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

Short running title: phylogeography of *Pararge aegeria*

Luca Livraghi[1†], Raluca Vodă[2†], Luke Christopher Evans[1,3], Melanie Gibbs[4], Vlad Dincă[5,6], Peter W.H. Holland[7], Tim G. Shreeve[8], Roger Vila[6], Leonardo Dapporto[9] and Casper J. Breuker[1*]

[1] 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

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

[3] Ecology and Evolutionary Biology Research Division, School of Biological Sciences, Harborne Building, University of Reading, Whiteknights, Reading, Berkshire, RG6 6AS, UK

[4] NERC Centre for Ecology & Hydrology, Maclean Building, Crowmarsh Gifford, Wallingford, OX10 8BB, UK

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

[6] Institute of Evolutionary Biology, Passeig Marítim de la Barceloneta 37, 08003 Barcelona, Spain

[7] Department of Zoology, University of Oxford, Oxford OX1 3PS, UK

[8] Centre for Ecology, Environment and Conservation, Oxford Brookes University, Gipsy Lane, Headington, Oxford, OX3 0BP, UK

[9] Dipartimento di Biologia, Università degli Studi di Firenze, Via madonna del piano 6, 50019 Sesto Fiorentino, Italy

† Authors contributed equally

* Correspondence

E-mail addresses (for submission purposes)

Luca Livraghi (luca.livraghi-2013@brookes.ac.uk), Raluca Vodă (raluvoda@gmail.com), Luke Christopher Evans (L.C.Evans@pgr.reading.ac.uk), Melanie Gibbs (mela1@ceh.ac.uk), Vlad Dincă (vlad.dinca@oulu.fi), Peter Holland (peter.holland@zoo.ox.ac.uk), Tim G. Shreeve, (tgshreeve@brookes.ac.uk), Roger Vila (roger.vila@csic.es), Leonardo Dapporto (leondap@gmail.com) and Casper J. Breuker (cbreuker@brookes.ac.uk)*

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).

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 *zerknullt*) 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.

Main conclusions

We propose that following the post-glacial recolonisation of Europe, the ancestral *COI* lineage of *P. aegeria* was maintained in North Africa and Mediterranean islands, while a new lineage colonised from Eastern Europe, replacing and apparently outcompeting 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.

Editor Simone Fattorini

Keywords: *Pararge aegeria*, Speckled Wood butterfly, phylogeography, barcoding, pre- and postzygotic barriers, *wingless*, *zerknullt*, life-history variation, selection, and gene flow

Introduction

Climatic fluctuations during the Pleistocene period in Europe had a tremendous impact on the emergence of different lineages for many temperate species (Cooper *et al.,* 1995; Taberlet *et al.,* 1998; Seddon *et al.,* 2001). During cold periods most European species were presumably restricted to Mediterranean areas. Due to the geographic configuration of the Mediterranean region, a series of areas separated by mountain chains and sea channels have been hypothesised to function as differentiation centres for many organisms (e.g. Hewitt, 1999; Hewitt, 2000; Schmitt, 2007). In Europe, such areas have been typically identified in the Iberian and Italian Peninsulas and the Balkans, which were isolated from each other to various degrees during the long cold periods that characterized the Pleistocene. The large Mediterranean islands, Maghreb and Asia Minor represented further important refugia and centres of differentiation for species living in the Mediterranean area (Habel *et al.,* 2009; Habel *et al.,* 2011; Dapporto *et al.,* 2011, 2012).

Following isolation, populations of many species in glacial refugia evolved into distinct lineages and (sub-)species (Ribera and Volger, 2004). During the relatively short warm periods, thermophilic species that were constrained to these areas began northwards expansions and recolonised previously glaciated regions. It has been inferred on a number of occasions that although lineages and sister species can (post-glacially) expand over thousands of kilometres in Europe, they tend to form only very limited areas of overlap or they establish contact zones along even narrow sea straits when they meet in re-colonized areas (Waters, 2011 for a review, Dapporto *et al*., 2011, 2012, 2017; Vodă *et al*., 2015a,b; Habel *et al.*, 2017 for Mediterranean butterflies). Several explanations for this have been proposed including density dependent phenomena, climatic and environmental preferences, reproductive interference, dispersal limitation and/or competitive exclusion (Waters, 2011; Vodă *et al.,* 2015b; 2016). Due to the large number of potential mechanisms that can determine patterns of mutual exclusion,

114 understanding the processes responsible for the observed distributions requires a
115 multidisciplinary approach (Vodă *et al*., 2015b for Mediterranean butterflies). Studying the
116 spatial distribution of highly diverging genetic lineages and their tendency to form extended
117 parapatric areas of contact has fundamental implications in understanding the onset of the
118 speciation process (e.g. Arias *et al.,* 2008, Habel *et al.,* 2017 for butterflies in particular).

119 The Speckled Wood butterfly, *Pararge aegeria* (Linnaeus, 1758) has been widely used as a
120 model system to study the distribution of genetic lineages and their spatial segregation; it is
121 an ubiquitous species with a widespread distribution (ranging from the Maghreb, throughout
122 Europe and reaching western Asia), experiencing various environmental settings from cold
123 and wet conditions in northern Europe to hot and dry conditions in southern Europe and north
124 Africa (Weingartner *et al.,* 2006; Habel *et al.,* 2013; Tison *et al.,* 2014). Moreover, this
125 species occurs in many Mediterranean islands and the Atlantic island of Madeira, thus also
126 allowing the study of dispersal both on mainland and across sea straits (Dapporto *et al*.,
127 2017). These attributes mean the species is highly suitable for studying the distribution of
128 genetic lineages and their spatial segregation and giving insight into broad biogeographical
129 patterns associated with responses to both biotic and abiotic factors and into the evolution of
130 different lineages. The model species also has potential to provide valuable insights on how
131 species react in time and space to environmental pressures across large geographic ranges
132 (Parmesan, 1999; Oliver *et al.,* 2015).

133 Using variation in the mitochondrial cytochrome *c* oxidase subunit I (*COI*) gene, the nuclear
134 *wingless* gene and in microsatellites, two main lineages of *P. aegeria* have been identified
135 (Weingartner *et al.,* 2006; Habel *et al.*, 2013; Dapporto *et al*., 2017), in agreement with the
136 subdivision of this species into two subspecies (ssp. *aegeria* and ssp. *tircis*). The *aegeria*
137 lineage occurs in Maghreb, the Balearic Islands and in Sardinia, and the *tircis* lineage in
138 mainland Europe and Asia. Accordingly, the two lineages are separated by three sea channels

– the Gibraltar strait between Morocco and Spain, the strait of Sicily between Sicily and Tunisia, and the strait of Bonifacio between Sardinia and Corsica (Vodă *et al.,* 2016; Dapporto *et al.*, 2017). The differentiation between Corsican and Sardinian populations of *P. aegeria* is also evident at the morphological level, with a divergence in male genital shape between the two lineages (Dapporto *et al.,* 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 (Wilfert & Jiggins, 2014; Longdon *et al.,* 2017), 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 less than 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 *et al.,* 2005). It is highly unlikely that climatic differences alone prevent the European lineage from establishing populations on Sardinia and *vice versa*, but local adaptation may reduce the likelihood of colonization (cf. Richter-Boix *et al.,* 2013). Climatic factors and their effects on host plants have indeed imposed strong selection pressures in *P. aegeria* that influence egg-laying

strategies (Hill *et al.,* 1999; Gibbs & Van Dyck, 2010; Gibbs *et al.,* 2012). 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 *et al.,* 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 and 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.

Material and Methods

188 **DNA extraction, amplification, sequencing, and phylogenetic analyses**

189 Using PCR and direct Sanger sequencing (see Appendix S1 in Supporting Information,
190 including details on primers and cycling conditions), we obtained sequences of; the 658 bp
191 barcoding region of *COI* for 345 individuals spanning from North Africa to northern Europe,
192 of a 411 bp region of the gene *wingless* (*wg*) for a subset of 87 individuals, and of the entire
193 coding region of *zen* (1599 bp) for 79 individuals. We further obtained two outgroup
194 sequences of *wg* and *zen* belonging to the closely related species *Pararge xiphioides*
195 (Staudinger, 1871) (cf. Weingartner *et al.,* 2006). Bayesian inference (BI) for the three genes
196 was employed to infer phylogenetic relationships with BEAST 1.8.0 (Drummond *et al.,*
197 2012). Patterns of genetic variation were inferred by maximum parsimony haplotype
198 networks using the program TCS 1.21, with a 95% connection limit (Clement *et al.,* 2000).
199 Representations of genetic diversity were created for the three genetic markers by calculating
200 matrices of p-distances for each of them, and subsequently analysing and plotting these using
201 R and QGIS 2.0.1. (QGIS Development Team 2009). Further details on sequencing and the
202 phylogenetic, haplotype network and genetic distance analyses can be found in Appendix S1.

203

204 **Sardinian and Corsican samples**

205 Between the 6th and the 12th of May 2014 *P. aegeria* females were collected in the field from
206 11 different localities in: Sardinia (Aritzo, Desulo and Tempio Pausania), Corsica (Asco,
207 Zonza, Bavella, Bonifacio, Solenzara, Cavallo Morto and Pietralba) and La Maddalena
208 (Sualeddu), a smaller island off the north coast of Sardinia. In total 32 females were caught:
209 18 from Corsica, 13 from Sardinia and 1 from La Maddalena. Eggs from collected females
210 were obtained *in-situ* and brought to the laboratory in Oxford, UK. Upon hatching these eggs
211 were put on host plants for this species in Europe (a mix of *Poa trivialis* and *Dactylis*

212 *glomerata)* and reared at 21 ±2°C (60% RH, 16L:8D) (cf. Gibbs *et al.,* 2010b). Rearing and
213 mating conditions in this study included full daylight spectrum lamps, including UV-light
214 (Osram Biolux). The females collected in the field laid readily on these plant species, as did
215 all the females used in this experiment. Pupae were sexed and kept individually, to ensure
216 virgin adults were available for setting up crosses. The offspring of a total of 22 of the field-
217 collected females provided the adults to perform the crosses detailed below (Appendix S2 in
218 Supporting Information for details on collection).

219

220 **Pre-zygotic reproductive barriers: courtship behaviour in Sardinian and Corsican**
221 ***Pararge aegeria***

222 Crosses were performed with the offspring of the wild-caught females. Four types of crosses
223 were set up: Corsican male/Corsican female (CC), Corsican male/Sardinian female (CS),
224 Sardinian male/Sardinian female (SS) and Sardinian male/Corsican female (SC). In total 72
225 crosses generated data to be used in the analyses (CC=27, CS=20, SC=11, SS=14). To
226 perform the crosses, newly eclosed virgin females were placed in cages along with an
227 artificial flower containing 10% honey solution (Gibbs *et al.,* 2012). A newly eclosed virgin
228 male was then introduced into a female's cage and the total courtship duration (seconds) was
229 timed using stopwatches. The primary aim of these crosses was to establish how willing
230 males and females of the two islands were to mate with each other, and having done so, what
231 the reproductive output was. Thus, no mate choice experiments *per se* were conducted (i.e.
232 where by a female needed to choose between a male from her own island or the other one).
233 *Pararge aegeria* have a courtship very similar to that described for the grayling where
234 courtship is initiated by a wing flick used by males, to the front and side of the female
235 (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 8am and 6pm, it was removed and replaced with a new virgin male the following morning (8am). After mating had finished an egg laying plant was placed in the cage and the male was removed.

Reproductive barriers

After mating the female was left to oviposit for six 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 *et al.,* 2010b). Female age throughout the experiments was recorded as it affects willingness to mate, and reproductive output (Gibbs *et al.,* 2010a,b). After six 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 sylavticum* 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 (mm^2) of both forewings was measured using ImageJ software (Abramoff *et al.,* 2004; Breuker *et al.,* 2010), and the

260 average forewing area was used as a proximate measure of individuals' size (cf. Merckx &
261 Van Dyck, 2006), and included as a covariate in the models.

262

263 **Backcrosses**

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

276

277 ***Wolbachia* presence**

278 The wild-collected Sardinian and Corsican females whose offspring were used as parents in
279 the crosses (with the exception of females 3 and 14; Appendix S2) were screened for the
280 presence of *Wolbachia*, as this has been shown to sometimes affect reproductive output and
281 fertility in insects, and the presence of this endosymbiont has been reported in *P. aegeria*
282 ovaries (reviewed in Carter *et al.,* 2013). In order to screen for *Wolbachia*, we PCR amplified
283 *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 et al 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.

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 variability 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 willingness and effectiveness of the male (e.g. Holveck *et al*., 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 Akaike Information Criterion (AIC) value as a guideline, and these are the models presented in this study. This means that non-significant fixed covariates 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 *et al.,* 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 *et al*., 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).

Results

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 characterised 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 (Figure 1, 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=0.05), signifying expansion and/or recent selective sweep, but not for the North African one (Tajima's D = -0.91, p> 0.10). 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 0.2x0.2 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.

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, Appendix 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).

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=0.0068), but Sardinian females also mated more readily with Corsican males, than Corsican females did (CC versus CS t=-3.61, df=59, p<<0.001). Sardinian males also mated more readily with Corsican females than Corsican males did (CC versus SC t=-2.18, df=59, p=0.033). Female age and size, male size or temperature did not improve the model.

Reproductive barriers

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 six days following mating than those that mated young, having presumably stored mature eggs for fertilisation (t=3.25, df=71.90, p=0.0018). Larger females laid significantly more eggs (t=2.88, df=67.48, p=0.0053). 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=0.0034; CS versus CC t=-3.87, df=15.23, p=0.0015). There was no significant difference between CC and SC (t=-1.75, df=71.82, p=0.085).

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-square 1.58, df=3, and p=0.66). 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=0.018).

There were no significant differences (i.e. P>>0.05) 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).

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-square=0.12, df=1, p=0.73), CS (chi-square = 0.017, df=1, p=0.90), SC (chi-square = 0.059, df=1, p=0.81) or SS (chi-square = 0.76, df=1, p=0.78). 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.

Sterility of F_1 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=0.13).

Discussion

Corsica and Sardinia are characterised by the occurrence of a variety of endemic populations for various butterfly species (Aubert *et al.*, 1997, Grill *et al.,* 2002; Dapporto, 2010). This is likely to be the result of the long-term isolation of these islands since the early or late Miocene (Ketmaier *et al,.* 2006). Mutually exclusive pairs of cryptic butterfly species such as *Aricia agestis* and *A. cramera* or *Polyommatus icarus* and *P. celina* have been shown to occur on Corsica and Sardinia (Dincă *et al.,* 2011, Vodă *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 six 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 five km) (Tison *et al.,* 2014). Even though 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 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.*,

427 2012). The conservative nature of nuclear markers (Wahlberg and Wheat, 2008) was most
428 notably exemplified between Corsica and Sardinia, given the similarity in nuclear sequences
429 despite the occurrence of different lineages.

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

452 However, it must be noted that *COI* is maternally inherited and it can only trace the dynamics
453 of females. Nuclear genes show a general correspondence into two main southern and
454 northern groups but also areas of discrepancy where the northern *COI* lineage is associated to
455 southern *wg*/*zen* genes. Our data suggest that hybrid sterility and hybrid-purebred
456 incompatibilities do not limit introgression between these islands, and there appear to be no
457 obvious pre- or postzygotic barriers. Moreover, we observed that the two *COI* lineages are
458 highly inter-fertile and also that there are temperature-related differences across types in both
459 female fecundity and offspring fitness during the egg stage, indicating possible effects of
460 local adaptation to temperature during oviposition and embryogenesis. Other Speckled wood
461 populations across Europe show significant and distinct population structuring, evidenced by
462 sequence analyses of the *P. aegeria* specific Rhabdovirus PAegRV (Longdon *et al.,* 2017)
463 and population genetic analyses (Tison *et al.,* 2014). For the UK in particular, this is
464 remarkable, given the relatively recent contraction and subsequent expansion of *P. aegeria* in
465 the UK (Hill *et al.,* 1999; Longdon *et al.,* 2017). A nuclear gene such a *zen* evolves relatively
466 slowly, not least as it has an important developmental role in the specification and
467 functioning of the serosa, an extra-embryonic tissue involved in drought resistance (Ferguson
468 *et al*., 2014), and has been shown to be under negative selection in *P. aegeria* (Livraghi *et al*.,
469 unpubl). Viral genes evolve much faster, showing a higher propensity to population
470 structuring (of their hosts; see Longdon *et al*., 2015). The differences in the spatial patterning
471 of nuclear and *COI* as well as PAegRV variation might thus reflect complex patterns of past
472 and current selection, past isolation and recolonisation events, in theory including sex-biased
473 dispersal (Toews & Brelsford, 2012).

474 Although dispersal may be more costly to female *P. aegeria*, often lowering reproductive
475 output (Hughes *et al*., 2003), females have been shown to be the most dispersive sex in
476 typical metapopulation dynamics in for example the UK and Belgium (Hughes *et al*., 2003;

Bergerot *et al*., 2012). 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 of 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 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 hypothesise 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 *et al.,* 2010b). 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 in order 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 suggests that *P. aegeria* would be an excellent model for these kinds of studies.

Conflict of Interest

The authors declare no conflict of interests

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/10.5883/DS-PARARGE), and for *zen* sequences through a *P. aegeria hox3* sequence database (DOI 10.24384/000476).

References

Abramoff, M.D., Magalhaes, P.J. & Ram, S.J. (2004). Image Processing with ImageJ. *Biophotonics International,* **11**, 36-42.

Arias, C.F., Munoz, A.G., Jiggins, C.D., Mavarez, J., Bermingham, E. & Linares, M., (2008). A hybrid zone provides evidence for incipient ecological speciation in *Heliconius* butterflies. *Molecular ecology,* **17**, 4699-4712.

Aubert, J., Barascud, B., Descimon, H., & Michel, F. (1997). Ecology and genetics of interspecific hybridization in the Swallowtails, *Papilio hospiton* Géné and *P. machaon* l., in Corsica (Lepidoptera: Papilionidae). *Biological Journal of the Linnean Society* **60**, 467–492.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effect models using lme4. *Journal of Statistical Software* **67**, 1-48.

Bergerot, B., Merckx, T., Van Dyck, H., & Baguette, M. (2012). Habitat fragmentation impacts mobility in a common and widespread woodland butterfly: do sexes respond differently? *BMC Ecology*, **12**:5.

Breuker, C.J., Gibbs, M., Merckx, T., Van Dongen, S. & Van Dyck, H. (2010). The use of geometric morphometrics in studying butterfly wings in an evolutionary ecological context. In A.M.T. Elewa (ed.), *Morphometrics for non-morphometricians* (DOI 10.1007/978-3-540-95853-6_12). Springer-Verlag, Berlin Heidelberg.

Carter, J.-M., Baker, S.C., Pink, R., Carter, D.R.F., Collins, A., Tomlin, J., Gibbs, M. & Breuker, C.J. (2013). Unscrambling butterfly oogenesis. *BMC Genomics,* **14**, 283.

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

544 Cooper, S.J.B., Ibrahim, K.M. & Hewitt, G.M. (1995). Postglacial expansion and genome
545 subdivision in the European grasshopper *Chorthippus parallelus*. *Molecular Ecology,*
546 **4**, 49-60.
547 Dapporto, L. (2008). Geometric morphometrics reveal male genitalia differences in the
548 *Lasiommata megera/paramegaera* complex (Lepidoptera, Nymphalidae) and the lack
549 of a predicted hybridization area in the Tuscan Archipelago. *Journal of Zoological*
550 *Systematics and Evolutionary Research,* **46**, 224-230.
551 Dapporto, L. (2010). Satyrinae butterflies from Sardinia and Corsica show a kaleidoscopic
552 intraspecific biogeography (Lepidoptera, Nymphlidae). *Biological Journal of the*
553 *Linnean Society,* **100**, 195-212.
554 Dapporto, L., Habel, J.C., Dennis, R.L.H. & Schmitt, T. (2011). The biogeography of the
555 western Mediterranean: elucidating contradictory distribution patterns of
556 differentiation in *Maniola jurtina* (Lepidoptera: Nymphalidae). *Biological Journal of*
557 *the Linnean Society,* **103**, 571-577.
558 Dapporto, L., Bruschini, C., Dincă, V., Vila, R. & Dennis, R.L. (2012). Identifying zones of
559 phenetic compression in West Mediterranean butterflies (Satyrinae): refugia, invasion
560 and hybridization. *Diversity and Distributions,* **18**, 1066-1076.
561 Dapporto, L., Cini, A., Menchetti, M., Vodă, R., Bonelli, S., Casacci, L.P.,…..Vila, R.
562 (2017). Rise and fall of island butterfly diversity: Understanding genetic
563 differentiation and extinction in a highly diverse archipelago. *Diversity and*
564 *Distributions*, **23**, 169-1181.
565 Davies, N.B. (1978). Territorial defence in the speckled wood butterfly (*Pararge aegeria*):
566 The resident always wins. *Animal Behaviour,* **26**, 138-147.

567 Dennis, R.L.H., Shreeve, T.G., Olivier, A. & Coutsis, J.G. (2000). Contemporary geography
568 dominates butterfly diversity gradients within the Aegean archipelago (Lepidoptera :
569 Papilionoidea, Hesperioidea). *Journal of Biogeography,* **27**, 1365-1383.
570 deWaard, J.R., Ivanova, N.V., Hajibabaei, M. & Hebert, P.D.N. (2008). Assembling DNA
571 Barcodes: Analytical Protocols. *Methods in Molecular Biology,* **410**, 275-293.
572 Dincă, V., Dapporto, L. & Vila, R. (2011). A combined genetic-morphometric analysis
573 unravels the complex biogeographic history of *Polyommatus icarus* and *P. celina*
574 Common Blue butterflies. *Molecular Ecology,* **20**, 3921-3935.
575 Dincă, V., Wiklund, C., Lukhtanov, V.A., Kodandaramaiah, U., Norén, K., Dapporto L.,
576 .....Friberg, M. (2013). Reproductive isolation and patterns of genetic differentiation
577 in a cryptic butterfly species complex. *Journal of Evolutionary Biology,* **26,** 2095-
578 2106.
579 Dobson, S.L., Bourtzis, K., Braig, H.R., Jones, B.F., Zhou, W.G., Rousset, F. & O'Neill, S.L.
580 (1999). *Wolbachia* infections are distributed throughout insect somatic and germ line
581 tissues. *Insect Biochemistry and Molecular Biology,* **29**, 153-60.
582 Dobzhansky, T. (1970). *Genetics and the origin of species*. Columbia University Press, New
583 York.
584 Drummond, A. J., Suchard, M. A., Xie, D., & Rambaut, A. (2012). Bayesian phylogenetics
585 with BEAUti and the BEAST 1.7. *Molecular biology and evolution*, **29**, 1969-1973.
586 Ferguson, L., Marletaz, F., Carter, J.-M., Taylor, W.R., Gibbs, M., Breuker, C.J. & Holland,
587 P.W.H. (2014). Ancient expansion of the Hox cluster in Lepidoptera generated four
588 Homeobox genes implicated in extraembryonic tissue formation. *PLoS Genetics,* **10**:
589 e1004698.

590 Gibbs, M., Breuker, C.J., Hesketh, H., Hails, R.S. & Van Dyck, H. (2010a). Maternal effects,
591 flight versus fecundity trade-offs, and offspring immune defence in the Speckled
592 Wood butterfly, *Pararge aegeria*. *BMC Evolutionary Biology,* **10**: 345.
593 Gibbs, M., Breuker, C.J. & Van Dyck, H. (2010b). Flight during oviposition reduces maternal
594 egg provisioning and influences offspring development in *Pararge aegeria* (L.).
595 *Physiological Entomology,* **35**, 29-39.
596 Gibbs, M. & Van Dyck, H. (2009). Reproductive plasticity, oviposition site selection, and
597 maternal effects in fragmented landscapes. *Behavioral Ecology and Sociobiology,* **64**,
598 1-11.
599 Gibbs, M., Van Dyck, H. & Breuker, C.J. (2012). Development on drought-stressed host
600 plants affects life history, flight morphology and reproductive output relative to
601 landscape structure. *Evolutionary Applications,* **5**, 66-75.
602 Grill, A., Crnjar, R., Casula, P. & Menken, S. (2002). Applying the IUCN threat categories to
603 island endemics: Sardinian butterflies (Italy). *Journal for Nature Conservation,* **10**,
604 51-60.
605 Habel, J.C., Dieker, P. & Schmitt T. (2009). Biogeographical connections between the
606 Maghreb and the Mediterranean peninsulas of southern Europe. *Biological Journal of*
607 *the Linnean Society,* **98**, 693-703.
608 Habel, J.C., Husemann, M., Schmitt, T., Dapporto, L., Rodder, D. & Vandewoestijne, S.
609 (2013). A forest butterfly in Sahara desert oases: Isolation does not matter. *Journal of*
610 *Heredity,* **104**, 234-47.
611 Habel, J.C., Lens, L., Rodder, D. & Schmitt, T. (2011). From Africa to Europe and back:
612 refugia and range shifts cause high genetic differentiation in the Marbled White
613 butterfly *Melanargia galathea*. *BMC Evolutionary Biology,* **11**: 215.

614 Habel, J.C., Schmitt, T. & Muller, P. (2005). The fourth paradigm pattern of post-glacial
615 range expansion of European terrestrial species: the phylogeography of the Marbled
616 White butterfly (Satyrinae, Lepidoptera). *Journal of Biogeography,* **32**, 1489-1497.
617 Habel, J.C., Vila, R., Vodă, R., Husemann, M., Schmitt, T. & Dapporto, L. (2017).
618 Differentiation in the marbled white butterfly species complex driven by multiple
619 evolutionary forces. *Journal of Biogeography,* **44**, 433–445.
620 Hewitt, G.M. (1999). Post-glacial re-colonization of European biota. *Biological Journal of*
621 *the Linnean Society,* **68**, 87-112.
622 Hewitt, G.M. (2000). The genetic legacy of the Quaternary ice ages. *Nature,* **405**, 907-913.
623 Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. & Jarvis, A. (2005). Very high
624 resolution interpolated climate surfaces for global land areas. *International Journal of*
625 *Climatology,* **25**, 1965-1978.
626 Hill, J.K., Thomas, C.D. & Huntley, B. (1999). Climate and habitat availability determine
627 20th century changes in a butterfly's range margin. *Proceedings of the Royal Society*
628 *of London Series B-Biological Sciences,* **266**, 1197-1206.
629 Holveck, M-J., Gauthier, A-L. & Nieberding, C.M. (2015). Dense, small and male-biased
630 cages exacerbate male–male competition and reduce female choosiness in *Bicyclus*
631 *anynana*. *Animal Behaviour*, **104**, 229-245.
632 Hughes, C.L., Hill, J.K., & Dytham, C. (2003). Evolutionary trade-offs between reproduction
633 and dispersal in populations at expanding range boundaries. *Proceedings of the Royal*
634 *Society of London B: Biological Sciences*, **270**, S147-S150.
635 Ketmaier, V., Giusti, F. & Caccone, A. (2006). Molecular phylogeny and historical
636 biogeography of the land snail genus *Solatopupa* (Pulmonata) in the peri-Tyrrhenian
637 area. *Molecular Phylogenetics and Evolution,* **39**, 439-451.

638 Kumar, S., Stecher, G. & Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics
639 Analysis Version 7.0 for Bigger Datasets. *Molecular Biology and Evolution,* **33**,
640 1870-1874.
641 Kuznetsova, A., Bruun Brockhoff, P. & Christensen, R.H.B. (2016). lmerTest: Tests in linear
642 mixed effect models. r package version 2.0-33. https://CRAN.R-
643 project.org/package=lmerTest
644 Leigh, J.W. & Bryant. D. (2015). POPART: full-feature software for haplotype network
645 construction. *Methods in Ecology and Evolution,* **6**, 1110-1116.
646 Librado, P. & Rozas, J. (2009). DnaSP v5: a software for comprehensive analysis of DNA
647 polymorphism data. *Bioinformatics,* **25**, 1451-1452.
648 Longdon, B., Day, J., Schulz, N., Leftwich, P.T., de Jong, M.A., Breuker, C.J., …..Jiggins,
649 F.M. (2017). Vertically transmitted rhabdoviruses are found across three insect
650 families and have dynamic interactions with their hosts. *Proceedings of the Royal*
651 *Society B-Biological Sciences,* **284**, 20162381.
652 Longdon, B., Murray, G.G., Palmer, W.J., Day, J.P., Parker, D.J., Welch, J.J., Obbard, D.J. &
653 Jiggins, F.M. (2015). The evolution, diversity, and host associations of rhabdoviruses.
654 *Virus Evolution,* **1**: vev014.
655 Mende, M.B., & Hundsdoerfer, A.K. (2013). Mitochondrial lineage sorting in action –
656 historical biogeography of the *Hyles euphorbiae* complex (Sphingidae, Lepidoptera)
657 in Italy. *BMC Evolutionary Biology*, **13**:83.
658 Merckx, T. & Van Dyck, H. (2006). Landscape structure and phenotypic plasticity in flight
659 morphology in the butterfly *Pararge aegeria*. *Oikos,* **113**, 226-232.
660 Okonechnikov, K., Golosova, O., Fursov, M. & the UGENE team. (2012). Unipro UGENE: a
661 unified bioinformatics toolkit. *Bioinformatics,* **28**, 1166-1167.

662 Oliver, T.H., Marshall, H.H., Morecroft, M.D., Brereton, T., Prudhomme, C. & Huntingford,
663 C. (2015). Interacting effects of climate change and habitat fragmentation on drought-
664 sensitive butterflies. *Nature Climate Change,* **5**, 941-945.
665 Parmesan, C. (1999). Metapopulation ecology. *Nature,* **399**, 747.
666 Pinzari, M. & Sbordoni, V. (2013). Species and mate recognition in two sympatric Grayling
667 butterflies: *Hipparchia fagi* and *H. hermione genava* (Lepidoptera). *Ethology Ecology*
668 *& Evolution,* **25**, 28-51.
669 Presgraves, D.C., Balagopalan, L., Abmayr, S.M. & Orr, H.A. (2003). Adaptive evolution
670 drives divergence of a hybrid inviability gene between two species of *Drosophila*.
671 *Nature,* **423**, 715-719.
672 QGIS Development Team (2009). QGIS Geographic Information System. Open Source
673 Geospatial Foundation. URL http://qgis.osgeo.org.
674 Quah, S., Breuker, C.J. & Holland, P.W. (2015). A diversity of conserved and novel ovarian
675 microRNAs in the Speckled Wood (*Pararge aegeria*). *PloS one*, **10**: e0142243.
676 R Development Core Team (2016). R: A language and environment for statistical computing.
677 R Foundation for Statistical Computing, Vienna.
678 Ribera, I., & Vogler, A.P. (2004). Speciation of Iberian diving beetles in Pleistocene refugia
679 (Coleoptera, Dytiscidae). *Molecular Ecology*, **13**, 179-193.
680 Richter-Boix, A., Quintela, M., Kierczak, M., Franch, M. & Laurila, A. (2013). Fine-grained
681 adaptive divergence in an amphibian: genetic basis of phenotypic divergence and the
682 role of nonrandom gene flow in restricting effective migration among wetlands.
683 *Molecular Ecology,* **22**, 1322-1340.
684 Rogers, S.M. & Bernatchez, L. (2006). The genetic basis of intrinsic and extrinsic post-
685 zygotic reproductive isolation jointly promoting speciation in the lake whitefish

686 species complex (*Coregonus clupeaformis*). *Journal of Evolutionary Biology* **19**,
687 1979-1994.
688 Schmitt, T. (2007). Molecular biogeography of Europe: Pleistocene cycles and postglacial
689 trends. *Frontiers in Zoology,* **4**, 11.
690 Seddon, J.M., Santucci, F., Reeve, N.J. & Hewitt, G.M. (2001). DNA footprints of European
691 hedgehogs, *Erinaceus europaeus* and *E. concolor*. Pleistocene refugia, postglacial
692 expansion and colonization routes. *Molecular Ecology,* **10**, 2187-2198.
693 Taberlet, P., Fumagalli, L., Wust-Saucy, A.G. & Cosson, J.F. (1998). Comparative
694 phylogeography and postglacial colonization routes in Europe. *Molecular Ecology,* **7**,
695 453-464.
696 Templeton, A.R., Routman, E. & Phillips, C.A. (1995). Separating population structure from
697 population history: a cladistic analysis of geographical distribution of mitochondrial
698 DNA haplotypes in the tiger salamander, *Ambystoma tigrinum*. *Genetics,* **140**, 767-
699 782.
700 Tison, J.-L., Edmark, V.N., Sandoval-Castellanos, E., Van Dyck, H., Tammaru, T., Välimäki,
701 P., Dalén, L. & Gotthard, K. (2014). Signature of post-glacial expansion and genetic
702 structure at the northern range limit of the speckled wood butterfly. *Biological*
703 *Journal of the Linnean Society,* **113**, 136-48.
704 Toews, D.P.L. & Brelsford, A. (2012). The biogeography of mitochondrial and nuclear
705 discordance in animals. *Molecular Ecology,* **21**, 3907-3930.
706 Vodă, R., Dapporto, L. Dincă, V. & Vila, R. (2015a). Cryptic matters: overlooked species
707 generate most butterfly beta-diversity. *Ecography*, **38**, 405-409.
708 Vodă, R., Dapporto, L., Dincă, V. & Vila, R. (2015b). Why do cryptic species tend not to co-
709 occur? A case study on two cryptic pairs of butterflies. *PLoS one,* **1**0: e0117802.

710 Vodă, R., Dapporto, L., Dincă, V., Shreeve, T.G., Khaldi, M., Barech, G., … Vila, R. (2016).
711 Historical and contemporary factors generate unique butterfly communities on
712 islands. *Scientific Reports*, **6**: 28828.
713 Wahlberg, N. & Wheat, C.W. (2008). Genomic outposts serve the phylogenomic pioneers:
714 Designing novel nuclear markers for genomic DNA extractions of Lepidoptera.
715 *Systematic Biology*, **57**, 2, 231–242.
716 Waters, J.M. (2011). Competitive exclusion: phylogeography's 'elephant in the room'?.
717 *Molecular Ecology*, **20**, 4388-4394.
718 Weingartner, E., Wahlberg, N. & Nylin, S. (2006). Speciation in *Pararge* (Satyrinae:
719 Nymphalidae) butterflies - North Africa is the source of ancestral populations of all
720 *Pararge* species. *Systematic Entomology,* **31**, 621-632.
721 Wilfert, L. & Jiggins, F.M. (2014). Flies on the move: an inherited virus mirrors *Drosophila*
722 *melanogaster*'s elusive ecology and demography. *Molecular Ecology,* **23**, 2093-2104.
723
724

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 synthesise 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.

Tables and Figures - Legends

Figure legends

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).

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).

749 Figure 3

750 Distribution of the genetic richness of *Pararge aegeria* in the study area based on 0.25x0.25
751 degree squares for which at least 4 specimens were sequenced in a 100km radius. Genetic
752 richness was calculated separately for the two lineages identified in this study for each of
753 these squares. The method involves calculating matrices of p-distances (proportions of
754 nucleotide differences), taking geographic distances into account. At the end, a single value,
755 indicating the genetic differentiation of four specimens closest to each other weighted for
756 their distance from the centre of their locations, is then plotted onto a map. This has been
757 represented here as a heat map of sequence variation across a wide geographical range (full
758 range 0% (green) to 1.6% (red); values indicated in figure)(for full details on the genetic
759 richness method see Supplementary File 1).

760 Figure 4

761

762 Proposed hypothesis for the historical biogeography of *Pararge aegeria*. The ancestral
763 lineage (blue circles) was present throughout the range of *P. aegeria* in Europe (A), without
764 substantial differentiation of the nDNA markers due to unrestricted dispersal between
765 populations. During the last glacial period (possibly also including previous series of glacial
766 events) (B) the range retracted southwards (red arrows), and gene flow was restricted
767 between the refugia due to the Alps and Pyrenees acting as barriers, which allowed for
768 periods of differentiation (yellow circles in C). Following the warming of the climate, the
769 eastern lineage spread northwards and westwards (red arrows in D), where it could have
770 introgressed with the nuclear genome of warm adapted populations in the Iberian peninsula
771 as well as the islands of Ibiza, Corsica and Sardinia resulting in the discordance between the
772 markers (indicated by blue and yellow circles in D). This introgression was presumably
773 hindered by sea straits, giving rise to the sharp boundary observed for the *COI* data.

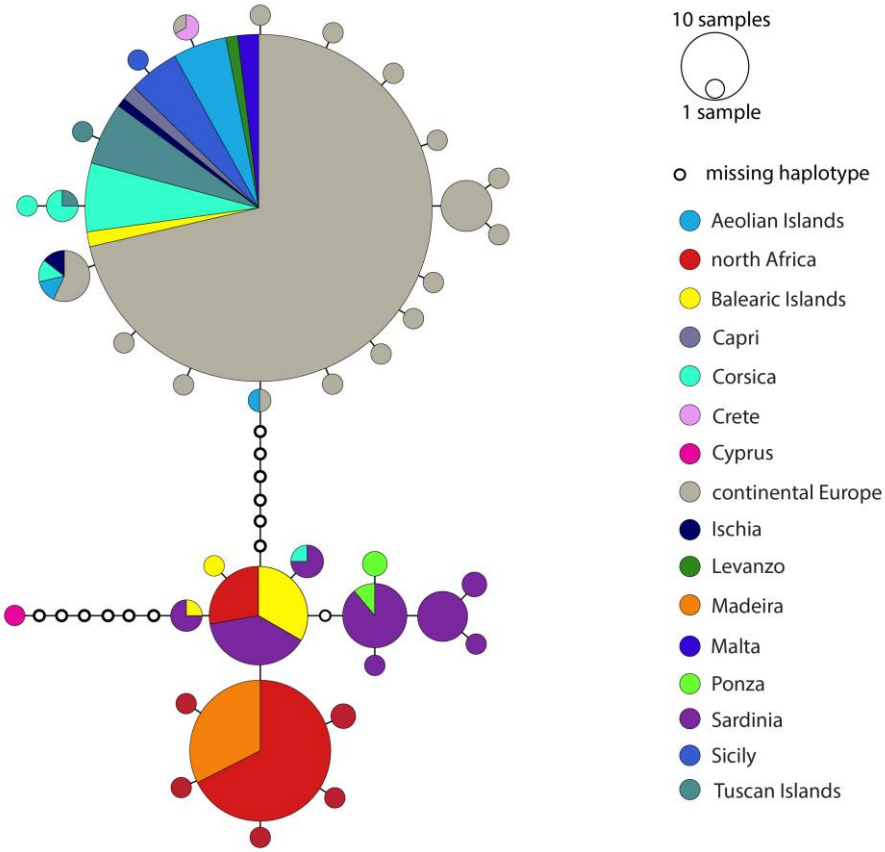
774

775

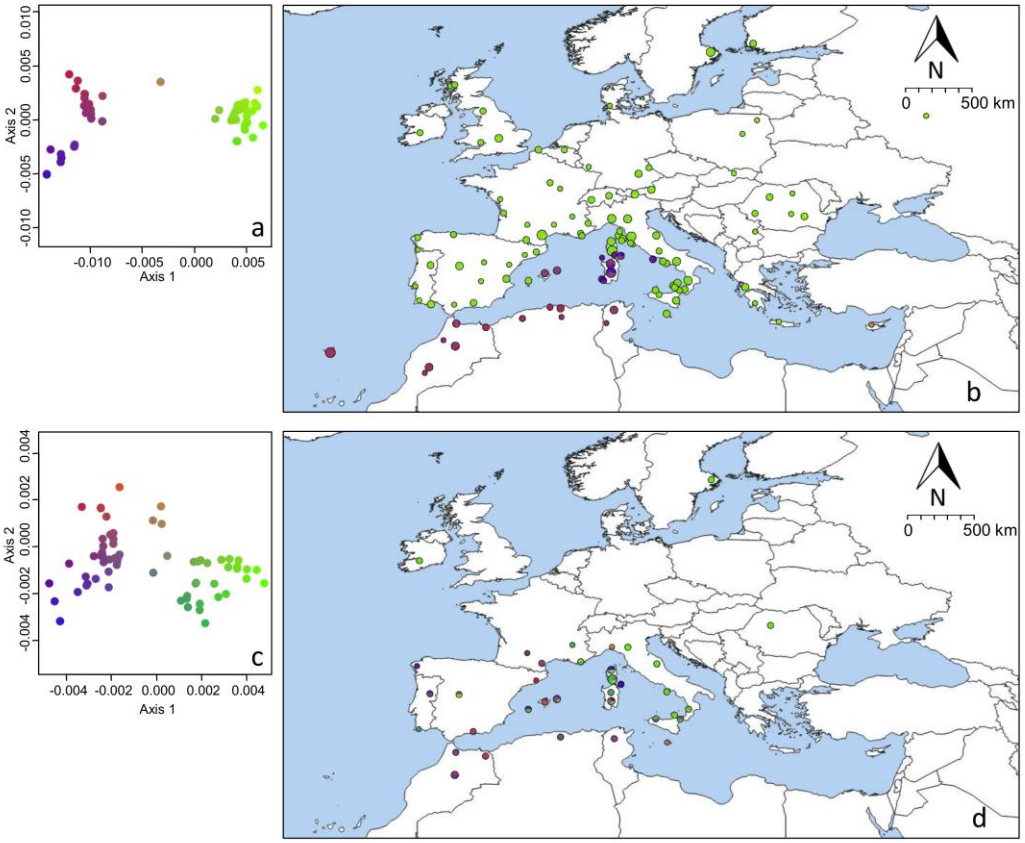
776

Figures

Figure 1



790 Figure 2



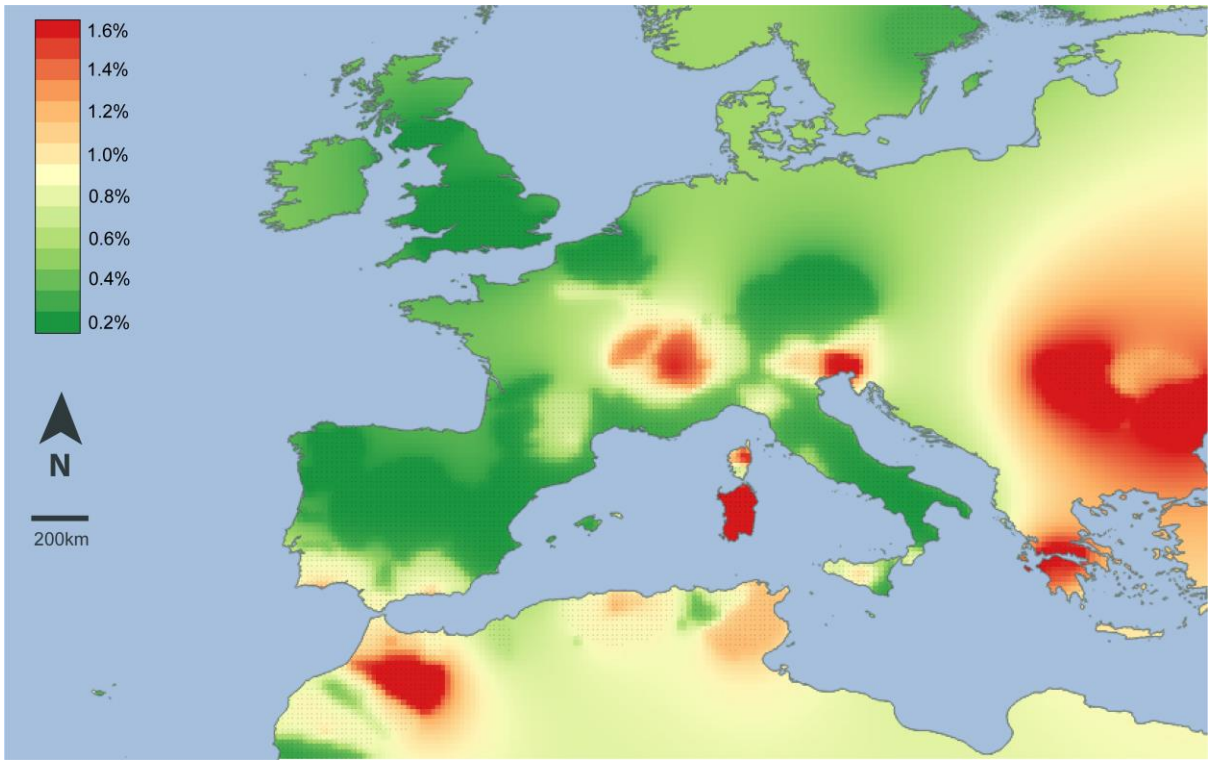
791

792

793

794

795 Figure 3



796

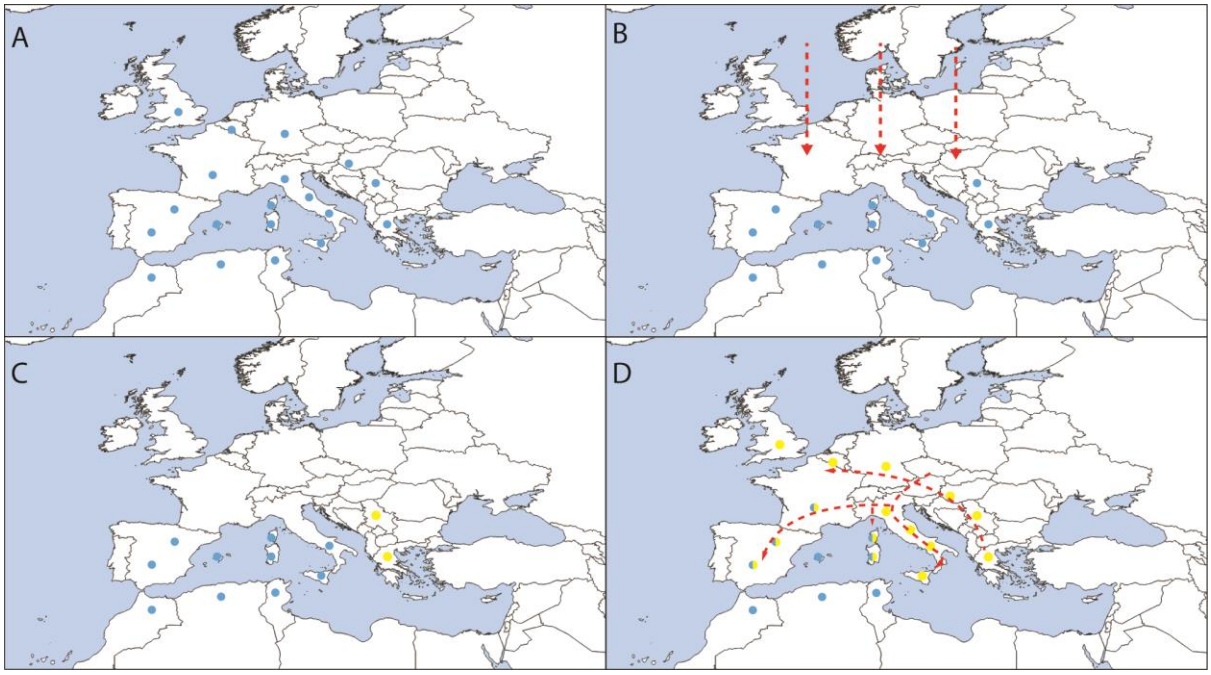
797

798

799

800

801 Figure 4



802

803

804