Historical and current patterns of gene flow in the butterfly *Pararge aegeria*

Luca Livraghi¹,*, Raluca Vodă²,*, Luke Christopher Evans¹,³ | Melanie Gibbs⁴ | Vlad Dincă⁵,⁶ | Peter W.H. Holland⁷ | Tim G. Shreeve⁸ | Roger Vila⁶ | Leonardo Dapporto⁹ | Casper J. Breuker¹

¹Evolutionary Developmental Biology Research Group, Department of Biological and Medical Sciences, Faculty of Health and Life Sciences, Oxford Brookes University, Headington, Oxford, UK
²DBIOS Department of Life Sciences and Systems Biology, University of Turin, Turin, Italy
³Ecology and Evolutionary Biology Research Division, School of Biological Sciences, University of Reading, Reading, Berkshire, UK
⁴Maclean Building, NERC Centre for Ecology & Hydrology, Wallingford, UK
⁵Department of Ecology and Genetics, University of Oulu, Oulu, Finland
⁶Institute of Evolutionary Biology, Barcelona, Spain
⁷Department of Zoology, University of Oxford, Oxford, UK
⁸Centre for Ecology, Environment and Conservation, Oxford Brookes University, Headington, Oxford, UK
⁹Dipartimento di Biologia, Università degli Studi di Firenze, Sesto Fiorentino, Italy

Correspondence
Casper J. Breuker, Evolutionary Developmental Biology Research Group, Department of Biological and Medical Sciences, Faculty of Health and Life Sciences, Oxford Brookes University, Gipsy Lane, Headington, Oxford OX3 0BP, UK. Email: cbreuker@brookes.ac.uk

Funding information
European Union’s Seventh Framework programme for research and innovation under the Marie Skłodowska-Curie grant agreement, Grant/Award Number: 609402 - 2020; Santander Student Research Grant; Nigel Groome PhD studentship; Marie Curie International Outgoing Fellowship within the 7th European Community Framework Programme, Grant/Award Number: 625997; Santander Research Scholarship

Editor: Simone Fattorini

Abstract

**Aim:** We have investigated the phylogeography and genetic structure of the Speckled Wood butterfly (*Pararge aegeria*) across its entire distribution range and studied its dispersal both on mainland and across sea straits. The apparent lack of gene flow between Sardinia and Corsica was further investigated by means of mating experiments.

**Location:** Europe and North Africa.

**Methods:** We sampled 345 individuals and sequenced one mitochondrial gene (*Cytochrome c Oxidase subunit I, COI*) for all samples and two nuclear genes (*wingless* and *zerknult*) for a subset of the specimens. A total of 22 females from Corsica and Sardinia were used to establish a series of crosses to investigate reproductive compatibility and were screened for the presence of *Wolbachia*. Bayesian inference (BI) and haplotype networks were employed to infer phylogenetic relationships and a Principal Coordinate Analysis (PCoA) was used to represent geographical patterns of genetic diversity. Mating and courtship data were analysed using linear mixed effect models.

**Results:** We detected two main COI lineages separated by the Mediterranean Sea and maintained over relatively short sea straits. While nuclear gene variation was generally in agreement with that of COI, this was not the case in all areas (e.g. Iberian Peninsula and Corsica/Sardinia). Mating experiments revealed no evidence of reproductive isolation between the lineages, nor clear relation to *Wolbachia* infection status.

*Authors contributed equally.*

DOI: 10.1111/jbi.13354
Main conclusions: We propose that following the post-glacial recolonization of Europe, the ancestral COI lineage of *P. aegeria* was maintained in North Africa and Mediterranean islands, while a new lineage colonized from Eastern Europe, replacing and apparently out-competing the ancestral variant. Several hypotheses are discussed that may explain the local discordance between the nuclear genes and COI, including sex-specific dispersal, selection and differential rates of gene evolution.

**KEYWORDS**

2012). A recent study by Longdon et al. (2017) examining the modes of transmission in a range of different Rhabdoviruses and their population genetics, which often reflect those of their hosts (Longdon et al., 2017; Wilfert & Jiggins, 2014), highlighted discrete Sardinian and Corsican populations of the P. aegeria specific Rhabdovirus PAegRV (for a detailed description of this recently discovered virus see Longdon et al., 2015). This strongly suggests limited dispersal between the islands.

The variation between populations on Corsica and Sardinia thus represents a particularly intriguing case, which is a focus of this study. These islands are separated by $<12$ km of sea straits in which several small adjacent islands could potentially act as stepping stones. Moreover, in contrast to the areas separated by the Gibraltar and Sicilian channels, these islands were connected during the Last Glacial Maximum (LGM) suggesting that the two different populations were established from different source populations following relatively recent post-glacial dynamics (Dapporto, 2010) and thereafter, there has been little or no dispersal over the Bonifacio strait.

Several explanations can be provided for the observed distributions of island populations. Corsica and Sardinia have different environmental settings, with considerable variation in temperature and rainfall (reflected in the vegetation) (Hijmans, Cameron, Parra, Jones, & Jarvis, 2005). It is highly unlikely that climatic differences alone prevent the European lineage from establishing populations on Sar- dinia and vice versa, but local adaptation may reduce the likelihood of colonization (cf. Richter-Boix, Quintela, Kierczak, Franck, & Laurila, 2013). Climatic factors and their effects on host plants have indeed imposed strong selection pressures in P. aegeria that influence egg-laying strategies (Gibbs & Van Dyck, 2010; Gibbs, Van Dyck, & Breuker, 2012; Hill, Thomas, & Huntley, 1999). Furthermore, it may be possible that reproductive isolation is emerging between the two lineages; female mate choice, in particular, has been recorded as a factor in maintaining reproductive isolation in several butterfly species (e.g. Dincă et al., 2013; Pinzari & Sbordoni, 2013).

Even in the absence of reproductive barriers, hybrid fitness could be reduced, thus explaining the mutual exclusion pattern. Although very little is understood about the reduction in hybrid fitness at the molecular level (Presgraves, Balagopalan, Abmayr, & Orr, 2003; Rogers & Bernatchez, 2006), three specific forms of post-zygotic isolation have been described: sterility of F1 hybrids, lethality of F1 hybrids and degeneracy of F2 hybrids (Dobzhansky, 1970; Dumas et al., 2015). Thus, it may be possible that no strict pre- or postzygotic barriers exist, but that immigrants and their (hybrid) offspring find themselves at a selective disadvantage compared to the endemics.

To address these issues, we sampled numerous populations of P. aegeria to cover as much as possible of their European and North African range. We specifically focused on Corsica and Sardinia, the closest areas where the two lineages can be found with apparent lack of gene flow and sequenced COI as well as two nuclear developmental genes for a subset of individuals. The transcription factor-encoding zerknullt (zen) (for a description of this gene in P. aegeria see Ferguson et al., 2014), was added to data on the traditionally used gene wingless (wg), encoding a signalling protein to increase the nuclear sequence depth (see also Wahlberg & Wheat, 2008). This allowed us to investigate, with high spatial resolution, the distribution of the two genetic lineages and their intra-lineage genetic diversity over the study area. Moreover, to test the hypothesis that pre- or postzygotic barriers affect gene flow between the two lineages over Sardinia and Corsica, the reproductive strategies of Sardinian and Corsican P. aegeria were examined using courtship and mating experiments.

2 | MATERIALS AND METHODS

2.1 | DNA extraction, amplification, sequencing and phylogenetic analyses

Using PCR and direct Sanger sequencing (see Appendix S1 in Supporting Information, including details on primers and cycling conditions), we obtained sequences of: the 658 bp barcoding region of COI for 345 individuals spanning from North Africa to northern Europe, of a 411 bp region of the gene wingless (wg) for a subset of 87 individuals, and of the entire coding region of zen (1,599 bp) for 79 individuals. We further obtained two outgroup sequences of wg and zen belonging to the closely related species Pararge xiphioides (Stau- dinger, 1871) (cf. Weingartner et al., 2006). Bayesian inference (BI) for the three genes was employed to infer phylogenetic relationships with BEAST 1.8.0 (Drummond, Suchard, Xie, & Rambaut, 2012). Patterns of genetic variation were inferred by maximum parsimony haplotype networks using the program tcs 1.21, with a 95% connection limit (Clement, Posada, & Crandall, 2000). Representations of genetic diversity were created for the three genetic markers by calculating matrices of p-distances for each of them, and subsequently analysing and plotting these using R and QGIS 2.0.1. (QGIS Development Team 2009). Further details on sequencing and the phylogenetic, haplotype network and genetic distance analyses can be found in Appendix S1.

2.2 | Sardinian and Corsican samples

Between the 6th and the 12th of May 2014 P. aegeria females were collected in the field from 11 different localities in: Sardinia (Arizto, Desulo and Tempio Pausania), Corsica (Asco, Zonza, Bavella, Bonifa- cio, Solenzara, Cavallo Morto and Pietralba) and La Maddalena (Sua- leddu), a smaller island off the north coast of Sardinia. In total 32 females were caught: 18 from Corsica, 13 from Sardinia and 1 from La Maddalena. Eggs from collected females were obtained in-situ and brought to the laboratory in Oxford, UK. Upon hatching these eggs were put on host plants for this species in Europe (a mix of Poa trivialis and Dactylis glomerata) and reared at 21 ± 2°C (60% RH, 16L:8D) (cf. Gibbs, Breuker, & Van Dyck, 2010). Rearing and mating conditions in this study included full daylight spectrum lamps, including UV-light (Osram Biolux). The females collected in the field laid readily on these plant species, as did all the females used in this experiment. Pupae were sexed and kept individually, to ensure virgin adults were available for setting up crosses. The offspring of a total
of 22 of the field-collected females provided the adults to perform the crosses detailed below (Appendix S2 in Supporting Information for details on collection).

2.3 | Pre-zygotic reproductive barriers: courtship behaviour in Sardinian and Corsican Pararge aegeria

Crosses were performed with the offspring of the wild-caught females. Four types of crosses were set up: Corsican male/Corsican female (CC), Corsican male/Sardinian female (CS), Sardinian male/Sardinian female (SS) and Sardinian male/Corsican female (SC). In total 72 crosses generated data to be used in the analyses (CC = 27, CS = 20, SC = 11, SS = 14). To perform the crosses, newly eclosed virgin females were placed in cages along with an artificial flower containing 10% honey solution (Gibbs et al., 2012). A newly eclosed virgin male was then introduced into a female’s cage and the total courtship duration (seconds) was timed using stopwatches. The primary aim of these crosses was to establish how willing males and females of the two islands were to mate with each other, and having done so, what the reproductive output was. Thus, no mate choice experiments per se were conducted (i.e. where a female needed to choose between a male from her own island or the other one). Pararge aegeria have a courtship very similar to that described for the graying where courtship is initiated by a wing flick used by males, to the front and side of the female (Davies, 1978 and references therein). We used this male wing flick as our cue for courtship initiation. If the male was unsuccessful at initiating mating after numerous bouts of courtship between 8 a.m. and 6 p.m., it was removed and replaced with a new virgin male the following morning (8 a.m.). After mating had finished an egg laying plant was placed in the cage and the male was removed.

2.4 | Reproductive barriers

After mating the female was left to oviposit for 6 days and all eggs laid in that period were collected. All females were allowed to oviposit on the exact same host plant species (a mix of P. trivialis and D. glomerata). Six days represent the period of peak egg laying in P. aegeria, usually followed by a rapid increase in mortality of both eggs and females (Gibbs, Breuker, & Van Dyck, 2010). Female age throughout the experiments was recorded as it affects willingness to mate, and reproductive output (Gibbs, Breuker, Hesketh, Hails, & Van Dyck, 2010; Gibbs, Breuker, & Van Dyck, 2010). After 6 days females were removed and used for wing measurements. From the collected eggs, the first eight larvae to hatch from a particular cross were each reared through on a mix of P. trivialis, D. glomerata, Brachypodium sylvaticum and Festuca rubra. The hatching success of the remaining eggs was noted and the remaining larvae sacrificed. Larvae placed on food plants were monitored to eclosion and the proportion of individuals surviving to adulthood and the sex ratio of the adults was recorded. Pupae were sexed and kept individually, to ensure that virgin adults were available to set-up mating pairs in backcrosses.

After the individuals used in the crosses had been sacrificed their forewings were removed and the dorsal side of the forewing was placed between glass slides and photographed using a Leica MZ6 dissection microscope with integrated camera (Leica IC80 HD camera with Las EZ software suite) under controlled light conditions. Wing area (mm²) of both forewings was measured using ImageJ software (Abramoff, Magalhaes, & Ram, 2004; Breuker, Gibbs, Merckx, Van Dongen, & Van Dyck, 2010), and the average forewing area was used as a proximate measure of individuals’ size (cf. Merckx & Van Dyck, 2006), and included as a covariate in the models.

2.5 | Backcrosses

The offspring resulting from the aforementioned crosses (both hybrids and pure-bred Sardinian and Corsican individuals; F1), were crossed amongst each other (see below; i.e. backcrossed) to generate an F2 (see also Longdon et al., 2017) under similar conditions. For the backcrosses hatching, success of a sample was assessed only for a minimum of 10 and a maximum of 20 eggs, as this was considered a representative sample size. The aim of the backcrosses was to test for fertility issues of the hybrids versus pure-breds. Thus, only those crosses for which a successful mating was observed were included in the analyses, and no behavioural data were collected. A hybrid male or female, was backcrossed to either a purebred Sardinian or Corsican specimen (Appendix S2, Table S2.2; a total of 54 successfully mated backcrosses were obtained). After the male had been removed females were provided with an egg laying plant and allowed to oviposit for 6 days (see also original crosses).

2.6 | Wolbachia presence

The wild-collected Sardinian and Corsican females whose offspring were used as parents in the crosses (with the exception of females 3 and 14; Appendix S2) were screened for the presence of Wolbachia, as this has been shown to sometimes affect reproductive output and fertility in insects, and the presence of this endosymbiont has been reported in P. aegeria ovaries (reviewed in Carter et al., 2013). To screen for Wolbachia, we PCR amplified Wolbachia specific sequences (Wolbachia surface protein—wsp) using previously described primers (Dobson et al., 1999). The PCR products were run on a gel and screened for the presence of amplification. The gene caudal was used as a positive control and absence of Wolbachia had been verified for a number of samples in a separate study using RNA sequencing (Quah, Breuker, & Holland, 2015). Individuals used in the crosses presented in this study were not tested for Wolbachia prior to mating, as that was not feasible given the design of the experiments, nor postmating. Whether or not Wolbachia infection was detected in the mothers of the animals used to establish the crosses was used as a fixed factor in the models described below.

2.7 | Statistical analyses of the courtship and mating experiments

Linear mixed effect models (fitted by maximum likelihood t tests use Satterthwaite approximations to degrees of freedom) were
constructed to investigate variabiliy amongst the crosses in reproductive output (both number of laid eggs and egg hatching success), larval survival and courtship duration. The latter can be considered largely the net result of the choosiness of the female, and the willingnessness of the male (e.g. Holveck, Gauthier, & Nieberding, 2015). Fixed effects tested for inclusion were age of both male and female at the time of mating, wing size (measured as wing size), Wolbachia infection and type of cross. Both male and female maternal origin were included as random factors. Model selection has been carried out based on Minimum models were constructed using AIC value as a guideline, and these are the models presented in this study. This means that non-significant fixed covaries and interactions were removed. Once model selection had been completed, significance of the remaining fixed effects was provided through use of the lmerTest package (Kuznetsova, Bruun Brockhoff, & Christensen, 2016) providing. All residuals for included effects were tested for normality and log and square root transformations were used where appropriate (e.g. courtship duration). Both male and female maternal origin were kept as random factors in all the models, and as the models tested the significance of differences between the various cross types, cross type was always included as a fixed effect. Analyses were performed using R (3.4.0) (R Development Core Team 2016) with packages ‘lme4’ (Bates, Maechler, Bolker, & Walker, 2015) ‘lmerTest’ (Kuznetsova et al., 2016). Chi-square tests were used to test for cross type and fertility associations; while for the backcrosses, the cross type was always included as a fixed effect. Analyses were performed using R (3.4.0) (R Development Core Team 2016) with packages ‘lme4’ (Bates, Maechler, Bolker, & Walker, 2015) ‘lmerTest’ (Kuznetsova et al., 2016). Chi-square tests were used to test for cross type and fertility associations; while for the backcrosses, the Fisher’s exact test for count data was used, as some counts were low (Appendix S2).

3 | RESULTS

3.1 | COI variation reveals the presence of two distinct lineages

We obtained 345 sequences with 27 haplotypes characterized by 28 variable nucleotide sites for the COI gene. Haplotype networks based on COI sequences show a discrete boundary between North African and European populations, forming two distinct lineages separated by a minimum of seven mutations (1.1%), with a single divergent specimen from Cyprus in evidence (Figure 1, Appendix S4). North African haplotypes show significant population structure, with several highly frequent haplotypes occurring throughout the areas analysed. In contrast, populations in continental Europe are characterized by one main haplotype, separated from several low frequency ones by a maximum of two mutations (Figure 1). The two haplogroups are also supported in our phylogenetic analyses (Appendix S4). Interestingly, the islands of Sardinia, Mallorca, Menorca and Ponza are all populated exclusively by North African haplotypes, even though they are in closer proximity to continental Europe (Figures 1 and 2a,b). Furthermore, we found evidence of only one individual carrying the Sardinian haplotype in Corsica (Bonifacio), suggesting very limited gene flow (of matrilines) between the two islands.

When splitting the populations based on the COI lineages, we observed a marginally significant negative Tajima's D for the European lineage (Tajima's D = -2.10, p = .05), signifying expansion and/or recent selective sweep, but not for the North African one (Tajima's D = -0.91, p > .10). Overall genetic diversity was also higher for the North African lineage compared to the European populations (average nucleotide diversity, π was 0.0025 and 0.0013 respectively). This was also evident when the genetic differences among the nearest four specimens to each 0.2 x 0.2 square of latitude and longitude is plotted on a map (Appendix S1; Figure 3). Average genetic divergence between the lineages is 0.30% for wg and 0.50% for zen (based on mutations in aligned sequences). Geographical locations corresponding to the North African lineage were shown to harbour more genetic heterogeneity. Interestingly, the populations in Romania and Alps are also more variable, suggesting increased genetic diversity for the European clade in central and Eastern Europe.

3.2 | Nuclear genes versus COI lineages

Sequence variation in the developmental genes wg and zen was generally in agreement, but locally discordant with the mtDNA, since the pattern of genetic clustering showed a south-western genotype mainly distributed across north Africa, Iberia, southern France, Sardinia and Sicily and a north-eastern genotype in the Italian Peninsula, north Europe and Eastern Europe (Figure 2b,c, Appendices S5 and S6). The Iberian Peninsula and Sicily were inhabited solely by COI haplotypes belonging to the European lineage, while nuclear sequences also belong to the south-western lineage (Figure 2b,c).

3.3 | Courtship behaviour

Females in pure-bred Corsican crosses were significantly slower in mating than any of the other crosses (full minimum mixed model AIC = 142.1, BIC = 156.6, df resid = 52). Not only did they take longer to mate compared to Sardinian females in pure-bred Sardinian crosses (CC versus SS; t = -2.80, df = 59, p = .0068), but Sardinian females also mated more readily with Corsican males than Corsican females did (CC versus CS t = -3.61, df = 59, p < .001). Sardinian males also mated more readily with Corsican females than Corsican males did (CC versus SC t = -2.18, df = 59, p = .033). Female age and size, male size or temperature did not improve the model.

3.4 | Reproductive barriers

3.4.1 | Female fecundity

Reproductive output (i.e. number of eggs laid) was significantly affected by female age and size, as well as cross type (AIC = 605.8, BIC = 626.3, df resid = 63). Females that were older at the time of mating laid more eggs in the 6 days following mating than those that mated young, having presumably stored mature eggs for fertilization (t = 3.25, df = 71.90, p = .0018). Larger females laid significantly more eggs (t = 2.88, df = 67.48, p = .0053). Sardinian females (i.e.
SS \[9.57 ± 4.83\] and CS \[16.69 ± 4.25\]) laid significantly less eggs than Corsican females (i.e. CC \[36.72 ± 3.96\] and SC \[24.45 ± 3.55\]), regardless whom they mated with (SS versus CC \(t = -3.31, df = 20.53, p = .0034\); CS versus CC \(t = -3.87, df = 15.23, p = .0015\)). There was no significant difference between CC and SC (\(t = -1.75, df = 71.82, p = .085\)).

### 3.4.2 Offspring fitness and the effect of temperature on egg hatching success

All four types of crosses were similar in terms of infertile (i.e. egg hatching success \(= 0\%), or no eggs laid, despite having been observed to mate successfully) versus fertile (i.e. egg hatching success \(>0\%\)) crosses per se (\(\chi^2 = 1.58, df = 3, p = .66\)). Egg hatching success was very high, with no significant differences in hatching success between the different cross types (CC 94.38 ± 2.62%, CS 92.53 ± 2.83%, SC 92.85 ± 4.23%, and SS 99.1 ± 0.59%). Hatching success was only affected by temperature, but not female age at mating or female size (AIC = 506.6, BIC = 523.9, df resid = 57). Within the temperature range used (range: 22.1–25.4°C), more eggs hatched successfully at higher temperatures (\(t = 2.43, df = 60.82, p = .018\)).

There were no significant differences (i.e. \(p > .05\)) in survival of the offspring (i.e. from larval hatching to eclosion as an adult) between the crosses (full model with only cross type AIC = 32.7, BIC = 48.0, df resid = 58).

### 3.4.3 Wolbachia infection status

The majority of the field-collected females tested for Wolbachia were found to be infected, with the exception of five females: three from Aritzo (Sardinia), one from Desulo (Sardinia), and one from Bonifacio (Corsica). However, Aritzo is not a location free from Wolbachia, as other females collected there were infected (Appendix S3). We cannot rule out Wolbachia presence in populations from Desulo and Bonifacio as only a single specimen was collected in each of these localities. The Wolbachia infection status of the mothers of the specimens used to establish the crosses was not a factor that significantly improved the statistical models reported earlier, and therefore not included in the reported final models. Finally, for each of the four cross types chi-squared tests were used to evaluate the presence of sex ratio distortion in the surviving offspring. No significant sex ratio distortion was found in any of cross types: CC (\(\chi^2 = 0.12, df = 1, p = .73\)), CS (\(\chi^2 = 0.017, df = 1, p = .90\)), SC (\(\chi^2 = 0.059, df = 1, p = .81\)) or SS (\(\chi^2 = 0.76, df = 1, p = .78\)). The lack of sex ratio distortion and the absence of fertility problems suggests that cytoplasmic incompatibility does not explain the lack of gene flow between Sardinia and Corsica.

**FIGURE 1** Haplotype network based on COI sequences of *Pararge aegeria* from the study area. Each colour indicates the geographic location of the haplotypes, as indicated in the legend, and the size of the circle corresponds to the frequency of a haplotype. The number of nucleotide changes at each node is shown as white circles (putative ancestral haplotypes).
3.4.4 | Sterility of F₁ hybrids

F₁ hybrids were backcrossed to either pure-bred Sardinians or Corsicans (Appendix S2). There were no differences between the 10 types of crosses in terms of fertility (Fisher’s exact test for count data, \( p = .13 \)).

5 | DISCUSSION

Corsica and Sardinia are characterized by the occurrence of a variety of endemic populations for various butterfly species (Aubert, Barascud, Descimon, & Michel, 1997; Dapporto, 2010; Grill, Crnjar, Casula, & Menken, 2002). This is likely to be the result of the long-term isolation of these islands since the early or late Miocene (Ketmaier, Giusti, & Caccone, 2006). Mutually exclusive pairs of cryptic butterfly species such as Aricia agestis and A. cramera or Polyommatinus icarus and P. celina have been shown to occur on Corsica and Sardinia (Dinca, Dapporto, & Vila, 2011; Vodà et al., 2015a,b). Such divergence between Corsican and Sardinian populations is in many ways unexpected as the islands are separated by a narrow sea strait (approximately 12 km wide, while the shortest distance between Corsica and Sardinia, represented by the small islands in between is about 6 km), and were connected during the last glaciation period (Dapporto, 2010 and references therein). Similarly, in Sweden, P. aegeria revealed little to no gene flow between the populations of the island Öland and the near-by mainland (separated by 5 km) (Tison et al., 2014). Although the data presented in this study confirm that even short sea straits can provide a strong barrier to the dispersal of P. aegeria, we observed some markedly discordant patterns between the nuclear and mitochondrial genes. For instance, the Iberian Peninsula is inhabited solely by COI haplotypes belonging

**FIGURE 2** A principal coordinates analysis projection of the p-distances genetic variation in COI, among the Pararge aegeria specimens (dots), in the bidimensional red, green, blue (RGB) space (a), spatial distribution of genetic variants of COI (b), RGB PCoA projection of p-distances genetic variation in concatenated nuclear dataset (c), and spatial distribution of nuclear genes (d)
to the European lineage, but the nuclear markers at the same locations clustered together with North Africa and Sicily. This pattern is reinforced by the geometric morphometric split observed for male genitalia shape between populations of *P. aegeria* where the same east–west differentiation pattern is observed (Dapporto et al., 2012). The conservative nature of nuclear markers (Wahlberg & Wheat, 2008) was most notably exemplified between Corsica and Sardinia, given the similarity in nuclear sequences despite the occurrence of different lineages.

The presence of the North African COI lineage on several Mediterranean islands is intriguing (Dapporto et al., 2017; Voda et al., 2015b, 2016), as they are in closer proximity to the European mainland and in this region wind generally blows from west-northwest (Dapporto et al., 2012). Thus, one would expect them to be more easily colonized from either the Italian Peninsula (in the case of Ponza and Sardinia) or the Iberian Peninsula (in the case of Mallorca and Menorca). The higher genetic heterogeneity observed in the Maghreb lineage (Figure 2), suggests the presence of ancestral populations not only in North Africa, as suggested by Weingartner et al. (2006), but also in other Mediterranean islands. This is in stark contrast to the reduced genetic variation observed in the European clade in the circum-mediterranean populations, suggestive of a recent colonization and population expansion from Eastern continental areas. The significant negative Tajima’s D for European populations also supports this hypothesis, because low frequency variants segregating at high frequencies can indicate population expansion by founder effect and gene surfing (Waters, 2011). Given the higher genetic variation found in the Alps and Romania (Figure 3) one could propose a putative centre of origin for the European populations further east, and then, as found in other Lepidopteran species (Mende & Hundsdoerfer, 2013), the contact zone among genetic variants has likely shifted to the west (Figure 4). This could have occurred as a phalanx-like colonization over the mainland, which was impeded at sea straits, resulting in the island lineages being unexpectedly similar to the North African populations (Dapporto et al., 2012). The populations in Sardinia, Mallorca, Menorca and Ponza might thus represent “relict” populations harbouring the ancestral COI haplotypes, which have not been replaced due to the physical barriers imposed by the sea straits.

However, it must be noted that COI is maternally inherited and it can only trace the dynamics of females. Nuclear genes show a general correspondence into two main southern and northern groups.

**Figure 3** Distribution of the genetic richness of *Pararge aegeria* in the study area based on 0.25° × 0.25° squares for which at least four specimens were sequenced in a 100 km radius. Genetic richness was calculated separately for the two lineages identified in this study for each of these squares. The method involves calculating matrices of p-distances (proportions of nucleotide differences), taking geographic distances into account. At the end, a single value, indicating the genetic differentiation of four specimens closest to each other weighted for their distance from the centre of their locations, is then plotted onto a map. This has been represented here as a heat map of sequence variation across a wide geographical range (full range 0% [green] to 1.6% [red]; values indicated in figure) (for full details on the genetic richness method see Supplementary File 1).
but also areas of discrepancy where the northern COI lineage is associated to southern wg/zen genes. Our data suggest that hybrid sterility and hybrid-purebred incompatibilities do not limit introgression between these islands, and there appear to be no obvious pre- or postzygotic barriers. Moreover, we observed that the two COI lineages are highly inter-fertile and also that there are temperature-related differences across types in both female fecundity and offspring fitness during the egg stage, indicating possible effects of local adaptation to temperature during oviposition and embryogenesis. Other Speckled wood populations across Europe show significant and distinct population structuring, evidenced by sequence analyses of the *P. aegeria* specific Rhabdovirus PAegRV (Longdon et al., 2017) and population genetic analyses (Tison et al., 2014). For the UK in particular, this is remarkable, given the relatively recent contraction and subsequent expansion of *P. aegeria* in the UK (Hill et al., 1999; Longdon et al., 2017). A nuclear gene such a zen evolves relatively slowly, not least as it has an important developmental role in the specification and functioning of the serosa, an extra-embryonic tissue involved in drought resistance (Ferguson et al., 2014), and has been shown to be under negative selection in *P. aegeria* (Livraghi et al., unpubl). Viral genes evolve much faster, showing a higher propensity to population structuring (of their hosts; see Longdon et al., 2015). The differences in the spatial patterning of nuclear and COI as well as PAegRV variation might thus reflect complex patterns of past and current selection, past isolation and recolonization events, in theory including sex-biased dispersal (Toews & Brelsford, 2012).

Although dispersal may be more costly to female *P. aegeria*, often lowering reproductive output (Hughes, Hill, & Dytham, 2003), females have been shown to be the most dispersive sex in typical metapopulation dynamics in, for example, the UK and Belgium (Bergéro, Merckx, Van Dyck, & Baguette, 2012; Hughes et al., 2003). Male-biased dispersal would not satisfactorily explain the Sardinia-Corsica results since PAegRV is transmitted to offspring by both males and females. Thus, in the case of male-biased dispersal, genetic variation observed for PAegRV genetic variation would reflect nuclear variation; instead it reflects the observed COI pattern, arguing against male-biased dispersal. Consequently, the similarity between the islands in terms of variation in slowly evolving nuclear genes between the islands is likely to be historical, rather than the result of an ongoing process of male-specific dispersal.

At present, we do not know enough about the differences between populations across the whole of the geographical range of *P. aegeria* in terms of the selection pressures operating on dispersal.

**FIGURE 4** Proposed hypothesis for the historical biogeography of Pararge aegeria. The ancestral lineage (blue circles) was present throughout the range of *P. aegeria* in Europe (a), without substantial differentiation of the nDNA markers due to unrestricted dispersal between populations. During the last glacial period (possibly also including previous series of glacial events) (b) the range retracted southwards (red arrows), and gene flow was restricted between the refugia due to the Alps and Pyrenees acting as barriers, which allowed for periods of differentiation (yellow circles in c). Following the warming of the climate, the eastern lineage spread northwards and westwards (red arrows in d), where it could have introgressed with the nuclear genome of warm adapted populations in the Iberian peninsula as well as the islands of Ibiza, Corsica and Sardinia resulting in the discordance between the markers (indicated by blue and yellow circles in d). This introgression was presumably hindered by sea straits, giving rise to the sharp boundary observed for the COI data.
propensity, reproductive strategies and the trade-offs made between reproduction and dispersal. Strong differences between *P. aegeria* populations are not only evident on the basis of sequence variation, but also in terms of expression patterns of specific miRNA genes (Quah et al., 2015). This has been shown for egg production in Corsican (specifically Zonza) and Belgian populations (Quah et al., 2015). This leads one to hypothesize that female reproductive strategies, and the genes involved therein, are very likely to be under selection in response to habitat variation (e.g. temperature and oviposition plants) with significant population differences, as observed in other *P. aegeria* populations across Europe (Gibbs & Van Dyck, 2009; Gibbs, Breuker, & Van Dyck, 2010). Such differences may possibly also exist in our study area since Sardinian and Corsican females significantly differed in reproductive output.

*Pararge aegeria* is confirmed as a highly suitable model to study the distribution of genetic lineages and their spatial segregation to reveal broad biogeographical patterns associated with responses to both biotic and abiotic factors and to the evolution of different lineages. Open questions to pursue are whether the historical polymorphisms of nuclear genes are: actively maintained by selection in the areas of discordance, simply the result of different evolutionary rates of nuclear genes versus COI per se (i.e. neutral variation; when genes likely to be under different selection pressures show similar patterns) and/or whether sex-biased dispersal underpins observed patterns of discordance between nuclear genes and COI. The wider availability of other molecular techniques such as RAD-seq and genome-wide association study (GWAS) for non-model organisms now provides the opportunity for more in-depth analyses of population genetics and the adaptive nature of particular SNPs across different selective environments. Studies on gene flow and local adaptation in a life-history context are now more pertinent than ever given that most species are facing rapid environmental changes (e.g. climate and land use), and our data suggest that *P. aegeria* would be an excellent model for these kinds of studies.

CONFLICT OF INTEREST

The authors declare no conflict of interests.

ACKNOWLEDGEMENTS

Support for this research was provided by a Nigel Groome PhD studentship awarded to Breuker for Livraghi, a Santander Student Research Grant to Livraghi, a Santander Research Scholarship to Breuker, a Marie Curie International Outgoing Fellowship within the 7th European Community Framework Programme to Dincă (project no. 625997), an European Union’s Seventh Framework programme for research and innovation under the Marie Curie grant agreement No 609402—2020 researchers: Train to Move (T2M) to Vodă, the projects “Barcoding Italian Butterflies” and “Barcoding Butterflies of the Tuscan Archipelago National Park”, and the Spanish Plan Nacional I+D+I CGL2016-76322-P (AEI/FEDER, UE).

ORCID

Casper J. Breuker http://orcid.org/0000-0001-7909-1950

REFERENCES


BIOSKETCH

Members of the research team are actively engaged in studying: (1) life history evolution and maternal effects in response to environmental variation, aiming to synthesize life history models with developmental genetic models of evolution, and (2) insect biogeography, systematics and conservation, with a specific interest in unravelling the historical and present-day factors responsible for species distributions across mainland Europe and Mediterranean islands.

DATA AVAILABILITY

Sequence data are publicly available via GenBank (MH090747–MH090823; dedicated databases are publicly available for COI and wg sequences through the Barcode of Life Data (BOLD) system (dx.doi.org/10.5838/DS-PARARGE), and for zen sequences through a P. aegeria hox3 sequence database (https://doi.org/10.24384/000476).

SUPPORTING INFORMATION

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