ORIGINAL ARTICLE

MOLECULAR ECOLOGY WILEY

Hybridization fuelled diversification in *Spialia* butterflies

Joan C. Hinojosa[1] | Leonardo Dapporto[2] | Camille Pitteloud[3] |
Darina Koubínová[4] | Juan Hernández-Roldán[5] | Juan Carlos Vicente[6] |
Nadir Alvarez[3,7] | Roger Vila[1]

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

[2]ZEN Laboratory, Biology Department, Università degli Studi di Firenze, Sesto Fiorentino, Italy

[3]Geneva Natural History Museum, Geneva, Switzerland

[4]Laboratory of Evolutionary Genetics, Institute of Biology, University of Neuchâtel, Neuchâtel, Switzerland

[5]Departamento de Biología, Centro de Investigación en Biodiversidad y Cambio Global (CIB-UAM), Madrid, Spain

[6]Asociación Española para la Protección de las Mariposas y su Medio (ZERYNTHIA), Madrid, Spain

[7]Department of Genetics and Evolution, University of Geneva, Geneva, Switzerland

Correspondence
Roger Vila, Institut de Biologia Evolutiva (CSIC-UPF), Passeig Marítim de la Barceloneta 37-49, Barcelona, Spain.
Email: roger.vila@csic.es

Abstract

The importance of hybridization and introgression is well documented in the evolution of plants but, in insects, their role is not fully understood. Given the fact that insects are the most diverse group of organisms, assessing the impact of reticulation events on their evolution may be key to comprehend the emergence of such remarkable diversity. Here, we used an insect model, the *Spialia* butterflies, to gather genomic evidence of hybridization as a promoter of novel diversity. By using double-digest RADseq (ddRADseq), we explored the phylogenetic relationships between *Spialia orbifer*, *S. rosae* and *S. sertorius*, and documented two independent events of interspecific gene flow. Our data support that the Iberian endemism *S. rosae* probably received genetic material from *S. orbifer* in both mitochondrial and nuclear DNA, which could have contributed to a shift in the ecological preferences of *S. rosae*. We also show that admixture between *S. sertorius* and *S. orbifer* probably occurred in Italy. As a result, the admixed Sicilian populations of *S. orbifer* are differentiated from the rest of populations both genetically and morphologically, and display signatures of reproductive character displacement in the male genitalia. Additionally, our analyses indicated that genetic material from *S. orbifer* is present in *S. sertorius* along the Italian Peninsula. Our findings add to the view that hybridization is a pervasive phenomenon in nature and in butterflies in particular, with important consequences for evolution due to the emergence of novel phenotypes.

KEYWORDS
butterflies, ddRADseq, hybridization, phylogeography, reproductive character displacement

1 | INTRODUCTION

Reticulation events, namely hybridization (the production of viable offspring from distinct species) and introgression (the transfer of genes between species as a result of hybridization and repeated backcrossing), are increasingly recognized as drivers of evolution. One of the most important outcomes of reticulation is that the resulting hybrids can display distinct phenotypes compared to the parental species,

including novelties in morphology (Dasmahapatra et al., 2012), behaviour (Pärssinen et al., 2020) or habitat preferences (Rieseberg et al., 2007). This can increase their fitness in particular environments (Hedrick, 2013; Suarez-Gonzalez et al., 2018) and/or allow to colonize previously unavailable niches (Gross & Rieseberg, 2005). In this scenario, several authors (e.g., Abbott et al., 2013; Hedrick, 2013; Suarez-Gonzalez et al., 2018) regard interspecific gene flow not only as a unifying force, but rather as a source of adaptation and divergence between populations. Abbott et al. (2013) argued that, considering how common hybridization is and how rare mutations are, introgression could be a major source of adaptive genetic variation, and adaptations would, in turn, enhance differences between populations.

Despite the relevant impact that reticulation processes may have in evolution, its role as a source of new diversity has not been clarified for many groups of organisms. This issue has been extensively assessed in plants (Hegarty & Hiscock, 2005), but its importance in other eukaryotic organisms, including animals (Mavárez & Linares, 2008), was mostly neglected until the current century. Even assuming that hybridization is more widespread in plants than in animals—25% and 10% of plant and animal species are known to hybridize, respectively (Mallet, 2005)—still more species would be affected by reticulate evolution in animals in absolute numbers. This is particularly striking in insects, since they constitute the most diverse group of organisms. Nevertheless, the introduction of genetic techniques catalysed a paradigm shift and insect groups such as butterflies (superfamily Papilionoidea) are becoming a popular model to assess reticulate evolution.

Butterflies have two characteristics that make them particularly suitable to study reticulation processes. First, a robust taxonomic framework exists for them and large-scale studies in systematics have been conducted, especially in Europe (Dincă et al., 2021; Wiemers et al., 2018, 2020). The presence of a reliable classification of the study species is an essential step that facilitates the design of genetic studies and the interpretation of genetic results. Second, the chromosomes of Lepidoptera are holocentric (Mandrioli & Manicardi, 2020; Melters et al., 2012), that is, they lack a centromeric region that concentrates all kinetochores, and the possibility of inverted meiosis has been documented for some species of butterflies (Lukhtanov et al., 2018, 2020). These characteristics facilitate chromosome pairing during meiosis, so that hybrid offspring may be produced to a certain degree by parents with different chromosome numbers—a situation that can be even found between individuals of the same species (de Lesse, 1969; Lukhtanov et al., 2011). As a consequence, butterflies (superfamily Papilionoidea) are becoming models for blooming research fields such as hybrid speciation (e.g., Capblancq et al., 2015; Nice et al., 2013), hybrid zones (e.g., Mallet et al., 2011; Mullen et al., 2008) and adaptive introgression (e.g., Enciso-Romero et al., 2017; Pardo-Diaz et al., 2012). However, the evolutionary outcomes of interspecific gene flow and its weight along all the butterfly tree of life remain widely unclear and more research efforts covering all the families are necessary.

The genus *Spialia* (Hesperiidae) is one of the butterfly groups in which interspecific gene flow could have played a key role in shaping its diversity. Hernández-Roldán et al. (2016) revised this butterfly genus and described *Spialia rosae*, a species endemic to the Iberian Peninsula. Based on morphology and on two nuclear DNA (nuDNA) markers (wingless and the internal transcribed spacer 2), *S. rosae* appears virtually indistinguishable from *S. sertorius* (Hernández-Roldán et al., 2018), a species distributed in western Europe (Tolman & Lewington, 2008; Tshikolovets, 2011). Nevertheless, these closely related and widely sympatric species harbour important differences in ecology, in *Wolbachia* infection rates and in their mitochondrial DNA (mtDNA). Concerning ecology, they differ in larval host plant and habitat: *S. rosae* feeds on *Rosa* spp. and is associated with mountain habitats, while *S. sertorius* feeds on *Poterium* spp. and can be found both in lowland and highland (Hernández-Roldán et al., 2016). Regarding *Wolbachia* infection rates, all *S. rosae* individuals studied by Hernández-Roldán et al. (2016) were infected by *Wolbachia*, while all the *S. sertorius* were uninfected. *Wolbachia* is a maternally inherited endosymbiont that may cause cytoplasmic incompatibility and thus constitute a partial reproductive barrier between infected and noninfected organisms (Hurst & Jiggins, 2005; Jiggins, 2003; Werren et al., 2008). In the mtDNA, the cytochrome c oxidase subunit 1 (*COI*) of *S. rosae* formed a lineage clustered within *S. orbifer*, a species distributed from Sicily and eastern Europe to central Asia (Tolman & Lewington, 2008; Tshikolovets, 2011), with populations frequently infected with *Wolbachia*, and with an ecology similar to *S. sertorius* (Hernández-Roldán et al., 2016). In summary, morphology and the nuDNA suggested that *S. rosae* and *S. sertorius* had a relatively recent last common ancestor, whereas the mtDNA and the *Wolbachia* infection pattern suggested potential introgression from *S. orbifer* to *S. rosae*—albeit their distributions are currently separated by a large geographical distance.

In this study, we hypothesized that the genus *Spialia* experienced hybridizations at the contact zones, which could have influenced evolution and fostered diversification. Special attention was given to the Sicilian population of *S. orbifer*, as it constitutes an isolated population spatially close to *S. sertorius*. To shed light on the potential role of hybridization on the diversification of the group, we used genetic data obtained through double-digest RADseq (ddRADseq) to (1) retrieve the phylogenetic relationships between *S. rosae*, *S. sertorius*, *S. orbifer* and their closest relatives, (2) examine their genetic structure, and (3) test for the existence of reticulation processes. Finally, we analysed whether these events left a measurable effect in the morphology of male wings and genitalia, as these are characters frequently involved in sexual selection and reproductive isolation in butterflies.

2 | MATERIALS AND METHODS

2.1 | Sampling

We analysed DNA data from 67 samples including 18 *S. orbifer* (4 from Sicily), 15 *S. rosae*, 23 *S. sertorius*, seven *S. ali* and four *S. therapne*. Morphological data of genitalia was retrieved from 22 *S. orbifer*

(five from Sicily), 28 *S. rosae* and 48 *S. sertorius* male specimens. Data for morphometric analysis of wing patterns was obtained from 31 *S. orbifer* (six from Sicily), 28 *S. rosae* and 60 *S. sertorius* specimens. Many of the specimens used for DNA analyses were also used for the morphological measurements (Table S1). Butterfly bodies were stored in 99% ethanol at −20°C and wings were kept separately as vouchers.

2.2 | Analyses of mitochondrial DNA sequences

The barcode fragment of the *COI* was used for molecular dating analyses, specifically to estimate the time when the mtDNA lineage of *S. rosae* split from the current *S. orbifer* clades. The barcode sequences available in GENBANK for the studied species (Table S1) were downloaded and aligned in GENEIOUS version 11.0.5 (Kearse et al., 2012). The best fitting model was found using JMODELTEST version 2.1.7 (Darriba et al., 2012) under the Bayesian information criterion and a Bayesian phylogeny was reconstructed in BEAST version 2.6.1 (Bouckaert et al., 2014). Base frequencies were estimated, four gamma rate categories were selected, and a randomly generated initial tree was used. Estimates of node ages were obtained by applying a strict clock and a normal prior distribution centred on the mean between two substitution rates for invertebrates: 1.5% and 2.3% uncorrected pairwise distance per million years—Quek et al. (2004) and Brower (1994), respectively. The standard deviation was modified so that the 95% confidence interval of the posterior density coincided with the 1.5% and 2.3% rates. Albeit these substitution rates provide very approximate divergence estimates, better calibrations are, as far as we know, unavailable for these taxa due to the absence of fossils or other phylogenetically close calibration points. Parameters were estimated using two independent runs of 20 million generations each, and convergence was checked using TRACER 1.7.1 (Rambaut et al., 2018) with a 10% burnin applied.

2.3 | ddRADseq library preparation

DNA was extracted from half thoraxes using a Qiagen DNeasy Blood & Tissue Kit (Qiagen) following the protocol recommended by the manufacturer, and it was eluted in 50 μl of Buffer EB (Qiagen). DNA concentrations ranged between 24 and 114 ng/μl as quantified with a Qubit 2.0 fluorometer (Life Technologies). Six microlitre of each sample were digested with MseI and SbfI restriction enzymes, and P1 (24 different barcodes) + P2 (the same sequence for all samples) adaptors were ligated. Samples were purified with 0.8× AMPure XP magnetic beads (Beckman-Coulter). Following a double indexing approach, samples were individually amplified and tagged using Illumina sequencing primers. Two replicates per sample were used to reduce PCR bias, and corresponding PCR products were then pooled. All the individual libraries were further pooled in equimolar ratio and cleaned with 1× AMPure XP beads. After quantification and quality check of the libraries pool with Fragment

Analyser (Agilent), fragments at around 325 bp were selected using Pippin Prep (Sage Science). A final cleaning step was conducted with 1× AMPure XP beads. Finally, the libraries were sequenced using HiSeq 2500 100 bp paired-end sequencing at Lausanne Genomic Technologies Facility (Lausanne, Switzerland).

2.4 | Read assembly and SNP calling

Demultiplexing of the raw reads was performed using the process_radtags tool of the Stacks software PIPELINE version 2.2 (Catchen et al., 2011). Further read processing (quality filtering, de novo assembly, construction of reference sequence, read mapping and SNP calling) was done using the DDOCENT version 2.2.25 (Puritz et al., 2014) pipeline since, compared to other pipelines, it was faster and recovered more usable loci (Thallinger, 2017). Contaminant DNA fragments matching noninsect organisms (such as microbionts) were detected in the reference sequence with CENTRIFUGE version 1.0.3 (Kim et al., 2016). All contaminant fragments were then removed from the data, thus conserving only fragments assigned to Insecta or with no match in the GENBANK database. A second Centrifuge run was performed to verify that no contaminant fragments remained in the reference after removal. Both outputs (pre- and post-removal) were visualized in the R package PAVIAN version 1.0.0 (Breitwieser & Salzberg, 2020). Potential mitochondrial loci were identified using BLAST+ version 2.7.1 (Camacho et al., 2009) and discarded, since several were detected (up to 49) and hence their inclusion could have heavily influenced the results. The resulting SNP data set was filtered to remove the low-quality loci and/or individuals with a large amount of missing data using VCFTOOLS version 0.1.15 (Danecek et al., 2011) based on recommendations in the dDocent filtering tutorial with slight modifications, that is, by adding a second round of additional filtering of individuals and loci showing still high number of missing data at the end of the process.

Raw reads were additionally processed using IPYRAD version 0.7.23 (Eaton & Overcast, 2016) to obtain loci alignments, essential for some analyses. The software was run for two data sets: one with all the samples, and another one with only *S. rosae* and *S. sertorius*, in order to optimize the number of shared loci between the two species. The following parameters were changed from the default settings for both data sets: datatype was set to ddRADseq, restriction overhang to TGCAG, TAA, clustering threshold to 0.88, minimum trimmed length to 60, maximum Ns to 2, maximum heterozygous bases to 5, minimum number of samples with a given locus to 6, maximum SNPs per locus to 15 and maximum indels per locus to 5. Contaminant loci/SNPs were also removed with CENTRIFUGE version 1.0.3 as described above. Mitochondrial loci were identified using BLAST+ version 2.7.1 but not excluded, since there were only two in the raw data set.

2.5 | Phylogenetic analyses of the ddRADseq data

A phylogeny based on ddRADseq data was used to infer the relationships between *S. rosae* and the other taxa. It was constructed using a

maximum likelihood inference with RAXML version 8.2.4 (Stamatakis, 2014) and with the dDocent SNPs alignment (752 SNPs). The Lewis correction, a GTRGAMMA model and 1,000 bootstrap replicates were applied. We visualized and exported the resulting phylogeny using FIGTREE version 1.4.2 (Rambaut, 2015).

The same data set was employed in a species-based phylogeny performed with SNAPP version 1.5.0 (Bryant et al., 2012), a package implemented in BEAST version 2.6.1, with the aim of dating the divergence time between *S. rosae* and *S. sertorius*, and between *S. orbifer* and its Sicilian population. Thus, we used only the individuals of *S. orbifer*, *S. sertorius* and *S. rosae*, and we introduced an age of 2.55 million years for the crown of the tree, i.e. the divergence time calculated between *S. orbifer* and *S. sertorius* (Ebdon et al., 2021). We selected an MCMC chain length of 250,000 and sampled every 250 generations. Two independent runs were carried out and convergence was checked in TRACER version 1.7.1. A 10% burnin was applied and results from both runs were merged.

2.6 | Genetic structuring and genetic distances

To gather detailed information on the genetic structure in the target species, both a STRUCTURE analysis and a PCA were performed. In STRUCTURE version 2.3.4 (Pritchard et al., 2000), we tested *K* values from 1 to 10 using the dDocent-filtered SNPs data set (752 SNPs) including all the species and from 1 to 6 in the same data set but including only *S. orbifer*, *S. rosae* and *S. sertorius*. The selected burnin was 100,000, followed by 2,000,000 MCMC replicates. Ten runs were done for each *K*, which were afterwards combined in one per *K* and plotted with CLUMPAK version 1.1 (Kopelman et al., 2015) and DISTRUCT version 1.1 (Rosenberg, 2004). The best *K* was obtained from STRUCTURE HARVESTER version 0.6.94 (Earl, 2012).

We further performed two PCA using the R software package ADEGENET version 1.4-1 (Jombart et al., 2010) with the dDocent-filtered data set (752 SNPs). In the first one, all the samples were included; in the second, in order to visualize better potential differences between *S. rosae* and *S. sertorius*, only these two species were included.

Along with information about the genetic structure, the estimation of genetic distances was used to gather evidence related to the evolutionary history of populations. Average genetic distances (d_{XY}) were calculated with GENEIOUS version 11.0.5 using the ipyrad loci alignment (1,918 concatenated loci, two mitochondrial); from these data, average intraspecific distances (d_X) and net genetic distances (d_A) were also calculated. The results were plotted with the R package COMPLEXHEATMAP version 2.7.6.1004 (Gu et al., 2016).

2.7 | Gene flow calculations

Gene flow events were explored in TREEMIX version 1.13 (Pickrell & Pritchard, 2012) using the dDocent data set and only including *S. ali* as outgroup. This approach allows inferring historical relationships between populations, based on covariance of allele frequencies and Gaussian approximation to genetic drift. The Python script vcf2treemix.py (available in https://github.com/CoBiG2/RAD_Tools/blob/master/vcf2treemix.py) was used to generate the TreeMix input file. Interspecific introgression was tested using 1, 2, 3, and 4 migration edges (m) and using 11 SNPs windows (k) per migration edge, from 10 to 20. TreeMix additionally calculates the weight of the migration events which is, if admixture occurs in a single generation, the fraction of alleles in the descendant population that originated in each parental population. The R package OPTM (Fitak, 2021) was used to select the optimal number of migration edges.

Additional potential reticulation events were investigated using the Julia package PHYLONETWORKS version 0.12.0 (Solís-Lemus et al., 2017). Here, the ipyrad data set without contaminants was used since loci alignments are required. This data set was filtered to include the species *S. rosae*, *S. sertorius* and *S. orbifer* and a minimum of 12 individuals; the final data set included one mitochondrial locus of 88 bp. This method uses a maximum pseudolikelihood estimator applied to quartet concordance factors (CF), that is, gene tree frequencies, of 4-taxon trees under the coalescent model, incorporating incomplete lineage sorting (ILS) and reticulation events (Solís-Lemus et al., 2017). The observed CFs are used to estimate reticulation events and inheritance values (γ) indicating the proportion of ancestral contribution to the hybrid lineage genome. To estimate individual gene trees for each locus, we used RAXML version 8.2.4 using the GTRGAMMA model and 100 bootstrap replicates per locus. Then, we used PhyloNetworks to read all RAxML gene-trees (506 trees including a minimum of six individuals) and estimate CFs, with all individuals per clade mapped as alleles to species. As starting topology, we used a tree built in ASTRAL version 5.7.5 (Rabiee et al., 2019; Zhang et al., 2018) using the RAxML loci trees and with the default parameters. We tested values for h (number of reticulations) from 0 to 4, assessing maximum support using a slope heuristic for the increase in likelihood plotted against h (Solís-Lemus & Ané, 2016). We ran 50 independent runs per h-value to ensure convergence on a global optimum.

We used the ABBA-BABA analysis to detect significant introgression (Durand et al., 2011) and infer admixture rates of the reticulations indicated by TreeMix and Phylonetworks. Thus, we sought testing the presence of admixture between the pairs (1) *S. rosae*-*S. orbifer* and (2) the Italian *S. sertorius*-Sicilian *S. orbifer*. In (1), only the non-Italian *S. sertorius* were selected as P1 and the non-Sicilian *S. orbifer* as P3, since the Sicilian *S. orbifer* and the Italian *S. sertorius* showed signatures of introgression. In (2), we selected all the non-Italian *S. sertorius* specimens as P1, the Italian *S. sertorius* as P2 and the Sicilian *S. orbifer* as P3. *Spialia ali* was used as outgroup in all the scenarios. Calculations were done with the software DTRIOS, included in DSUITE version 0.4 (Malinsky et al., 2020), with the dDocent-filtered data set (752 SNPs). Dtrios calculates the Patterson's D (Patterson et al., 2012) and admixture fraction (f_4-ratio) of the mixing event (Patterson et al., 2012). To assess whether D is significantly different from zero, Dtrios uses a standard block-jackknife procedure. The size of blocks should be chosen to be larger than the extent

of linkage disequilibrium (Durand et al., 2011). Here, every dDocent loci contributed with an average of 3.3 SNPs and we thus tested blocks of 10 and 50 SNPs.

2.8 | Detecting the most divergent loci between *S. rosae* and *S. sertorius*

From the ipyrad raw output of the data set including only *S. rosae* and *S. sertorius* (1482 loci), we searched for the most divergent loci between both taxa by applying BAYESCAN version 2.1 (Foll & Gaggiotti, 2008). The default parameters were used. We selected the loci with highest F_{ST} values that were statistically significant (the outliers), and analysed these with two procedures. First, with the aim of finding candidate genes involved in the differential traits between *S. rosae* and *S. sertorius*, we manually revised (using GENEIOUS) the outlier loci and searched for those that displayed species-specific alleles in at least one position. These loci were identified in NCBI BLAST web server (Johnson et al., 2008). For the second procedure, we additionally searched for loci with nearly-exclusive haplotypes—allowing a maximum of five individuals with shared haplotypes between species and representing less than 15% of the samples—between *S. rosae* and *S. sertorius*. We searched for both loci type, with exclusive and nearly-exclusive haplotypes, whether they also had sequences from *S. orbifer* in the ipyrad data set including all the species. With these loci, we constructed a phylogeny using RAxML with a GTRGAMMA model and 1,000 bootstrap replicates. The mitochondrial locus was discarded from this phylogenetic analysis. This second procedure was done to assess if the most distinct loci between *S. rosae* and *S. sertorius* were especially similar between *S. rosae* and *S. orbifer*. The alignment used to construct this phylogeny has been deposited in figshare (https://doi.org/10.6084/m9.figshare.16810459).

2.9 | *Wolbachia* infection analysis

Wolbachia bacteria are maternally inherited endosymbionts responsible for several effects on hosts, such as feminization, induced parthenogenesis, male killing and cytoplasmic incompatibility (Hurst & Jiggins, 2005; Jiggins, 2003; Werren et al., 2008). As a consequence, the infection may trigger selective sweeps resulting in the association between a mitochondrial genome and a *Wolbachia* strain. Thus, similarly to mtDNA, *Wolbachia* genomes might reflect the phylogeographic history of their host species and are useful to detect vicariance and introgression, for instance. *Wolbachia* sequences were identified from the raw ddRADseq loci by IPYRAD using CENTRIFUGE version 1.0.4 and extracted using a custom script.

2.10 | Morphological analyses

A combination of landmarks and sliding semi-landmarks (Bookstein, 1997) was applied to the outline of the left valva of the male genitalia and to the vein junctions and the outline of the white spots of the male hindwing underside. Points that could be precisely identified were considered as landmarks, whereas other points were allowed to slide along the outline trajectory (sliding semi-landmarks). Following a previous study on the same species group (Hernández-Roldán et al., 2016), we identified eight landmarks and 24 semi-landmarks on the valva and 35 landmarks on the hindwings (Figure S1). A generalized Procrustes analysis (GPA) was applied to the landmark data to remove nonshape variation and to superimpose the objects in a common coordinate system. Partial warps were calculated using the shape residuals from GPA. Applying PCA to partial warps, relative warps (PCs) were obtained and used as variables in subsequent analyses (Bookstein, 1997).

Bidimensional representations for morphology variation of wings and genitalia were obtained by applying partial least square discriminant analyses (PLSDA) to the relative warps. As a grouping variable, we aggregated specimens based on species membership as identified by a previous study (Hernández-Roldán et al., 2016). Since relative warps can be particularly numerous, overfitting had to be avoided. We thus applied a sparse PLSDA (Lê Cao et al., 2011) only selecting the most influential five variables for each resulting component. To evaluate the degree of diversification as a percentage of cases that can be blindly attributed to their group, we applied a jackknife (leave-one-out) algorithm and individually classified each specimen. These analyses were carried out with the splsda and predict functions in the MIXOMICS version 6.15.0 R package (Rohart et al., 2017). We evaluated if specific populations—namely the Sicilian *S. orbifer*, the *S. sertorius* sympatric with *S. rosae* (collection sites closer than 20 km), and *S. sertorius* from the contact with *S. orbifer* at the Messina strait—were particularly divergent with respect to individuals from other localities. For this purpose, we obtained the distance between each specimen to the PLSDA barycentre of its species and this value was then compared among populations by using Mann–Whitney and Kruskal–Wallis tests.

To evaluate the relationship between morphological and spatial distances between specimens belonging to different species, we calculated overground distances among pairs of non-conspecific specimens. With this aim, we obtained the average temperature climatic layer from WORDCLIM database (Fick & Hijmans, 2017; http://worldclim.org) at a resolution of 5 min; then, we attributed a permeability of 1 to sea areas and a permeability of 1,000 to ground areas, and distances were obtained by calculating the length of minimum cost paths among pairs of specimens using the COSTDISTANCE function implemented in the GDISTANCE version 1.2-1 R package (van Etten, 2017). We calculated, in STATS version 3.6.2 R package (R Core Team, 2021), the Spearman's ρ between morphological Procrustes distances for wings and genitalia between specimens belonging to different species and the overground distances. Since pairwise data are not independent, we obtained *p*-values by resampling distances 1,000 times and calculating the frequency of absolute values of ρ being higher than the absolute number of observed ρ. Other than Spearman's ρ, we calculated a generalized additive model (GAM) for morphological distances against a smooth factor ($k = 4$) of spatial distances using

the gam function of the MGCV version 1.8-33 R package. This allowed detecting curvilinear trends in the relationships.

3 | RESULTS

3.1 | ddRADseq data sets

The number of loci and SNPs of each data set and the analyses in which they have been used are indicated in Table S2. All the data sets were deposited in figshare (https://doi.org/10.6084/m9.figsh are.16810459).

3.2 | Discordant mitochondrial and RADseq phylogenies

The ddRADseq phylogeny (Figure 1) retrieved all the species as monophyletic. Supports were high except for *S. rosae* (bootstrap = 53) and *S. sertorius* (bootstrap = 58). These two species were recovered as sister species (bootstrap = 100), thus contrasting with previous results based on the *COI* mtDNA marker in which *S. rosae* was embedded within *S. orbifer*. In the mitochondrial phylogeny (Figure S2), the split between the lineage of *S. rosae* and *S. orbifer* occurred c. 0.61 (0.37–0.89) million years ago (Ma). The application of a molecular clock to the ddRADseq data set using SNAPP estimated a split between *S. rosae* and *S. sertorius* at c. 0.28 Ma. The specimens of *S. orbifer* from Sicily formed a clade well differentiated from the rest of conspecific specimens (bootstrap = 100). SNAPP dated the divergence between Sicilian and mainland *S. orbifer* at c. 1.15 Ma. These divergence times are, however, expected to be affected by interspecific gene flow (see more details in the Discussion). Within *S. sertorius*, the individuals from the Italian Peninsula grouped together and formed a relatively long branch (bootstrap = 100).

3.3 | Genetic structuring of the focal species highlights the Sicilian *S. orbifer* as a diverged taxon

When applied to the data set with all the specimens (dDocent data set), K = 3 had the highest ΔK value (8496.1) in STRUCTURE HARVESTER. In this K, three groups emerged (Figure 2): *S. ali* + *S. therapne*, *S. orbifer* and *S. rosae* + *S. sertorius*. When looking for higher values of K, these groupings remained firm while indicating further diverging layers: for instance, when applying K = 4, the new genetic cluster was shared almost exclusively by Sicilian *S. orbifer* (with a small portion in the Italian *S. sertorius*) and, with K = 5, the new cluster represented a component present in all specimens of *S. rosae*. Distinct clusters for *S. ali* and *S. therapne* were only given in one of the clusters appearing with K = 9 (Figure S3). This result is not surprising given that STRUCTURE may have problems to separate poorly sampled outgroup taxa, even if they are well diverged according to phylogenetic analyses (e.g., Dupuis & Sperling, 2020; Quattrini et al., 2019). In the data set

with only *S. orbifer*, *S. rosae* and *S. sertorius* (Figure S4), K = 2 demonstrated the highest ΔK (23314.9). Here, the distinction between *S. rosae* and *S. sertorius* appeared in K = 3 in four out of 10 runs (lnP = −22876.7); in the other six runs, the Sicilian *S. orbifer* were differentiated from the rest of *S. orbifer* specimens (lnP = −22865.8). These two alternative divisions in K = 3 were retrieved simultaneously in K = 4, and in K = 5 the Italian *S. sertorius* formed an independent cluster.

The PCA (Figure 3) performed with all the species (dDocent data set) showed similar results. The PC1 vs. PC2 projection clearly distinguished the three STRUCTURE groups defined with K = 3. *Spialia ali* and *S. therapne* were highly differentiated in PC3 (Figure 3b) and just slightly in PC2 (Figure 3a). The Sicilian *S. orbifer* remained distinct to the rest in PC2. Although when all the species were included *S. rosae* and *S. sertorius* appeared mixed, they separated well in PC1 when only specimens of these two species were used in the PCA (Figure 3c,d).

Pairwise comparison of genetic distances (Figure S5) highlighted *S. therapne* as the most distinct species of the group in both d_{XY} and d_A parameters, with values that surpassed 1.5% and 1%, respectively, in all the comparisons. By contrast, *S. rosae* and *S. sertorius* was the closest species pair (d_{XY} = 0.88%; d_A = 0.02%). *Spialia orbifer* appeared to be much more differentiated from *S. rosae* (d_{XY} = 1.33%; d_A = 0.52%). The genetic distance of the Sicilian *S. orbifer* with respect to the rest of *S. orbifer* individuals (d_{XY} = 1.04%; d_A = 0.44%) was similar to those found between other species pairs. *Spialia therapne* and the Sicilian *S. orbifer* had the lowest average intragroup distances (0.41% and 0.4%, respectively) while the other taxa displayed distances at about 0.8% (Figure S5c).

3.4 | Gene flow between *Spialia* species

OptM estimated that the optimal number of migration edges was two, that is, the number of migration edges that explain 99.8% of the observed variation (Figure S6a). For two migration edges, TreeMix suggested in all the runs (hence, for all the different windows of SNPs) the occurrence of an admixture event between the Italian *S. sertorius* and the Sicilian *S. orbifer* (mean weight = 0.292, mean p-value = .005), and another one between *S. orbifer* and *S. rosae* (mean weight = 0.104, mean p-value = .020; Figure 4a).

PhyloNetworks analyses revealed that all models involving reticulation events (h > 0) fit our data better than models considering strict bifurcating trees (h = 0; Figure S6b). The best phylogenetic network inferred by PhyloNetworks identified one introgression event (hmax = 1, −loglik = 0.037) from the Sicilian *S. orbifer* to the Italian *S. sertorius* (γ = 0.168; Figure 4b).

The ABBA-BABA tests (Figure 5) showed excess of shared alleles in the two reticulation events pinpointed by TreeMix and PhyloNetworks: between *S. rosae* and *S. orbifer* (Figure 5a: D = 0.076, f_4-ratio = 0.056, p-value = .027–.097) and between the Sicilian *S. orbifer* and *S. sertorius* (Figure 5b: D = 0.228, f_4-ratio = 0.175, p-value = .0007–.004).
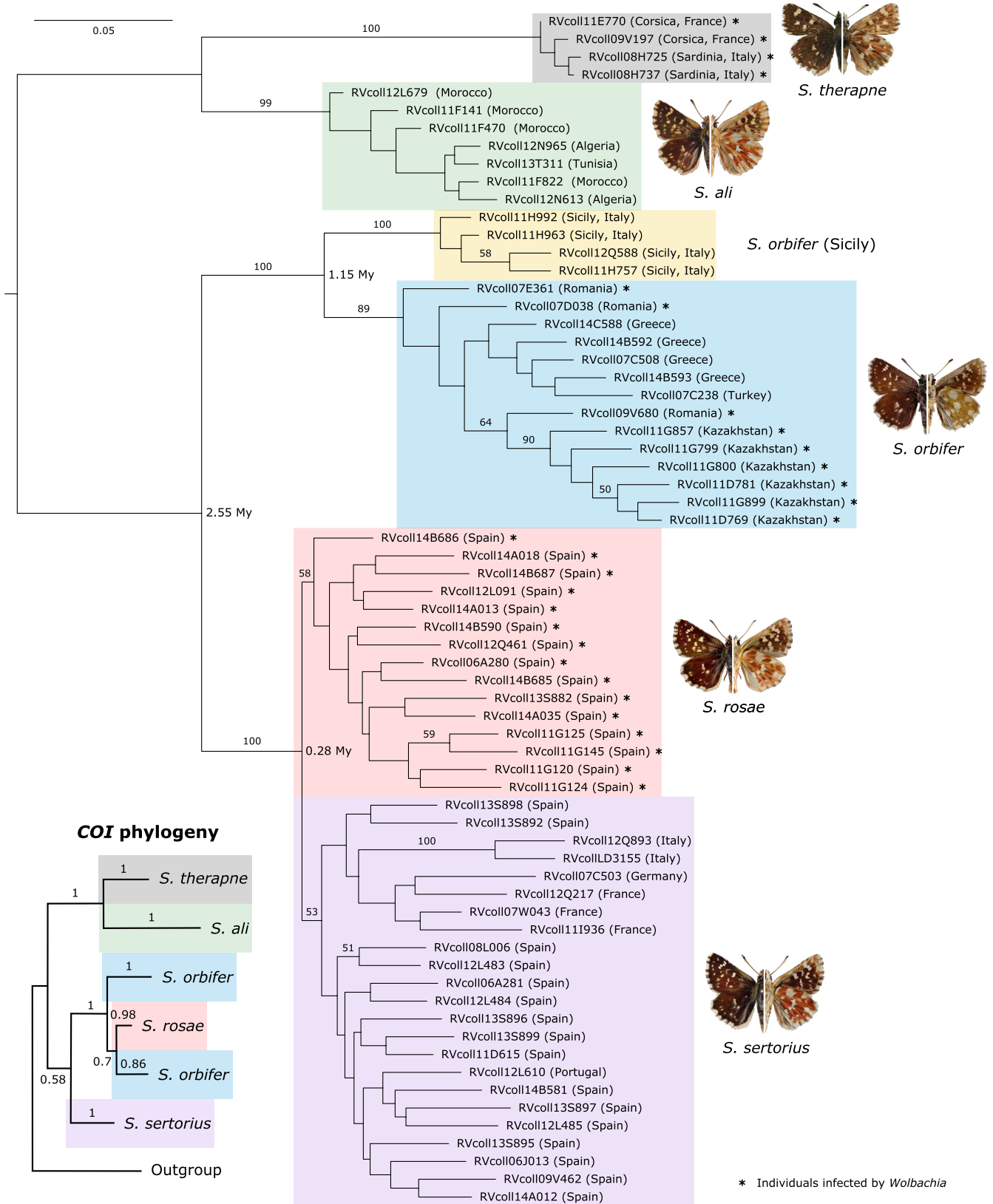
FIGURE 1 Maximum likelihood inference tree based on 752 SNPs; mitochondrial loci were not included. Node bootstrap supports are indicated >50. Individuals found to be infected by *Wolbachia* are marked with an asterisk. Scale represents 0.05 substitutions per site. The phylogeny of the *COI* gene is based on that published in Hernández-Roldán et al. (2016)
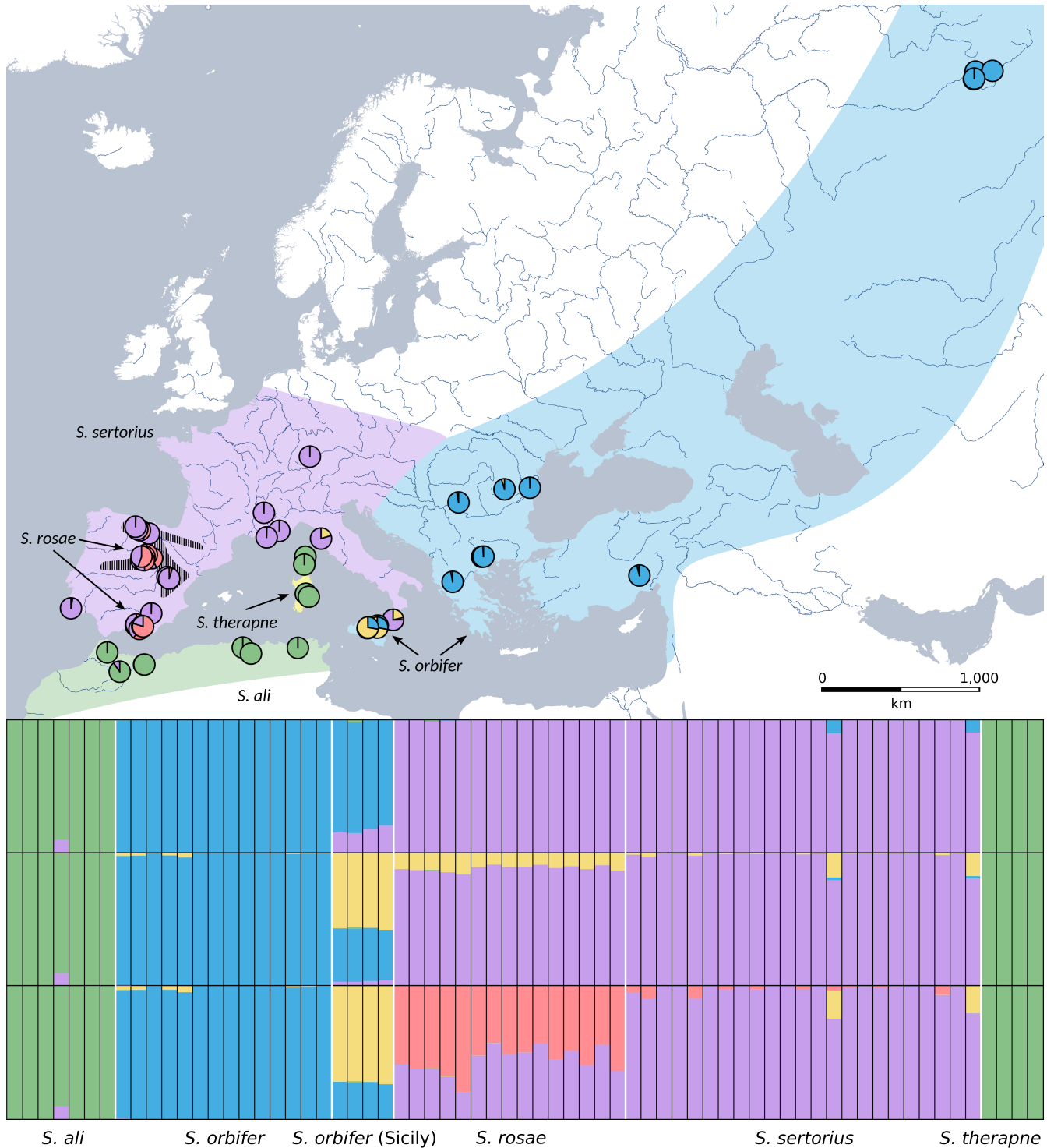
FIGURE 2 STRUCTURE results (K = 3–5) calculated with 752 SNPs; mitochondrial loci were not included. In the map, K = 5 is represented as pie charts. Approximate ranges of *Spialia ali* (green), *S. rosae* (black stripes), *S. sertorius* (purple), *S. therapne* (yellow), and *S. orbifer* (blue) are plotted

3.5 | Divergent loci between *S. rosae* and *S. sertorius*

BayeScan identified 51 outliers between *S. rosae* and *S. sertorius*, eight with species-specific exclusive haplotypes. All eight loci have been deposited in figshare (https://doi.org/10.6084/m9.figshare.16810459).

One was identified with NCBI BLAST as a fragment of the mitochondrial *COI*. Another matched (max score >50 and identities above 80%) with several unidentified sequences and with the gene dynein heavy chain 6 (*DNAH6*); when translated to amino acid, in the frame corresponding to the DNAH6 protein, all changes were synonymous and only one amino acid sequence was obtained for both species. A
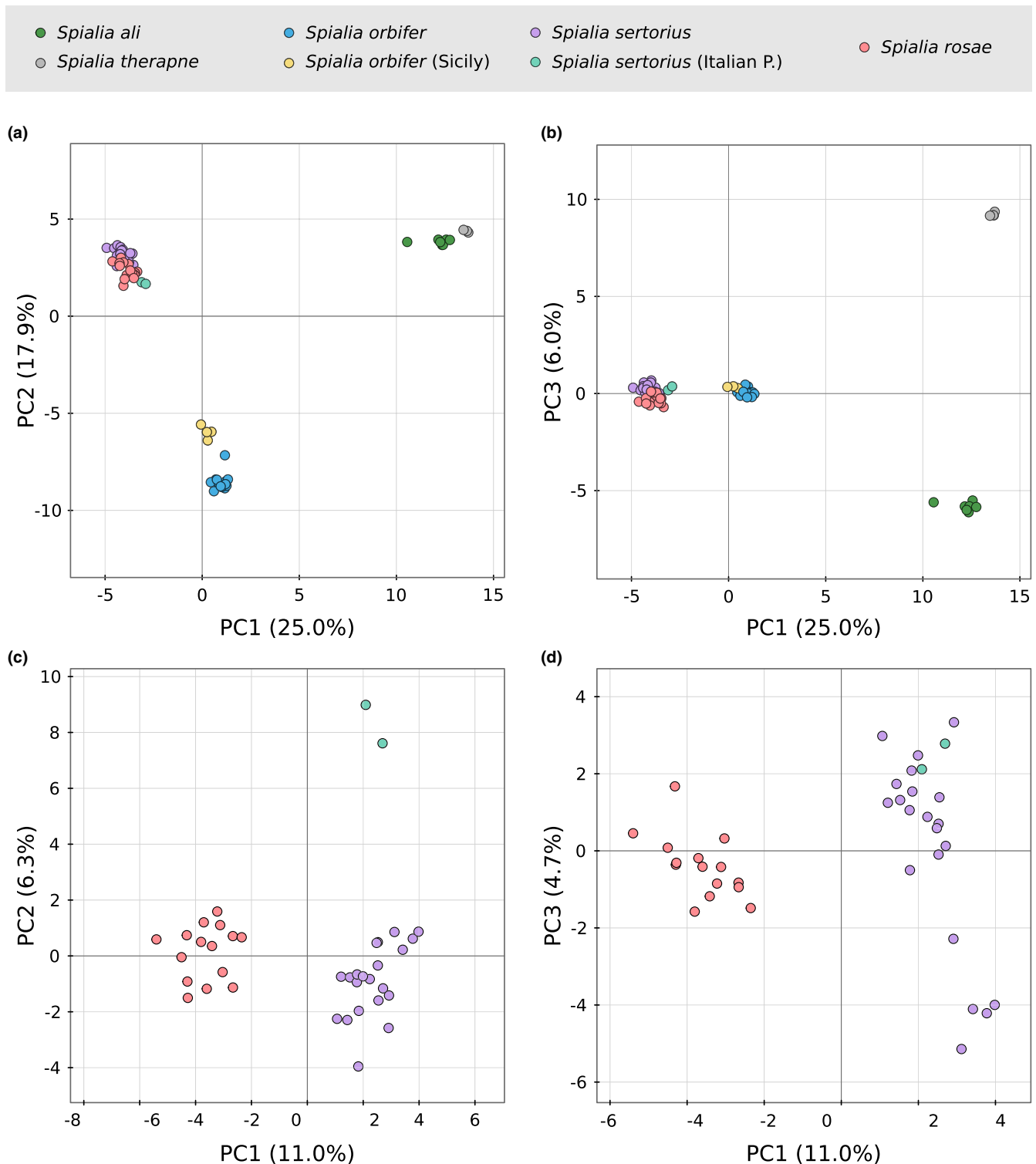
FIGURE 3 Principal component analysis (PCA) including all the samples (a, b) and including only *Spialia orbifer*, *S. rosae* and *S. sertorius* (c, d). The 752 SNPs alignment was used; mitochondrial loci were not included

third locus also matched with several unidentified sequences and with various aldehyde dehydrogenase 1 (*ALDH1*), including, in score order, *ALDH1B1*, *ALDH1A1* and *ALDH1A3*; when translated to amino acid, in the frame corresponding to the ALDH1 protein, it displayed one change in one amino acid that differentiated all *S. rosae* individuals from *S. sertorius*.

Spialia orbifer sequences were available in seven loci with exclusive and nearly-exclusive *S. rosae* haplotypes with respect to *S. sertorius*. The phylogeny constructed with these loci (in which the mitochondrial fragment was excluded) retrieved *S. rosae* and *S. orbifer* in a clade well diverged from *S. sertorius* (Figure S7). Regarding the potential *ALDH1* gene, 12 out of 18 individuals of *S. orbifer* were homozygous for the
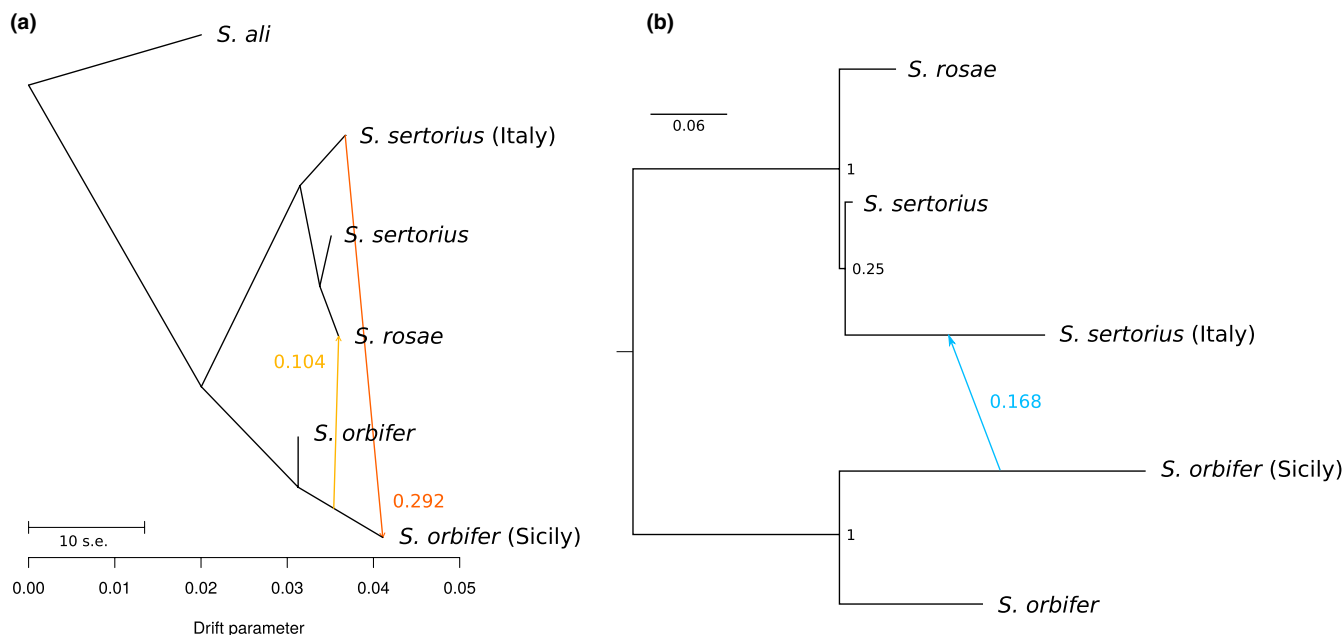
FIGURE 4 (a) Population relationships and migration edges inferred by TreeMix with the optimal number of migration edges identified by OptM. The colour indicates the weight of migration edges. The drift parameter is a relative temporal measure and the scale bar indicates 10 times the average standard error of the relatedness among populations based on the variance-covariance matrix of allele frequencies. The 752 SNP alignment was used; mitochondrial loci were not included. (b) ASTRAL phylogeny with the optimal reticulation events calculated in PhyloNetworks. Branch lengths are expressed in coalescent units and support values are expressed as local posterior probabilities. A total of 506 loci were used, including one mitochondrial locus

same amino acid sequence present in *S. rosae*, while the rest were heterozygous or homozygous for the sequences found in *S. sertorius*.

3.6 | *Wolbachia* infection rate

We extracted 59 loci identified as belonging to *Wolbachia*. We considered an individual to be infected when we retrieved at least 10% of the loci (here, six loci), which was the case for 28 specimens. Among them, all *S. rosae*, all *S. therapne* and some *S. orbifer* samples were infected, while all those of *S. sertorius* and *S. ali* were uninfected (Figure 1). In *S. orbifer*, only individuals from Romania and Kazakhstan were infected. Interestingly, 59 out of the 67 samples tested here were already employed for PCR-based *Wolbachia* screening in Hernández-Roldán et al. (2016) and the results matched.
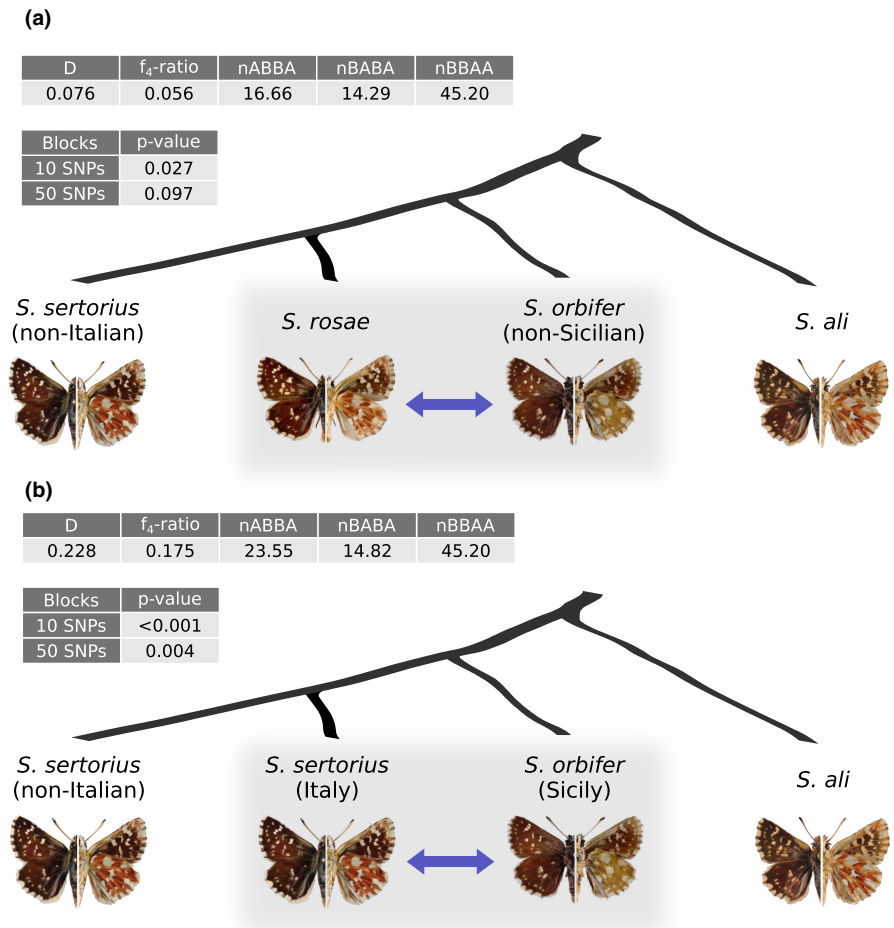
3.7 | Morphological analyses

The PLSDA showed an incomplete separation between species based on male genitalia shape (Figure 6a), but the jackknife procedure attributed most specimens to their correct species (87.0% for *S. orbifer*, 56.3% for *S. sertorius* and 78.6% for *S. rosae*). A comparison of distances from the species shape barycentre between Sicilian *S. orbifer* and continental *S. orbifer* (Figure 6b) revealed that the Sicilian *S. orbifer* was characterized by a significantly higher distance than the continental *S. orbifer* (Figure 6b; Mann–Whitney test, $W = 80$, p-value = .006).

In *S. sertorius* (Figure 6c), specimens sympatric with *S. rosae*, specimens from the contact zone with *S. orbifer* (southern tip of Italian Peninsula) and specimens from the rest of Europe only showed a trend in the differences between the distances to their species barycentre (Kruskal–Wallis chi-squared = 5.682, df = 2, p = .0583).

The analysis of Procrustes morphological distances for male genitalia among *S. sertorius* and *S. orbifer* against overground distances showed an overall negative correlation ($n = 1104$, Spearman's $\rho = -0.209$, $p < .001$). The GAM revealed a nonlinear trend (edf ~3) with high morphological distances between Sicilian and southern Italian Peninsula specimens steeply decreasing and attaining a minimum at around 3,000 km, then increasing at higher distances (Figure 6d). An important part of deviance (>18%) is explained by this relationship (Table S3). *Spialia rosae* and *S. sertorius* showed the opposite trend (Figure 6e) with a positive correlation ($n = 1344$, Spearman's $\rho = -0.209$, $p < .001$) and a curvilinear (edf ~3) trend that, however, explained a lower fraction of deviance (4.4%; Table S3).

The PLSDA for wing morphology showed a better separation between species compared to genitalia (Figure S8a) and a jackknife procedure attributed more specimens to their correct species (96.8% for *S. orbifer*, 76.7% for *S. sertorius* and 66.9% for *S. rosae*). A comparison of distances from their species shape barycentre between Sicilian *S. orbifer* and continental *S. orbifer* (Figure S8b) did not reveal any difference (Mann-Whitney test, $W = 70$, p-value = .826). Similarly, *S. sertorius* specimens sympatric with *S. rosae*, specimens from the contact zone with *S. orbifer* (southern tip of Italian Peninsula) and specimens from the rest of Europe did not show

FIGURE 5 ABBA-BABA scenarios and results. Populations are represented in order (P1, P2, P3 and outgroup). The 752 SNPs alignment was used; mitochondrial loci were not included

(a)

D	f_4-ratio	nABBA	nBABA	nBBAA
0.076	0.056	16.66	14.29	45.20

Blocks	p-value
10 SNPs	0.027
50 SNPs	0.097

S. sertorius (non-Italian) S. rosae S. orbifer (non-Sicilian) S. ali

(b)

D	f_4-ratio	nABBA	nBABA	nBBAA
0.228	0.175	23.55	14.82	45.20

Blocks	p-value
10 SNPs	<0.001
50 SNPs	0.004

S. sertorius (non-Italian) S. sertorius (Italy) S. orbifer (Sicily) S. ali

differences in the distances to their species barycentre (Kruskal–Wallis chi-squared = 2.604, df = 2, p = .272; Figure S8c).

The analysis of Procrustes distances for wing morphology among *S. sertorius* and *S. orbifer* against overground distances showed a positive correlation (n = 1860, Spearman's ρ = 0.101, p < .001). The GAM revealed a nonlinear trend (edf ~3) with low morphological distances between Sicilian *S. orbifer* and S. Italian Peninsula *S. sertorius* (Figure S8d). Differently from genitalia, a low fraction of deviance (<2%) is explained by this relationship (Table S3). *Spialia rosae* and *S. sertorius* also showed a generalized positive correlation with distance (n = 1680, Spearman's ρ = 0.086, p < .001) and a curvilinear (edf ~3) trend that, however, showed an initial decrease at shortest distances followed by an increase for distances longer than 500 km (Figure S8e). Nonetheless, this relationship only explained a very low fraction of deviance (1.5%; Table S3).

4 | DISCUSSION

4.1 | The influence of *Spialia orbifer* in the evolution of *S. rosae*

The phylogeny inferred by Hernández-Roldán et al. (2016) based on poorly-resolved nuclear but well-discriminating mitochondrial markers showed that *S. orbifer* was sister to *S. rosae*. In contrast,

our genetic (ddRADseq data based on nuDNA loci) and morphological data provided evidence showing that *S. rosae* and *S. sertorius* are very closely related. First, *S. rosae* and *S. sertorius* formed a highly supported clade and they were reciprocally monophyletic (Figure 1). SNAPP calculated a divergence time between *S. rosae* and *S. sertorius* of c. 0.28 Ma, which is one of the lowest interspecific divergence times documented in European butterflies (Ebdon et al., 2021; Wiemers et al., 2020; but see Capblancq et al., 2015). Second, in the PCA (Figure 3), differences between *S. rosae* and *S. sertorius* were only clearly visible in the data set that only included both species, and were grouped in the data set with all the taxa. Third, they displayed very low interspecific genetic distances even if the mtDNA data was included (Figure S5a,b), especially for net genetic distances (d_A = 0.02%). Fourth, similarly to Hernández-Roldán et al. (2016), these species neither completely separate based on male genitalia (Figure 6a), nor based on wing morphology (Figure S8a).

Among the biological processes that may account for discrepancies between mtDNA and nuDNA, ILS of ancestral polymorphism seems unlikely in our case. The higher mutation rates and smaller population sizes of mitochondrial markers should ensure sorting mtDNA previous to most nuDNA loci. Regarding the relationship between *S. orbifer* and *S. rosae*, our data displayed the opposite pattern (unsorted mtDNA loci and sorted nuDNA loci). This supported
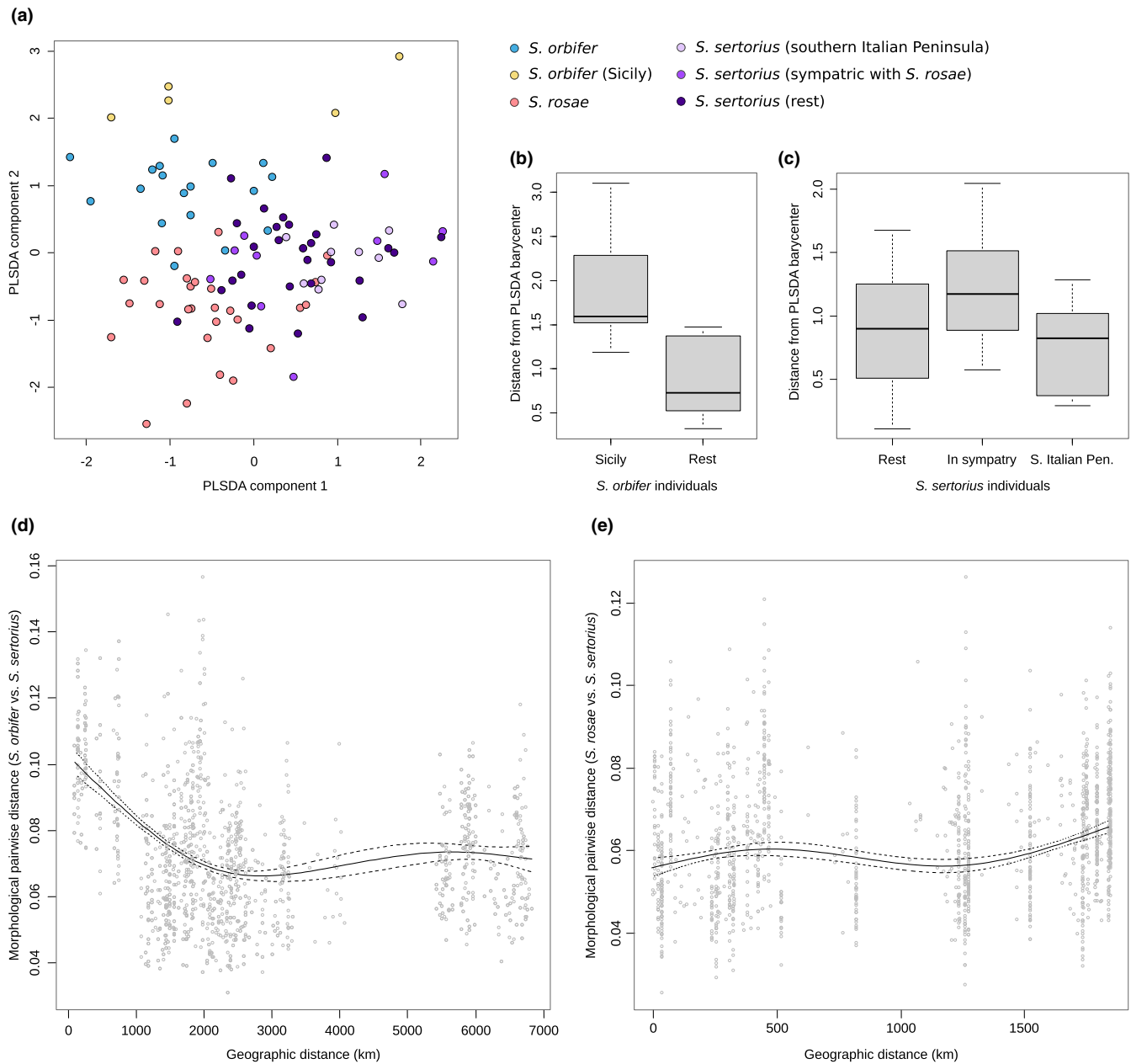
FIGURE 6 Geometric morphometry analyses of the male genitalia including *Spialia orbifer*, *S. rosae* and *S. sertorius* specimens. (a) PLSDA scatterplot representing the first two components, which was used to calculate species morphological barycentres. (b) Distances in PLSDA configuration of Sicilian *S. orbifer* specimens compared to rest of *S. orbifer* specimens. (c) Distances in PLSDA configuration among *S. sertorius* sympatric with *S. rosae*, from southern Italian Peninsula and from the rest of Europe. Relationship between overground geographic distances and morphological distances in genitalic shape (d) comparing *S. orbifer* versus *S. sertorius* and (e) comparing *S. rosae* versus *S. sertorius*

the hypothesis of mtDNA introgression from *S. orbifer* to *S. rosae*, a phenomenon that is relatively widespread in butterflies (e.g., Cong et al., 2017; Talavera et al., 2013). The mtDNA data can be used to infer when the introgression between *S. orbifer* lineage and *S. rosae* occurred. The estimated divergence time between the *S. rosae* haplotypes and other lineages of *S. orbifer* based on *COI* was 0.61 (0.37–0.89) Ma (Figure S2). Further hints of mtDNA introgression can be obtained from the bacterial endosymbiont *Wolbachia* infection rates. Since it is also maternally inherited, along with the transmission of the mtDNA from *S. orbifer* to *S. rosae*, *Wolbachia* may have

also been transferred. Indeed, the infection affects *S. orbifer* (some European populations) and *S. rosae* (all individuals), while *S. sertorius* was uninfected (Figure 1).

Although introgression of mtDNA associated with *Wolbachia* is commonly found in insects (Toews & Brelsford, 2012) and butterflies (e.g., Gaunet et al., 2019; Miyata et al., 2020), it is not known to which extent the same hybridization events may affect and leave a durable signal in the nuDNA. In *Spialia*, we detected also signs of introgression in the nuDNA. The phylogeny based on the nuDNA loci that displayed the highest F_{ST} values between *S. sertorius* and

S. rosae grouped *S. rosae* and *S. orbifer* in a well-differentiated clade (Figure S7). Indeed, haplotypes of these loci were widely shared between *S. orbifer* and *S. rosae*, and they could represent introgressed fragments between these species, although ILS cannot be fully discarded. To discriminate between ILS and gene flow, we used TreeMix and PhyloNetworks. TreeMix indicated the existence of gene flow from *S. orbifer* to *S. rosae* with a transmitted fraction of 10.4%, but this event was not detected in PhyloNetworks. The ABBA-BABA test confirmed an excess of shared alleles between *S. rosae* and *S. orbifer*, although the result was significant in only the smallest SNPs block; here, the percentage of admixture was 5.6%. Overall, the signal of introgression from *S. orbifer* to *S. rosae* is present but weak (affecting between 5.6% and 10.4% of the nuDNA) and analyses using whole-genome sequencing would be optimal to fully confirm these results.

The possibility of the presence of introgressed genetic material in *S. rosae* questions whether the *S. orbifer*-like genes had a role in the speciation of *S. rosae*. One of the differences between *S. rosae* and *S. sertorius* involves the host plant. *Spialia rosae* is a specialist on *Rosa* spp. and *S. sertorius* feeds on *Poterium* spp. Interestingly, one of the genes with private alleles in *S. orbifer* and *S. rosae* (a potential *ALDH1*) could be involved in detoxification (Sobreira et al., 2011), a key metabolic process that might have helped the host shift. However, the fact that *S. orbifer* partially shares this haplotype with *S. rosae* but feeds on *Poterium* spp. put in doubt the link between the haplotype and the host plant shift. The other main ecological difference is the habitat. *Spialia rosae* occupies significantly higher altitudes compared to *S. sertorius* (Hernández-Roldán et al., 2016). *Spialia orbifer* inhabits altitudes similar to *S. sertorius*, but it is important to notice that winters reach much colder temperatures in most of its present range. Hence, we cannot dismiss the potential relationship between the introgression and the habitat tied to mountains of *S. rosae*.

4.2 | Signatures of introgression between *Spialia orbifer* and *S. sertorius* in Italy

Spialia orbifer ranges from the Balkan Peninsula to central Asia (Tolman & Lewington, 2008; Tshikolovets, 2011). It is also present in Sicily, where it lives isolated from the rest of populations since in the Italian Peninsula only *S. sertorius* is present. Our results highlighted the Sicilian populations as a very divergent group of *S. orbifer*. The RAD phylogeny (Figure 1) retrieved the Sicilian *S. orbifer* as a clade sister to all other *S. orbifer* specimens. These clades were estimated to split c. 1.15 Ma, but note that this phylogenetic reconstruction does not consider gene flow. Genetic distances between these two main lineages of *S. orbifer* were similar to those found between the other species analysed (Figure S5a,b). In addition, the Sicilian population demonstrated a private genetic cluster in the STRUCTURE results ($K \geq 4$, Figure 2) and formed a group well separated from the rest of *S. orbifer* in the PCA (Figure 3).

Founder effect and/or genetic drift due to small population size and isolation may have played a role in genetic differentiation.

This is suggested by the low intragroup distances, about half of the value calculated for the rest of *S. orbifer* (Figure S5c). Nevertheless, hybridization between the Sicilian *S. orbifer* and *S. sertorius* seems widespread. Several analyses supported this scenario, including: (1) STRUCTURE, in which the Italian *S. sertorius* and the Sicilian *S. orbifer* individuals had admixed genetic clusters (in $K = 3$, Figure 2)—here, 17.5% of the Sicilian *S. orbifer* was attributed to the genetic cluster of *S. sertorius* and 9.5% of the Italian *S. sertorius* structure corresponded the genetic cluster of *S. orbifer*—(2) the PCA (Figure 3), as in all principal component axes the Sicilian population is shifted from the rest of *S. orbifer* exactly in the direction of *S. sertorius*, while the Italian *S. sertorius* were also shifted to the position of *S. orbifer*; (3) TreeMix (Figure 4a) indicated that 29.2% of the alleles of Sicilian *S. orbifer* would have introgressed from the Italian *S. sertorius* and (4) PhyloNetworks (Figure 4b) detected gene flow in the opposite sense and estimated that 16.8% of the Italian *S. sertorius* would have introgressed from the Sicilian *S. orbifer*. Under this scenario of gene flow, the c. 1.15 Ma of divergence between Sicilian *S. orbifer* and the rest of *S. orbifer* populations is expected to be overestimated.

Hernández-Roldán et al. (2016) argued that the current distribution pattern of *S. orbifer* would be the result of the colonization of the Italian Peninsula by *S. sertorius*, that confined *S. orbifer* to Sicily. Our results are perfectly compatible with this hypothesis. During this process a part of the ancestral *S. orbifer* population of the Italian Peninsula, now surviving in Sicily, could have admixed with the expanding *S. sertorius* until the shifting hybrid zone was spatially fixed by the presence of the Messina strait (3 km wide). In fact, Sicily often harbours relic lineages (Dincă et al., 2011, 2019; Scalercio et al., 2020), which shows that the demographic dynamics explained for the Italian *S. orbifer* may apply in many other cases, even if most represent intraspecific lineage replacement. This situation may be equivalent to that occurring in America around the river Amazonas, which has a similar width than this strait, and it has been recently identified as a barrier spatially fixing a hybrid zone of *Heliconius* butterflies (Rosser et al., 2021). As mentioned above, this scenario could also have been at work if one considers that an ancestral *S. orbifer* population might have been present in or near Iberia, before being swept by *S. sertorius*, only leaving a trace in the part of the genome that was apparently transferred to *S. rosae*.

The Sicilian *S. orbifer* specimens, which are the geographically closest to *S. sertorius* in our data, had the most distinct genitalia among all the *S. orbifer* and compared with *S. sertorius* (Figure 6a,b,d). The south Italian *S. sertorius* did not exhibit differences compared to the rest of *S. sertorius* (Figure 6c). If distinct genitalia shapes increase reproductive isolation, as expected under the lock-and-key hypothesis, this morphological pattern could indicate reproductive character displacement affecting the Sicilian *S. orbifer*, produced by the contact with expanding *S. sertorius*. In Lepidoptera, the lock-and-key hypothesis has been questioned (Shapiro & Porter, 1989) and evidence of character displacement in the genitalia are very scarce, but our results suggest that this phenomenon may exist in shifting hybrid zones, particularly affecting the taxon that is retreating. It is worth noting that *Wolbachia*-mediated cytoplasmic incompatibility

would not limit gene flow at this hybrid zone, because the Sicilian populations of *S. orbifer* seem to be uninfected, as is also the case for all *S. sertorius*. Regarding the ventral hindwing morphology, the Sicilian *S. orbifer* exhibited intermediate patterns between *S. orbifer* and *S. sertorius* (Figure S8a), which could be a possible by-product of past hybridization. In fact, *S. orbifer* individuals with the typical coloration of *S. sertorius* have been documented in Sicily (de Jong, 1974), while *S. sertorius* specimens with the colours of *S. orbifer* have been found in the Italian Peninsula (Pinzari et al., 2010).

The Sicilian populations of *S. orbifer* are included within the nominal subspecies by current authors (Tolman & Lewington, 2008; Tshikolovets, 2011) but, due to their genetic and morphological—in the ventral hindwing patterns (see also in de Jong, 1974) and in the valvae of the genitalia—differences with respect to the rest of *S. orbifer*, we suggest at least a subspecific treatment. As far as we know, the oldest available name for these populations is *tesselloides* (type locality: Sicily), given by Herrich-Schäffer (1845) as a variety of *S. sertorius* (original name: *Hesperia eucrate* var. *tesselloides*). Thus, we propose to name this taxon *Spialia orbifer tesselloides* (Herrich-Schäffer, [1845]) comb. & stat. nov.

Our analyses retrieved the role of hybridization and character displacement in shifting hybrid zones across sympatric lineages of *Spialia* butterflies, two factors that arise as important drivers of the evolution of these butterflies. We predict that the pattern retrieved here using a combination of mitochondrial and nuclear markers obtained by ddRADseq, as well as cutting edge morphometric analyses, are far more common than currently considered in butterflies. These findings encourage further research on hybrid zones to better apprehend the mechanisms and the prevalence of hybridization processes in the evolutionary history of species.

ACKNOWLEDGEMENTS

AUTHOR CONTRIBUTIONS

Joan C. Hinojosa, Nadir Alvarez and Roger Vila conceived, designed and coordinated the study. Funding was secured by Roger Vila. Laboratory protocols were conducted by Joan C. Hinojosa and Camille Pitteloud. Genetic and morphological data were analysed by Joan C. Hinojosa, Darina Koubínová and Leonardo Dapporto. The manuscript was initially written by Joan C. Hinojosa and R.V., with significant contributions from the rest of the authors.

CONFLICT OF INTEREST

The authors have no conflicts of interest.

DATA AVAILABILITY STATEMENT

Raw ddRADseq FASTQ reads: NCBI Bioproject PRJNA771264. Data sets used: figshare (https://doi.org/10.6084/m9.figshare.16810459). The divergent loci between *S. rosae* and *S. sertorius*: figshare (https://doi.org/10.6084/m9.figshare.16810459).

ORCID

Joan C. Hinojosa 🄳 https://orcid.org/0000-0002-6318-4252
Leonardo Dapporto 🄳 https://orcid.org/0000-0001-7129-4526
Darina Koubínová 🄳 https://orcid.org/0000-0002-1854-7675
Nadir Alvarez 🄳 https://orcid.org/0000-0002-0729-166X
Roger Vila 🄳 https://orcid.org/0000-0002-2447-4388

REFERENCES

Abbott, R., Albach, D., Ansell, S., Arntzen, J. W., Baird, S. J. E., Bierne, N., Boughman, J., Brelsford, A., Buerkle, C. A., Buggs, R., Butlin, R. K., Dieckmann, U., Eroukhmanoff, F., Grill, A., Cahan, S. H., Hermansen, J. S., Hewitt, G., Hudson, A. G., Jiggins, C., … Zinner, D. (2013). Hybridization and speciation. *Journal of Evolutionary Biology*, 26(2), 229–246. https://doi.org/10.1111/j.1420-9101.2012.02599.x

Bookstein, F. L. (1997). Landmark methods for forms without landmarks: Morphometrics of group differences in outline shape. *Medical Image Analysis*, 1(3), 225–243. https://doi.org/10.1016/s1361-8415(97)85012-8

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

Breitwieser, F. P., & Salzberg, S. L. (2020). Pavian: Interactive analysis of metagenomics data for microbiome studies and pathogen identification. *Bioinformatics*, 36(4), 1303–1304. https://doi.org/10.1093/bioinformatics/btz715

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

Bryant, D., Bouckaert, R., Felsenstein, J., Rosenberg, N. A., & RoyChoudhury, A. (2012). Inferring species trees directly from biallelic genetic markers: Bypassing gene trees in a full coalescent analysis. *Molecular Biology and Evolution*, 29(8), 1917–1932. https://doi.org/10.1093/molbev/mss086

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: Architecture and applications. *BMC Bioinformatics*, 10(1), 1–9. https://doi.org/10.1186/1471-2105-10-421

Capblancq, T., Després, L., Rioux, D., & Mavárez, J. (2015). Hybridization promotes speciation in *Coenonympha* butterflies. *Molecular Ecology*, 24(24), 6209–6222. https://doi.org/10.1111/mec.13479

Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W., & Postlethwait, J. H. (2011). Stacks: Building and genotyping loci de novo from short-read sequences. *G3: Genes, Genomes, Genetics*, 1(3), 171–182. https://doi.org/10.1534/g3.111.000240

Cong, Q., Shen, J., Borek, D., Robbins, R. K., Opler, P. A., Otwinowski, Z., & Grishin, N. V. (2017). When *COI* barcodes deceive: Complete genomes reveal introgression in hairstreaks. *Proceedings of the Royal Society B: Biological Sciences*, *284*(1848), 20161735. https://doi.org/10.1098/rspb.2016.1735

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., & Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

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

Dasmahapatra, K. K., Walters, J. R., Briscoe, A. D., Davey, J. W., Whibley, A., Nadeau, N. J., Zimin, A. V., Hughes, D. S. T., Ferguson, L. C., Martin, S. H., Salazar, C., Lewis, J. J., Adler, S., Ahn, S.-J., Baker, D. A., Baxter, S. W., Chamberlain, N. L., Chauhan, R., Counterman, B. A., … Jiggins, C. D., Heliconius Genome Consortium. (2012). Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature*, *487*(7405), 94–98. https://doi.org/10.1038/nature11041

de Jong, R. (1974). Systematics and evolution of the Palaearctic *Spialia* species (Lepidoptera, Hesperiidae). *Tijdschrift Voor Entomologie*, *117*(6), 225–271.

de Lesse, H. (1969). Les nombres de chromosomes dans le groupe *Lysandra coridon* (Lepidoptera, Lycaenidae). *Annales De La Société Entomologique De France*, *5*, 469–522.

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

Dincă, V., 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*(18), 3921–3935. https://doi.org/10.1111/j.1365-294X.2011.05223.x

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*(1911), 20191311. https://doi.org/10.1098/rspb.2019.1311

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

Durand, E. Y., Patterson, N., Reich, D., & Slatkin, M. (2011). Testing for ancient admixture between closely related populations. *Molecular Biology and Evolution*, *28*(8), 2239–2252. https://doi.org/10.1093/molbev/msr048

Earl, D. A., & vonHoldt, B. M. (2012). STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, *4*(2), 359–361. https://doi.org/10.1007/s12686-011-9548-7

Eaton, D. A. R., & Overcast, I. (2016). *ipyrad: interactive assembly and analysis of RADseq data sets*. Retrieved from http://ipyrad.readthedocs.io/

Ebdon, S., Laetsch, D. R., Dapporto, L., Hayward, A., Ritchie, M. G., Dincă, V., & Lohse, K. (2021). The Pleistocene species pump past its prime: Evidence from European butterfly sister species. *Molecular Ecology*, *30*(14), 3575–3589. https://doi.org/10.1111/mec.15981

Enciso-Romero, J., Pardo-Díaz, C., Martin, S. H., Arias, C. F., Linares, M., McMillan, W. O., Jiggins, C. D., & Salazar, C. (2017). Evolution of novel mimicry rings facilitated by adaptive introgression in tropical butterflies. *Molecular Ecology*, *26*(19), 5160–5172. https://doi.org/10.1111/mec.14277

Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: New 1 km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, *37*(12), 4302–4315. https://doi.org/10.1002/joc.5086

Fitak, R. R. (2021). optM: An R package to optimize the number of migration edges using threshold models. *Biology Methods and Protocols*, *6*(1), 1–6. https://doi.org/10.1093/biomethods/bpab017

Foll, M., & Gaggiotti, O. E. (2008). A genome scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. *Genetics*, *180*(2), 977–993. https://doi.org/10.1534/genetics.108.092221

Gaunet, A., Dincă, V., Dapporto, L., Montagud, S., Vodă, R., Schär, S., & Vila, R. (2019). Two consecutive *Wolbachia*-mediated mitochondrial introgressions obscure taxonomy in Palearctic swallowtail butterflies (Lepidoptera, Papilionidae). *Zoologica Scripta*, *48*(4), 507–519. https://doi.org/10.1111/zsc.12355

Gross, B. L., & Rieseberg, L. H. (2005). The ecological genetics of homoploid hybrid speciation. *Journal of Heredity*, *96*(3), 241–252. https://doi.org/10.1093/jhered/esi026

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

Hedrick, P. W. (2013). Adaptive introgression in animals: Examples and comparison to new mutation and standing variation as sources of adaptive variation. *Molecular Ecology*, *22*(18), 4606–4618. https://doi.org/10.1111/mec.12415

Hegarty, M. J., & Hiscock, S. J. (2005). Hybrid speciation in plants: New insights from molecular studies. *New Phytologist*, *165*(2), 411–423. https://doi.org/10.1111/j.1469-8137.2004.01253.x

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. https://doi.org/10.1111/mec.13756

Hernández-Roldán, J. L., Vicente, J. C., Vila, R., & Munguira, M. L. (2018). Natural history and immature stage morphology of *Spialia* Swinhoe, 1912 in the Iberian Peninsula (Lepidoptera, Hesperiidae). *Nota Lepidopterologica*, *41*(1), 1–22. https://doi.org/10.3897/nl.41.13539

Herrich-Schäffer, G. A. W. (1845). *Systematische Bearbeitung der Schmetterlinge von Europa. Erster band. Die Tagfalter*. G. J. Manz.

Hurst, G. D. D., & Jiggins, F. M. (2005). Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: The effects of inherited symbionts. *Proceedings of the Royal Society B: Biological Sciences*, *272*(1572), 1525–1534. https://doi.org/10.1098/rspb.2005.3056

Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S., & Madden, T. L. (2008). NCBI BLAST: A better web interface. *Nucleic Acids Research*, *36*(S2), W5–W9. https://doi.org/10.1093/nar/gkn201

Jombart, T., Devillard, S., & Balloux, F. (2010). Discriminant analysis of principal components: A new method for the analysis of genetically structured populations. *BMC Genetics*, *11*(1), 94. https://doi.org/10.1186/1471-2156-11-94

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

Kim, D., Song, L., Breitwieser, F. P., & Salzberg, S. L. (2016). Centrifuge: Rapid and sensitive classification of metagenomic sequences. *Genome Research*, *26*(12), 1721–1729. https://doi.org/10.1101/gr.210641.116

Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., & Mayrose, I. (2015). Clumpak: A program for identifying clustering modes and packaging population structure inferences across

K. *Molecular Ecology Resources*, 15(5), 1179–1191. https://doi.org/10.1111/1755-0998.12387

Lê Cao, K. A., Boitard, S., & Besse, P. (2011). Sparse PLS discriminant analysis: Biologically relevant feature selection and graphical displays for multiclass problems. *BMC Bioinformatics*, 12(1), 1–17. https://doi.org/10.1186/1471-2105-12-253

Lukhtanov, V. A., Dincă, V., Friberg, M., Šíchová, J., Olofsson, M., Vila, R., Marec, F., & Wiklund, C. (2018). Versatility of multivalent orientation, inverted meiosis, and rescued fitness in holocentric chromosomal hybrids. *Proceedings of the National Academy of Sciences*, 115(41), E9610–E9619. https://doi.org/10.1073/pnas.1802610115

Lukhtanov, V. A., Dincă, V., Friberg, M., Vila, R., & Wiklund, C. (2020). Incomplete sterility of chromosomal hybrids: Implications for karyotype evolution and homoploid hybrid speciation. *Frontiers in Genetics*, 11, 1205. https://doi.org/10.3389/fgene.2020.583827

Lukhtanov, V. A., Dincă, V., Talavera, G., & Vila, R. (2011). Unprecedented within-species chromosome number cline in the Wood White butterfly *Leptidea sinapis* and its significance for karyotype evolution and speciation. *BMC Evolutionary Biology*, 11(1), 1–11. https://doi.org/10.1186/1471-2148-11-109

Malinsky, M., Matschiner, M., & Svardal, H. (2020). Dsuite-fast D-statistics and related admixture evidence from VCF files. *Molecular Ecology Resources*, 21(2), 584–595. https://doi.org/10.1111/1755-0998.13265

Mallet, J. (2005). Hybridization as an invasion of the genome. *Trends in Ecology & Evolution*, 20(5), 229–237. https://doi.org/10.1016/j.tree.2005.02.010

Mallet, J., Wynne, I. R., & Thomas, C. D. (2011). Hybridisation and climate change: Brown argus butterflies in Britain (*Polyommatus* subgenus *Aricia*). *Insect Conservation and Diversity*, 4(3), 192–199. https://doi.org/10.1111/j.1752-4598.2010.00122.x

Mandrioli, M., & Manicardi, G. C. (2020). Holocentric chromosomes. *PLOS Genetics*, 16(7), e1008918. https://doi.org/10.1371/journal.pgen.1008918

Mavárez, J., & Linares, M. (2008). Homoploid hybrid speciation in animals. *Molecular Ecology*, 17(19), 4181–4185. https://doi.org/10.1111/j.1365-294x.2008.03898.x

Melters, D. P., Paliulis, L. V., Korf, I. F., & Chan, S. W. (2012). Holocentric chromosomes: Convergent evolution, meiotic adaptations, and genomic analysis. *Chromosome Research*, 20(5), 579–593. https://doi.org/10.1007/s10577-012-9292-1

Miyata, M. N., Nomura, M., & Kageyama, D. (2020). *Wolbachia* have made it twice: Hybrid introgression between two sister species of *Eurema* butterflies. *Ecology and Evolution*, 10(15), 8323–8330. https://doi.org/10.1002/ece3.6539

Mullen, S. P., Dopman, E. B., & Harrison, R. G. (2008). Hybrid zone origins, species boundaries, and the evolution of wing pattern diversity in a polytypic species complex of North American admiral butterflies (Nymphalidae: *Limenitis*). *Evolution*, 62(6), 1400–1417. https://doi.org/10.1111/j.1558-5646.2008.00366.x

Nice, C. C., Gompert, Z., Fordyce, J. A., Forister, M. L., Lucas, L. K., & Buerkle, C. A. (2013). Hybrid speciation and independent evolution in lineages of alpine butterflies. *Evolution*, 67(4), 1055–1068. https://doi.org/10.1111/evo.12019

Pardo-Diaz, C., Salazar, C., Baxter, S. W., Merot, C., Figueiredo-Ready, W., Joron, M., McMillan, W. O., & Jiggins, C. D. (2012). Adaptive introgression across species boundaries in *Heliconius* butterflies. *PLOS Genetics*, 8(6), e1002752. https://doi.org/10.1371/journal.pgen.1002752

Pärssinen, V., Hulthén, K., Brönmark, C., Skov, C., Brodersen, J., Baktoft, H., Chapman, B. B., Hansson, L.-A., & Nilsson, P. A. (2020). Maladaptive migration behaviour in hybrids links to predator-mediated ecological selection. *Journal of Animal Ecology*, 89(11), 2596–2604. https://doi.org/10.1111/1365-2656.13308

Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., & Reich, D. (2012). Ancient admixture in human history. *Genetics*, 192(3), 1065–1093. https://doi.org/10.1534/genetics.112.145037

Pickrell, J., & Pritchard, J. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. *PLOS GENETICS*, 8(11), e1002967. https://doi.org/10.1371/journal.pgen.1002967

Pinzari, M., Pinzari, M., & Zilli, A. (2010). Deep lepidopterological exploration of Mt Cagno and surroundings (Central Italy), a restricted mountain massif and hotspot for butterfly and moth diversity (Lepidoptera). *Bollettino Dell'associazione Romana Di Entomologia*, 65(1–4), 3–383.

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

Puritz, J. B., Hollenbeck, C. M., & Gold, J. R. (2014). dDocent: A RADseq, variant-calling pipeline designed for population genomics of non-model organisms. *PeerJ*, 2, e431. https://doi.org/10.7717/peerj.431

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

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

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

Rabiee, M., Sayyari, E., & Mirarab, S. (2019). Multi-allele species reconstruction using ASTRAL. *Molecular Phylogenetics and Evolution*, 130, 286–296. https://doi.org/10.1016/j.ympev.2018.10.033

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

Rambaut, A., Drummond, A. J., Xie, D., Baele, G., & Suchard, M. A. (2018). Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Systematic Biology*, 67(5), 901–904. https://doi.org/10.1093/sysbio/syy032

Rieseberg, L. H., Kim, S. C., Randell, R. A., Whitney, K. D., Gross, B. L., Lexer, C., & Clay, K. (2007). Hybridization and the colonization of novel habitats by annual sunflowers. *Genetica*, 129(2), 149–165. https://doi.org/10.1007/s10709-006-9011-y

Rohart, F., Gautier, B., Singh, A., & Lê Cao, K. A. (2017). mixOmics: An R package for 'omics feature selection and multiple data integration. *PLOS Computational Biology*, 13(11), e1005752. https://doi.org/10.1371/journal.pcbi.1005752

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

Rosser, N., Shirai, L. T., Dasmahapatra, K. K., Mallet, J., & Freitas, A. V. (2021). The Amazon river is a suture zone for a polyphyletic group of co-mimetic heliconiine butterflies. *Ecography*, 44(2), 177–187. https://doi.org/10.1111/ecog.05282

Scalercio, S., Cini, A., Menchetti, M., Vodă, R., Bonelli, S., Bordoni, A., & Dapporto, L. (2020). How long is 3 km for a butterfly? Ecological constraints and functional traits explain high mitochondrial genetic diversity between Sicily and the Italian Peninsula. *Journal of Animal Ecology*, 89(9), 2013–2026. https://doi.org/10.1111/1365-2656.13196

Shapiro, A. M., & Porter, A. H. (1989). The lock-and-key hypothesis: Evolutionary and biosystematic interpretation of insect genitalia. *Annual Review of Entomology*, 34(1), 231–245. https://doi.org/10.1146/annurev.en.34.010189.001311

Sobreira, T. J. P., Marlétaz, F., Simões-Costa, M., Schechtman, D., Pereira, A. C., Brunet, F., Sweeney, S., Pani, A., Aronowicz, J., Lowe, C. J., Davidson, B., Laudet, V., Bronner, M., de Oliveira, P. S. L., Schubert, M., & Xavier-Neto, J. (2011). Structural shifts of aldehyde dehydrogenase enzymes were instrumental for the early evolution of

retinoid-dependent axial patterning in metazoans. *Proceedings of the National Academy of Sciences*, 108(1), 226–231. https://doi.org/10.1073/pnas.1011223108

Solís-Lemus, C., & Ané, C. (2016). Inferring phylogenetic networks with maximum pseudolikelihood under incomplete lineage sorting. *PLOS Genetics*, 12(3), e1005896. https://doi.org/10.1371/journal.pgen.1005896

Solís-Lemus, C., Bastide, P., & Ané, C. (2017). PhyloNetworks: A package for phylogenetic networks. *Molecular Biology and Evolution*, 34(12), 3292–3298. https://doi.org/10.1093/molbev/msx235

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

Suarez-Gonzalez, A., Lexer, C., & Cronk, Q. C. (2018). Adaptive introgression: A plant perspective. *Biology Letters*, 14(3), 20170688. https://doi.org/10.1098/rsbl.2017.0688

Talavera, G., Lukhtanov, V. A., Rieppel, L., Pierce, N. E., & Vila, R. (2013). In the shadow of phylogenetic uncertainty: The recent diversification of *Lysandra* butterflies through chromosomal change. *Molecular Phylogenetics and Evolution*, 69(3), 469–478. https://doi.org/10.1016/j.ympev.2013.08.004

Thallinger, G. (2017). *Comparison of ddRAD Analysis Pipelines* (Unpublished master's thesis). Graz University of Technology.

Toews, D. P., & Brelsford, A. (2012). The biogeography of mitochondrial and nuclear discordance in animals. *Molecular Ecology*, 21(16), 3907–3930. https://doi.org/10.1111/j.1365-294X.2012.05664.x

Tolman, T., & Lewington, R. (2008). *Collins butterfly guide*. HarperCollins Publishers.

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

van Etten, J. (2017). R package gdistance: Distances and routes on geographical grids. *Journal of Statistical Software*, 76(1), 1–21. https://doi.org/10.18637/jss.v076.i13

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

Wiemers, M., Balletto, E., Dincă, V., Fric, Z. F., Lamas, G., Lukhtanov, V., Munguira, M. L., van Swaay, C. A. M., Vila, R., Vliegenthart, A., Wahlberg, N., & Verovnik, R. (2018). An updated checklist of the European butterflies (Lepidoptera, Papilionoidea). *ZooKeys*, 811, 9–45. https://doi.org/10.3897/zookeys.811.28712

Wiemers, M., Chazot, N., Wheat, C. W., Schweiger, O., & Wahlberg, N. (2020). A complete time-calibrated multi-gene phylogeny of the European butterflies. *ZooKeys*, 938, 97–124. https://doi.org/10.3897/zookeys.938.50878

Zhang, C., Rabiee, M., Sayyari, E., & Mirarab, S. (2018). ASTRAL-III: Polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics*, 19(6), 15–30. https://doi.org/10.1186/s12859-018-2129-y

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.