Methods

Sampling and dataset

We gathered 15,609 COI sequences belonging to 307 species occurring in Western Europe (Spain, Portugal, Andorra, France, United Kingdom, Belgium, Germany, Italy, Switzerland, Austria, Sweden, Norway, Denmark, Belgium, Netherlands). 5,380 COI sequences have been generated for this study by using standard procedures (see Supplementary Methods and Results), and the rest have been obtained from BOLD (http://www.boldsystems.org/) and GenBank (10,229). Sequences have been screened to verify that i) they had a length of at least 500 bp, ii) they were georeferenced and iii) they were assigned to the correct species. The recent check list of European butterflies (Wiemers et al., 2018) has been used as a reference for taxonomy, but a series of species sharing DNA barcodes according to previous studies (Dincă et al., 2015; Dincă, Zakharov, Hebert, & Vila, 2011) have been merged into a single entity because they share mitochondrial history (Supplementary Methods and Results). Sequences have been grouped into spatial units as follows: islands have been treated as individual units and sequences for the European mainland have been divided into areas of 2.5x2.5 degrees of latitude and longitude, resulting in 123 spatial units with at least one sequence (Fig. 1).

Indices of genetic differentiation

For species occurring in at least 4 spatial units and with a minimum of 15 sequences in total we calculated three indices of genetic differentiation. The first, haplotype diversity (Hd), was calculated as the average of p-distance matrices among haplotypes using the "nuc.div" function of the "pegas" R package. This index is only dependent on the degree of differentiation among haplotypes regardless of their spatial distribution and frequency, and it is typically used to measure mtDNA polymorphism (Nabholz et al., 2009).

The second is the absolute differentiation among populations (Nei, 1987), which is given by:

$$Dst = Ht - Hs$$

where Ht represents the average p-distances for all specimens of a given species, and Hs is the average of the intra-unit p-distances. Thus, Dst represents the average genetic differentiation among areas in p-distance units. Species showing a higher differentiation among haplotypes (high Ht) and a spatial segregation (low Hs) have a maximum value for this index. Negative Dst values (intra-area differentiation higher than inter-area differentiation) can have different subtle meanings, but are most often generated as artefacts due to relatively small sample sizes; usually they are set to zero (Meirmans & Hedrick, 2011) and we applied this solution.

The third measure was the widely used standardized index of population differentiation (Nei, 1987) defined as:

$$Gst = Dst/Ht$$

which represents the fraction of the total genetic differentiation encompassed by the differentiation among areas (Nei, 1987). This index ranges from negative values to 1 (complete differentiation) and is independent of the number of changes exhibited by the

different haplotypes of a given species. Negative values have been set to zero (see above). The use of Gst has been debated as a measure of population diversification for extremely variable markers (which is usually not the case for mitochondrial markers) as it tends to underestimate differentiation among populations and to strongly depend on intra-population variability (Jost, 2008; Whitlock, 2011). For this reason, we also applied both D and G'st indices, which are less affected by high values of Hs (see Supplementary Methods and Results for their formulation).

Species traits

For the selected species we gathered a series of traits (Dapporto et al., 2017) representing four (morphology, feeding, life history and physiology) of the five groups identified by Moretti et al. (2017) to cover the primary functions of invertebrates: a) trophic generalism (feeding trait), was identified as i) the number of host plant genera reported in two literature sources (Table S2); b) mobility measured by the ii) wingspan proxy morphological trait as indicated by Sekar (2012) and assessed as the average of minimum and maximum wingspan reported for each species in Higgins & Riley (1970); c) phenology (life history trait) identified as iii) the number of months during which adults occur in Europe, iv and v) the first and the last month when adults fly, and vi) voltinism, i.e. the maximum number of generations per year recorded in Europe (Tolman & Lewington, 2008). We also included a series of variables describing d) the climatic preference and tolerance (physiological trait) according to Schweiger, Harpke, Wiemers, & Settele (2014). Although these climatic niche indices cannot be considered as functional traits (Moretti et al., 2017), they are widely used as proxies for the traits responsible for eco-physiological responses to climate (Dapporto et al., 2017; Devictor et al., 2012). The variables we included are: vii) mean annual temperature,

viii) mean annual precipitation, ix) standard deviation of mean temperature, x) standard deviation of mean precipitation, xi) upper 95% confidence limit of temperature mean, and xii) lower 95% confidence limit of precipitation mean. Although direct information about effective population size for all species over the entire continent is unavailable, their occurrence in Europe is well assessed and range size, calculated as xiii) the number of 30x30km squares occupied (Schweiger et al., 2014), is used here as a proxy for population size for at least two reasons: 1) the species showing wider distributions can be expected to have a higher total of individuals across their range, 2) butterfly species with larger ranges also tend to have more numerous local populations (Brändle, Öhlschläger, & Brandl, 2002). Another distributional trait has been included as xiv) the maximum altitude to which a species lives in Europe (Table S3).

Butterfly traits are usually highly inter-correlated and can be reduced to factors by using ordination methods (Dapporto et al., 2017; Middleton-Welling, Wade, Dennis, Dapporto, & Shreeve, 2018). Principal Component Analysis was applied to life history and physiology traits using the R function "rda" and the components with eigenvalues higher than one have been retained as variables.

As a reference phylogeny, we constructed a phylogenetic tree for all 496 species of European butterflies based on 14 genes (1 mitochondrial and 13 nuclear). The complete alignment was made with ClustalW as implemented in BioEdit 7.2.5 (Hall, 1999) and consisted of 496 sequences (one for each species) with a total length of 15,741 sites, 5,214 of them parsimony-informative, containing the following genes: COI (covering 496 species with a total length of 1,532 sites and a mean number of 1,087 nucleotides per species), wingless (283/467/386), EF1α (282/1,725/1,022), rPS5 (143/760/594), GAPDH (137/714/651), CAD (103/2,928/946), MDH (67/750/590), IDH (65/710/681), H3 (57/329/328), RpS2 (42/862/454), DDC (27/2,012/689), HCL (21/633/623), Thiolase (21/1,020/1,020), and CAT (20/1,299/1,292). A

maximum likelihood tree was estimated with IQ-TREE using the above alignment partitioned by genes and codon positions, with the substitution model option set to “Auto”, applying the FreeRate model with 4 rate categories, and default settings for branch support analysis and search parameters. The existence of a phylogenetic signal for species traits and for Hd, Dst, D, G’st and Gst was tested with Pagel’s lambda index by applying the “phylosig” R function of the "phytools" package.

Assessing predictions 1 and 2: Determinants for mtDNA differentiation

The relationships between species traits and their Hd, Dst, D, Gst and G’st were assessed using phylogenetic regression. We used Pagel's lambda as a model for the phylogenetic covariance of residuals as implemented in the function “pgls” of the R package “caper”. To avoid model overfitting and to provide a better parameterization of variables, we used the framework of multi-model inference of Generalized Linear Models through Information-Theoretic Approach (Burnham & Anderson, 2002) to select a set of “best models” by using the “MuMIn” R package. This approach allows selection of the best combination of predictors from the global model including all possible combinations. The model comparisons were performed adopting the corrected Akaike Information Criterion (AICc), and the model choice was done based on ΔAICc (which represents the difference between each model and the most parsimonious model). We selected all models with ΔAICc values < 4, considered to be equally parsimonious (Burnham & Anderson, 2002). According to this procedure only a small subset of predictors is selected as significantly affecting the response variable. The correlation coefficients of each predictor are averaged among the selected best-fitting models. The significance of the estimated coefficient is calculated with a z Wald test.

Assessing prediction 3: Overall phylogeographic structure

To provide a zoogeographic regionalization of South-Western Europe based on intraspecific diversification of COI sequences we applied the most recent procedures used in zoological regionalization based on a combination of hierarchical tree analysis to define clusters and unconstrained ordination to describe their relationships (Holt et al., 2013). At the basis of the procedure, a distance matrix among units was produced using pairwise Gst among pairs of units for each species, using the following formula:

$$Gst_{i,j} = Dst_{i,j}/Ht$$

This represents the fraction of the overall genetic differentiation (Ht) expressed by the population differentiation between a given pair of units (i and j).

Using the Gst pairwise matrices for each species, we then calculated the mean of the available values of the corresponding cells of the matrix. We retained a series of units that shared at least 10 species among each other to produce a final mean Gst matrix, representing the degree of genetic differentiation among selected units based on all species. We then applied a Ward hierarchical clustering to this matrix. By using the "recluster" R package the tree was cut at different levels returning a series of clustering solutions. Then, a Principal Coordinates Analysis (PCoA) was applied to the dissimilarity matrix and we projected the configuration in the RGB space using the R package "recluster" (Dapporto et al., 2013). The colour resemblance of the resulting dots is proportional to the genetic similarity among the units. For each cut of the tree we attributed colours to the areas belonging to each cluster. These colours corresponded to the barycentre of area positioning in the RGB space. This

"average colour" for each region has been used for mapping the zoogeographic regions as done by Holt et al. (2013).

Results

Genetic dataset and species traits

The 15,609 COI sequences belonging to 307 species have been grouped into 123 areas identified as islands or into areas of 2.5x2.5 degrees of latitude and longitude for the European mainland (Fig. 1). Among the 307 species for which at least one sequence was available, 224 fulfilled a minimum requirement set for the assessment of population diversification (at least 4 areas and 15 specimens). A full set of 14 traits describing feeding ecology, mobility, phenology, climatic tolerance and demography was available for 214 of these species. A PCA carried on four traits defining butterfly phenology identified only one component with an eigenvalue higher than 1 and it was mostly positively represented by voltinism and length of the flight period (Table S3, Fig. S1). For eco-physiological traits defining climatic tolerance, two components had eigenvalues >1 (Table S3, Fig. S2): the first component ordered species from those adapted to colder climates to those living in warmer areas; the second component ordered species mostly according to the precipitation they experienced (Table S3, Fig. S2). A phylogenetic tree was obtained by using 14 markers and all 496 species of butterflies occurring in Europe (Wiemers et al., 2018) (Fig. 2) and showed an almost complete topological agreement with recent global phylogenies (Espeland et al., 2018). The tree allowed us to verify that the indices for population diversification (haplotype diversity, Dst, D, Gst, G'st) did not show any phylogenetic signal, while most functional traits did (Table 1).

Prediction 1. mtDNA polymorphism and species traits

As done in recent studies correlating genetic variability with species features (Fujisawa et al., 2015; Leffler et al., 2012; Romiguier et al., 2014), we performed phylogenetic regressions to model the three indices of genetic diversification against species trait controlling for phylogenetic signal. Phylogenetic regressions revealed that none of the selected traits significantly explained mtDNA polymorphism measured as haplotype diversity (Table 2), although ecophysiology PC1, related to temperatures of the locations occupied was close to the significance threshold; species in current warmer areas tended to have higher mtDNA diversity than those of cold areas (Table 2).

Prediction 2. Population differentiation and species traits

When the spatial information was added to the genetic differentiation in the assessment of Dst, phenology significantly explained the variation in overall population differentiation, with species characterized by longer flight periods and higher number of generations showing a lower level of differentiation (Fig. 3a, Table 3). An almost identical result was obtained by using D (Table S2) probably because for mtDNA the 1-Hs denominator tends to 1 due to the low intra-population differentiation in species showing spatial structure. Several species traits significantly correlated with Gst, only measuring the spatial segregation of haplotypes regardless of their degree of differentiation (Table 4). Species characterized by smaller wings and those exploiting a lower number of host plants (specialists), had higher population differentiation, with a similar, near significant, relationship for species with short flight periods. Widespread species also had high population differentiation (Table 4, Fig. 3 b-d). G'st showed very similar results to Gst (Table S1).

Prediction 3: Zoogeographic region with intra-specific differentiation

Using the 224 species with sufficient data previously selected, we calculated a pairwise Gst matrix for each species among the areas where it has been found, and then an average Gst matrix has been calculated among areas. A Ward hierarchical clustering produced from the average pairwise Gst distance matrix was sliced at different levels as usually done for zoogeographic regionalization (Fig. 4a). Due to higher species richness and a higher sampling effort in southern European regions, a subset of units having at least 10 shared species was concentrated in the Mediterranean area (Fig. 4b). The different hierarchical clustering solutions from K = 2 to K = 6 provided regionalization results that link nearby areas and that were highly coherent with existing geographic barriers (Fig. 4b-f). The first node separated Iberia (except for Catalonia), the Balearics and Sardinia from the other areas (Fig. 3b). A solution with three clusters divided Sicily from the Pyrenees, Alps and Italian peninsula, according to a well-known efficient barrier to dispersal represented by the narrow Messina strait (Dapporto, Bruschini, Dincă, Vila, & Dennis, 2012; Vodă, Dapporto, Dincă, & Vila, 2015) (Fig. 4c). The fourth cluster recognised Sardinia and the Balearics as a unit, according to a well-known and still largely unexplained similarity between these areas that may be explained by a refugium hypothesis (Dincă et al., 2011; Vodă et al., 2015) (Fig. 4d). The fifth cluster separated the Alps and Pyrenees from the Italian peninsula, with Corsica and circum-Italian islands resembling more the Pyrenees-Alps, reflecting a recurrent phylogeographic pattern found in several butterfly species (Dapporto et al., 2012) (Fig. 4e). The sixth cluster produced the expected division between the Alps and Pyrenees with Corsica, Elba and Giglio resembling more the Pyrenees than the spatially closer Alps (Dapporto et al., 2017) (Fig 4f).

Discussion

The three comprehensive resources (COI sequence dataset, species traits and phylogenetic tree) here presented for butterflies of Western Europe allowed us to test three specific predictions regarding mtDNA intraspecific genetic differentiation. First, we confirmed that overall intraspecific genetic variation in mtDNA (i.e. COI polymorphism) cannot be explained by any of the selected species traits, which were chosen to cover most of the main functions of invertebrates; dispersal, feeding, natural history, ecophysiology and distribution, the last being a proxy for population size (prediction 1). Second, when a spatially-explicit framework was applied, genetic differentiation among populations showed an effect for species traits, mostly when the influence of absolute genetic divergence is removed, as done by using Gst (and G'st). This indicates that the emergence and maintenance of mtDNA differentiation is, at least in part, deterministically shaped by species ecology and by historical factors. Our result also supports the use of COI as a marker to understand ecological fingerprints in mtDNA spatial differentiation (prediction 2). Finally, the zoogeographic regionalization based on average COI population differentiation agreed with the main European biogeographical paradigm, which suggests that most butterfly species were isolated in restricted southern European areas during glacial periods, where they differentiated. From these refugia, the different lineages of most species experienced pulses of poleward expansion during the warmer interglacial periods, producing the observed recurrent suture zones along physical barriers (Alps, Pyrenees).

Prediction 1. mtDNA polymorphism and species traits

Substitution rate and degree of polymorphism of mitochondrial and nuclear DNA are known to vary largely among taxa (Allio et al., 2017; Bazin et al., 2006; Fujisawa et al., 2015;

Leffler et al., 2012; Nabholz et al., 2009; Romiguier et al., 2014; Welch, Bininda-Emonds, & Bromham, 2008). In general, animal groups show differences in DNA substitution rates that are linked to their population size or related traits (species range or body mass) or to other pressures acting on mtDNA (e.g. generation time, fecundity, homeo-heterothermic physiology) (Allio et al., 2017; Nabholz et al., 2009; Pentinsaari et al., 2016). On the other hand, there is mixed evidence for the expected relationship between intraspecific mtDNA polymorphism and species traits (Bazin et al., 2006; Fujisawa et al., 2015; Nabholz et al., 2009; Romiguier et al., 2014).

Despite the theoretical expectation that levels of neutral genetic variation should increase with effective population size, no relationships between range size (a likely good proxy for population size) and mitochondrial DNA diversity has emerged in our extensive dataset. Similarly, other species traits determining butterfly climatic and feeding generalism (ecophysiological traits and number of host plants), dispersal capability (length of flight period and wingspan) as well as the number of generations are expected to influence the degree of gene flow and thus the level of genetic differentiation. Nevertheless, no significant relationships emerged between these traits and mitochondrial diversity. As previously suggested, the causes for the absence of these correlations likely reside in the non-neutral nature of mtDNA variation, in its apparently erratic mutation rate and in the strong fluctuations affecting both mtDNA polymorphism and population sizes in historical times (Bazin et al., 2006; Nabholz et al., 2009, 2008; Romiguier et al., 2014). The absence of any phylogenetic signal in mtDNA overall intraspecific genetic diversity, as well as in the indices of spatial differentiation, combined with most species traits being similar among closely related species supports the erratic behaviour of mtDNA polymorphism and its highly stochastic determinants.

mtDNA variants can determine different respiration performances (Toews, Mandic, Richards, & Irwin, 2014) and adaptive mutations are expected to rapidly spread in a population and even across populations, producing selective sweeps (Bazin et al., 2006; Galtier et al., 2009). Similarly, the maternally-inherited symbiotic bacterium *Wolbachia* can induce reproductive alterations (e.g. feminization, male killing, cytoplasmatic incompatibility) which are adaptive for the bacterium by enhancing the production of infected females (Werren et al., 2008). *Wolbachia* infection can result in a single haplotype (or haplogroup) dominating an entire population and there is growing evidence in European butterflies that different mtDNA lineages are associated with different infection status or strains of this bacterium (Dincă et al., 2018; Hernández-Roldán et al., 2016; Ritter et al., 2013). Due to the non-recombinant transmission of mtDNA, genetic sweeps strongly affect the entire mtDNA genome, thus drastically lowering or eliminating former genetic diversity. This phenomenon can occur more frequently in species with a large effective population size, thus counterbalancing the emergence of a larger number of genetic variants (genetic draft) (Gillespie, 2000, 2001). In-depth studies of mtDNA polymorphism in vertebrates have provided evidence that selective sweeps and genetic drift are less frequent than previously hypothesized (Allio et al., 2017; Karl, Toonen, Grant, & Bowen, 2012), but this may not be the case in insects, where effective population sizes are orders of magnitude higher.

In addition, demographic stochasticity and historical changes in population size may be fundamental factors explaining the absence of a relationship between range size and mtDNA polymorphism in European butterflies. Nevertheless, due to the rarity of fossil data for butterflies, ranges and population sizes can only be calculated in the contemporary climatic conditions, which represent a warm interglacial period after a series of longer cold periods (Augustin et al., 2004). Given the rapidity with which butterflies can shift their distribution tracking suitable climatic conditions with recent climate change (Devictor et al., 2012), it can

be expected that most species now having restricted distribution on mountains were more widely distributed during cold periods of the Pleistocene, while many warm loving species were much more restricted to southern refugia, thus making contemporary population size uncorrelated with current mtDNA polymorphism (Nabholz et al., 2009). The fact that species now occurring in Mediterranean areas, which had likely been restricted to isolated southern refugia during the longer cold periods, showed a trend for higher mtDNA polymorphism could support this hypothesis. Finally, traits may not be fixed in time, but only their current states are known, and analyses using current traits may thus not fully reveal the historical relationships between mtDNA diversification and traits.

Prediction 2. Population differentiation and species traits

Although the genetic variation of mtDNA can be wiped out by selective sweeps or mixed among populations after dispersal events in highly mobile taxa, most species show strong genetic structuring among more or less isolated populations, strongly supporting the use of mtDNA as a marker in phylogeography (Avise, 2000, 2009). Indeed, in case such events lead to the elimination of genetic differentiation, a spatial structure can be re-established in a relatively short time by the emergence of new haplotypes and lineages or by population dynamics such as gene surfing (Waters, 2011). For example, a number of butterfly populations inhabiting Mediterranean islands that were connected to the mainland until the end of the last glacial period (about 15ka) show highly diverged mtDNA compared to populations inhabiting the mainland (Dapporto et al., 2017). This might be due to the recent occurrence of selective sweeps or shifts of lineages along the mainland (Mallet, 2010; Moritz et al., 2009), which have not yet reached insular populations due to sea barriers (Dapporto et al., 2017; Livraghi et al., 2018). These mechanisms could maintain spatially structured populations even if their overall degree of divergence strongly changes in time. In this case,

we can expect that populations separated by the same semi-permeable barriers (mountain chains and relatively narrow sea channels) do not show homogeneous divergence times. This is a recurrent finding in comparative phylogeography as found e.g. in butterflies of the Tuscan Archipelago, Sardinia and Corsica (Dapporto et al., 2017), in Neotropical birds (Burney & Brumfield, 2009) and in many Australian taxa (Moritz et al., 2009).

When the spatial information is added to genetic differentiation, correlations between mtDNA differentiation and species traits emerged. Dst and D are still largely dependent on overall diversification being measured in terms of percentage divergence (Nei, 1987; Whitlock, 2011). The variation in these indices was significantly explained by phenology traits, with species characterized by a longer flight period and a higher number of generations showing a lower population differentiation. This is a highly expected result for butterflies, since the winged adults are the dispersive stage. Accordingly, species showing a shorter flight period have lower possibilities to cross physical barriers (Dapporto & Dennis, 2009), resulting in higher probabilities to diverge in allopatry, with longer times required to attain secondary sympatry among lineages and in a slower propagation of genetic sweeps.

Contrary to Dst and D, Gst and G'st are pure numbers that can reach the maximum value of one even if a single mtDNA substitution is completely segregated among populations. Compared to D and Dst, Gst and G'st showed a significant correlation with a higher number of species traits. Smaller species, species exploiting less genera of food plants, and species with a larger range size showed a significantly higher COI genetic spatial structure – and species with a shorter flight period also showed a strong trend in the same direction. Range size is expected to correlate with genetic spatial structure since species with a larger range are expected to have more possibilities for divergence as barriers to gene flow can occur within their distribution and to thus comprise different lineages. In turn, wingspan is a well-known correlate of dispersal capability in butterflies (Dennis, Hardy, & Dapporto, 2012; Sekar,

2012). It has been found to correlate with genetic divergence and occupancy in different insular populations of European butterflies (Dennis et al., 2012). Similarly, exploiting a limited range of host plants (specialism) may mean that resources are not generally widespread and thus dispersal is limited, promoting population segregation (Dennis et al., 2012).

This is in line with results obtained in comparative studies on birds, where tendency to secondary sympatry was positively correlated with a characteristic of wing morphology determining dispersal capabilities (Pigot & Tobias, 2014) and genetic differentiations across south American barriers correlated with life history traits (Burney & Brumfield, 2009). It must be noted that although traits had a significant effect on Dst and Gst and most traits show a clear phylogenetic association, there is no phylogenetic signal in any index we measured, as was found in birds (Burney & Brumfield, 2009). This could be because erratic mutation rate and genetic sweeps can produce strong contrasts between closely related taxa even when the processes generating and maintaining genetic structure are facilitated by species traits (Allio et al., 2017; Nabholz et al., 2009).

Prediction 3. Zoogeographic region with intraspecific differentiation

To the best of our knowledge, we here present the first zoogeographic regionalization at the sub-continental level based on intraspecific genetic diversification. Zoogeographic regions have been assessed so far by comparing faunistic communities, i.e. they were based on species distribution/occurrence differences (Holt et al., 2013). However, it is expected that a zoogeographic assessment based on averaging intraspecific population differentiation among hundreds of species should i) reflect the main physical barriers and paleogeographic history of the study area; and ii) correlate with a zoogeographic regionalization based on species

composition, because the barriers separating species distributions are also expected to limit gene flow.

During the long Quaternary cold periods, most of the current European species were likely limited to the southern regions of the continent, represented by three peninsulas (Iberia, Italy, Balkans) and by several Mediterranean islands (Hewitt, 2000). Accordingly, the existence of different lineages of butterflies in these areas with narrow suture zones at the physical barriers among them (Alps, Pyrenees and sea channels) represents a pervasive pattern in European phylogeography (Bowen et al., 2016; Hewitt, 2000; Schmitt, 2007). According to this paradigm, we found evidence for six different zoogeographic regions that largely agree with the hypothesis of divergence in different Pleistocene refugia.

A similar regionalization was obtained by comparing butterfly communities in the same area (Dapporto, Fattorini, Vodă, Dincă, & Vila, 2014). The main difference between the assessment at community and intraspecific differentiation level refers to Sardinia, Corsica and the Balearic islands. These differences are rooted in the fundamental differences of the two assessments. In the analysis at the community level, the considerable number of endemic species from Sardinia and Corsica determined a high contrast and resulted in a highly distinct group; the Balearics without any endemic butterfly species, appeared very similar to Iberia (Dapporto et al., 2014). In the assessment of genetic diversity, the Sardo-Corsican endemics can only generate contrasts between Sardinia and Corsica (and Tuscan Islands) when they show distinct populations between these areas, as is often the case (Dapporto et al., 2014). Most of the pattern obtained with intraspecific genetic variation is instead encompassed by widespread species responsible for determining similarity/dissimilarity patterns among islands and mainland. Previous studies on butterflies showed that in many cases Sardinia and Corsica differ in their populations: those populations from Corsica and Tuscan Islands mostly resemble those occurring in the Alps and the Pyrenees, whilst populations from Sardinia are

often similar to those occurring in Iberia and the Balearics (e.g. *Callophrys rubi*, *Maniola jurtina*, *Pararge aegeria*, *Coenonympha pamphilus*) (Dapporto et al., 2017; Dincă et al., 2015; Livraghi et al., 2018). The analysis of intraspecific genetic divergence captured this main pattern, the determinants of which are still largely unexplained. The existence of distinct lineages from Iberia, Sicily and the Italian Peninsula is a very common pattern in butterfly phylogeography, while the diversification in the regions of the Alps and the Pyrenees is mainly determined by a different admixture of lineages (Dincă et al., 2018, 2015; Hernández-Roldán et al., 2016; Schmitt, 2007).

Conclusion

Among the challenges imposed by the current and accelerating biodiversity loss (Dirzo et al., 2014), understanding the dynamics of biodiversity is critical for predicting future scenarios and undertaking effective conservation measures (Hoffmann et al., 2015; Joly et al., 2014; Pacifici et al., 2015). This endeavour requires cheap, fast and reliable approaches to map and track changes of biological diversity over spatial and temporal scales, as well as linking them with species functional traits (Joly et al., 2014; Kress et al., 2015; Pacifici et al., 2015; Stein, Martinez, Stiles, Miller, & Zakharov, 2014). In the last decades, mitochondrial DNA has become increasingly prominent in biodiversity research, notably for phylogenetics, phylogeography and in the study of divergence processes (Avise, 2009; Burney & Brumfield, 2009; Cameron, 2014; Dincă et al., 2015; Galtier et al., 2009; Hernández-Roldán et al., 2016; Joly et al., 2014; Kress et al., 2015).

Nevertheless, the validity of many of the claimed advantages of using mtDNA as a marker in molecular ecology has been questioned in the last decade (Galtier et al., 2009; Stein et al., 2014). In fact, the assumption of neutrality of mtDNA has been weak and, in addition,

mtDNA variation is not necessarily representative of genomic variation, as it is subjected to different determinants and inheritance mechanisms. We provide evidence that mtDNA spatial differentiation has a deterministic fingerprint, being correlated with species traits known to determine the dispersal capability and colonization ability of butterflies (Dennis et al., 2012). Thus, we argue that mtDNA should be still considered as a fundamental marker for the understanding of ecological and evolutionary processes (such as demographic changes and dispersal patterns) that affect the historical and contemporary spatial ecology of species (or at least of their female populations). Moreover, the fall of the neutrality assumption for mtDNA implies the importance of mtDNA variation in influencing functional traits. Recent evidence proved that these mitochondrial-derived traits are involved, among others, in local adaptation to climatic conditions (Toews et al., 2014). Under this perspective, the strong variation in mtDNA sequences among populations and the relatively fast shifts in their distributions demonstrated by direct and indirect evidence, indicate that spatial mtDNA variation represents a source of differential adaptive optima, which can sweep across populations and preserve species in a rapidly changing environment (de Lafontaine, Napier, Petit, & Hu, 2018). The macroscopic consequence is the preservation of species diversity and of the related ecosystem functioning. In this vein, information about mtDNA variation has been recommended to be taken into consideration in conservation plans and reintroductions of butterflies (Dincă et al., 2018). The results and the extensive datasets provided here can constitute a basis to produce genetically informed conservation plans for a highly charismatic group in a continent where flying insects have been proven to be under incessant decline (Hallmann et al., 2017).

Acknowledgements

Funding for this research came from the Spanish MINECO and AEI/FEDER, UE (CGL2013-48277-P and CGL2016-76322-P to RVi), Marie Skłodowska-Curie Train2Move to RVo (grant 609402-2020), Marie Sklodowska-Curie IOF grant (project 625997) to VD, Marie Sklodowska-Curie Action, (grant 706208 SocParPhenoEvol) to AC and from the project "Barcoding-Italian-Butterflies". We thank many colleagues that collaborated in field collections and those who have generated the genetic data mined from BOLD and GenBank.

References

Allio, R., Donega, S., Galtier, N., & Nabholz, B. (2017). Large variation in the ratio of mitochondrial to nuclear mutation rate across animals: implications for genetic diversity and the use of mitochondrial DNA as a molecular marker. *Molecular Biology and Evolution*, *34*(11), 2762–2772.

Augustin, L., Barbante, C., Barnes, P. R., Barnola, J. M., Bigler, M., Castellano, E., … Delmonte, B. (2004). Eight glacial cycles from an Antarctic ice core. *Nature*, *429*, 623–628.

Avise, J. C. (2000). *Phylogeography: the history and formation of species*. Harvard university press.

Avise, J. C. (2009). Phylogeography: retrospect and prospect. *Journal of Biogeography*, *36*(1), 3–15.

Avise, J. C., Arnold, J., Ball, R. M., Bermingham, E., Lamb, T., Neigel, J. E., … Saunders, N. C. (1987). Intraspecific phylogeography: the mitochondrial DNA bridge between

population genetics and systematics. *Annual Review of Ecology and Systematics*, *18*(1), 489–522.

Bazin, E., Glémin, S., & Galtier, N. (2006). Population size does not influence mitochondrial genetic diversity in animals. *Science*, *312*(5773), 570–572.

Bowen, B. W., Gaither, M. R., DiBattista, J. D., Iacchei, M., Andrews, K. R., Grant, W. S., … Briggs, J. C. (2016). Comparative phylogeography of the ocean planet. *Proceedings of the National Academy of Sciences*, *113*(29), 7962–7969.

Brändle, M., Öhlschläger, S., & Brandl, R. (2002). Range sizes in butterflies: correlation across scales. *Evolutionary Ecology Research*, *4*(7), 993–1004.

Burney, C. W., & Brumfield, R. T. (2009). Ecology predicts levels of genetic differentiation in Neotropical birds. *The American Naturalist*, *174*(3), 358–368.

Burnham, K., & Anderson, D. (2002). Model selection and multimodel inference New York. *NY: Springer*.

Cameron, S. L. (2014). Insect mitochondrial genomics: implications for evolution and phylogeny. *Annual Review of Entomology*, *59*, 95–117.

Coyne, J. A., & Orr, H. A. (2004). *Speciation*. Sunderland, Mass: Sinauer Associates.

Dapporto, L., Bruschini, C., Dincă, V., Vila, R., & Dennis, R. L. H. (2012). Identifying zones of phenetic compression in West Mediterranean butterflies (Satyrinae): refugia, invasion and hybridization: Phenetic ‘compression’ in Western Mediterranean. *Diversity and Distributions*, *18*(11), 1066–1076. doi: 10.1111/j.1472-4642.2012.00903.x

Dapporto, L., Cini, A., Menchetti, M., Vodă, R., Bonelli, S., Casacci, L. P., … Vila, R. (2017). Rise and fall of island butterfly diversity: Understanding genetic differentiation and extinction in a highly diverse archipelago. *Diversity and Distributions*, *23*(10), 1169–1181. doi: 10.1111/ddi.12610

Dapporto, L., & Dennis, R. L. H. (2009). Conservation biogeography of large Mediterranean islands. Butterfly impoverishment, conservation priorities and inferences for an ecologica "island paradigm." *Ecography*, *32*(1), 169–179. doi: 10.1111/j.1600-0587.2008.05600.x

Dapporto, L., Fattorini, S., Vodă, R., Dincă, V., & Vila, R. (2014). Biogeography of western Mediterranean butterflies: combining turnover and nestedness components of faunal dissimilarity. *Journal of Biogeography*, *41*(9), 1639–1650.

Dapporto, L., Ramazzotti, M., Fattorini, S., Talavera, G., Vila, R., & Dennis, R. L. (2013). recluster: an unbiased clustering procedure for beta-diversity turnover. *Ecography*, *36*(10), 1070–1075.

Dawson, M. N. (2014). Natural experiments and meta-analyses in comparative phylogeography. *Journal of Biogeography*, *41*(1), 52–65.

de Lafontaine, G., Napier, J. D., Petit, R. J., & Hu, F. S. (2018). Invoking adaptation to decipher the genetic legacy of past climate change. *Ecology*.

Dennis, R. L., Hardy, P. B., & Dapporto, L. (2012). Nestedness in island faunas: novel insights into island biogeography through butterfly community profiles of colonization ability and migration capacity. *Journal of Biogeography*, *39*(8), 1412–1426.

Devictor, V., van Swaay, C., Brereton, T., Brotons, L., Chamberlain, D., Heliölä, J., … Jiguet, F. (2012). Differences in the climatic debts of birds and butterflies at a continental scale. *Nature Climate Change*, *2*(2), 121–124. doi: 10.1038/nclimate1347

Dincă, V., Bálint, Z., Vodă, R., Dapporto, L., Hebert, P. D., & Vila, R. (2018). Use of genetic, climatic, and microbiological data to inform reintroduction of a regionally extinct butterfly. *Conservation Biology*.

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

Dincă, V., Zakharov, E. V., Hebert, P. D. N., & Vila, R. (2011). Complete DNA barcode reference library for a country's butterfly fauna reveals high performance for temperate Europe. *Proceedings of the Royal Society B: Biological Sciences*, *278*(1704), 347–355. doi: 10.1098/rspb.2010.1089

Dirzo, R., Young, H. S., Galetti, M., Ceballos, G., Isaac, N. J., & Collen, B. (2014). Defaunation in the Anthropocene. *Science*, *345*(6195), 401–406.

Espeland, M., Breinholt, J., Willmott, K. R., Warren, A. D., Vila, R., Toussaint, E. F., … Eastwood, R. (2018). A comprehensive and dated phylogenomic analysis of butterflies. *Current Biology*, *28*(5), 770–778.

Fujisawa, T., Vogler, A. P., & Barraclough, T. G. (2015). Ecology has contrasting effects on genetic variation within species versus rates of molecular evolution across species in water beetles. *Proc. R. Soc. B*, *282*(1799), 20142476.

Galtier, N., Nabholz, B., GléMin, S., & Hurst, G. D. D. (2009). Mitochondrial DNA as a marker of molecular diversity: a reappraisal. *Molecular Ecology*, *18*(22), 4541–4550. doi: 10.1111/j.1365-294X.2009.04380.x

Gillespie, J. H. (2000). Genetic drift in an infinite population: the pseudohitchhiking model. *Genetics*, *155*(2), 909–919.

Gillespie, J. H. (2001). Is the population size of a species relevant to its evolution? *Evolution*, *55*(11), 2161–2169.

Hall, T. A. (1999). *BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT*. *41*, 95–98. [London]: Information Retrieval Ltd., c1979-c2000.

Hallmann, C. A., Sorg, M., Jongejans, E., Siepel, H., Hofland, N., Schwan, H., … Hörren, T. (2017). More than 75 percent decline over 27 years in total flying insect biomass in protected areas. *PloS One*, *12*(10), e0185809.

Hausmann, A., Haszprunar, G., Segerer, A. H., Speidel, W., Behounek, G., & Hebert, P. D. (2011). Now DNA-barcoded: the butterflies and larger moths of Germany. *Spixiana*, *34*(1), 47–58.

Hebert, P. D. N., Cywinska, A., Ball, S. L., & deWaard, J. R. (2003). Biological identifications through DNA barcodes. *Proceedings of the Royal Society B: Biological Sciences*, *270*(1512), 313–321. doi: 10.1098/rspb.2002.2218

Hernández-Roldán, J. L., Dapporto, L., Dincă, V., Vicente, J. C., Hornett, E. A., Šíchová, J., … Vila, R. (2016). Integrative analyses unveil speciation linked to host plant shift in *Spialia* butterflies. *Molecular Ecology*, *25*(17), 4267–4284. doi: 10.1111/mec.13756

Hewitt, G. (2000). The genetic legacy of the Quaternary ice ages. *Nature*, *405*(6789), 907.

Hewitt, G. M. (2004). Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *359*(1442), 183–195.

Higgins, L., & Riley, N. D. (1970). *Field Guide to the Butterflies of Britain and Europe*. London: Collins.

Hoffmann, A., Griffin, P., Dillon, S., Catullo, R., Rane, R., Byrne, M., … Joseph, L. (2015). A framework for incorporating evolutionary genomics into biodiversity conservation and management. *Climate Change Responses*, *2*(1), 1.

Holt, B. G., Lessard, J.-P., Borregaard, M. K., Fritz, S. A., Araújo, M. B., Dimitrov, D., … Jønsson, K. A. (2013). An update of Wallace's zoogeographic regions of the world. *Science*, *339*(6115), 74–78.

Huemer, P., Mutanen, M., Sefc, K. M., & Hebert, P. D. (2014). Testing DNA barcode performance in 1000 species of European Lepidoptera: large geographic distances have small genetic impacts. *PLoS One*, *9*(12), e115774.

Hughes, A. R., Inouye, B. D., Johnson, M. T., Underwood, N., & Vellend, M. (2008). Ecological consequences of genetic diversity. *Ecology Letters*, *11*(6), 609–623.

Joly, S., Davies, T. J., Archambault, A., Bruneau, A., Derry, A., Kembel, S. W., … Wheeler, T. A. (2014). Ecology in the age of DNA barcoding: the resource, the promise and the challenges ahead. *Molecular Ecology Resources*, *14*(2), 221–232.

Jost, L. (2008). GST and its relatives do not measure differentiation. *Molecular Ecology*, *17*(18), 4015–4026.

Karl, S. A., Toonen, R., Grant, W., & Bowen, B. (2012). Common misconceptions in molecular ecology: echoes of the modern synthesis. *Molecular Ecology*, *21*(17), 4171–4189.

Kress, W. J., García-Robledo, C., Uriarte, M., & Erickson, D. L. (2015). DNA barcodes for ecology, evolution, and conservation. *Trends in Ecology & Evolution*, *30*(1), 25–35. doi: 10.1016/j.tree.2014.10.008

Leffler, E. M., Bullaughey, K., Matute, D. R., Meyer, W. K., Segurel, L., Venkat, A., … Przeworski, M. (2012). Revisiting an old riddle: what determines genetic diversity levels within species? *PLoS Biology*, *10*(9), e1001388.

Lewontin, R. C. (1974). *The genetic basis of evolutionary change* (Vol. 560). Columbia University Press New York.

Livraghi, L., Vodă, R., Evans, L. C., Gibbs, M., Dincă, V., Holland, P. W., … Breuker, C. J. (2018). Historical and current patterns of gene flow in the butterfly Pararge aegeria. *Journal of Biogeography*.

Mallet, J. (2010). Shift happens! Shifting balance and the evolution of diversity in warning colour and mimicry. *Ecological Entomology*, *35*(s1), 90–104.

Meirmans, P. G., & Hedrick, P. W. (2011). Assessing population structure: FST and related measures: Invited Technical Reviwe. *Molecular Ecology Resources*, *11*(1), 5–18. doi: 10.1111/j.1755-0998.2010.02927.x

Middleton-Welling, J., Wade, R. A., Dennis, R. L. H., Dapporto, L., & Shreeve, T. G. (2018). Optimising trait and source selection for explaining occurrence and abundance changes: A case study using British butterflies. *Functional Ecology*. doi: 10.1111/1365-2435.13082

Moretti, M., Dias, A. T. C., de Bello, F., Altermatt, F., Chown, S. L., Azcárate, F. M., … Berg, M. P. (2017). Handbook of protocols for standardized measurement of terrestrial invertebrate functional traits. *Functional Ecology*, *31*(3), 558–567. doi: 10.1111/1365-2435.12776

Moritz, C., Hoskin, C. J., MacKenzie, J. B., Phillips, B. L., Tonione, M., Silva, N., … Graham, C. H. (2009). Identification and dynamics of a cryptic suture zone in tropical rainforest. *Proceedings of the Royal Society of London B: Biological Sciences*, rspb. 2008.1622.

Nabholz, B., Glémin, S., & Galtier, N. (2009). The erratic mitochondrial clock: variations of mutation rate, not population size, affect mtDNA diversity across birds and mammals. *BMC Evolutionary Biology*, *9*(1), 54.

Nabholz, B., Mauffrey, J.-F., Bazin, E., Galtier, N., & Glemin, S. (2008). Determination of mitochondrial genetic diversity in mammals. *Genetics*, *178*(1), 351–361.

Nei, M. (1987). *Molecular evolutionary genetics*. Columbia University Press.

Pacifici, M., Foden, W. B., Visconti, P., Watson, J. E., Butchart, S. H., Kovacs, K. M., … Akcakaya, H. R. (2015). Assessing species vulnerability to climate change. *Nature Climate Change*, *5*(3), 215.

Papadopoulou, A., & Knowles, L. L. (2016). Toward a paradigm shift in comparative phylogeography driven by trait-based hypotheses. *Proceedings of the National Academy of Sciences*, *113*(29), 8018–8024.

Pentinsaari, M., Salmela, H., Mutanen, M., & Roslin, T. (2016). Molecular evolution of a widely-adopted taxonomic marker (COI) across the animal tree of life. *Scientific Reports*, *6*, 35275.

Pigot, A. L., & Tobias, J. A. (2014). Dispersal and the transition to sympatry in vertebrates. *Proceedings of the Royal Society B: Biological Sciences*, *282*(1799), 20141929–20141929. doi: 10.1098/rspb.2014.1929

Ratnasingham, S., & Hebert, P. D. (2007). BOLD: The Barcode of Life Data System (http://www. barcodinglife. org). *Molecular Ecology Resources*, *7*(3), 355–364.

Ritter, S., Michalski, S. G., Settele, J., Wiemers, M., Fric, Z. F., Sielezniew, M., … Durka, W. (2013). Wolbachia infections mimic cryptic speciation in two parasitic butterfly species, Phengaris teleius and P. nausithous (Lepidoptera: Lycaenidae). *PLoS One*, *8*(11), e78107.

Romiguier, J., Gayral, P., Ballenghien, M., Bernard, A., Cahais, V., Chenuil, A., … Faivre, N. (2014). Comparative population genomics in animals uncovers the determinants of genetic diversity. *Nature*, *515*(7526), 261.

Schmitt, T. (2007). Molecular biogeography of Europe: Pleistocene cycles and postglacial trends. *Frontiers in Zoology*, *4*(1), 11.

Schweiger, O., Harpke, A., Wiemers, M., & Settele, J. (2014). CLIMBER: Climatic niche characteristics of the butterflies in Europe. *ZooKeys*, *367*, 65–84. doi: 10.3897/zookeys.367.6185

Sekar, S. (2012). A meta-analysis of the traits affecting dispersal ability in butterflies: can wingspan be used as a proxy?: Factors affecting dispersal ability in butterflies. *Journal of Animal Ecology*, *81*(1), 174–184. doi: 10.1111/j.1365-2656.2011.01909.x

Smith, M. A., Bertrand, C., Crosby, K., Eveleigh, E. S., Fernandez-Triana, J., Fisher, B. L., … Hind, K. (2012). Wolbachia and DNA barcoding insects: patterns, potential, and problems. *PloS One*, *7*(5), e36514.

Stein, E. D., Martinez, M. C., Stiles, S., Miller, P. E., & Zakharov, E. V. (2014). Is DNA barcoding actually cheaper and faster than traditional morphological methods: results from a survey of freshwater bioassessment efforts in the United States? *PloS One*, *9*(4), e95525.

Toews, D. P., Mandic, M., Richards, J. G., & Irwin, D. E. (2014). Migration, mitochondria, and the yellow-rumped warbler. *Evolution*, *68*(1), 241–255.

Tolman, T., & Lewington, R. (2008). *Butterflies of Europe*. Harper Collins.

Vodă, R., Dapporto, L., Dincă, V., & Vila, R. (2015). Why Do Cryptic Species Tend Not to Co-Occur? A Case Study on Two Cryptic Pairs of Butterflies. *PLOS ONE*, *10*(2), e0117802. doi: 10.1371/journal.pone.0117802

Waters, J. M. (2011). Competitive exclusion: phylogeography's 'elephant in the room'?: NEWS AND VIEWS: OPINION. *Molecular Ecology*, *20*(21), 4388–4394. doi: 10.1111/j.1365-294X.2011.05286.x

Welch, J. J., Bininda-Emonds, O. R., & Bromham, L. (2008). Correlates of substitution rate variation in mammalian protein-coding sequences. *BMC Evolutionary Biology*, *8*(1), 53.

Werren, J. H., Baldo, L., & Clark, M. E. (2008). Wolbachia: master manipulators of invertebrate biology. *Nature Reviews Microbiology*, *6*(10), 741.

Whitlock, M. C. (2011). and D do not replace FST. *Molecular Ecology*, *20*(6), 1083–1091.

Wiemers, M., Balletto, E., Dincă, V., Fric, Z. F., Lamas, G., Lukhtanov, V., … Verovnik, R. (2018). An updated checklist of the European Butterflies (Lepidoptera, Papilionoidea). *ZooKeys*, (811), 9.

Author contributions

LD designed the paper framework; LD, AC, RVo, VD, MM, LPC, SS, TS, RVi, collected the specimens, identified them at species level and carried out DNA sequencing and mined data from DNA repositories (BOLD and GenBank); LD, AC, GM, MM, EB, SB, LPC, gathered the trait dataset; MW gathered nuclear and mitochondrial markers and built the phylogenetic tree; LD carried out comparative phylogeography analyses; all authors participated in interpreting the results, in writing and editing the manuscript.

Data Accessibility

- DNA sequences and specimen information: The DNA sequences, the information about collection data, taxonomy and GenBank and BOLD accession codes are available in the BOLD project DS-WEUP and from Dryad: doi: 10.5061/dryad.2q76p8f.

Butterfly traits are on Dryad: doi: 10.5061/dryad.2q76p8f

Phylogenetic tree: available on Dryad doi: 10.5061/dryad.2q76p8f

R scripts and file used to carry out the analyses available on Dryad: doi: 10.5061/dryad.2q76p8f

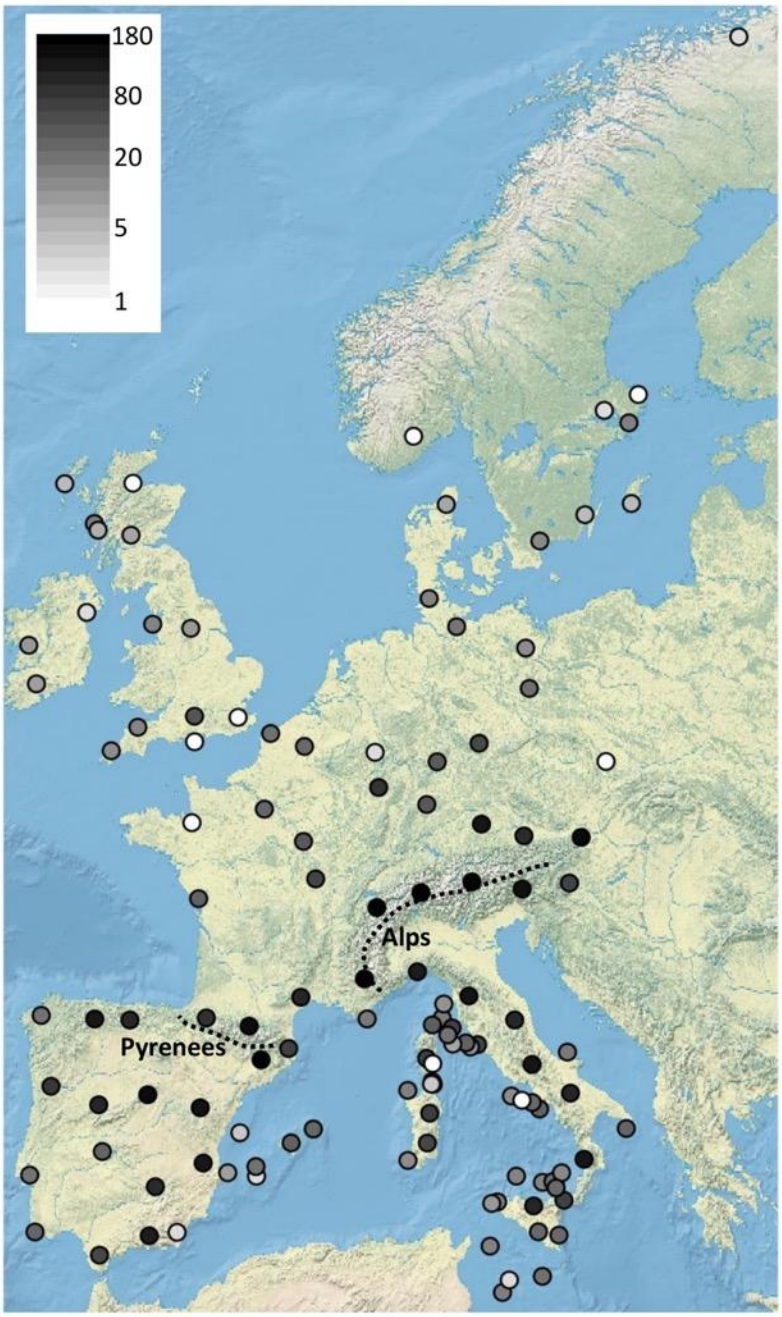

Figure 1. Map of the study area and areas used for the assessment of spatial genetic variation. Dots represent barycenters of sequences collapsed to squares of 2.5 degrees of latitude and longitude, and to small islands, which are treated independently. The colour of the dots is proportional to the log of the number of species analysed in each area (see legend). The mountain chains (Alps and Pyrenees) separating the two main southern peninsulas (Iberian and Italian Peninsulas) are highlighted.

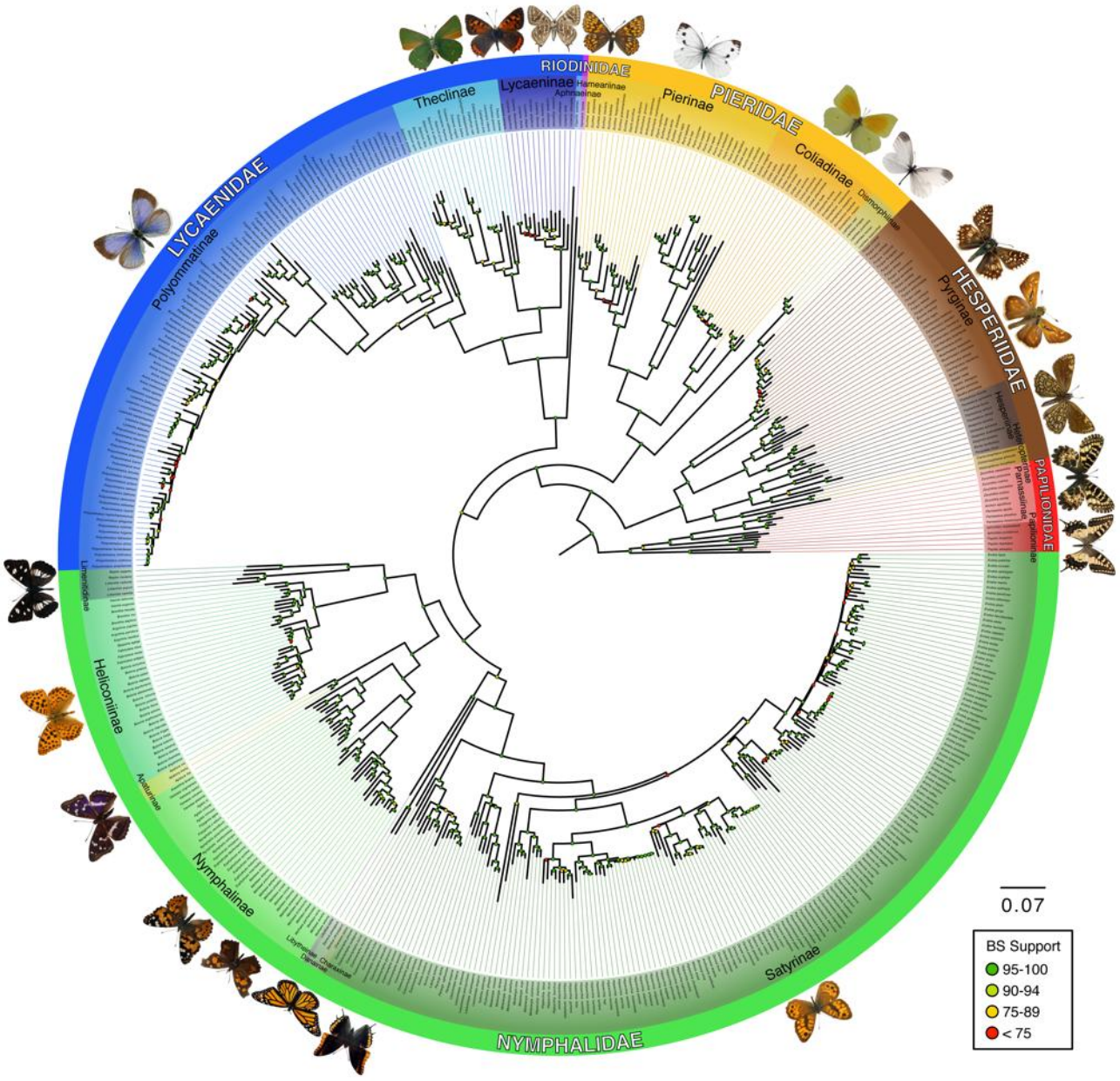

Figure 2. The multi-locus phylogenetic tree for all European butterfly species. Families and subfamilies are indicated, as well as supports for nodes (BS support, see legend).

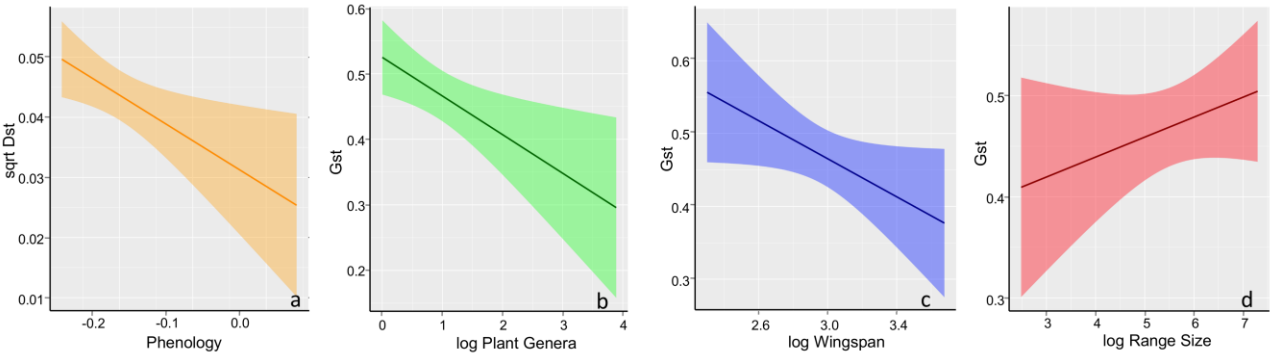

Figure 3. The significant univariate linear relationships between (a) square root-transformed Dst and phenology, (b) Gst and log-transformed number of plant genera used by species, (c) Gst and wingspan, and (d) Gst and range size. Shaded areas represent 95% confidence regions.

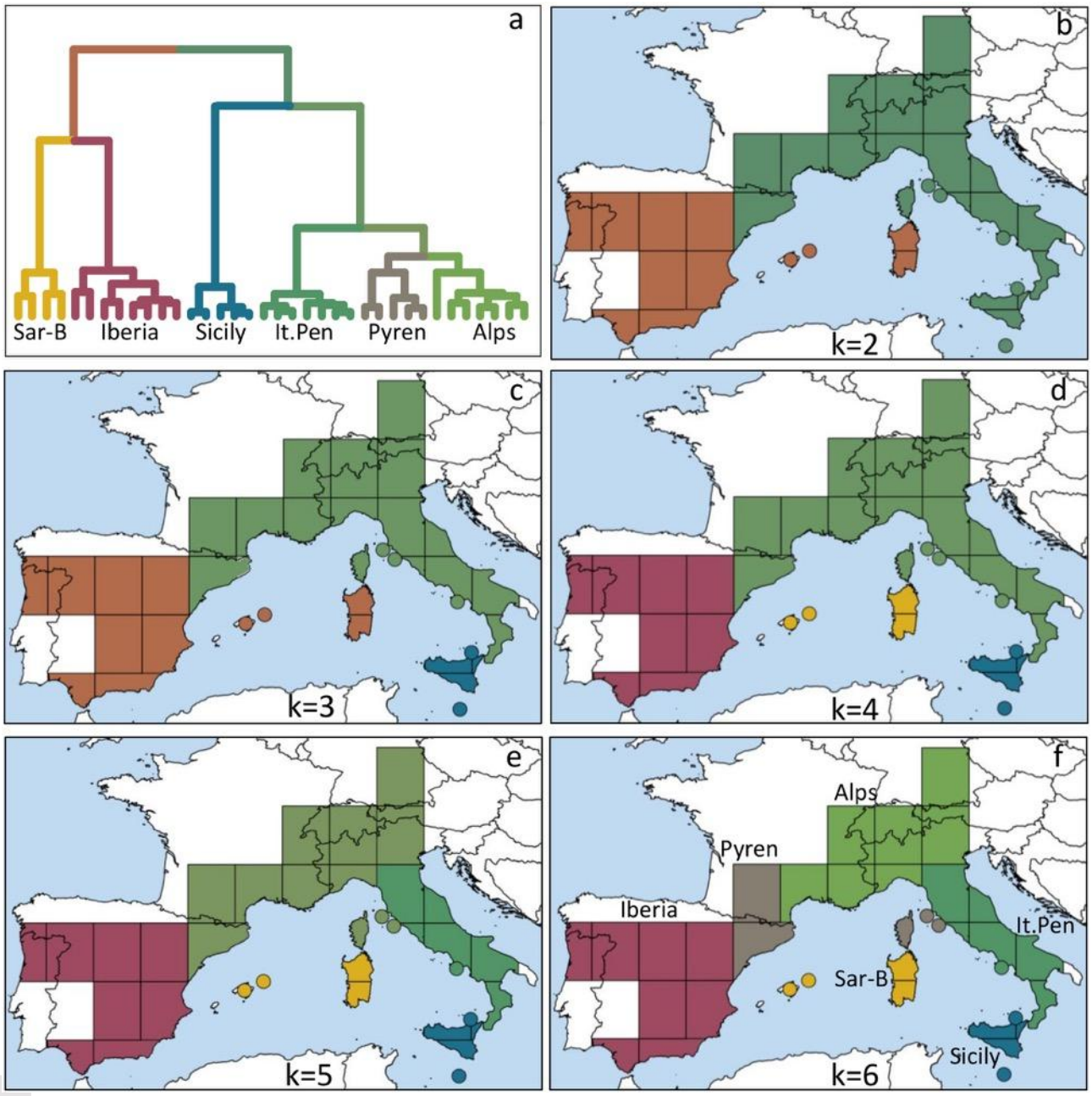

Figure 4. The tree obtained by applying the Ward algorithm to the average Gst distances based on 226 species in the North-Western Mediterranean among a series of 36 areas (a). The tree is cut at different nodes to obtain five solutions from 2 to 6 clusters (b-f). For each solution, the tree branches are represented by using the colours obtained by projecting the bidimensional representation of the original dissimilarity matrix in RGB space and then by calculating the barycentres of the dots belonging to each subtree. The same colours are used in the maps to visualize the different zoogeographic regions. The regions identified by the solution for k=6 are reported in the tree and in figure 3f; Iberia, Iberian Peninsula; Sar-B, Sardinia and Balearics; Sicily, Sicily and circum-Sicilian islands (Malta and Vulcano); It.Pen, Italian Peninsula and Capri; Pyren; Pyrenees, Corsica and surrounding islands (Elba and Giglio); Alps, Alps.

Table 1. Phylogenetic signal for the three indexes of COI genetic diversification and for the examined species traits. Variables highlighted in bold showed a significant effect.

	Pagel´s lambda	P
Gst	<0.0001	1.000
G'st	<0.0001	1.000
Dst	<0.0001	1.000
D	<0.0001	1.000
Nucleotide Diversity	0.021	0.550
PC1 phenology	**0.589**	**<0.001**
Range size	**0.336**	**<0.001**
PC2 ecophysiology	**0.183**	**<0.001**
Max Altitude	<0.0001	1.000
Host plants	**0.362**	**<0.001**
PC1 ecophysiology	0.011	0.710
Wing size	**0.988**	**<0.001**

Table 2. Conditional average results among the selected models in a phylogenetic regression comparing haplotype diversity with butterfly traits. Prediction 1. Variables highlighted in bold showed a significant effect.

HD	Estimate	Std.Error	Z	P	models(29)
PC1 ecophysiology	0.175	0.092	1.894	0.058	21
Phenology	-0.149	0.085	1.758	0.079	15
Max Altitude	0.147	0.084	1.746	0.081	15
Range size	0.114	0.084	1.360	0.174	9
PC2 ecophysiology	0.042	0.084	0.505	0.613	8
Wing size	0.029	0.068	0.428	0.669	6
Host plants	0.026	0.072	0.357	0.721	4

Table 3. Conditional average results among the selected models in a phylogenetic regression comparing Dst diversity with butterfly traits. Prediction 2. Variables highlighted in bold showed a significant effect.

Dst	Estimate	Std.Error	Z	P	models(45)
Phenology	**-0.206**	**0.081**	**2.547**	**0.011**	**44**
Range size	0.132	0.084	1.571	0.116	23
PC1 ecophysiology	0.122	0.084	1.460	0.144	23
Host plants	-0.076	0.072	1.047	0.295	19
Max Altitude	0.094	0.084	1.125	0.261	19
Wing size	-0.033	0.068	0.481	0.631	15
PC2 ecophysiology	0.016	0.095	0.167	0.868	17

Table 4. Conditional average results among the selected models in a phylogenetic regression comparing Gst with butterfly traits. Prediction 2. Variables highlighted in bold showed a significant effect.

Gst	Estimate	Std.Error	Z	P	Models(17)
Host plants	**-0.194**	**0.071**	**2.742**	**0.006**	**17**
Range size	**0.195**	**0.083**	**2.341**	**0.019**	**17**
Wing size	**-0.135**	**0.067**	**2.025**	**0.043**	**12**
PC1 Phenology	-0.141	0.074	1.922	0.055	11
Max Altitude	-0.076	0.070	1.093	0.274	7
PC2 ecophysiology	-0.061	0.093	0.655	0.512	6
PC1 ecophysiology	0.021	0.080	0.261	0.794	5