How long is 3 kilometres for a butterfly? Ecological constraints and functional traits explain high mitochondrial genetic diversity between Sicily and the Italian Peninsula

Running title: Genetic differentiation across a sea strait

Stefano Scalercio[1*], Alessandro Cini[2, 3*], Mattia Menchetti[2*], Raluca Vodă[4*], Simona Bonelli[4], Adele Bordoni[2], Luca Pietro Casacci[4,5], Vlad Dincă[6], Emilio Balletto[4], Roger Vila[7], Leonardo Dapporto[2]**

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

[2]Dipartimento di Biologia, Università degli Studi di Firenze, Via Madonna del Piano 6, 50109, Sesto Fiorentino, Florence, Italy.

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

[4]Dipartimento di Scienze della Vita e Biologia dei Sistemi, Università degli Studi di Torino, Via Accademia Albertina 13, 10123, Turin, Italy

[5]Muzeum i Instytut Zoologii, Polskiej Akademii Nauk, ul. Wilcza 64, 00-679, Warsaw, Poland

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

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

* These authors equally contributed to the paper

** Corresponding author: Leonardo Dapporto leondap@gmail.com

Abstract

1. Populations inhabiting Mediterranean islands often show contrasting genetic lineages, even on islands that were connected to the mainland during glacial maxima. This pattern is generated by forces acting in historical and contemporary times. Understanding these phenomena requires comparative studies relating genetic structure, functional traits and dispersal constraints.

2. Using as a model the butterfly species living across the Messina strait separating Sicily from the Italian Peninsula (3 km wide), we aimed to unravel the mechanisms limiting the dispersal of matrilines across a narrow sea strait and producing genetic differentiation.

3. We analysed the mitochondrial COI gene of 84 butterfly species out of 90 documented in Sicily and compared them with populations from the neighbouring southern Italian Peninsula (1398 sequences) and from the entire Palearctic region (8093 sequences). For each species, we regressed 13 functional traits and two ecological constraints to dispersal (winds experienced at the strait and climatic suitability) against genetic differentiation between Sicily and Italian Peninsula to understand the factors limiting dispersal.

4. More than a third of the species showed different haplogroups across the strait and most of them also represented endemic haplogroups for this island. One fifth of Sicilian populations (and 32.3% of endemic lineages) had their closest relatives in distant areas, instead of the neighbouring Italian Peninsula, which suggests high relictuality. Haplotype diversity was significantly explained by length of the flight period, an intrinsic phenology trait, while genetic differentiation was explained by both intrinsic traits (wingspan and degree of generalism) and contemporary local constraints (winds experienced at the strait and climatic suitability).

5. A relatively narrow sea strait can produce considerable differentiation among butterfly matrilines and this phenomenon showed a largely deterministic fingerprint. Because of

unfavourable winds, populations of the less dispersive Sicilian butterflies tended to differentiate into endemic variants or to maintain relict populations. Understanding these phenomena required the integration of DNA sequences, species traits and physical constraints for a large taxon at continental scale. Future studies may reveal if the patterns here shown for mitochondrial DNA are also reflected in the nuclear genome or, alternatively, are the product of limited female dispersal.

Keywords: butterflies, COI differentiation, environmental constraints, dispersal, functional traits, Messina strait

Introduction

There is extensive evidence that most genetic divergence and speciation events arise between isolated populations (Coyne & Orr, 2004). Unsurprisingly, islands have become valuable models to study the evolutionary forces leading to the emergence of species and genetic lineages (Whittaker, Fernández-Palacios, Matthews, Borregaard, & Triantis, 2017). The degree of isolation, the age of insular systems and the dispersal capabilities of organisms are the main factors determining the characteristics of island populations (Whittaker & Fernández-Palacios, 2007). As a rule, the spectacular adaptive radiations are typical of low dispersive taxa living on highly isolated oceanic islands, while the large and less isolated continental islands usually host subsets of mainland populations and relictual elements (Whittaker & Fernández-Palacios, 2007; Whittaker, Triantis, & Ladle, 2008).

Contrasting with this model, in the Mediterranean region, the highly mobile butterflies often show faunistic and genetic contrasts even across narrow sea straits. There are several documented cases for the Gibraltar strait (14.3 km wide, Husemann, Schmitt, Zachos, Ulrich, & Habel, 2014; Weingartner, Wahlberg, & Nylin, 2006), for the strait of Bonifacio between Sardinia and Corsica (11 km wide, Dapporto et al., 2017; Vodă, Dapporto, Dincă, & Vila, 2015b) and for the sea straits separating Sicily (3 km wide) and the smaller circum-Italian islands from the Italian mainland (Cesaroni, Lucarelli, Allori, Russo, & Sbordoni, 1994; Dapporto et al., 2017; Vodă et al., 2016, 2015b). The mechanisms limiting gene flow across relatively weak barriers compared to the dispersal capability of a given organism are still largely unknown and several contemporary factors (e.g., climatic and environmental preferences, reproductive interference, dispersal limitations and/or competitive exclusion) limit the admixture of lineages and cryptic species (Moritz et al., 2009; Pigot & Tobias, 2013, 2014; Vodă, Dapporto, Dincă, & Vila, 2015a; Waters, Fraser, & Hewitt, 2013). There is also evidence that the distribution of butterfly lineages in the Mediterranean and, consequently, their contact zones can change in relatively short times. While direct evidence has been provided by the analysis of historical samples in hawk moths (Mende & Hundsdoerfer, 2013), indirect examples are represented by islands harboring different lineages and morphotypes compared to the nearest areas to which they were connected during the last glacial maximum (LGM). This has been clearly documented for the straits separating Sardinia, Corsica, Tuscan islands and Italian mainland (Dapporto et al., 2017; Vodă et al., 2015a, 2015b). If contact zones between lineages can shift in relatively short and recent times, it is plausible that current ecological impediments to dispersal, besides historical factors, explain the observed genetic differentiation (Mallet, 2010; Moritz et al., 2009).

Phylogeographic exploration plays a crucial role in understanding the emergence of genetic differentiation and the establishment of contact zones (Avise, 2000; Marske, Rahbek, & Nogués-Bravo, 2013; Waters et al., 2013). However, phylogeographic studies are rarely projected into unified ecological and evolutionary frameworks and many examine only a limited number of species (Marske et al., 2013). Recently, a shift toward multitaxon comparative phylogeography relating genetic variation with environmental features and functional traits has been recommended (Papadopoulou & Knowles, 2016). This approach may disentangle the role played by deterministic processes with respect to stochasticity in structuring the spatial genetic variation and shed light on the historical/contemporary and biotic/abiotic processes that enabled the establishment and maintenance of the observed phylogeographic patterns. Comparative phylogeography has been facilitated by advancements in DNA sequencing and the availability of large genetic datasets in public repositories (GenBank and BOLD) (Burney & Brumfield, 2009; Dapporto et al., 2017, 2019; Fujisawa, Vogler, & Barraclough, 2015; Moritz et al., 2009; Pigot & Tobias, 2014), mostly in the form of mitochondrial DNA (mtDNA) cytochrome *c* oxidase subunit 1 (COI) sequences.

The particularly fast mutation rate and the short coalescence time characterising mtDNA produce frequent cases of genetic differentiation among allopatric populations (Allio, Donega, Galtier, & Nabholz, 2017), thus making mtDNA a widely used marker in phylogeography (Avise, 2000).

The butterfly populations occurring across the 3-km-wide Messina strait, which separates the Italian mainland from Sicily Island, represent an excellent model to study genetic diversification over narrow sea straits. First, this channel is one of the narrowest separating a large island from continental Europe. Second, Sicily was connected to the Italian Peninsula during the LGM for at least 1.5 kiloannum (ka), from 21.5 to 20 ka before present (Antonioli et al., 2016). Finally, Sicily is the largest island in the Mediterranean and due to its environmental heterogeneity, it hosts about 90 butterfly species which offer the opportunity to compare genetic differentiation within a large set of species.

We hypothesised that: 1) a fraction of the species occurring on Sicily could represent different genetic lineages compared to the nearest Italian mainland, and 2) the genetic differentiation and the occurrence of endemic lineages could correlate with a series of species traits and environmental constraints to dispersal. With the aim of testing these main hypotheses, we sequenced a fragment of the COI region of 84 butterfly species living in Sicily and quantified their genetic structure across the Messina channel, as well as along the entire Palearctic region.

We set three main goals: i) determine if a fraction of the Sicilian populations represent different lineages compared to the Italian Peninsula and/or endemic lineages; ii) test if the genetic contrasts are correlated with functional traits determining dispersal capabilities and with contemporary environmental constraints to dispersal (winds and climatic conditions); iii) determine to what extent Sicilian lineages are genetically closer to populations occurring on the Italian Peninsula, or if some of them likely represent relict populations having their closest relatives in distant Palearctic areas as shown for several butterfly populations in the Mediterranean islands (Dapporto & Bruschini, 2012; Dapporto et al., 2017; Vodă et al., 2015b).

Dissecting historical and ecological correlates of genetic differentiation between Sicily and the Italian mainland for insects having good dispersal capabilities can contribute to a better understanding of the processes generating and maintaining genetic contrasts across apparently weak barriers.

Material and methods

During the years 2008-2016, we collected butterflies in two main areas: the Southern Italian Peninsula - Calabria, Basilicata and Campania (SIP) - and Sicily Island. We retrieved COI sequences using standard sequencing procedures (Appendix S1) and added other COI data available in GenBank and BOLD. We obtained 659 sequences for specimens from Sicily belonging to 84 species out of 90 reported for this island (Sicily barcoding dataset). Among these, 81 species also occur in SIP and 739 specimens were sequenced from this area, which together with the Sicilian specimens represent the zoogeographic dataset, 1398 specimens. We used a subset of 73 species, with at least eight specimens per species and a minimum of three specimens for both Sicily and SIP to model genetic differentiation against geographic location and species traits (comparative dataset, 1308 specimens). To test if Sicilian populations have their closest relatives in the Italian mainland or in more distant regions, we also retrieved 8093 sequences from the Palearctic area (139 sequenced for this study, Palearctic Dataset). Species and specimen information is available in Table S1 and Appendix S2.

Identification of sub-areas and genetic differentiation

Using the *costDistance* function of the *gdistances* R package (van Etten, 2017), we identified the minimum over-land distance between each of the 5x5 minute cells in the study area and the centre of the narrowest area of the Messina strait (Fig. S1). Within Sicily and SIP, we identified a set of six sub-areas increasingly distant from the Messina strait (0-100 km, 100-200 km, 200-300 km, Fig. S1). We attributed the sequenced specimens to the six areas based on their collection site.

Recent studies indicate that uncorrected p-distances are the best option for analysing COI data (Collins, Boykin, Cruickshank, & Armstrong, 2012; Srivathsan & Meier, 2012). We thus calculated the genetic p-distances among specimens for each of the 84 species of the Sicily barcoding dataset. We also verified if the indices of genetic differentiation (Gst and Dst, see below) showed considerable differences when genetic distances were computed with different substitution models. For this reason, we estimated the best substitution model by using *MEGA X*, recomputed the differentiation indices and correlated the values obtained with p-distances with those obtained by the selected substitution model. For each species, we also calculated pairwise differentiation among the six sub-areas (Appendix S1). We applied Principal Coordinates Analysis (PCoA) to these matrices and projected the configuration in Red-Green-Blue graphs using the R package *recluster* (Appendix S1). In these plots, the most similar specimens (dots) are closer to each other and they are visualised with similar colours. The colour resemblance among dots facilitated mapping the genetic differentiation of each species in the study area. For the 73 species belonging to the comparative dataset we also calculated haplotype diversity (Hd) as the average distances among haplotypes using the *nuc.div* function of the *pegas* R package, the absolute differentiation (Dst) and the standardised differentiation (Gst) (Nei, 1987) (Appendix S1) between Sicily and SIP.

Using the Gst and Dst pairwise matrices for each species, we calculated the mean of the available values of the corresponding cells to generate the average Gst and Dst matrices, representing the overall genetic differentiation among sub-areas based on shared species. We applied a PCoA to these matrices and projected the configuration in the Red-Green-Blue space to map the overall differentiation among sub-areas. Moreover, for each species, we extrapolated pairwise Gst and Dst among nearby sub-areas. By using the *glmer* function of the *lme4* R package, we carried out two generalised linear models (GLMMs) comparing pairwise Gst and Dst between the pairs of connected sub-areas, including species as random factor. This was done to detect if the two areas facing the Messina strait showed higher Gst and Dst compared to other pairs of areas. We calculated ANOVA tables with the *Anova* function of the *car* R package and obtained pairwise comparisons by using Tukey test, as implemented in the *glht* function of the *multcomp* R package.

Patterns of genetic similarity between populations

We divided the Palearctic region into 13 regions (Fig. 4d). To inspect the phylogenetic relationships, we inferred haplotype networks for each species of the Sicily barcoding dataset, using *TCS 1.21* (Clement, Posada, & Crandall, 2000). We imposed a 95% connection limit (11 steps) except for a series of species for which a higher number of steps was necessary (Appendix S1). Haplotype networks and Red-Green-Blue maps allowed recognising which species of the zoogeographic dataset had different lineages between Sicily and SIP, and which species in the Sicily barcoding dataset show endemic haplogroups and lineages.

Moreover, we identified sectors of 5x5 degrees of latitude per longitude inside the Palearctic region and grouped the sequences according to this division. For all the species included in the zoogeographic dataset, we created a list of haplotypes recorded for Sicily and SIP and for each sector of each Palearctic area. We calculated the genetic distance of Sicily and of SIP with respect to each Palearctic sector as:

$$DS = \sum_{h=1}^{n} \min\,(pdist_h)$$

Where min($pdist_h$) represents the p-distance between the haplotype h and the closest haplotype occurring in a given sector. For each species, we recorded the sector showing the lowest DS with respect to Sicily and SIP and calculated their wolf distances. In cases of ties in DS, we recorded the closest sector.

Modelling genetic differentiation

We verified the dependency between the three indices (Hd, Dst and Gst) and sample size (number of sequences obtained) for the species in the comparative dataset by constructing rarefaction curves. For each species, we calculated the value of each index (Hd, Dst and Gst) for one-step sequential sample sizes starting from a minimum of four specimens to the total number of specimens sequenced. We re-sampled the order 100 times, obtained 100 rarefaction curves and then calculated the average curve. We used the smoothed averaged curve to compute a horizontal asymptote by modelling a general y=a+(b/x) function using the *nls* R function. Then, we obtained the difference between the observed values of each index and the asymptotic ones (Hda, Gsta and Dsta). This allowed to evaluate the influence of sample size on similarity between observed and asymptotic values. Finally, we correlated the number of sequenced specimens with the observed and asymptotic values.

Moretti and co-workers (2017) identified 29 functional traits covering the primary functions of invertebrates, divided into five major groups: morphology, feeding, life history, physiology and behaviour. We assessed 13 species traits representing four groups: morphology, feeding, life history and physiology. We excluded behavioural traits because they are unavailable for most European butterfly species. We formulated functional hypotheses for each group of traits (Dapporto et al., 2017) (Table S4): a) We measured trophic generalism as i) the number of host plant genera reported in two literature sources (feeding trait); b) Mobility was represented by ii) wingspan (Sekar, 2012) (morphological trait), obtained as the average between minimum and maximum wingspan reported in four literature sources. For each studied species wingspan was available in at least three sources and we imputed the missing values by using the *mice* function of the *mice* R package. The algorithm assigns an incomplete variable by generating plausible values based on other variables in the dataset by Multivariate Imputations by Chained Equations (MICE). Then we applied a Principal Component Analysis (PCA) using the values from different sources to obtain a single (PC1) measure of size; c) We assessed phenology as iii) the number of months during which adults occur in the study area, iv) the first month when adults emerge, v) and vi) the first and the last month when adults fly, and vii) voltinism (life history traits). Data belong to LD, AB and SS personal observations collected from 2009 to 2018 and from the database of the Butterfly Diversity and Evolution Lab. Finally, d) we included a series of variables describing climate experienced by each species in its area of distribution in Europe (Schweiger et al. 2014). These features have been obtained by modelling species distribution in Europe based on occurrence data, and then by averaging temperature and precipitation among the spatial cells where each species is predicted to occur. Although these indices cannot be considered as functional traits (Moretti et al., 2017), they are widely used to describe climatic preference and tolerance and are considered proxies for physiological traits (Dapporto et al., 2017; Devictor et al., 2012). The variables included were: mean annual temperature and precipitation (viii, ix); standard deviation of mean temperature and precipitation (x, xi); upper 95% confidence limit of temperature mean and of precipitation mean (xii, xiii).

Butterfly traits are usually highly inter-correlated but they can be reduced to factors by using ordination methods (Dapporto et al., 2017; Middleton-Welling, Wade, Dennis, Dapporto, & Shreeve, 2018). We applied a PCA to life history and physiology traits using the R function *rda* and we retained as variables the components with eigenvalues higher than one.

For each species, we calculated two kinds of environmental variables as: i) the relative geographic distances between Sicilian and SIP specimens and ii) the tendency of winds to facilitate or hamper the crossing of the Messina strait. The first measure has been obtained for two distances, i) the length of the shortest path over land (wolf distance) and ii) the total cost of a random walk between pairs of points based on species-specific landscape permeability (drunk distance) (van Etten, 2017) (Appendix S1 and S2). Then, we applied the same formula used to obtain Dst:

$$Geodist = Geodist_{all} - Geodist_{intra}$$

where $Geodist_{all}$ represents the average distance of the specimens of a given species and $Gdist_{intra}$ the average intra-area distance.

To calculate the winds experienced by each species we downloaded the monthly data of winds for Reggio Calabria and Messina (the closest areas to the strait) from http://mesonet.agron.iastate.edu/sites/locate.php?network=IT__ASOS. For each month, we used the direction of winds (with 36 classes of 10 degree intervals; from class 0, 0-9 degrees to class 35, 350-359 degrees), their average frequencies between 2010 and 2017 and their strength divided into five classes (2-5, 5-7, 7-10, 10-15, 15-20, over 20 miles per hour).

Using a high-resolution geographic layer (30s) we obtained a series of 23 points, each 2 km apart and located at the same minimum distances from the two sides of the strait (Fig. 3c). Each pair of successive points identified a mid-point located along the strait and the segment connecting them identified a direction locally tangential to the strait (Fig. 3a). Therefore, for any butterfly crossing the strait over any of these points, a direction of winds parallel to the segment had the most negative impact in reaching one of the two landmasses, while an orthogonal wind had the highest potential to facilitate dispersal. The resultant effect of winds for each month and each segment (*WRm*), has been calculated as:

$$WRm = -\sum_{j=0}^{35}\sum_{i=1}^{6} freq_{ij}\ str_{ij} \cos\left(2\alpha + \mathrm{j}\frac{\pi}{9}\right)$$

Where $freq_{ij}$ is the frequency and str_{ij} the strength of a given wind of direction j (0-35 each representing 10 degrees) and strength of the i^{th} class. We scored strength as the central value for each of the six strength classes, evaluated as 25 mph for the strongest; α is the difference between the angle of the segment tangential to the strait and the angle of the wind blowing form the north (j=0). A wind with angle j having the same direction of the segment ($\alpha=0$ or $\alpha=\pi$) is scored $-freq_{jj}*str_{ij}$ (the most negative effect), while an orthogonal wind ($\alpha=\pi/2$ or $\alpha=3\ \pi/2$) will score the most positive effect of $+freq_{jj}*str_{ij}$. Winds with intermediate directions will score a zero contribution to *WRm* (Fig. 3a). For each month, we averaged the wind data for Reggio Calabria and Messina. We recorded, for each month, the average *WRm* value among the 22 segments along the strait. The potential effect of winds on each butterfly species has been finally obtained by scoring the average values for the months included in their flight period. Months with more rainy days are expected to have a lower influence on butterfly dispersal, as butterflies are not expected to fly on rainy days. We obtained the average number of rainy days in Messina and Reggio Calabria from the website https://www.woitalia.it, and *WRm* values have been weighted for the number of non-rainy days.

We tested the existence of a phylogenetic signal for species traits with Pagel's lambda index by applying the *phylosig* function of the *phytools* R package. As a reference phylogeny, we used the recently published phylogenetic tree based on 14 genes (1 mitochondrial and 13 nuclear) of all European butterflies (Dapporto et al., 2019).

As done in previous studies, we assessed the relationships between traits and environmental constraints for species and their Hd, Dst and Gst (using both asymptotic and observed values) using individual and multiple phylogenetic regressions (Fujisawa et al., 2015; Nabholz, Mauffrey, Bazin, Galtier, & Glemin, 2008; Romiguier et al., 2014). We employed Pagel's lambda as a model for the phylogenetic covariance of residuals as implemented in the function *pgls* of the R package *caper*. We square-root transformed Dst to improve its normality and standardised the values of the traits with zeta-scores. In multiple regressions, to avoid model overfitting and to provide a better parameterisation of variables, we applied multi-model inference through Information-Theoretic Approach to select a set of "best models" using the *MuMIn* R package. We selected the model with ΔAICc values <2, considered to be equally parsimonious (Burnham & Anderson, 2002).

Results

Spatial patterns of genetic differentiation

The comparison of Gst and Dst, based on p-distances and genetic distances calculated with the TN93 model, showed a very high correlation (Gst, R=0.99998; Dst, R=0.99878). We thus used p-distances as suggested for COI data (Collins et al., 2012; Srivathsan & Meier, 2012) and because they correspond to the distances represented in the haplotype networks employed for assessing endemic lineages (see below). Accumulation curves revealed that the differences between asymptotic and observed values were in most cases lower than 0.01% of divergence for Hd (which showed a range among species of 0-2.13%), 0.04% for Dst (range

0-1.69%) and 0.08 for Gst (range 0-1) (Table S2). None of the observed and asymptotic indices showed a significant correlation with sample size (Table S7).

PCoA representations based on pairwise Gst and haplotype networks revealed that 42 of the 81 species occurring in Sicily and SIP did not show any differentiation between the two areas and had Gst values usually lower than 0.2 (Fig. 1a,b; Tables S2 and S3). The second largest group (30 taxa) was represented by species with completely segregated haplogroups between Sicily and SIP and Gst values higher than 0.5 (Fig. 1c,d; Tables S2 and S3). A small number of species (7) showed admixtures of different haplotypes across the strait (Fig. 1e,f, Tables S2 and S3) or possible phenomena of founder effect or genetic bottlenecks (e.g. *Melitaea phoebe*, Appendix S1). Finally, only two species had mutually excluding lineages among sub-areas that were not defined by the Messina strait (*Aglais urticae* and *Anthocharis damone,* Fig. 1g,h).

A projection of the average genetic differentiation revealed the existence of two main groups of sub-areas (Sicily vs SIP, Fig. 2a,b). GLMMs indicated that the five segments connecting the six sub-areas exhibited different Dst (ChiSq.=33.629; df=4; $P<0.001$) and Gst (ChiSq.=32.276; df=4; $P<0.001$). In pairwise comparisons, the segment corresponding to the Messina strait had higher Gst and Dst values compared to the other four segments (Fig. 2c, Table S9 and S10).

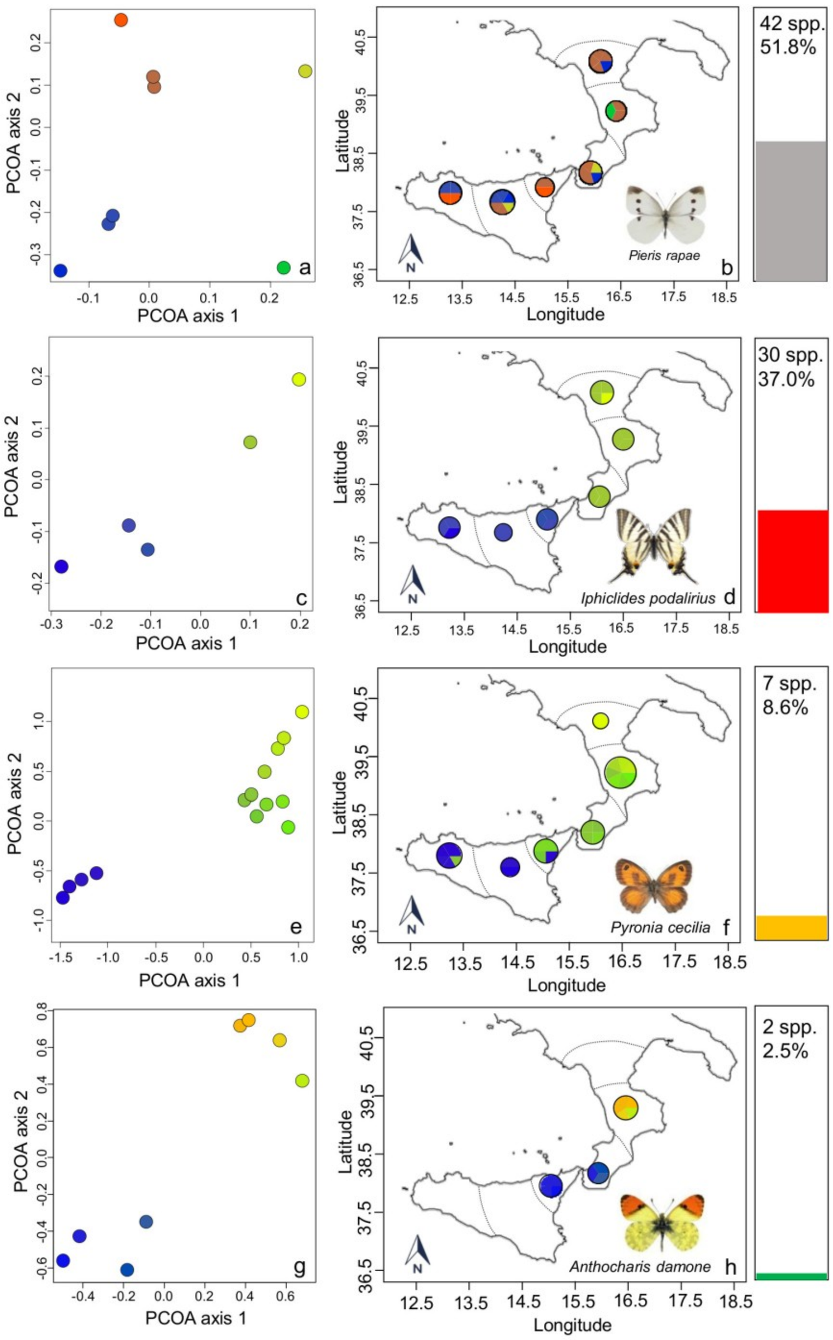

Figure 1. Red-Green-Blue projection of PCoA configurations based on COI p-distances (a,c,e,g) and mapping of resulting colours for specimens (b,d,f,h) for four species representative of different distribution patterns: *P. rapae* (complete admixture of haplogroups), *I. podalirius* (complete segregation), *P. cecilia* (incomplete segregation) and *A. damone* (main break outside the Messina strait). Histograms on the right represent the percentage of species showing a similar pattern in the zoogeographic dataset.

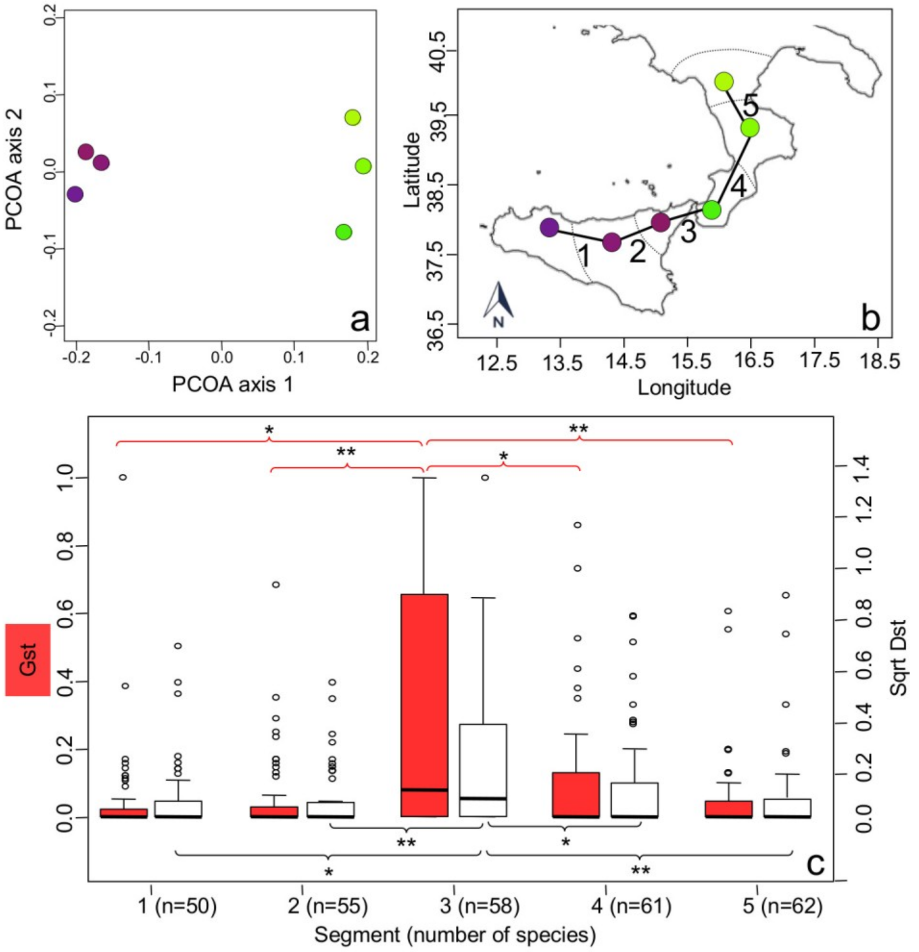

Figure 2. Red-Green-Blue projection of PCoA among the six sub-areas based on average pairwise Gst values for all 73 species in the comparative dataset (a) and projection of the colours on the map (b). The six areas are connected by five segments (b). Boxplots showing medians and quartiles for pairwise Gst (red boxes) and Dst (white boxes) of species sequenced for pairs of nearby areas (c). Significant differences in pairwise comparisons after a GLMM are indicated with asterisks (*, 0.05>P>0.01; **, 0.01>P>0.001).

Species traits, environmental variables and phylogenetic regressions

For phenology traits one component had an eigenvalue higher than one and it was mostly represented by voltinism and length of the flight period (Table S4, Fig. S4). For climatic preference traits, two components had eigenvalues higher than one (Table S4, Fig. S3): the first ordered species from those experiencing low temperatures, high precipitation and high tolerance (temperature and precipitation standard deviation) to those living in warmer and drier areas also showing a lower tolerance; the second ordered species mostly according to precipitation.

The dominant winds at the strait are directed from north to south (Fig. 3b). Accordingly, most segments scored negative values (wind direction hampering dispersal) during all months (Fig. 3d). The only exception was represented by the extreme northern part of the strait, which is also the narrowest sector, with positive values (Fig. 3c). Winter and autumn months showed less negative winds to dispersal than spring and summer ones (Fig. 3d, Fig. S5).

The five intrinsic traits (host plants, wingspan, phenology PC1, climatic preference PC1 and PC2) had Pearson R lower than 0.300 in all cases (Table S6). Conversely, the two geographic distances among areas showed high intercorrelations (Table S6). Due to the seasonal character, winds experienced by species showed high correlations with phenology PC1 (Table S6) and species with a longer flight period also experienced more favourable winds (Pearson R=0.84).

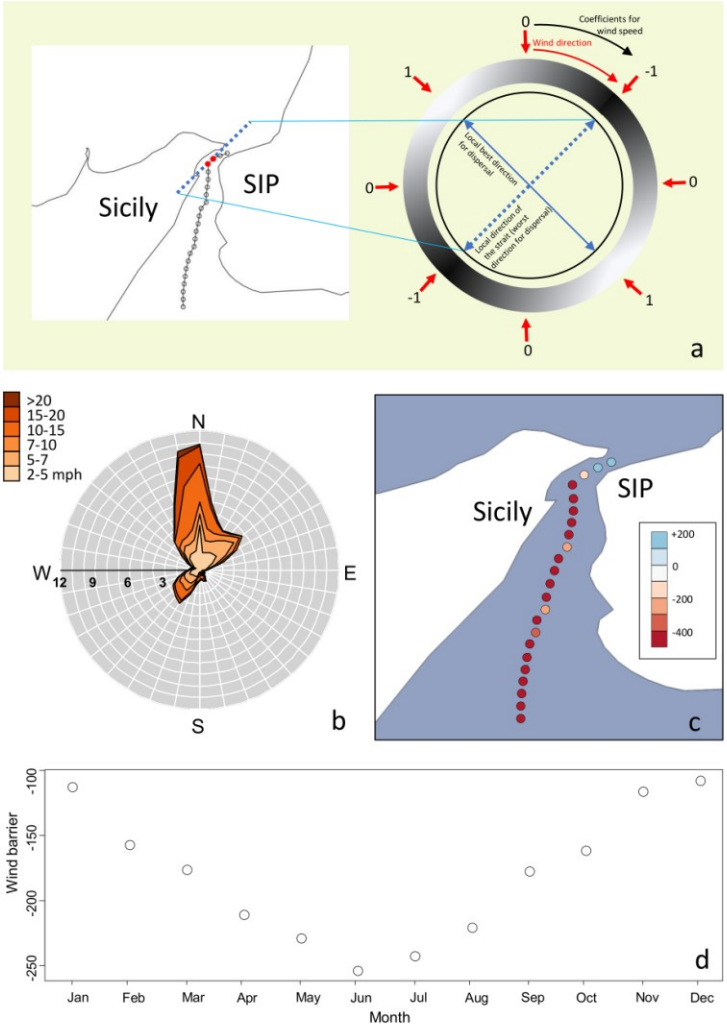

Figure 3. The method used to evaluate the effects of winds over the Messina strait. (a) Winds tangential to the direction of each segment (dotted line) have a coefficient of -1 (most negative value) to be multiplied by frequency and wind speed; winds orthogonal to the segment have a maximum favourable coefficient of 1. (b) The average wind rose measured for Messina and Reggio Calabria. (c) The values of wind suitability for dispersal over the 22 points identified along the Messina channel. (d) The average values of overall wind suitability in different months (*WRm*).

Among species traits, wingspan, host plant genera, number of flight months, first month of emergence and first and last month of flight showed a significant phylogenetic signal (Table S5). However, probably due to the low phylogenetic signal of voltinism, the phenology PC1 did not show a significant phylogenetic signal (Table S5).

Using the asymptotic values for mtDNA polymorphism (Hda), phenology PC1 and the correlated mean wind showed a significant effect in individual regressions with the former showing a higher R^2; accordingly, in the multi-model phylogenetic regressions (R^2 of the full model=0.238, F-statistic=2.493 on 8 and 64 DF, P=0.020) only phenology PC1 had a negative significant effect (species flying for shorter periods had higher haplotype diversity; Table 1). When the spatial structure was included in the indices (Dsta and Gsta), more variables significantly entered the models. In individual regressions, drunk distances were positively correlated with Gsta and Dsta explaining more variance than the analogous wolf distances (Table 1); similarly, mean wind experienced at the strait explained more variance for both indices than the correlated phenology PC1 (Table 1). In both indices, smaller species showed a higher genetic differentiation (Table 1). Species exploiting a lower number of host plants also had a higher Gsta. In the multiple model for Dsta (R^2=0.357, F-statistic=4.447 on 8 and 64 DF, P=0.0003) drunk distances, wingspan and mean wind had a significant effect, while for Gsta (R^2=0.362, F-statistic=4.551 on 8 and 64 DF, P=0.0002) a significant effect was found for mean wind, host plants and wingspan (Table 1). Similar results were obtained for the observed indices (Hd, Gst and Dst) (Table S8).

Table 1. Results for multiple and individual regressions for HDa, Gsta and Dsta between Sicily and SIP. The "models" column indicates the number of models with ΔAICc values <2 in which each variable entered. Significant effects are highlighted in bold.

	Multiple regression					Individual regression	
HDa	Estimate	St. Error	z	P	models	R^2	P
Phenol. PC1	**-0.394**	**0.112**	**3.524**	**<0.001**	**9/9**	**0.173**	**<0.001**
Clim. pr. PC2	-0.177	0.108	1.648	0.099	6/9	0.038	0.100
Wing Size	-0.126	0.107	1.171	0.242	3/9	0.021	0.223
Wolf dist.	0.121	0.115	1.052	0.293	3/9	**0.056**	**0.043**
Drunk dist.	0.094	0.114	0.824	0.410	1/9	0.052	0.054
Host plants	-0.069	0.113	0.612	0.540	1/9	0.024	0.187
Clim. pr. PC1	-	-	-	-	0/9	0.016	0.288
Mean wind	-	-	-	-	0/9	**0.143**	**<0.001**
Dsta	Estimate	St. Error	z	P	models	R^2	P
Wing size	**-0.230**	**0.102**	**2.265**	**0.024**	**10/10**	**0.086**	**0.012**
Drunk dist.	**0.241**	**0.110**	**2.199**	**0.028**	**10/10**	**0.161**	**<0.001**
Mean wind	**-0.320**	**0.124**	**2.585**	**0.010**	**7/10**	**0.211**	**<0.001**
Phenol. PC1	-0.279	0.148	1.881	0.060	4/10	**0.176**	**<0.001**
Host plants	-0.136	0.105	1.300	0.194	4/10	**0.053**	**0.050**
Clim. pr. PC2	-0.097	0.101	0.963	0.336	3/10	0.018	0.259
Clim. pr. PC1	0.116	0.105	1.103	0.270	2/10	0.003	0.658
Wolf dist.	-	-	-	-	0/10	**0.080**	**0.015**
Gsta	Estimate	St. Error	z	P	models	R^2	P
Mean wind	**-0.337**	**0.111**	**3.038**	**0.002**	**6/6**	**0.207**	**<0.001**
Host plants	**-0.225**	**0.103**	**2.186**	**0.029**	**6/6**	**0.080**	**0.015**
Wing Size	**-0.218**	**0.101**	**2.159**	**0.031**	**6/6**	**0.072**	**0.021**
Drunk dist.	0.214	0.138	1.551	0.121	4/6	**0.134**	**0.001**
Clim. pr. PC2	-0.158	0.101	1.571	0.116	3/6	0.029	0.151
Wolf dist.	-0.151	0.152	0.993	0.321	2/6	0.037	0.104
Phenol. PC1	-	-	-	-	0/6	**0.135**	**0.001**
Clim. pr. PC1	-	-	-	-	0/6	0.002	0.744

Patterns of genetic similarity between populations

Out of the 84 species of the Sicily barcoding dataset, 31 (36.9%) had endemic lineages in Sicily (Appendix S1). Most species collected in SIP (73/81) had their closest relative in Sicily or in other sectors of the Italian Peninsula and the Alps (Fig. 4b). The Sicilian populations showing medium and low Gst values with respect to population in SIP (in general lower than 0.6), more often had their closest relatives in SIP, the Alps and Apennines (61/84). Conversely, several endemic lineages and species with a Gst higher than 0.6 had their closest relatives in areas far from the Alps and the Italian Peninsula (Fig. 4a, Table S3). The species with Sicilian endemic lineages tended to show a low differentiation with respect to their closest relative, when that relative occurred in the Italian Peninsula (size of circles in Fig. 4b and Table S3). The Sicilian lineages having the closest relative in another region ranged from very low to very high differentiation (a maximum of 2.1% for *M. didyma* Fig. 4c).

Discussion

About 40% of the butterfly species living on the two sides of the Messina strait show a considerable genetic differentiation across this narrow channel, contrasting with the well-known dispersal capacity of butterflies and their ability to track available habitat (Devictor et al., 2012; Parmesan et al., 1999; Sekar, 2012). Across the channel, different species traits and ecological constraints correlate with various indexes of genetic differentiation. The largest part of the intraspecific differentiation across the Messina strait is encompassed by the existence of Sicilian endemic lineages, found in 36.9% of species. Among these endemic lineages, 41.9% do not have their closest relatives in the nearby Italian Peninsula, underlying the influence of complex processes in determining the genetic structure of Sicilian butterflies.

We rule out that a larger sample size would have drastically changed the results for the following reasons: i) the rarefaction curves indicate that a higher number of sequences would only slightly change the observed Gst, Dst and Hd values, ii) the absence of negative correlations between the three indices and sample size indicates that the observed differentiation also occurs in well sampled species and is not likely due to incomplete sampling of lineages across the two areas, iii) the observed genetic structure was rather dichotomic, with 89.3% of species showing either an almost complete admixture of haplotypes (51.8%, group 2 in Fig. 1) or a complete spatial segregation (37.0+2.5%, group 2 and 4 in Fig. 1). In the case of a few species showing an incomplete segregation (8.6%, group 3 in Fig. 1) the pattern was never caused by a single specimen typical of one side of the strait being detected in the other one. As for the comparison with other Palearctic regions, iv) butterflies are probably the best studied model organisms for mtDNA and the extensive spatial coverage currently available makes it unlikely that further increasing the data would render a considerable number of Sicilian lineages as non-endemic.

Although largely available in repositories, easily and cheaply sequenced, mitochondrial markers are almost strictly maternally inherited. Consequently, they can only describe the genetic history of female populations (Galtier et al., 2009). Since in several species, but not all, females are less dispersive than males, further research analysing the nuclear genome is key to determine to what extent patterns of differentiation across the Messina strait are reflected in both mitochondrial and nuclear DNA. Available research for some species suggests that patterns across Mediterranean Sea straits may vary, with cases of agreement between the two genomes (Dincă, Dapporto, & Vila, 2011; Dincă, Lee, Vila, & Mutanen, 2019) as well as disagreement (Livraghi et al. 2018) being reported.

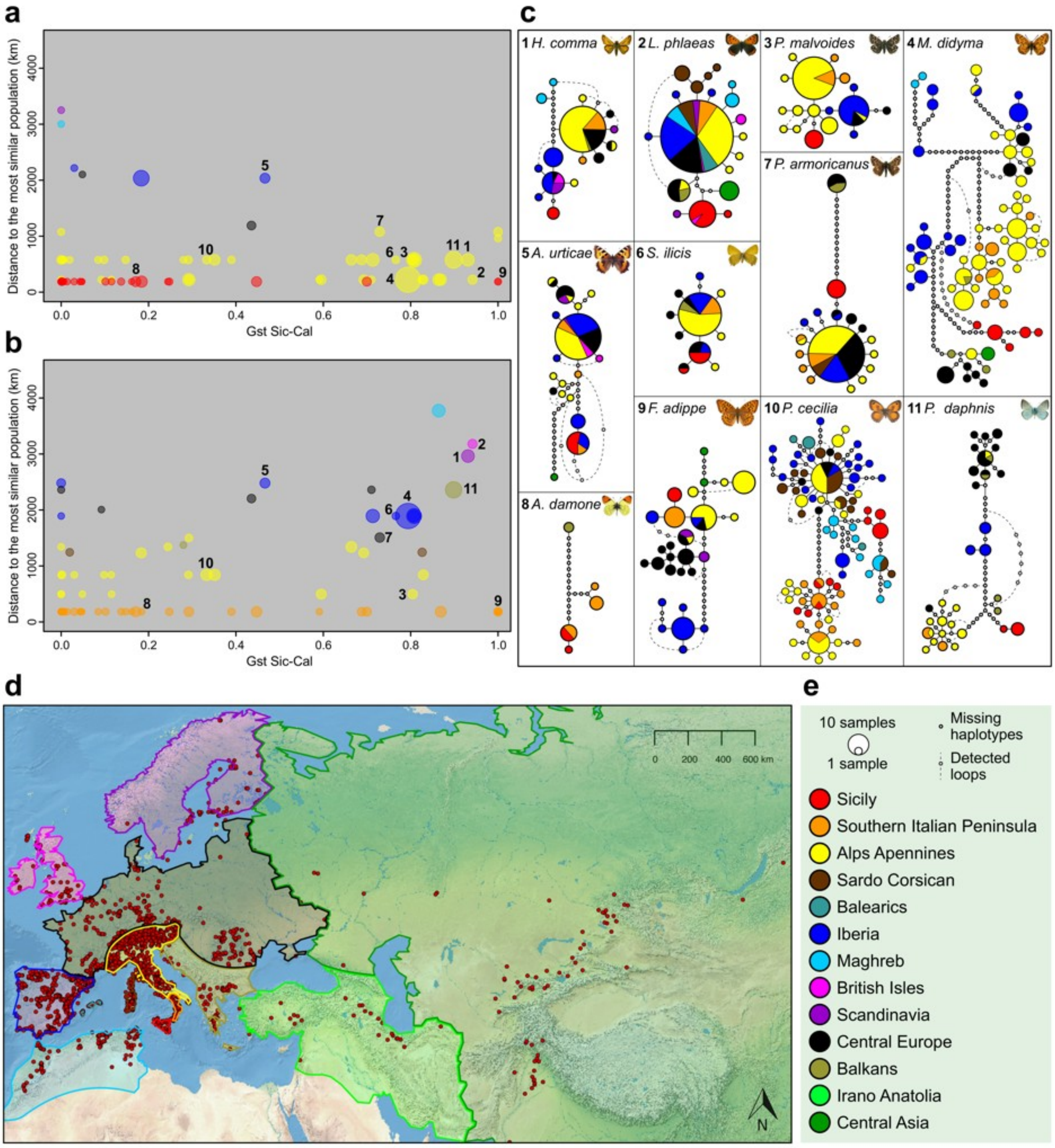

Figure 4. The relationships of Gst between Sicily and SIP and the ground distance to the genetically closest population from Sicily (a) and SIP (b). The radius of the circles represents the overall p-distance between the populations from Sicily (a) or SIP (b) and the genetically closest populations. The colours of the dots in (a) and (b) indicate the geographic region (d,e) of the genetically closest population. Haplotype networks for 11 species (c), the colours of the sectors indicate the geographic region of specimens described in panels (d) and (e). The numbers in graphs (a) and (b) correspond to the numbers indicated in the haplotype networks

in (c). The red dots on the map (d) indicate the sites from which COI sequences were available.

Genetic differentiation and its correlation with functional traits and ecological constraints

Following recent climate change, poleward distribution shifts of magnitudes of km/year have been documented for many butterfly species (Devictor et al., 2012; Parmesan et al., 1999). This has been detected even in highly fragmented areas of central-northern Europe where, for example, the expansion of the small species *Hesperia comma* has progressed with single jumps of up to 28.8 km (Wilson, Davies, & Thomas, 2009). Yet, a large fraction of the butterfly species analysed here showed notable population differentiation across the 3-km-wide Messina channel. Phylogenetic regressions indicated that this genetic pattern had a deterministic relationship with some functional traits and current ecological constraints. It is not surprising that lineages of more dispersive species and of species subjected to weaker ecological constraints can more easily admix across a barrier (Burney & Brumfield, 2009; Dapporto et al., 2017). Nevertheless, the three indices (Hd, Dst, Gst) showed correlations with different traits and constraints. Moving from the index measuring the overall genetic variation among haplotypes (Hd) to the indices determined by their distribution across the strait (Dst and Gst), the influence of intrinsic species traits became weaker in favour of extrinsic environmental constraints and of traits determining the interaction with the local environment.

Comparative studies indicate that different species attributes (e.g. population size, fecundity, body size, dispersal capability) explain mtDNA polymorphism and its spatial differentiation, (Allio et al., 2017; Burney & Brumfield, 2009; Dapporto et al., 2017, 2019; Fujisawa et al., 2015; Nabholz et al., 2008). Adult is the most mobile life stage in butterflies and the length of the flight period largely determines the potential dispersal events and the consequent gene-flow. Accordingly, in our study area, species with a shorter flight period and lower number of generations have accumulated a higher mtDNA polymorphism (Hd) in historical times.

Different variables explained Dst and Gst. Wingspan, a key factor in butterfly mobility (Dennis, Hardy, & Dapporto, 2012; Sekar, 2012), had a significant effect for both indices. Similarly, a comparative study on bird sister species found that the tendency to secondary sympatry was positively correlated with a characteristic of wing morphology determining dispersal capabilities (Pigot & Tobias, 2014). Extrinsic environmental constraints also entered the model. The first are winds experienced at the strait during the flight period. Due to the local distribution of winds throughout the year, the species flying for more months also experience more favourable winds. However, individual and multiple regressions showed that the information provided by wind suitability explained more Gst and Dst variance than phenology, thus indicating the impact of current wind direction, together with a reduced flight period, in hampering the admixture of butterfly populations. Due to the Pleistocene climatic oscillations, ecological settings have undoubtedly changed over the historical time required for the differentiation of the observed Sicilian lineages, in some cases exceeding 2% of COI variation. In this context, much of the observed genetic differentiation might be explained by historical but unknown variation in ecological settings that occurred over evolutionary time.

Two species traits determining interactions between butterflies and the environment (the degree of generalism and the distances among populations based on climatic suitability) also significantly explained Gst and Dst variance. The first represents the cost of dispersal between SIP and Sicily in terms of environmental permeability. The second determines the possibility to exploit multiple resources in different environmental settings, which could be crucial during a dispersal event.

Endemic and relict lineages in Sicily

Most of the differentiation across the Messina strait is encompassed by the existence of Sicilian endemic haplogroups. Conversely, almost no endemic lineage occurred in SIP and most butterfly populations inhabiting this area were identical to those occurring in Sicily and/or in the rest of the Italian Peninsula. The existence of endemic butterfly lineages on Mediterranean islands is well known (Cesaroni et al., 1994; Dapporto et al., 2017; Vodă et al., 2016), and in Sicily this phenomenon involved 36.9% of species. The most parsimonious hypothesis is that diverging lineages in Sicily originated from populations belonging to the nearby Italian Peninsula. Nevertheless, 28.5% of Sicilian populations and 41.9% of endemic lineages have their genetically closest population in distant regions, and in three cases the species are even absent from the Italian Peninsula (*Spialia orbifer*, *Polyommatus celina*, *Melanargia pherusa*). These cases likely represent relict populations that often characterize large continental islands (Whittaker & Fernández-Palacios, 2007). This phenomenon has been proven for the vertebrates of Mentawai islands, which, in many cases, do not have their closest relatives on the nearest island of Sumatra, but rather on the more distant Borneo, Java or Peninsular Malaysia (Wilting, Sollmann, Meijaard, Helgen, & Fickel, 2012).

Sicily hosts a large fraction of endemic lineages although it has been connected to SIP for 1.5 ka during the last glacial maximum (Antonioli et al., 2016), which probably facilitated the admixture of many butterfly populations. In Lepidoptera there is evidence from historical DNA assessment (Mende & Hundsdoerfer, 2013) and from paleogeographic, morphologic and phylogeographic data (Dapporto & Bruschini, 2012; Dapporto et al., 2017), that the distribution of lineages of several species has changed in the last thousands of years in the Italian Peninsula. The rapid substitutions of genetic variants have probably been fostered by a continuously changing environment (Dapporto & Bruschini, 2012) as found in other organisms and other geographic regions (Mallet, 2010; Moritz et al., 2009; Toews, Mandic, Richards, & Irwin, 2014). The strong genetic differentiation for butterflies existing at the Messina strait could be, at least in part, determined by secondary contacts among recently shifted populations (Mallet, 2010; Moritz et al., 2009). Moritz and co-workers (2009) predicted that in such cases, the barrier should segregate lineages with different degrees of divergence and should be located in areas of low population density. The latter clearly applies for a sea strait and our results agree with the former hypothesis. In fact, the lineages showing a complete segregation between Sicily and Calabria encompass a divergence ranging from a single mutation to more than 2% of the 658 bp of the COI marker. Previous studies indicated that circum-Italian islands host ancestral populations of butterflies once occurring on the mainland, where they have been replaced by lineages shifting their distribution and mostly belonging to Eastern Europe (Dapporto et al., 2012). In some species, the Sicilian lineages differing from lineages inhabiting the Italian Peninsula are genetically identical (*Lycaena phlaeas*, *Satyrium ilicis*, *A. urticae*) or show a few mutations (*Thymelicus acteon*, *Hesperia comma*) compared to populations inhabiting other areas (mainly Iberia). In these cases, the replacement on the Italian Peninsula of a supposedly ancestral haplotype has been likely recent. In other cases, the Sicilian lineages showed a higher genetic divergence (>0.5%) compared to the closest relatives, suggesting that the replacement has likely been more ancient, as for *Carcharodus alceae*, *Thymelicus lineola*, *Melitaea didyma*, *Polyommatus daphnis*. In a few cases the replacement can still be incomplete (*A. urticae*) and the Sicilian haplotypes also occur at the tip of the Italian Peninsula; alternatively, the Sicilian populations could have recently colonised the mainland area. In other species (*M. jurtina* and *P. cecilia*), the supposedly expanding lineages from the Italian Peninsula partially colonised the eastern area of Sicily (Vodă et al., 2016). Based on estimated substitution rates for invertebrates, of about 2.3% uncorrected pairwise distance per million years (Brower, 1994), the Sicilian endemic and relict populations have likely established during the Quaternary period.

Conclusion

We provided a comprehensive assessment of mtDNA diversity for most butterfly species recorded in the largest Mediterranean island, with comparisons from surrounding areas. The accumulation of similar datasets in public repositories opens the door to perform large-scale comparisons of population differentiation and to correlate genetic patterns with species functional traits and ecological constraints. When deterministic fingerprints explain a large fraction of variance, as in our phylogenetic regressions, mechanisms regulating phylogeographic processes can be unveiled.

Despite the limitations of using a single mitochondrial marker in phylogeography and in the assessment of genetic diversity (Galtier, Nabholz, GléMin, & Hurst, 2009; Nabholz, Glémin, & Galtier, 2009; Papadopoulou & Knowles, 2016), we demonstrated that butterfly matrilines do not easily cross and establish across the Messina strait. Regardless of how and when the genetic divergence emerged, the main drivers determining the admixture of mtDNA haplotypes represent a combination of intrinsic traits and environmental constraints. Our findings confirm that strong genetic differentiation and endemicity can occur in mobile taxa even in barely isolated islands, and that they are produced by a complex combination of ecological forces and historical events acting differently on the species of a given community. Notably, we showed how understanding the pattern of genetic differentiation among two close areas requires the use of different types of data (DNA sequences, climatic and occurrence data, species traits) and that some of these data involve an assessment at a much larger scale. This was required by the assessment of climatic preferences and relict populations, the latter entailing a comparison among specimens from three continents (Europe, Asia and Africa).

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 (609402-2020), Marie Sklodowska-Curie IOF grant (625997) and the Academy of Finland (decision no. 328895) to VD, and from the project "Barcoding-Italian-Butterflies". We thank Gianni Ciolli for his help with the wind analysis, Camille Pitteloud, Leonardo Platania, Charline Masson, Marco Infusino and Enrico Schifani for collaborating in field collections, and Isaac Kay-Lavelle for revising the English. We thank Simone Fattorini, Niklas Wahlberg and two anonymous referees for their constructive suggestions.

Author Contributions

LD, SS, AC, AB, MM and RVi designed the research, LD, SS, AC, MM and RVo performed the research, LD, SS, AC, AB, MM, VD, RVo, RVi and LPC collected and analysed data, all the authors wrote the paper.

Data accessibility

All the data and scripts used for this paper are available in Appendix S2 and in Dryad DOI: https://doi.org/10.5061/dryad.5qfttdz1x. The COI dataset is also available in the BOLD dataset DS-MESSINA.

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*, 2762–2772.

Antonioli, F., Presti, V. L., Morticelli, M. G., Bonfiglio, L., Mannino, M. A., Palombo, M. R., … Lambeck, K. (2016). Timing of the emergence of the Europe–Sicily bridge (40–17 cal ka BP) and its implications for the spread of modern humans. *Geological Society, London, Special Publications*, *411*, 111–144.

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

Brower, A. V. (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*, *91*, 6491–6495.

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

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

Cesaroni, D., Lucarelli, M., Allori, P., Russo, F., & Sbordoni, V. (1994). Patterns of evolution and multidimensional systematics in graylings (Lepidoptera: *Hipparchia*). *Biological Journal of the Linnean Society*, *52*, 101–119.

Clement, M., Posada, D., & Crandall, K. A. (2000). TCS: a computer program to estimate gene genealogies. *Molecular Ecology*, *9*, 1657–1659.

Collins, R. A., Boykin, L. M., Cruickshank, R. H., & Armstrong, K. F. (2012). Barcoding's next top model: an evaluation of nucleotide substitution models for specimen identification. *Methods in Ecology and Evolution*, *3*, 457–465.

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

Dapporto, L., & Bruschini, C. (2012). Invading a refugium: post glacial replacement of the ancestral lineage of a Nymphalid butterfly in the West Mediterranean. *Organisms Diversity & Evolution*, *12*, 39–49.

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*, 1169–1181.

Dapporto, L., Cini, A., Vodă, R., Dincă, V., Wiemers, M., Menchetti, M., … Bonelli, S. (2019). Integrating three comprehensive datasets shows that mitochondrial DNA variation is linked to species traits and paleogeographic events in European butterflies. *Molecular Ecology Resources*, 19, 1623-1636.

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*, 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*, 121–124.

Dincă, V., Dapporto, L., & Vila, R. (2011). A combined genetic-morphometric analysis unravels the complex biogeographical history of *Polyommatus icarus* and *Polyommatus celina* Common Blue butterflies. *Molecular Ecology*, *20*, 3921–3935.

Dincă, V., Lee, K. M., Vila, R., & Mutanen, M. (2019). The conundrum of species delimitation: a genomic perspective on a mitogenetically super-variable butterfly. *Proceedings of the Royal Society B*, *286*, 20191311.

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. *Proceedings of the Royal Society B*, *282*, 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*, 4541–4550.

Husemann, M., Schmitt, T., Zachos, F. E., Ulrich, W., & Habel, J. C. (2014). Palaearctic biogeography revisited: evidence for the existence of a North African refugium for Western Palaearctic biota. *Journal of Biogeography*, *41*, 81–94.

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

Marske, K. A., Rahbek, C., & Nogués-Bravo, D. (2013). Phylogeography: spanning the ecology-evolution continuum. *Ecography*, *36*, 1169–1181.

Mende, M. B., & Hundsdoerfer, A. K. (2013). Mitochondrial lineage sorting in action– historical biogeography of the *Hyles euphorbiae* complex (Sphingidae, Lepidoptera) in Italy. *BMC Evolutionary Biology*, *13*, 83.

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*, *32*, 1609-1619.

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*, 558–567.

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*, *276*, 1235-1244.

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*, 54.

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

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*, 8018–8024.

Parmesan, C., Ryrholm, N., Stefanescu, C., Hill, J. K., Thomas, C. D., Descimon, H., … Tammaru, T. (1999). Poleward shifts in geographical ranges of butterfly species associated with regional warming. *Nature*, *399*, 579–583.

Pigot, A. L., & Tobias, J. A. (2013). Species interactions constrain geographic range expansion over evolutionary time. *Ecology Letters*, *16*, 330–338.

Pigot, A. L., & Tobias, J. A. (2014). Dispersal and the transition to sympatry in vertebrates. *Proceedings of the Royal Society B*, *282*, 20141929.

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*, 261.

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

Srivathsan, A., & Meier, R. (2012). On the inappropriate use of Kimura-2-parameter (K2P) divergences in the DNA-barcoding literature. *Cladistics*, *28*, 190–194.

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

van Etten, J. (2017). *R* Package gdistance : Distances and Routes on Geographical Grids. *Journal of Statistical Software*, *76*.

Vodă, R., Dapporto, L., Dincă, V., Shreeve, T. G., Khaldi, M., Barech, G., … Vila, R. (2016). Historical and contemporary factors generate unique butterfly communities on islands. *Scientific Reports*, *6, 28828.*

Vodă, R., Dapporto, L., Dincă, V., & Vila, R. (2015a). Cryptic matters: overlooked species generate most butterfly beta-diversity. *Ecography*, *38*, 405–409.

Vodă, R., Dapporto, L., Dincă, V., & Vila, R. (2015b). Why do cryptic species tend not to co-occur? A case study on two cryptic pairs of butterflies. *PLOS ONE*, *10*, e0117802.

Waters, J. M., Fraser, C. I., & Hewitt, G. M. (2013). Founder takes all: density-dependent processes structure biodiversity. *Trends in Ecology & Evolution*, *28*, 78–85.

Weingartner, E., Wahlberg, N., & Nylin, S. (2006). Speciation in *Pararge* (Satyrinae: Nymphalidae) butterflies–North Africa is the source of ancestral populations of all *Pararge* species. *Systematic Entomology*, *31*(4), 621–632.

Whittaker, R. J., & Fernández-Palacios, J. M. (2007). *Island biogeography: ecology, evolution, and conservation*. Oxford University Press.

Whittaker, R. J., Fernández-Palacios, J. M., Matthews, T. J., Borregaard, M. K., & Triantis, K. A. (2017). Island biogeography: Taking the long view of nature's laboratories. *Science*, *357*, eaam8326.

Whittaker, R. J., Triantis, K. A., & Ladle, R. J. (2008). A general dynamic theory of oceanic island biogeography. *Journal of Biogeography*, *35*, 977–994.

Wilson, R. J., Davies, Z. G., & Thomas, C. D. (2009). Modelling the effect of habitat fragmentation on range expansion in a butterfly. *Proceedings of the Royal Society B*, *276*, 1421–1427.

Wilting, A., Sollmann, R., Meijaard, E., Helgen, K. M., & Fickel, J. (2012). Mentawai's endemic, relictual fauna: is it evidence for Pleistocene extinctions on Sumatra? *Journal of Biogeography*, *39*, 1608–1620.