Received: 23 December 2023 | Accepted: 4 March 2024

DOI: 10.1111/syen.12631

ORIGINAL ARTICLE

Systematic Entomology

Can species endure massive introgression? Genomic evidence of asymmetric gene flow in *Melitaea* butterflies

Joan C. Hinojosa[1] | Valéria Marques[1] | Luis Sánchez Mesa[2] | Leonardo Dapporto[3] | Vlad Dincă[4] | Roger Vila[1]

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

[2]Independent Researcher, Granada, Spain

[3]ZEN Lab, Department of Biology, University of Florence, Sesto Fiorentino, Italy

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

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

Funding information
Academy of Finland, Grant/Award Numbers: 324988, 352652; European Science Foundation; Ministerio de Ciencia, Innovación y Universidades; Generalitat de Catalunya; MICIU/AEI, Grant/Award Numbers: PID2019-107078GB-I00, PID2022-139689NB-I00, PRE2020-094870; NBFC (Ministero Italiano dell'Università e della Ricerca, PNRR), Grant/Award Number: Investimento1.4, Project CN00000033; Direttiva Biodiversità, Ministero Italiano dell'Ambiente e della Sicurezza Energetica

Abstract

Hybridisation and introgression are increasingly seen as important drivers of the evolution of organisms, particularly in Lepidoptera. One group that is gaining attention due to recently published cases of interspecific gene flow is the genus *Melitaea* Fabricius (Nymphalidae). In this study, we used genomics to investigate the role of hybridisation in the evolution of the western Palearctic species of the *Melitaea phoebe* group *M. ornata* Christoph, the recently described *M. pseudornata* Muñoz Sariot & Sánchez Mesa, *M. phoebe* (Denis & Schiffermüller), *M. punica* Oberthür, and *M. telona* (Fruhstorfer). We provide evidence of asymmetric gene flow from *M. phoebe* to both *M. ornata* and *M. pseudornata*. Gene flow from *M. phoebe* to *M. pseudornata* was very high (25.0%–31.9%), widespread throughout the distribution of the latter, and not equally distributed along the genome. The Z chromosome showed patterns compatible with the large-Z effect, which were mimicked by two autosomes. *Melitaea pseudornata* endured massive introgression while remaining a separate entity from *M. phoebe*, although gene flow may have altered its phenotype, including its voltinism and the morphology of the adults and caterpillars. These findings suggest that hybridisation may be pervasive in this genus and highlight its key role in the evolution of butterflies, emphasising the need for further research on this topic.

KEYWORDS
ddRADseq, hybridisation, introgression, Lepidoptera, phylogeography

INTRODUCTION

The study of hybridisation (the production of viable offspring from distinct species) and introgression (the transfer of genes between species as a result of hybridisation and repeated backcrossing) in Lepidoptera has attracted the interest of scientists since early times. Initially, it was mostly restricted to experimental crosses in laboratory conditions and anecdotal observations in the wild, compiled by Tutt (1899–1906) on his landmark series *A natural history of the British Lepidoptera: a text-book for students and collectors*. The detection of hybrids in nature nonetheless depended on the phenotype of the hybrid individuals, which had to be intermediate between the parental species; and even in these cases, doubts about a hybrid origin could persist. This changed drastically with the introduction of molecular techniques, such as allozymes, single-marker sequencing and, more recently, genomics. They allowed tracing and even quantifying gene flow from wild specimens, thus greatly advancing the understanding of hybridisation processes (Arnold, 1997; Goulet et al., 2017). Nevertheless, the incidence of reticulate evolution is unknown for most of the lepidopteran groups, leading to an incomplete knowledge of hybridisation in these insects, including its recurrence and outcomes.

The global frequency of hybridisation for lepidopterans is unknown, but it can be estimated for specific geographic areas and groups, such as the European butterflies. Descimon and Mallet (2009) stated that interspecific mating occurs in roughly 16% of the species and that at least 8% can produce fertile offspring (or hybridise under the given definition). This proportion is similar to the average for animals, estimated to be about 10% (Mallet, 2005). However, hybridisation has not been thoroughly studied for many of the European species, hence this percentage is expected to be a low estimate. To complement these data, mitochondrial DNA (mtDNA) can be a valuable source of information because shared haplotypes may be the result of hybridisation. The data provided by a recent high-resolution DNA barcode library of the European butterflies revealed that 69 out of 459 species share cytochrome c oxidase subunit I (*COI*) haplotypes (Dincă et al., 2021). This may suggest that about 15% of the species have potentially hybridised in relatively recent times, but the precision of this estimation is questioned by several shortcomings. For example, sharing mtDNA haplotypes can be due to flawed taxonomy, lack of independent lineage sorting of the current variation or, in very young species, the apparent lack of variation in this marker—for example, in the *Erebia tyndarus* group (Dincă et al., 2021; Litman et al., 2018). Furthermore, given that Lepidoptera obey Haldane's rule (Haldane, 1922; Presgraves, 2002), that is, the preferential sterility or inviability of the heterogametic sex (here, females), mating resulting in infertile females but fertile males are expected to be frequent. As a result, two species could hybridise without exchanging mtDNA—likely examples are the pairs *Spialia orbifer* Hübner—*S. sertorius* Hoffmannsegg (Hinojosa, Dapporto, et al., 2022) and *Papilio hospiton* Gené—*P. machaon* Linnaeus (Cianchi et al., 2003).

Considering that interspecific gene flow in butterflies seems to be at least as frequent as the animal average, it might have played a substantial role in their evolution. Hybridisation acts not only as a mere merging force, but also as a source of new diversity, as it can contribute to diversification by locally forming hybrid populations or by the acquisition of novel phenotypes via introgression. These two processes lead to divergence between populations and can ultimately cause speciation (Abbott et al., 2013). Cases of hybrid populations or even hybrid species have been regularly documented in lepidopterans in recent years (Capblancq et al., 2015; Hinojosa, Dapporto, et al., 2022; Nice et al., 2013). On the other hand, phenotypic changes driven by introgression have been poorly studied and only a few features were proposed to have been transferred between species through this mechanism, such as wing colouration and patterns (Pardo-Diaz et al., 2012; The Heliconius Genome Consortium, 2012), or insecticide-resistant genes (Valencia-Montoya et al., 2020).

In the last years, amid the expansion of genomic techniques, non-model organisms have increasingly gained popularity. One of them is the genus *Melitaea* Fabricius (Nymphalidae), a diverse butterfly genus that comprises about a hundred species distributed in the Palearctic (van Oorschot & Coutsis, 2014). Cases of high intraspecific diversity (Dincă et al., 2019) and hybridisation (Pazhenkova & Lukhtanov, 2021; Tahami et al., 2021) have highlighted them as valuable to study speciation and reticulate evolution. Additional examples of repeated hybridisation may involve one of the most widely distributed groups of the genus, the *Melitaea phoebe* group. Five of its species inhabit the western Palearctic region: *M. phoebe* (Denis and Schiffermüller), widely distributed across Europe, *M. punica* Oberthür, a north-western African endemic, *M. ornata* Christoph, present from Provence to Central Asia, *M. telona* Fruhstorfer, exclusive to the Middle East, and the recently described (Hinojosa, Tóth, et al., 2022; Muñoz Sairot & Sánchez Mesa, 2019a; Muñoz Sairot & Sánchez Mesa, 2019b) *M. pseudornata* Muñoz Sariot and Sánchez Mesa, an Iberian endemic. Shared *COI* barcode haplotypes have been documented between *M. phoebe*, *M. ornata* and *M. pseudornata* (Hinojosa, Tóth, et al., 2022; Tóth et al., 2017); the mitogenome of *M. pseudornata* is apparently fully replaced by that of *M. phoebe* while in *M. ornata* it is only partially replaced. Additionally, the morphology of the caterpillars (in northern populations) and adults of *M. pseudornata* is very similar to *M. phoebe*. These clues hinted at the presence of interspecific gene flow within the European species of the group, especially between *M. pseudornata* and *M. phoebe*.

In this study, we hypothesised that the evolution of the *M. phoebe* group in Europe could have been influenced by hybridisation. To test this, we used genomic data obtained through double-digest RADseq (ddRADseq) to (1) retrieve the evolutionary relationships between the western Palearctic species of the *M. phoebe* group, (2) examine their population genetic structure across their distribution ranges and (3) test for the existence of interspecific gene flow. Since evidence of hybridisation was found, its weight on different parts of the genome was assessed, and the possibility that phenotypic traits were transferred between species is discussed.

MATERIALS AND METHODS

Sampling

We analysed DNA data from 67 samples of the five species belonging to the *M. phoebe* group from the western Palearctic, including 10 *M. ornata*, 32 *M. phoebe*, 18 *M. pseudornata*, 3 *M. punica* and 2 *M. telona* (Table S1). Butterfly bodies were stored in 99% ethanol at −20°C and wings were kept separately as vouchers.

ddRADseq library preparation

Genomic DNA (gDNA) was extracted from half of the thorax using the DNeasy Blood & Tissue Kit (Qiagen). The amount of gDNA per sample was measured using the PicoGreen kit (Molecular Probes) according to the manufacturer's instructions. Using the REPLI-g Mini Kit (Qiagen), a whole genome amplification step was performed to increase gDNA quantity and quality and we estimated again the concentration of the amplified gDNA with the PicoGreen kit. For every sample, 500 ng of DNA were digested in a reaction consisting of 1 µL *Pst*I (20,000 units/mL, New England Biolabs), 2 µL *Mse*I (10,000 units/mL, New England Biolabs), 5 µL of CutSmart Buffer (New England

Biolabs) and ultra-pure (HPLC quality) water up to 50 μL. Afterwards, the samples were incubated for 2 h at 37°C, and enzymes were deactivated by freezing. They were then purified with AMPure XP x1 magnetic beads (Agencourt) using a Biomek automated liquid handler (Beckman Coulter), selecting a final elution in 40 μL. The DNA concentration was measured with PicoGreen; this value was used later for the pooling step. The following buffers and enzymes were added to every sample to enable ligation between the DNA and the adapters: 5 μL T4 DNA Ligase Buffer (New England Biolabs), 1 μL T4 DNA Ligase (2,000,000 units/mL, New England Biolabs), 0.6 μL rATP (100 mM, Promega), 5 μL P1 adapter (50 nM), 5 μL P2 adapter (50 nM) and 2.4 μL water. As a result, each sample was tagged with unique P1 adapters—which included a TGCA overhang on the top strand to match the sticky end left by *Pst*I—that differed in a 5-nt barcode sequence. The P2 adapters contained the Illumina sequencing primer sequences that are compatible with paired-end runs, and AT overhangs on the top strand to match the sticky end left by *Mse*I; it also incorporated a 'divergent-Y' (Baird et al., 2008) to prevent amplification of fragments with *Mse*I cut sites at both ends. This ligation step was performed for 1 h at 22°C, and enzymes were deactivated at 65°C for 20 min. Since the DNA concentration was measured previously, 200 ng from each individual were pooled independently in three tubes with a final volume of ~450 μL each. The pools were purified with AMPure XP magnetic beads and eluted in 40 μL and were size-selected at 300 bp with BluePippin (Sage Science) using the cassette type '2% DF Marker V1' and the 'tight' option. A final PCR amplification step was introduced with primers RAD1.F (5′-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACG-3′) and RAD2.R (5′-CAAGCAGAAGACGGCATACGAGATCGTGATGTG ACTGGAGTTCAGACGTGTGC-3′). Each pool was amplified in 60 μL volume reactions: 9 μL water, 30 μL Phusion High-Fidelity PCR Master Mix (Finnzymes), 3 μL of each primer (10 mM) and 15 μL of DNA. Reaction conditions comprised a first denature step at 98°C for 30 s, then 98°C for 10 s, 60°C for 30 s and 72°C for 40 s in 16 cycles with a final extension step at 72°C for 5 min. PCR products were purified with AMPure XP magnetic beads to remove the maximum of impurities prior to sequencing and DNA concentration was checked with PicoGreen. The size distribution and concentration of the three pools were measured with Bioanalyzer (Agilent Technologies). The three libraries were finally pooled into a single tube in equimolar amounts and sequenced on an Illumina HiSeq X by Macrogen Korea. The demultiplexed FASTQ files were archived in the NCBI SRA: SRR28153373-SRR28153439.

ddRADseq data set processing

Initial filtering steps, single nucleotide polymorphism (SNP) calling and alignment were carried out using the ipyrad v.0.6.15 (Eaton & Overcast, 2016) pipeline. After testing distinct parameter combinations, the following parameters were changed from the default settings: assembly method was set to 'reference' (*M. athalia* [Rottemburg] genome was used, submitted sequence: GenBank GCA_905220545.2), datatype to 'pairddradseq', restriction overhang to 'TGCAG, TAA', minimum length to 70, minimum samples per locus to 5, maximum SNP per locus to 0.1.

One SNP per locus was taken to create an unlinked data set. From this data set, a low missing data set was obtained using VCFtools v0.1.16 (Danecek et al., 2011), retaining only the SNPs present in 90% of the individuals (--max-missing 0.90). The mtDNA loci were removed from all the data sets used for this study.

Phylogenetic analyses of the ddRADseq data

An alignment was built by concatenating all the ddRADseq loci. This alignment was used to construct a phylogeny through maximum likelihood inference using RAxML v8.2.4 (Stamatakis, 2014). The GTRGAMMA model and 1000 bootstrap replicates were selected. The resulting phylogeny was visualised and exported using FigTree v.1.4.2 (Rambaut, 2015).

Genetic structuring

PCA and STRUCTURE analyses were applied to gather detailed information on the genetic structure of the target species. The PCA was performed using the R software package adegenet v1.4-1 (Jombart et al., 2010). For STRUCTURE v2.3.4 (Pritchard et al., 2000), an admixture model with correlated allele frequencies was chosen. The selected burn-in was 75,000, followed by 250,000 MCMC replicates. *K* (number of populations) values from 1 to 6 were tested, and 10 runs were done for each *K*, which were afterwards plotted with CLUMPAK (Kopelman et al., 2015). The best *K* according to the Evanno method (Evanno et al., 2005) was estimated in CLUMPAK.

Genetic differentiation

The fixation index (F_{ST}) was calculated with VCFtools using the weighted Weir and Cockerham's estimator. It was measured for windows of 200,000 base pairs (bp) as well as for chromosomes and between the species pairs *M. ornata*–*M. phoebe*, *M. ornata*–*M. pseudornata* and *M. phoebe*–*M. pseudornata*.

Gene flow calculations

Gene flow events were explored using TreeMix v1.13 (Pickrell & Pritchard, 2012). Based on covariance of allele frequencies and Gaussian approximation to genetic drift, this software infers historical relationships between populations. Additionally, TreeMix applies a methodology to approximate the fraction of admixture: it calculates the weight of the migration events which is, considering a situation in which admixture occurs in a single generation, the proportion of alleles in the descendant population that originated from each parental population. We used the Python script vcf2treemix.py (available at

https://github.com/CoBiG2/RAD_Tools/blob/master/vcf2treemix.py) to generate the input file. We ran TreeMix using 1, 2, 3 and 4 migration edges (m) and 11 SNPs windows (k) per migration edge, from 10 to 20. The optimal number of migration edges was estimated with the R package OptM (Fitak, 2021).

Additional interspecific gene flow events were investigated using the Julia package PhyloNetworks v0.12.0 (Solís-Lemus et al., 2017). To infer reticulations, this method uses a maximum pseudolikelihood estimator applied to quartet concordance factors (CFs), that is, gene tree frequencies, of 4-taxon trees under the coalescent model, incorporating incomplete lineage sorting and reticulation events (Solís-Lemus et al., 2017). As TreeMix, Phylonetworks can also estimate the fraction of the genome that has been potentially introgressed. Specifically, the observed CFs are used to retrieve the inheritance values (γ), the proportion of ancestral contribution to the hybrid lineage genome. We then used RaxML v8.2.4 to estimate a tree per locus; we chose the GTRGAMMA model and 100 bootstrap replicates per locus. Subsequently, CFs were estimated in PhyloNetworks from all RaxML gene-trees (13,419), with all individuals per clade mapped as alleles to species. We used a tree built in ASTRAL v5.7.5 (Rabiee et al., 2019; Zhang et al., 2018) as starting topology, obtained with the default parameters and using the RaxML loci trees. We tested values for h (number of gene flow events) from 0 to 4, assessing maximum support using a slope heuristic for the increase in likelihood plotted against h (Solís-Lemus & Ané, 2016). We performed 50 independent runs per h-value to ensure convergence on a global optimum.

We used the ABBA–BABA analysis (Durand et al., 2011) to further confirm the introgression event indicated by TreeMix and Phylonetworks, that is, between *M. phoebe* and *M. pseudornata*. Since *M. phoebe* and *M. ornata* also share mitochondrial haplotypes, the presence of shared alleles was also tested between these two species. Additionally, the same analysis was performed to assess if the northern or southern populations of *M. pseudornata* could have experienced different levels of gene flow. *Melitaea aetherie* Hübner was used as outgroup in all the analyses. Calculations were done in Dsuite v0.4 using program Dtrios (Malinsky et al., 2020). Dtrios calculates Patterson's D (Patterson et al., 2012) and uses a block-jackknife procedure to assess if D significantly differs from zero. The size of the selected blocks should be larger than the extent of linkage disequilibrium (Durand et al., 2011). We used block sizes of 100 SNPs, which, in the data set used for this analysis, surpassed the average length of chromosomes.

Wolbachia infection

Wolbachia is a genus of endosymbiotic bacteria thoroughly studied because of the wide range of effects they induce on their hosts, such as feminisation of males, induced parthenogenesis, male killing and cytoplasmic incompatibility, which may lead to selective sweeps (Hurst & Jiggins, 2005; Jiggins, 2003; Werren et al., 2008). Since *Wolbachia* is maternally inherited, the transmission of a *Wolbachia* strain is associated with a particular mitochondrial genome, hence selective sweeps can be inferred by the expansion of specific *COI* haplotypes.

In order to evaluate if the mtDNA pattern exposed in Hinojosa, Tóth, et al. (2022) was potentially caused by this microorganism, a search for *Wolbachia* sequences was performed using Centrifuge v1.0.4 on the ddRADseq loci obtained with the ipyrad 'de novo' method.

RESULTS

ddRADseq data sets

A total of 140,130,613 raw reads were assembled into 88,328 loci, of which the ipyrad pipeline retained 13,427 loci (Table S2) containing 123,028 SNPs; this SNP data set was used for the F_{ST} calculations, while the loci were used to build the phylogeny and for PhyloNetworks. The unlinked data set had 12,649 SNPs and was used for the STRUCTURE and ABBA–BABA analyses. The unlinked data set with low missing data had 239 SNPs and was used for the PCA and TreeMix analyses since these analyses have been suggested to be sensitive to missing data in population genetics studies (Pickrell & Pritchard, 2012; Yi & Latch, 2022). All these data sets were deposited on figshare (DOI: 10.6084/m9.figshare.22574470).

Phylogenetic reconstruction

The ddRADseq phylogeny (Figure 1a, S1) retrieved all the species as well-supported monophyletic groups (bootstrap = 100). Two main clades were obtained: one including *Melitaea pseudornata*, *M. ornata* and *M. telona* and another including *M. punica* and *M. phoebe*. Two well-supported lineages were recovered in *M. ornata*, separating the Italian individuals from the rest. In *M. phoebe*, the Iberian individuals were differentiated from the other Eurasian samples. In *M. pseudornata*, a main clade included samples from northern Iberia and another clade comprised samples from central and southern Iberia. *Melitaea pseudornata* appeared as sister to a clade including *M. telona* and *M. ornata*, a pattern that was shared by 16 chromosomes. However, using data from the chromosomes 3, 14 or Z (Figure 1b, S2), *M. pseudornata* was sister to *M. ornata* (Figure 1c). The rest of the chromosomes produced other topologies or polytomies, that is, poorly supported nodes (bootstrap <60).

Genetic structuring

The PCA successfully clustered all the species based on PC1 and PC2, which accounted for 53.3% of the variability (Figure 2a). In PC1 (45.1%), two groups emerged: *M. telona* + *M. ornata* + *M. pseudornata* and *M. phoebe* + *M. punica*. In PC2 (8.2%), the different species of these two groups appeared disjunct. PC3 (5.1%) only separated *M. punica* and *M. telona* from the rest.

The STRUCTURE analysis (Figures 2b, S3) displayed a cluster for each species at $K = 6$, except for *M. phoebe*, in which a second one

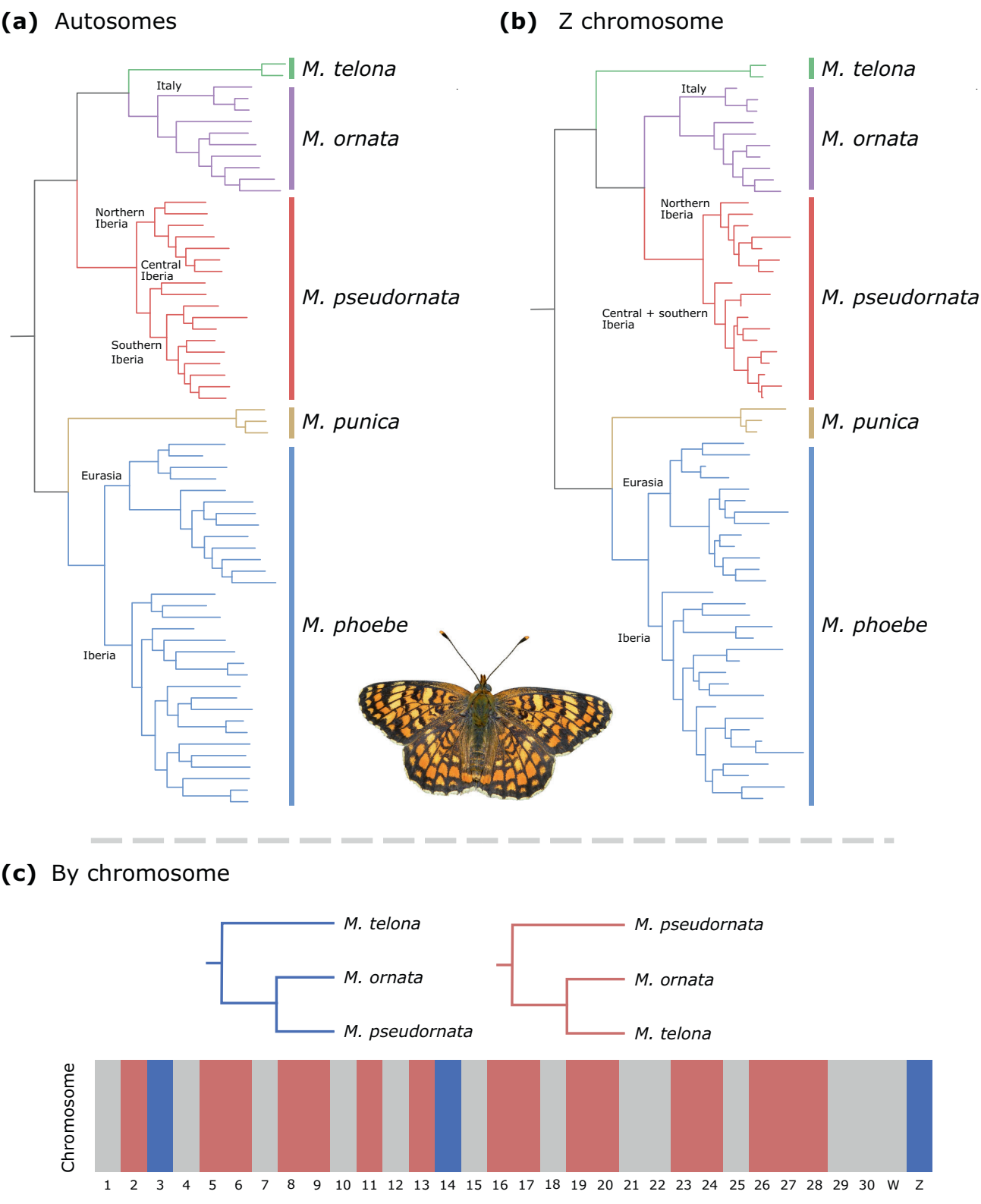

FIGURE 1 Maximum likelihood inference trees based on data from the (a) autosomes and (b) Z chromosome. (c) Relationships between *M. telona*, *M. ornata* and *M. pseudornata* recovered from loci in different chromosomes; grey colour indicates chromosomes that produced trees with poorly supported nodes (bootstrap <60).

appeared at $K > 3$. One of the *M. phoebe* clusters was present in the Iberian Peninsula (dark blue), while the other extended from the eastern Alps to Asia (light blue); individuals displaying an admixture of both clusters were found in the intermediate geographic regions. The only remaining individual with notable cluster admixture was a *M. ornata* from the Caucasus, which showed admixture with the

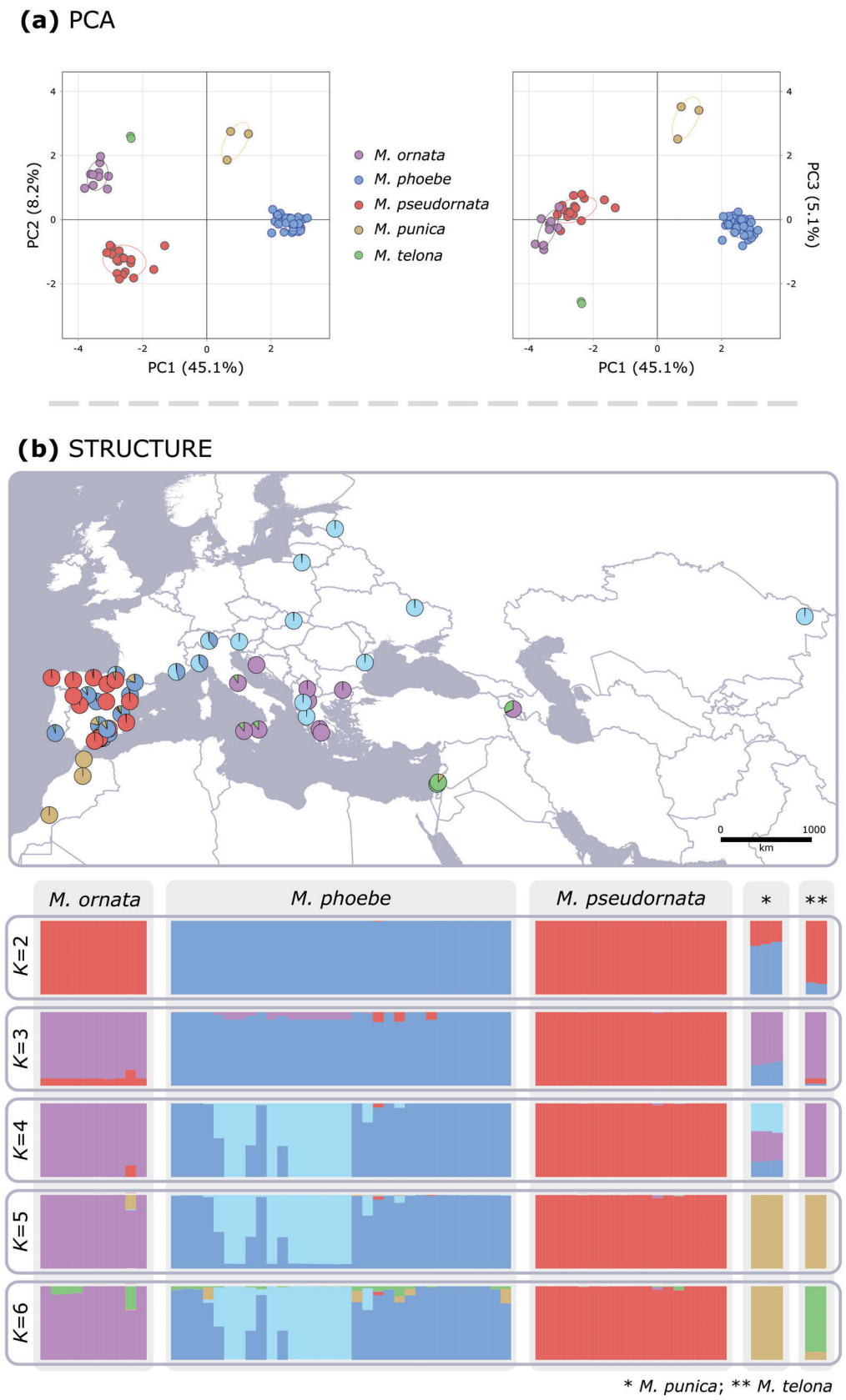

FIGURE 2 (a) Principal component analysis (PCA) and (b) STRUCTURE results ($K = 2$–6; $K = 6$ is represented as pie charts in the map).

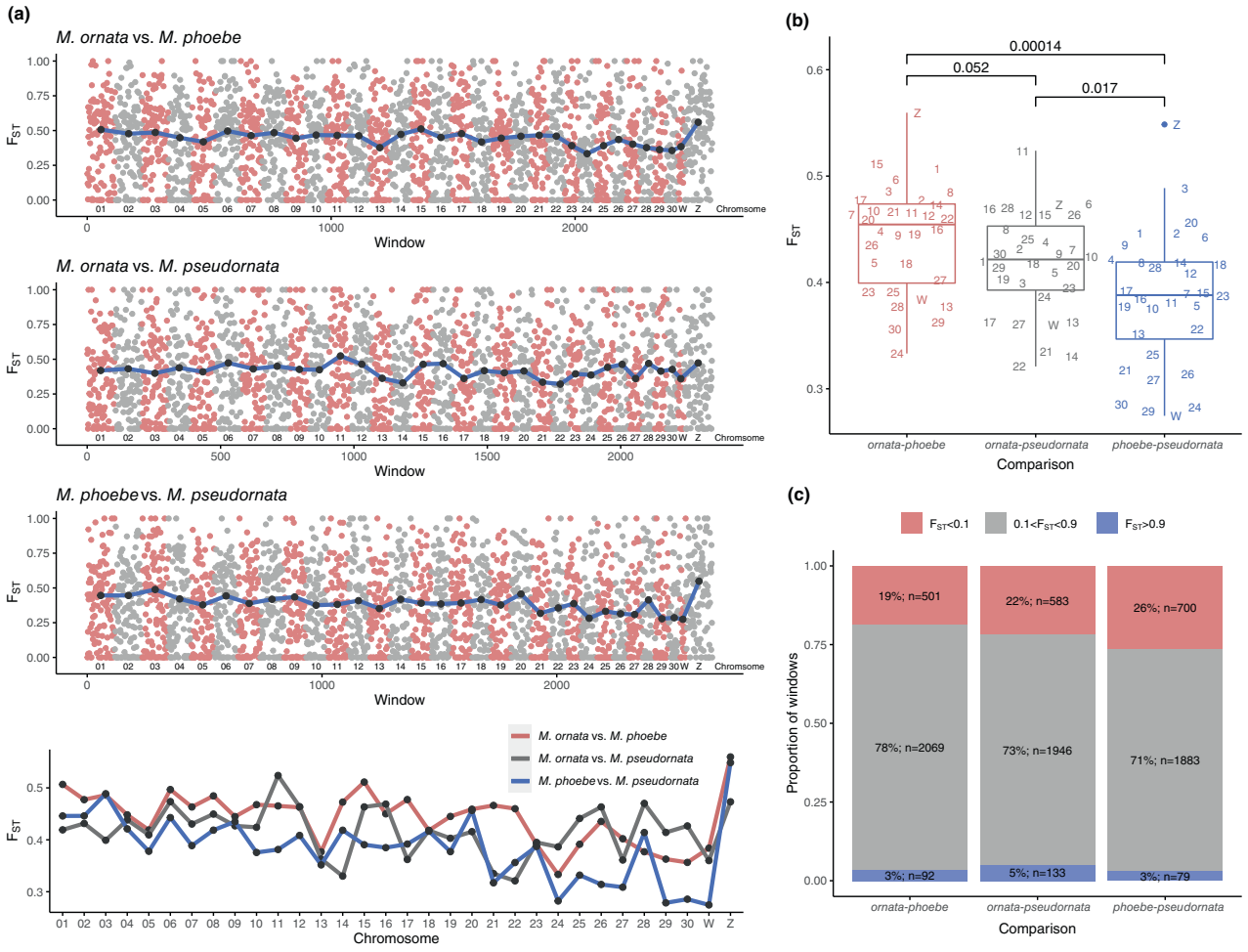

FIGURE 3 (a) F_{ST} values estimated in windows of 200,000 bp and mean F_{ST} per chromosome. (b) Boxplot of the F_{ST} per chromosome and per species pair. (c) Proportion of windows with low, intermediate and high F_{ST} values for each species pair.

Middle Eastern species *M. telona*. Despite including six species, the Evanno method chose $K = 2$ as the best K; this result may be attributed to the 'K = 2 conundrum' (Janes et al., 2017).

Genetic differentiation

The chromosome differentiation in the pair *M. phoebe*–*M. pseudornata* was significantly lower than in the pairs *M. ornata*–*M. phoebe* and *M. ornata*–*M. pseudornata* (Figure 3a,b). The Z chromosome was the most differentiated for the pairs *M. ornata*–*M. phoebe* and *M. phoebe*–*M. pseudornata* (in the latter, it was also an outlier in the boxplot) and was the third most divergent for the comparison *M. ornata*–*M. pseudornata*. The number of windows (Figure 3c) with low F_{ST} values (F_{ST} <0.1) was significantly higher in the pair *M. phoebe*–*M. pseudornata* compared to *M. ornata*–*M. phoebe* ($\chi^2 = 42.2$, $p = < 0.01$) and *M. ornata*–*M. pseudornata* ($\chi^2 = 13.8$, $p < 0.01$); the number of windows with high F_{ST} values (F_{ST} >0.9) was significantly lower in *M. phoebe*–*M. pseudornata* with respect to *M. ornata*–*M. pseudornata* ($\chi^2 = 13.8$, $p = < 0.01$), but not when compared to *M. ornata*–*M. phoebe* ($\chi^2 = 0.9$, $p = 0.35$).

Gene flow analyses

OptM estimated that the optimal number of migration edges (the highest Δm) was one (Figure S4a). Although at $m = 1$, the threshold of 99.8% of explained variance was not reached, when allowing for additional migration edges, multiple alternative gene flow combinations were obtained and no conclusion could be reached. For one migration edge (Figure 4a), TreeMix suggested in 6 out of 11 runs the occurrence of an admixture event from *M. phoebe* to *M. pseudornata* (mean weight = 0.319, mean $p < 0.01$); however, the other five runs indicated admixture in the opposite direction (mean weight = 0.458, mean $p < 0.01$).

PhyloNetworks analyses indicated that all models involving gene flow ($h > 0$) fit our data better than models considering strict bifurcating trees ($h = 0$; Figure S4b). The best phylogenetic network inferred by PhyloNetworks identified one introgression event (hmax = 1, loglik = 0.19) from *M. phoebe* to *M. pseudornata* (inheritance probability (γ) = 0.298; Figure 4b).

The ABBA–BABA tests (Figure 4c) showed an excess of shared alleles between *M. pseudornata* and *M. phoebe* (Figure 4c: $D = 0.57$, f4-ratio = 0.250, p-value <0.01; Figure S5a: $D = 0.27$, f4-ratio = 0.129,

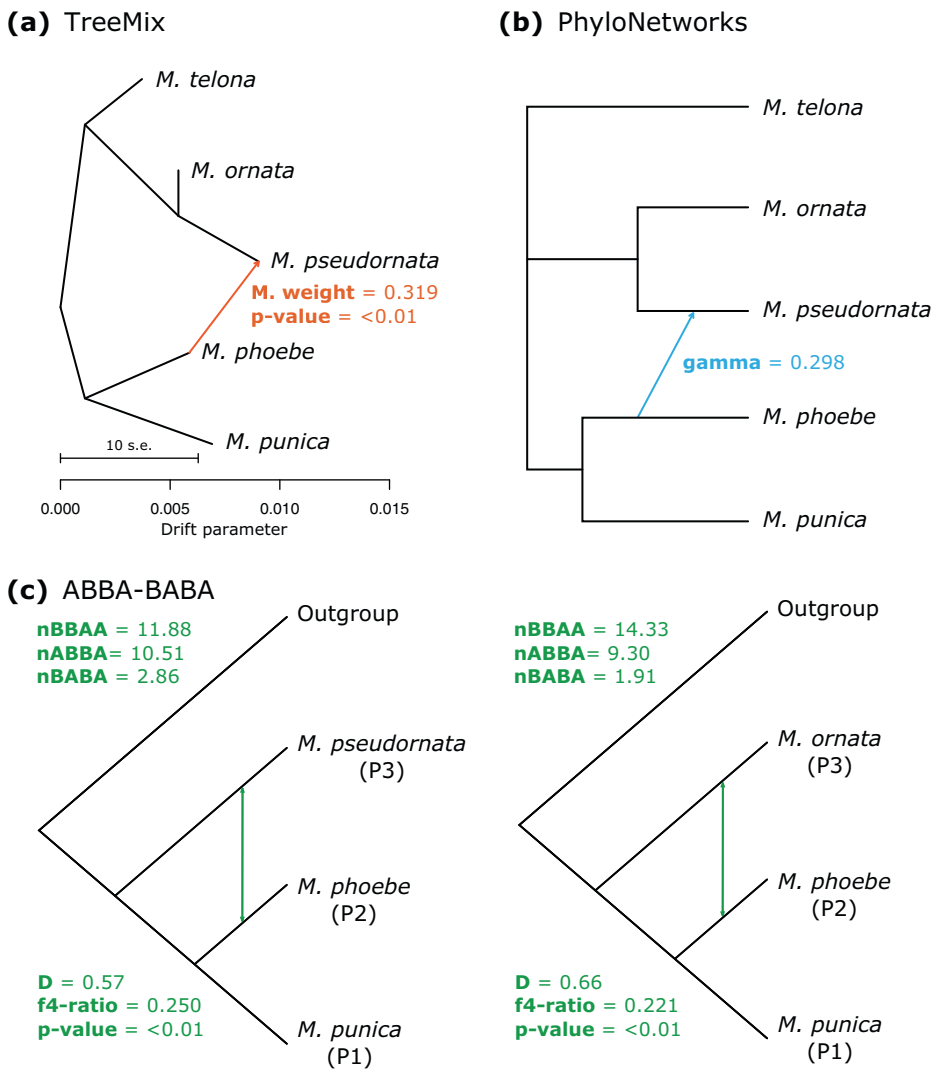

FIGURE 4 (a) Taxa relationships and migration edges inferred by TreeMix with the optimal number of migration edges identified by OptM. The colour indicates the weight of migration edges. The drift parameter is a relative temporal measure and the scale bar indicates 10 times the average standard error of the relatedness among taxa based on the variance–covariance matrix of allele frequencies. (b) PhyloNetworks phylogeny with the optimal reticulation. Branch lengths are expressed in coalescent units, and support values are expressed as local posterior probabilities. (c) Tested ABBA–BABA scenarios and results using blocks of 100 SNPs.

p-value = 0.02), between *M. ornata* and *M. phoebe* (Figure 4c: D = 0.66, f4-ratio = 0.221, *p*-value <0.01) and between the southern individuals of *M. pseudornata* and *M. phoebe* (Figure S5b: *D* = 0.31, f4-ratio = 0.087, *p*-value = 0.01).

Wolbachia infection

Multiple *Wolbachia* sequences were detected in 12 individuals of *M. phoebe* and 11 individuals of *M. pseudornata*, hence we considered these individuals as infected (Table S1). *Melitaea ornata*, *M. phoebe* and *M. pseudornata* share two *COI* barcode haplogroups (Hinojosa, Tóth, et al., 2022): Haplogroup 1, endemic to the Iberian Peninsula, and Haplogroup 2, with a Eurasian distribution and also present in Iberia. All infections were associated with Haplogroup 2, although not all the individuals within this haplogroup were infected (e.g., all the *M. ornata* specimens). The infected individuals were found across all Europe, from the Iberian Peninsula to eastern Europe.

DISCUSSION

Interspecific gene flow in the *M. phoebe* group

Our results revealed that hybridisation occurred between *M. phoebe* and *M. pseudornata*, as detected by Treemix (Figure 4a) and PhyloNetworks (Figure 4b). Additionally, these two species displayed an excess of shared alleles (Figure 4c). The gene flow between *M. phoebe* and *M. pseudornata* can be defined by three main characteristics, namely

(1) its asymmetry, (2) its magnitude and (3) its uneven distribution across the genome:

1. The transfer of genetic material is apparently asymmetric, from *M. phoebe* to *M. pseudornata*. This was highlighted by PhyloNetworks and TreeMix. Furthermore, the mtDNA was also apparently transferred in this direction (Hinojosa, Tóth, et al., 2022). The specific mechanisms causing asymmetric hybridisation in Lepidoptera are thought to be diverse but remain poorly understood. Some examples are reduced hybrid viability in backcrosses with one of the parental species (Havill et al., 2017) or asymmetric mating preferences (Deering & Scriber, 2002). In this case, the cause of the asymmetry remains unknown.
2. A considerable proportion of the genome of *M. pseudornata* was found to be introgressed (Figure 4). TreeMix indicated that the introgressed fraction of the genome was 31.9%, PhyloNetworks suggested a 29.8% and the F_4-ratio a 25.0%. These percentages of introgression are exceptional in animals (Ottenburghs, 2018; Payseur & Rieseberg, 2016), and only a few cases have been reported in insects. The hybrid butterfly species *Coenonympha darwiniana* Staudinger is considered a mixture between *C. arcania* Linnaeus (75% contribution) and *C. gardetta* Prunner (25%; Capblancq et al., 2015); 20%–40% of the genome of the butterfly *Heliconius melpomene* (Linnaeus) showed admixture with *H. cydno* (Doubleday) or *H. timareta* (Hewitson) in sympatry (Martin et al., 2013). These high values are surprising considering that the former pair are separated by ca. 3 million years (Kodandaramaiah & Wahlberg, 2009; Wiemers et al., 2020) and the latter species group by ca. 2 million years (Bull et al., 2006; Salazar et al., 2008). But the divergence of the parental species is even higher in our case, since *M. phoebe* and *M. pseudornata* split ca. 6 million years ago (Tóth et al., 2017). Hybridisation between old species has already been documented in *Melitaea*. In the pair *M. athalia*–*M. celadussa* Fruhstorfer, separated ca. 7 million years ago (Leneveu et al., 2009), widespread hybridisation was found across the contact zone (Tahami et al., 2021), and hybridisation between *M. acentria* Lukhtanov and *M. didyma* (Esper) occurred although they diverged ca. 5 million years ago (Pazhenkova & Lukhtanov, 2021). However, the case of *M. pseudornata* differs from all the above examples because the asymmetric introgression would have completely affected one of the parental species—the ancestral Iberian lineage from which the current *M. pseudornata* originated—across its range. This can be inferred given that: (1) *M. pseudornata* did not appear as an admixture of clusters in STRUCTURE, which suggests that gene flow affected all the individuals equally (Lawson et al., 2018); (2) no deeply diverged lineages of *M. pseudornata* have been recovered, as it might be expected if individuals with small admixture rates are present and (3) the mtDNA of *M. phoebe* is present in all the *M. pseudornata* individuals.
3. The transfer of genetic material from *M. phoebe* to *M. pseudornata* is apparently not uniformly distributed across the genome. The Z chromosome was highly differentiated compared to the autosomes (Figure 3b), which can be the result of the large Z-effect, that is, the disproportionately large role of the Z chromosome in reducing hybrid fitness (Coyne, 2018; Presgraves, 2018). As a result, high levels of divergence relative to autosomes are paired with lower levels of introgression due to the accumulation of incompatible alleles, as it has been reported in other insects (Fontaine et al., 2015) and butterflies (Cong et al., 2019; Martin et al., 2013; Presgraves, 2018), including *Melitaea* (Pazhenkova & Lukhtanov, 2021). Additionally, *M. pseudornata* placed as sister to *M. ornata* in the phylogeny based on the Z chromosome loci (Figure 1b,c), like in the trees obtained when gene flow is considered (Figure 4a,b), suggesting that introgression did not substantially affect this chromosome. Interestingly, the same phylogenetic relationships were inferred for two chromosomes, the 3rd and the 14th. This pattern may be caused by the presence of additional incompatibility genes in these autosomes, although stochasticity cannot be fully discarded.

TreeMix and PhyloNetworks suggested that *M. pseudornata* and *M. ornata* first diverged and only later gene flow occurred. Thus, *M. pseudornata* would not strictly be a hybrid species, but a differentiated entity that was able to overcome massive introgression from *M. phoebe*. Subsequently, the massive acquisition of genetic material from *M. phoebe* may have affected the phenotype of *M. pseudornata*. Gene flow could explain the morphological similarities between the adults of *M. pseudornata* and *M. phoebe*, and also why the body caterpillars of both species are almost identical in northern Iberia (Hinojosa, Tóth, et al., 2022). Besides morphology, gene flow could have contributed to shape other characteristics such as voltinism: in contrast to the exclusively univoltine *M. ornata*—second generations have been obtained only in captivity (Russell et al., 2014; Russell & Pateman, 2013)—at least some populations of *M. pseudornata* are bivoltine (Hinojosa, Tóth, et al., 2022) as is the case of *M. phoebe* in Iberia. These hypotheses, however, cannot be properly addressed with RADseq data and future studies using whole-genome sequencing will be necessary.

Evidence of hybridisation between *M. phoebe* and *M. ornata* was also obtained. There was an excess of shared alleles between these species, but Treemix and PhyloNetworks failed to detect gene flow, which may be the consequence of a lower proportion of the genome affected by gene flow compared to the pair *M. phoebe*–*M. pseudornata*. This is supported by the fact that there were more shared alleles between *M. phoebe* and *M. pseudornata* than between *M. phoebe* and *M. ornata* (Figure S5a), that the genetic differentiation (F_{ST}) of the chromosomes is significantly lower between *M. pseudornata* and *M. phoebe* compared to other species pairs (Figure 3b,c) and that the mtDNA is partially replaced in *M. ornata*, but fully replaced in *M. pseudornata* (Hinojosa, Tóth, et al., 2022). Nevertheless, the genetic exchange between *M. phoebe* and *M. ornata* still shares some characteristics with that between *M. phoebe* and *M. pseudornata*: the asymmetric transference of mtDNA, since some *M. ornata* populations carry *M. phoebe* mitogenomes (Hinojosa, Tóth, et al., 2022), and the uneven distribution across the genome, since

patterns compatible with the large-Z effect were also observed (Figure 3b). Tóth et al. (2017) stated that a single event of unidirectional hybridisation from *M. phoebe* to *M. ornata* occurred long ago in the Italian Peninsula—where all the mtDNA of *M. ornata* is replaced—but the high number of shared alleles between both species and the confirmation of the large Z-effect suggests that hybridisation could have been more widespread.

Patterns of intraspecific variation

The ddRADseq data revealed several phylogeographic patterns at the intraspecific level, probably products of present and/or past geographic isolation. The phylogeny displayed two main lineages within each species: *M. ornata* (Italy and Balkan Peninsula), *M. pseudornata* (north and central-south Iberia) and *M. phoebe* (Iberia and Eurasia). In the case of *M. pseudornata*, the distinction between the lineages was additionally influenced by differential gene flow with *M. phoebe*, since a slight excess of shared alleles was estimated between *M. phoebe* and the southern populations of *M. pseudornata* (Figure S5b).

Genetic differences between the lineages of *M. phoebe* appeared in the STRUCTURE results (Figure 2b). The two lineages partially match with the distribution of two mtDNA haplogroups (Hinojosa, Tóth, et al., 2022), although some Iberian individuals may have acquired Eurasian mtDNA haplotypes. Some of these haplotypes carry *Wolbachia*, hence their expansion could have been boosted by the presence of this endosymbiont. The distribution of the Iberian lineage matches with the subspecies *M. p. occitanica* Staudinger, currently considered to be endemic to the Iberian Peninsula (Russell et al., 2020; Tolman & Lewington, 2008), while the Eurasian lineage is represented by an array of subspecies, including the nominal. Similar phylogeographic patterns involving Iberia versus the rest of Europe genetically differentiated lineages have been described in butterflies and are a source of taxonomic debate. For example, this case presents analogies with that of *Euphydryas aurinia* (Rottemburg)—*E. beckeri* (Lederer) (Korb et al., 2016), treated as a single species in recent checklists (Dapporto et al., 2022; Wiemers et al., 2018). Besides, several individuals from intermediate geographic areas displayed an admixture of STRUCTURE clusters from both lineages, which is an expected output under a scenario of recent admixture (Lawson et al., 2018). These specimens constituted a distinct clade within the Eurasian lineage in the phylogeny, an effect that may be caused by phylogenetic inference methods not being able to properly reflect gene flow. The relatively low divergence and the abundance of admixed individuals suggests that no strong reproductive barriers isolate the two lineages, and hence we consider that they represent intraspecific variation.

CONCLUSIONS

We provided evidence of gene flow between *M. phoebe* and two other species, *M. pseudornata* and *M. ornata*. Gene flow between *M. phoebe* and *M. pseudornata* was characterised by (1) its asymmetry, with genetic transfer from *M. phoebe* to *M. pseudornata* being dominant; (2) its magnitude, with up to 31.9% of the genome of *M. pseudornata* potentially acquired from *M. phoebe* and (3) its uneven distribution across the genome, with most of the Z chromosome and two autosomes being less affected by introgression. Gene flow between *M. phoebe* and *M. ornata* could also be asymmetric (from *M. phoebe* to *M. ornata*) and patterns compatible with the large-Z effect were observed, but the proportion of the genome affected by hybridisation appears to be lower.

Melitaea pseudornata is unlikely to be a hybrid species. We argue that it first diverged from *M. ornata* and then overcame massive introgression from *M. phoebe*. Although *M. pseudornata* endured this process and remained a separate entity from *M. phoebe*, gene flow may explain the morphological similarities of their caterpillars (in northern Iberia) and adults, as well as the voltinism of *M. pseudornata*.

The intraspecific variation patterns of these three species were likely driven by present and/or past geographic isolation. Two *M. phoebe* lineages split early in the STRUCTURE analyses, but a wide area of admixed populations apparently exists, suggesting that no strong reproductive barriers occur and that these lineages of moderate divergence represent intraspecific variation. In the case of *M. pseudornata*, the distinction between the lineage from central and southern Iberia and the lineage from northern Iberia was also influenced by differential gene flow with *M. phoebe*, which was slightly higher in southern Iberia.

AUTHOR CONTRIBUTIONS

Joan C. Hinojosa: Writing – review and editing; conceptualization; investigation; writing – original draft; methodology; visualization; software; formal analysis; project administration; validation; data curation; resources. **Valéria Marques:** Writing – review and editing; data curation; formal analysis; software; methodology. **Luis Sánchez Mesa:** Writing – review and editing; investigation; resources. **Leonardo Dapporto:** Writing – review and editing; methodology; resources. **Vlad Dincă:** Writing – review and editing; methodology; resources. **Roger Vila:** Conceptualization; funding acquisition; supervision; project administration; resources; writing – review and editing; methodology.

ACKNOWLEDGEMENTS

We acknowledge M. G. Muñoz Sariot, Y. Monasterio, R. Vodă, S. Kunze, W. Opdekamp, M. Gascoigne-Pees, H. Anastassiu, P. Vantieghem, E. García Barros, J. Couckuyt, N. Micevski, E. Karolinskiy, S. Montagud, A. Sendra, J. Hernández-Roldán, F. González, M. López Munguira, H. Romo, S. Cuvelier, M. Tarrier and D. Benyamini for specimen collection; K. Lohse for his valuable comments on the study and L. Törmälä, C. Corbella, E. Sbrega and N. T. Lam for their help in the laboratory. We thank Marianne Espeland, Camille Cornet and one anonymous reviewer for insightful comments on an earlier version of the manuscript.

FUNDING INFORMATION

Financial support was provided by grants PID2019-107078GB-I00 and PID2022-139689NB-I00 (MICIU/AEI/10.13039/501100011033 and ERDF, EU) and 2021-SGR-00420 (Departament de Recerca i

Universitats, Generalitat de Catalunya) to Roger Vila, by grant BES-2017-080641, funded by MICIU/AEI/10.13039/501100011033 and by 'ESF Investing in your future' to Joan C. Hinojosa, by grant PRE2020-094870, funded by MICIU/AEI/10.13039/501100011033 and by the 'European Social Fund (ESF) Investing in your future' to Valéria Marques, by the Academy of Finland (Academy Research Fellow, decisions no. 324988 and 352652) to Vlad Dincă and by National Biodiversity Future Centre (Ministero dell'Università e della Ricerca, PNRR, project CN00000033) and by Direttiva Biodiversità del Ministero dell'Ambiente e della Sicurezza Energetica to Leonardo Dapporto.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

Raw ddRADseq FASTQ reads: NCBI Bioproject PRJNA1080223, SRA accession numbers SRR28153373-SRR28153439. Data sets used and output files: figshare (DOI: 10.6084/m9.figshare.22574470).

ORCID

Joan C. Hinojosa https://orcid.org/0000-0002-6318-4252
Valéria Marques https://orcid.org/0000-0002-9954-7027
Leonardo Dapporto https://orcid.org/0000-0001-7129-4526
Vlad Dincă https://orcid.org/0000-0003-1791-2148
Roger Vila https://orcid.org/0000-0002-2447-4388

REFERENCES

Abbott, R., Albach, D., Ansell, S., Arntzen, J.W., Baird, S.J., Bierne, N. et al. (2013) Hybridisation and speciation. *Journal of Evolutionary Biology*, 26(2), 229–246.

Arnold, M.L. (1997) *Natural hybridisation and evolution*. New York: Oxford University Press.

Baird, N.A., Etter, P.D., Atwood, T.S., Currey, M.C., Shiver, A.L., Lewis, Z.A. et al. (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLOS ONE*, 3(10), e3376. Available from: https://doi.org/10.1371/journal.pone.0003376

Bull, V., Beltrán, M., Jiggins, C.D., McMillan, W.O., Bermingham, E. & Mallet, J. (2006) Polyphyly and gene flow between non-sibling *Heliconius* species. *BMC Biology*, 4(1), 1–17.

Capblancq, T., Després, L., Rioux, D. & Mavárez, J. (2015) Hybridisation promotes speciation in *Coenonympha* butterflies. *Molecular Ecology*, 24(24), 6209–6222.

Cianchi, R., Ungaro, A., Marini, M. & Bullini, L. (2003) Differential patterns of hybridisation and introgression between the swallowtails *Papilio machaon* and *P. hospiton* from Sardinia and Corsica islands (Lepidoptera, Papilionidae). *Molecular Ecology*, 12(6), 1461–1471.

Cong, Q., Zhang, J. & Grishin, N.V. (2019) Genomic determinants of speciation. *bioRxiv*, 837666. Available from: https://www.biorxiv.org/content/10.1101/837666v1.abstract

Coyne, J.A. (2018) "Two rules of speciation" revisited. *Molecular Ecology*, 27(19), 3749–3752.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A. et al. (2011) The variant call format and VCFtools. *Bioinformatics*, 27(15), 2156–2158. Available from: https://doi.org/10.1093/bioinformatics/btr330

Dapporto, L., Menchetti, M., Vodă, R., Corbella, C., Cuvelier, S., Djemadi, I. et al. (2022) The atlas of mitochondrial genetic diversity for Western Palaearctic butterflies. *Global Ecology and Biogeography*, 31, 2184–2190.

Deering, M.D. & Scriber, M.J. (2002) Field bioassays show heterospecific mating preference asymmetry between hybridizing North American Papilio butterfly species (Lepidoptera: Papilionidae). *Journal of Ethology*, 20(1), 25–33.

Descimon, H. & Mallet, J. (2009) Bad species. In: Settele, J., Shreeve, T.G., Konvicka, M. & Van Dyck, H. (Eds.) *Ecology of butterflies in Europe*. Cambridge, UK: Cambridge University Press, pp. 219–249.

Dincă, V., Dapporto, L., Somervuo, P., Vodă, R., Cuvelier, S., Gascoigne-Pees, M. et al. (2021) High resolution DNA barcode library for European butterflies reveals continental patterns of mitochondrial genetic diversity. *Communications Biology*, 4(1), 1–11.

Dincă, V., Lee, K.M., Vila, R. & Mutanen, M. (2019) The conundrum of species delimitation: a genomic perspective on a mitogenetically supervariable butterfly. *Proceedings of the Royal Society B*, 286(1911), 20191311.

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. Available from: https://doi.org/10.1093/molbev/msr048

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

Evanno, G., Regnaut, S. & Goudet, J. (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, 14(8), 2611–2620. Available from: https://doi.org/10.1111/j.1365-294X.2005.02553.x

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

Fontaine, M.C., Pease, J.B., Steele, A., Waterhouse, R.M., Neafsey, D.E., Sharakhov, I.V. et al. (2015) Extensive introgression in a malaria vector species complex revealed by phylogenomics. *Science*, 347(6217), 1258524.

Goulet, B.E., Roda, F. & Hopkins, R. (2017) Hybridisation in plants: old ideas, new techniques. *Plant Physiology*, 173(1), 65–78.

Haldane, J.B. (1922) Sex ratio and unisexual sterility in hybrid animals. *Journal of Genetics*, 12(2), 101–109.

Havill, N.P., Elkinton, J., Andersen, J.C., Hagen, S.B., Broadley, H.J., Boettner, G.J. et al. (2017) Asymmetric hybridisation between non-native winter moth, *Operophtera brumata* (Lepidoptera: Geometridae), and native Bruce spanworm, *Operophtera bruceata*, in the northeastern United States, assessed with novel microsatellites and SNPs. *Bulletin of Entomological Research*, 107(2), 241–250.

Hinojosa, J.C., Dapporto, L., Pitteloud, C., Koubínová, D., Hernández-Roldán, J., Vicente, J.C. et al. (2022) Hybridisation fuelled diversification in *Spialia* butterflies. *Molecular Ecology*, 31(10), 2951–2967.

Hinojosa, J.C., Tóth, J.P., Monasterio, Y., Sánchez Mesa, L., Muñoz Sariot, M.G., Escobés, R. et al. (2022) Integrative taxonomy reveals a new *Melitaea* (Lepidoptera: Nymphalidae) species widely distributed in the Iberian Peninsula. *Insect Systematics and Diversity*, 6(2), 1.

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. Available from: https://doi.org/10.1098/rspb.2005.3056

Janes, J.K., Miller, J.M., Dupuis, J.R., Malenfant, R.M., Gorrell, J.C., Cullingham, C.I. et al. (2017) The K=2 conundrum. *Molecular Ecology*, 26(14), 3594–3602.

Jiggins, F.M. (2003) Male-killing *Wolbachia* and mitochondrial DNA: selective sweeps, hybrid introgression and parasite population dynamics. *Genetics*, 164(1), 5–12.

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. Available from: https://doi.org/10.1186/1471-2156-11-94

Kodandaramaiah, U. & Wahlberg, N. (2009) Phylogeny and biogeography of *Coenonympha* butterflies (Nymphalidae: Satyrinae)–patterns of colonization in the Holarctic. *Systematic Entomology*, 34(2), 315–323.

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

Korb, S.K., Bolshakov, L.V., Fric, Z.F. & Bartonova, A. (2016) Cluster biodiversity as a multidimensional structure evolution strategy: checkerspot butterflies of the group *Euphydryas aurinia* (Rottemburg, 1775) (Lepidoptera: Nymphalidae). *Systematic Entomology*, 41(2), 441–457.

Lawson, D.J., Van Dorp, L. & Falush, D. (2018) A tutorial on how not to over-interpret STRUCTURE and ADMIXTURE bar plots. *Nature Communications*, 9(1), 1–11.

Leneveu, J., Chichvarkhin, A. & Wahlberg, N. (2009) Varying rates of diversification in the genus *Melitaea* (Lepidoptera: Nymphalidae) during the past 20 million years. *Biological Journal of the Linnean Society*, 97(2), 346–361.

Litman, J., Chittaro, Y., Birrer, S., Praz, C., Wermeille, E., Fluri, M. et al. (2018) A DNA barcode reference library for Swiss butterflies and forester moths as a tool for species identification, systematics and conservation. *PLoS One*, 13(12), e0208639.

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. Available from: https://doi.org/10.1111/1755-0998.13265

Mallet, J. (2005) Hybridisation as an invasion of the genome. *Trends in Ecology & Evolution*, 20(5), 229–237.

Martin, S.H., Dasmahapatra, K.K., Nadeau, N.J., Salazar, C., Walters, J.R., Simpson, F. et al. (2013) Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. *Genome Research*, 23(11), 1817–1828.

Muñoz Sairot, M.G. & Sánchez Mesa, L. (2019a) Nueva subespecie de *Melitaea ornata* (Christoph, 1893), con la descripción de sus estadios preimaginales (Lepidoptera: Nymphalidae). *Arquivos Entomolóxicos*, 21, 5–20.

Muñoz Sairot, M.G. & Sánchez Mesa, L. (2019b) *Melitaea ornata pseudornata* nuevo nombre de reemplazo para *Melitaea ornata baetica* Muñoz Sariot & Sánchez Mesa, 2019 (Lepidoptera: Nymphalidae). *Arquivos Entomolóxicos*, 21, 83–84.

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: international journal of organic. *Evolution*, 67(4), 1055–1068.

Ottenburghs, J. (2018) Exploring the hybrid speciation continuum in birds. *Ecology and Evolution*, 8(24), 13027–13034.

Pardo-Diaz, C., Salazar, C., Baxter, S.W., Merot, C., Figueiredo-Ready, W., Joron, M. et al. (2012) Adaptive introgression across species boundaries in *Heliconius* butterflies. *PLoS Genetics*, 8(6), e1002752.

Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y. et al. (2012) Ancient admixture in human history. *Genetics*, 192(3), 1065–1093. Available from: https://doi.org/10.1534/genetics.112.145037

Payseur, B.A. & Rieseberg, L.H. (2016) A genomic perspective on hybridisation and speciation. *Molecular Ecology*, 25(11), 2337–2360.

Pazhenkova, E.A. & Lukhtanov, V.A. (2021) Genomic introgression from a distant congener in the Levant fritillary butterfly, *Melitaea acentria*. *Molecular Ecology*, 30(19), 4819–4832.

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

Presgraves, D.C. (2002) Patterns of postzygotic isolation in Lepidoptera. *Evolution*, 56(6), 1168–1183.

Presgraves, D.C. (2018) Evaluating genomic signatures of "the large X-effect" during complex speciation. *Molecular Ecology*, 27(19), 3822–3830.

Pritchard, J.K., Stephens, M. & Donnelly, P. (2000) Inference of population structure using multilocus genotype data. *Genetics*, 155(2), 945–959.

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

Rambaut, A. (2015) FigTree, v1.4.2: tree figure drawing tool. Molecular evolution, phylogenetics and epidemiology. Available from: http://tree.bio.ed.ac.uk/software/figtree/

Russell, P. & Pateman, J. (2013) Confirmation of the presence of *Melitaea ornata* Christoph, 1893 (Lepidoptera: Nymphalidae) on the eastern Aegean Island of Chíos, Greece, and its host-plant. *Entomologist's Gazette*, 64, 217–224.

Russell, P., Pateman, J. & Verovnik, R. (2014) First record of *Melitaea ornata* Christoph, 1893, from Slovenia, with notes on its confirmed distribution and hybridisation with *M. phoebe* ([Denis and Schiffermüller], 1775). *Entomologist's Gazette*, 65, 135–153.

Russell, P.J.C., Bartolozzi, L., Hawkins, R.L., Tennent, W.J. & Léger, T. (2020) Designation of lectotypes for some Spanish and other western European *Melitaea* taxa, some with mixed syntypic series of *M. phoebe* ([Denis & Schiffermüller], 1775) and *M. ornata* Christoph, 1893 (Lepidoptera: Nymphalidae). *SHILAP Revista de lepidopterología*, 48(191), 449–472.

Salazar, C., Jiggins, C.D., Taylor, J.E., Kronforst, M.R. & Linares, M. (2008) Gene flow and the genealogical history of *Heliconius heurippa*. *BMC Evolutionary Biology*, 8(1), 1–13.

Solís-Lemus, C. & Ané, C. (2016) Inferring phylogenetic networks with maximum pseudolikelihood under incomplete lineage sorting. *PLoS Genetics*, 12(3), e1005896.

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

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

Tahami, M.S., Dincă, V., Lee, K.M., Vila, R., Joshi, M., Heikkilä, M. et al. (2021) Genomics reveal admixture and unexpected patterns of diversity in a parapatric pair of butterflies. *Genes*, 12(12), 2009.

The Heliconius Genome Consortium. (2012) Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature*, 487, 94–98.

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

Tóth, J.P., Varga, Z., Verovnik, R., Wahlberg, N., Váradi, A. & Bereczki, J. (2017) Mito-nuclear discordance helps to reveal the phylogeographic patterns of *Melitaea ornata* (Lepidoptera: Nymphalidae). *Biological Journal of the Linnean Society*, 121(2), 267–281.

Valencia-Montoya, W.A., Elfekih, S., North, H.L., Meier, J.I., Warren, I.A., Tay, W.T. et al. (2020) Adaptive introgression across semipermeable species boundaries between local *Helicoverpa zea* and invasive *Helicoverpa armigera* moths. *Molecular Biology and Evolution*, 37(9), 2568–2583.

van Oorschot, H. & Coutsis, J.G. (2014) *The genus Melitaea Fabricius, 1807 (Lepidoptera: Nymphalidae, Nymphalinae). Taxonomy and systematics with special reference the male genitalia*. Pardubice: Tshikolovets Publications.

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

Wiemers, M., Balletto, E., Dincă, V., Fric, Z.F., Lamas, G., Lukhtanov, V. et al. (2018) An updated checklist of the European butterflies (Lepidoptera, Papilionoidea). *ZooKeys*, 811, 9–45.

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.

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

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.

SUPPORTING INFORMATION

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

Table S1. Samples used in this study.

Table S2. Number of raw reads obtained and loci retrieved after the ipyrad pipeline.

Figure S1. Maximum likelihood inference tree of nuDNA loci. Scale units are presented in substitutions per site.

Figure S2. Maximum likelihood inference tree of Z chromosome loci. Scale units are presented in substitutions per site.

Figure S3. STRUCTURE results represented as bars ($K = 2$–6) and ΔK values obtained after the Evanno method.

Figure S4. (a) Selection of the most likely number of migration edges by the OptM R package. Distribution of log likelihood and variance explained of TreeMix models with 0–4 migration edges. Standard deviation generated by repeating analyses by varying k (SNPs per window) from 10 to 20, with an increment of 1 per run. The optimal number of migration edges (m) correspond to the highest second-order rate of change (Δm), here 1. (b) Plot of pseudolikelihood score as a function of the number of reticulation events to estimate network complexity in PhyloNetworks. Here the slope heuristic suggests a single hybrid node (hmax = 1) since it is the lowest value and the score does not change beyond $h = 1$ node.

Figure S5. Tested ABBA-BABA scenarios and results using blocks of 100 SNPs.

How to cite this article: Hinojosa, J.C., Marques, V., Mesa, L.S., Dapporto, L., Dincă, V. & Vila, R. (2024) Can species endure massive introgression? Genomic evidence of asymmetric gene flow in *Melitaea* butterflies. *Systematic Entomology*, 1–13. Available from: https://doi.org/10.1111/syen.12631