Inferring the biogeography and demographic history of an endangered butterfly in Europe from multilocus markers

Laurence Després[1,*], Clément Henniaux[1], Delphine Rioux[1], Thibaut Capblancq[1], Sara Zupan[2], Tatjana Čelik[3], Marcin Sielezniew[4], Lucio Bonato[5], Gentile Francesco Ficetola[1,6]

[1]Univ. Grenoble Alpes, LECA UMR5553, CNRS, F-38000 Grenoble, France

[2] University of Primorska, Faculty of Mathematics, Natural Sciences and Information Technologies, Glagoljaška 8, SI-6000 Koper, Slovenia

[3] Research Centre of the Slovenian Academy of Sciences and Arts, Jovan Hadži Institute of Biology, Novi trg 2, SI-1000 Ljubljana, Slovenia

[4] Laboratory of Insect Evolutionary Biology and Ecology, Institute of Biology, University of Bialystok, 15-245 Białystok, Poland

[5] Department of Biology, Università degli Studi di Padova, Padova, Italy

[6] Department of Environmental Science and Policy, Università degli Studi di Milano, Milano, Italy

Date: May, 2018

* Corresponding author: Laurence Després, LECA, 2233 Rue de la Piscine, 38041 Grenoble Cedex 9, France.

Tel: +33 (0)4 76 63 56 99, E-mail: laurence.despres@univ-grenoble-alpes.fr

Running title: False ringlet population genomics in Europe

20 Abstract

21 The genetic structure of a species is influenced by its history and by current gene flow. Using a
22 population genomics approach, we inferred the demographic history of the False Ringlet
23 (*Coenonympha oedippus*) based on 1,594 genome-wide ddRADseq loci from 96 individuals (32
24 localities) sampled throughout the fragmented species range in Europe. In contrast with the lack of
25 geographical structure in mtDNA, a clear nuclear differentiation was observed between the
26 westernmost Atlantic populations, those from the western Alps, and all other sampled populations.
27 Mountain ranges were the main factor explaining population divergence at the European scale, while
28 isolation by distance was found at a regional scale. We applied Approximate Bayesian Computation in
29 a coalescent framework to infer past and contemporary demographic parameters. The best scenario
30 suggested a first divergence between French and all other European populations around 66,000 years
31 ago, so that the species survived the last glacial maximum in at least two distinct areas separated by
32 the Alps. This scenario fits species distribution modelling identifying variation of suitable areas with
33 past climatic modifications. The Atlantic and western Alps populations separated some 6,000 years
34 ago. Strong population decline was inferred in these populations during historical time, in agreement
35 with multiple records of recent decline of this species in Europe.

36 Key words: ddRADseq, mtDNA, demographic history, *Coenonympha oedippus,* glacial refugia,
37 genetic diversity, population size, species distribution modelling.

38

39

40 Introduction

41 The genetic structure of a species reflects both its history and ongoing gene flow. Characterizing
42 population histories and identifying the main environmental factors shaping genetic variation at
43 different spatial scales have been a major focus in evolutionary and conservation biology for decades.
44 The Pleistocene cold periods in the northern hemisphere have influenced the distribution of species
45 with range fluctuations in relation with climatic variations during the last 700 ky. During glaciations,
46 many temperate European taxa were restricted to southern ice-free refugia (Taberlet *et al.*, 1998). The
47 present distribution of most species in Europe result from a northward recolonization from those
48 southern refugia after the last glacial maximum (LGM), about 21 kya (Strandberg *et al.*, 2011). Under
49 this hypothesis, the southern part of Europe should present the highest genetic diversity, in contrast
50 with the recently recolonized northern part (e.g., Besold *et al.*, 2008; Patricelli *et al.* 2013). However,
51 genetic analyses have identified numerous additional extra-mediterranean refugia, thus strongly
52 modifying the biogeographical view of Europe (Schmitt & Varga, 2012; Kühne *et al.*, 2017). This
53 picture is complicated in species with a wide Eurasian distribution, where other potential eastern
54 refugia could have existed, with possible admixture occurring between diverging lineages during
55 postglacial recolonization (e.g., Grassi *et al.*, 2008). Geographical distribution and genetic structure
56 are affected not only by the species' evolutionary history but also by dispersal abilities, present
57 demographic characteristics – especially fluctuations in population size – and by habitat fragmentation
58 (Keyghobadi, 2007; Louy *et al.*, 2007). Many lowland insect species have been particularly affected
59 by human impact via the intensification of agriculture (insecticide spraying, land draining). Although
60 they were abundant a few decades ago, they now show highly fragmented populations with high
61 extinction risk (Hallmann *et al.*, 2017).

62 Analyses of genetic diversity within and between populations provide key information for
63 conservation of endangered species given that they allow inferring important demographic parameters
64 such as historical and contemporary effective population sizes, dispersal rates across populations, and
65 consanguinity levels. Such knowledge is necessary to guide conservation actions such as the creation
66 of corridors favoring the natural re-colonization of suitable habitats, or the best choice of individuals
67 for a successful re-location. To date most phylogeographical studies at the continental scale were
68 based on mitochondrial DNA, and population genetic analysis focused on a few allozyme or
69 microsatellite markers and required to analyze many individuals per population, which was not always
70 possible for endangered species. With the development of high throughput sequencing technologies, it
71 is now possible to infer the genetic diversity within and between populations with only few individuals
72 per population, because the low number of individuals sampled is partly compensated by a very high
73 number of loci genotyped (Nazareno *et al.*, 2017).

In this study we used high throughput genotyping besides the classical mitochondrial barcode (partial CO1) to uncover the past and current factors involved in shaping the genetic structure of one of the most endangered butterfly species in Europe, the False Ringlet, *Coenonympha oedippus*. Although the species is distributed across Eurasia from western France to Japan (Bozano, 2002), its range is today highly fragmented, especially throughout Europe (Kudrna *et al.*, 2011), because its habitat (mainly wetlands) has been significantly reduced and is still disappearing as a consequence of human activities (Lhonore & Lagarde, 1999). Despite the wide distribution range of this butterfly, from Atlantic to Pacific coast, very little information is available on the intra-specific pattern of genetic diversity in *C. oedippus*, and on the genetic connections between populations. *C. oedippus* is generally considered to be a monotypic species (Bozano, 2002), despite many subspecific and infrasubspecific taxa proposed by different taxonomists, but genetic studies have often found cryptic genetic structure within butterfly species (Hebert *et al.*, 2004; Dincă *et al.*, 2011; Ritter *et al.*, 2013).

We used double digest Restriction site Associated DNA sequencing (ddRADseq) to identify thousands of genetic markers without any prior knowledge on the *Coenonympha* genome (Peterson *et al.*, 2012). In contrast with the low mitochondrial variation detected by sequencing the cytochrome oxidase 1 mitochondrial (CO1) gene, which is routinely used as a barcode in butterflies, we found large variation in nuclear genetic diversity across Europe, and identified populations where loss of genetic diversity poses threats to species conservation. We used nuclear genetic diversity to test for alternative demographic histories (splits, expansions, recent declines) of European populations by Approximate Bayesian Computation (ABC) approach in a coalescent framework. Finally, we performed species distribution modelling (MaxEnt) to identify current and past climatically suitable areas for *C. oedippus* in Europe.

Materials and methods

Study species

The False Ringlet, *Coenonympha oedippus* (Fabricius, 1787) (Lepidoptera: Nymphalidae) is a univoltine Palearctic sedentary species flying mainly in June and July (Čelik *et al.*, 2009; Verovnik, Rebeušek & Jež, 2012; Bonato, Uliana & Beretta, 2014). It is a hygrophilous insect inhabiting mostly wet meadows and fens, where caterpillars feed on *Carex* spp. as well as on *Molinia caerulea*, but at the southern range limit in Slovenia it can be found also on abandoned drier grasslands, where caterpillars feed on other *Carex* species than in wet habitats, and also on *Festuca rupicola* (Čelik *et al.*, 2015).

C. oedippus is one of the most endangered butterfly species in Europe and listed in Annex II and IV of the Habitats Directive as well in the Appendix II of the Bern Convention (Van Swaay *et al.*, 2010). It

108 became extinct in three of the 14 countries where it had been recorded (Van Swaay & Warren, 1999),
109 i.e. in Slovakia (Pastoralis & Reiprich, 1995), Bulgaria (Staub & Aistleitner, 2006) and Switzerland
110 (Dušej *et al.*, 2010). In most of other countries *C. oedippus* is declining and during the last century it
111 has disappeared from many localities, e.g. in Germany, where only one meta-population is still present
112 in Bavaria (Bräu, Dolek & Stettmer, 2010), and France (Lhonore & Lagarde, 1999), where the species
113 went extinct in the Paris region, and is currently present in only two disconnected and distant regions:
114 between the Atlantic coast and the Pyrenees (SW France) and in the Rhône and Isère valleys in the
115 Western Alps (E France). In the former region, populations are locally abundant in marshes (Poitou-
116 Charente) and in managed maritime pine forests (Landes) (van Halder *et al.*, 2008), while in the latter
117 region the species is restricted to three protected marshes (Lavours-Ain, Chautagne-Savoie and
118 Montfort-Isère) (Varin, 1964). The *C. oedippus* range also contracted in Slovenia and now has a
119 disjunct distribution there (Čelik & Verovnik, 2010): the predominantly limestone region of SW
120 Slovenia, and marshy areas in central Slovenia south of Ljubljana. In contrast, over 100 populations
121 are known to occur in northern Italy, however often restricted to small isolated areas (Bonelli,
122 Canterino & Balletto, 2010; Bonato *et al.*, 2014). Knowledge about the past and present distribution of
123 *C. oedippus* in eastern Europe is still inadequate. For example, in Poland the species was considered
124 extinct in the 70's of 20th century, but over the last three decades several sites have been discovered in
125 the eastern part of the country as a result of intensification of inventory activities and therefore little is
126 known about recent trends (Sielezniew *et al.*, 2010; Sielezniew, 2012).

127 In the last two decades, given the dramatic decline of populations throughout the western part of
128 European range, several studies have investigated the factors limiting population viability (for review
129 see (Čelik *et al.*, 2015). Current threats include land reclamation for agriculture, land drainage and
130 urban expansion, but also natural reforestation of grasslands.

131

132 *Samples collection*

133 A total of 32 localities were sampled through most of the distribution range of *C. oedippus* in Europe,
134 from the westernmost populations on the Atlantic coast (SW France) to those in the eastern Polish
135 lowland, including many isolated populations around the Alpine mountain range (Figure 1 and Table
136 1). Pairwise distances between sampled localities ranged from 400 m to up to 2600 km. The 32
137 sampling localities where categorized into 5 geographical regions based on the presence of natural
138 barriers to dispersion (mountain range and distance): Atlantic (including populations from the Atlantic
139 coast to the Pyrenees foothills), Western Alps (including populations from the Rhône and Isère
140 valleys), Southern Alps (including populations from northern Italy to central Slovenia), Northern Alps
141 (including populations from Liechtenstein and Bavaria) and East European (including six Polish
142 populations). Because of the endangered status of *C. oedippus* and in order to have the lowest impact

143 as possible on the populations, only 2-5 males per sampled locality were caught using entomological
144 nets at the end of the flying period (in July) and were kept dry (<1 month). After wing removal, the
145 body was kept in ethanol 75° at -20°C for genetic analysis, except for samples from Slovenia, which
146 were kept at -80°C. To test whether even less invasive sampling could be performed on this
147 endangered species, we used only two legs from each of three specimens from Ger (Atlantic region).
148 The legs were kept at -80°C until extraction.

149 *DNA extraction*

150 DNA was extracted from the complete thorax of each individual, with the exception of the three
151 specimens from Ger, using the DNeasy Blood and Tissue Kit (QIAgen, Germany) according to the
152 manufacturer's instructions and stored at -20°C. For the specimens from Ger, DNA was extracted
153 from two legs using cetyl trimethyl ammonium bromide (CTAB) chloroform/isoamyl alcohol protocol
154 (Doyle & Doyle, 1987).
155

156 *ddRADseq library preparation and SNP calling*

157 A double-digested RAD (Restriction site Associated DNA) experiment was conducted on 104 samples
158 (98 specimens and 6 replicates; Table 1) in 3 libraries using a modified version of the protocol
159 previously described (Capblancq *et al.*, 2015; Peterson *et al.*, 2012). Briefly, 200 ng of DNA template
160 from each individual were double-digested with 10 units each of *Sbf*I–HF and *Msp*I (New England
161 Biolabs Inc.) at 37°C for one hour using the CutSmart buffer provided with the enzymes. Digestion
162 was further continued together with the ligation of P1 (individually indexed) and P2 adapters by
163 adding 10 units of T4 DNA ligase (New England Biolabs Inc.), adapters P1 and P2 and 1µl of 10mM
164 ribo-ATP (New England Biolabs Inc.) in each sample. The digestion-ligation was performed in a
165 thermocycler (60 cycles of 2 min digestion at 37°C and 2 min ligation at 16°C, followed by final heat
166 inactivation of the enzymes at 65°C for 10 min). An equal volume of all the digested-ligated
167 individuals was pooled and purified with Agencourt AMPure XP beads (Beckman Coulter, France).
168 After migration on 1.6% agarose gel, fragments between 250 and 500bp were excised and purified
169 with QIAquick Gel Extraction Kit (Qiagen, Germany). Each ddRAD library was amplified in ten
170 independent replicates of 15 PCR cycles (initial denaturation 10 min, 98°C; 15 cycles of 98°C for 10 s,
171 66°C for 30 s and 72°C for 1 min; followed by a final 10 min extension period at 72°C) in a final
172 volume of 20µl with 1µl of DNA template, 10 mM of dNTPs, 10µM of each PCR primers (Peterson *et*
173 *al.*, 2012) and 2U/µl of *Taq* Phusion-HF (New England Biolabs Inc.). The ten PCR products were
174 pooled and purified with QIAgen MinElute PCR Purification Kit (Qiagen, Germany). Each library
175 was sequenced on an Illumina Hi-Seq 2500 Illumina sequencer (1/10 lane per library, paired-end 2 x
176 125 bp, Fasteris SA, Switzerland). Sequencing errors per lane (PhiX control) were very low (0.26%,
177 0.82% and 0.27% for three libraries, respectively) which means that convergent sequencing errors (the

178 same error occurring independently at the same nucleotide position in the same read) are very
179 unlikely. Reads with depth coverage < 5 were excluded from further analyses. Genotyping errors
180 (locus and allelic dropout) were estimated by comparing 6 replicate pairs (3 inter-libraries and 3 intra-
181 library replicate pairs).
182 The ~68 million DNA reads obtained were used to call SNP genotypes with the *STACKS*
183 pipeline (Catchen *et al.*, 2013) as follows: the *process_radtags* function was first run to demultiplex
184 the data and filter the reads on their quality. We removed reads with length < 100 nucleotides and cut
185 all reads to this value, resulting in more than 92% of total reads retained. On average, we retained 89%
186 of total reads by individual after removing reads of low quality or with uncalled bases (options -q, -c
187 and -r). Individuals with <100,000 reads were discarded (n=2, Tables 1 and S1). Only *Sbf*I reads were
188 retained for *de novo* assembly on each individual using a maximum of 7 mismatches to merge two
189 stacks into a polymorphic locus (-M; *ustacks* function). This threshold was chosen after inspecting the
190 effect of increased values of M on the proportion of polymorphic loci (Figure S1). Highly-repetitive
191 stacks and over merged tags were dropped using the "Removal" (-r) and the "Deleveraging" (-d)
192 options. A catalog of the loci from all the individuals was built, with a maximum of 9 mismatches for
193 merging two individual loci (-n; *cstacks* function). Loci within each individual were searched against
194 the catalog (*sstacks* function) and a SNP dataset was produced with the genotype of each individual
195 for every polymorphic position (*populations* function).
196
197 *Genetic diversity and genetic structure estimation*
198 For genetic diversity indices and analysis of population structure, only SNPs present in more
199 than 60% of the whole sampling were retained to avoid an excess of missing data. SNPs with a
200 minimum allele frequency lower than 5% were removed from the data set and only one polymorphic
201 site was kept for each RAD-tag ('write_random_snp' option) in order to analyze only unlinked
202 polymorphisms.
203 Genetic diversity by individual (individual heterozygosity, or observed heterozygosity Ho),
204 within each sampled locality (population genetic diversity, or expected heterozygosity He), and
205 between all population pairs was assessed using hierfstat R package. F_{IS} and F_{ST} estimates were
206 calculated according to Weir & Cockerham (1984). Confidence intervals (95%) for Ho, He and F_{IS}
207 were assessed by 1,000 bootstraps across loci.
208 Clustering of individuals into homogenous genetic clusters ranging from K=1 to K=32 was
209 tested using Structure 2.3.4 (Falush, Stephens & Pritchard, 2003). For each run, a burn-in period of
210 5,000 steps was followed by 20,000 iterations under the admixture model and the assumption of
211 correlated allele frequencies among populations. For each K, 10 runs were performed. Estimated log
212 probabilities (Ln P(D)) were averaged across runs and compared to determine the posterior probability
213 of each K using Clumpak (Kopelman *et al.*, 2015). The best K was selected using the ΔK method
214 (Evanno, Regnaut & Goudet, 2005) in Structure Harvester (Earl, 2012). As one single K value only

215 provides an incomplete picture of overall population structure, we explored the pattern of population
216 structure within the main clusters detected (Janes *et al.*, 2017).
217 In order to assess the respective roles of geographic distance and orographic barriers in
218 population genetic differentiation of this species, which is depending on low-altitude habitats (see
219 Introduction), we performed a multiple linear regression on distance matrix (MRM) in package R
220 'ecodist' where genetic distance ($F_{ST}/(1 - F_{ST})$ was treated as a response matrix. The straight-line
221 geographic distances (square-root transformed) and presence of mountain ranges higher than 1100 m
222 a.s.l. were set as the explanatory matrices. The available data on historic and present distribution of the
223 species in Europe showed that 700 m a.s.l. is the highest altitude for the most localities of the species
224 (Verovnik *et al.* 2012; Bonato *et al.*, 2014). The exceptions are scarce localities on southern foothills
225 of the Alps where species was found on semi-open dry grasslands at 750–1100 m a.s.l. (e.g. Čelik &
226 Rebeušek, 1996).
227 For phylogenetic inference, we used every locus present in at least 59 individuals of the whole
228 sampling (N=96), including invariant positions. Heterozygote positions were coded with IUPAC code.
229 We used full sequences rather than just SNPs because it was shown to be preferable from the
230 perspectives of branch length and topological accuracy (Leaché *et al.*, 2015). The maximum likelihood
231 phylogenetic tree was generated using RAxML (Stamatakis, 2014) with 100 rapid bootstrap inferences
232 following search for the best ML tree using the GTR + G model for rate heterogeneity.

233 *CO1 sequencing and phylogeographic analysis*

234 We sequenced also a mitochondrial marker for 38 individuals representative of 24 localities
235 from all the main regions studied (Table 1). CO1 was amplified using the primer pairs LCO–HCO and
236 Jerry–Pat (Wahlberg & Freitas, 2007) (PCR protocol: 95°C for 10 min; 35 cycles of 95°C for 30 s,
237 50°C for 30 s and 72°C for 30 s; followed by a final extension period of 72°C for 7 min) and
238 sequenced by Genewiz Company, UK. The resulting chromatograms were visualized in the software
239 BIOEDIT ver. 7.2.5 (Hall, 1999) and aligned using ClustalW and by eye. Five sequences of *C.*
240 *oedippus* available in BOLD were added to the multiple alignment. They originate from the following
241 localities: Ruggell, Liechtenstein (BOLD accession code: PHLAF624-11); Munich, Oberbayern
242 (GenBank code: GU707147); Romano d`Ezzelino, Vicenza province, Italy (BOLD code: PHLSA390-
243 11); Obluchye, European Russia (GenBank code: EU920755); Tavolzhanka, Kazakhstan (BOLD
244 code: LOWA191-06). Additionally, sequences from eight outgroup species were chosen based on the
245 most recent phylogeny of Coenonymphina butterflies published (Kodandaramaiah & Wahlberg, 2009),
246 including six *Coenonympha* species (*C. tullia*, *C. hero*, *C. glycerion*, *C. nolckeni*, *C. phryne*
247 (previously under *Triphysa*), *C. myops* (previously under *Lyela*); Genbank codes EU920762,
248 EU920750, EU920749, EU920754, EU920739, EU920741), and two species of strictly related genera
249 (*Heteronympha merope* and *Mydosama terminus*; EU92073, DQ338765). The maximum likelihood

250 phylogenetic tree was generated after selecting for the best model of molecular evolution using Mega7
251 version 7.0.14 (Kumar, Stecher & Tamura, 2016).
252
253 *Demographic scenarios and population size inferences*

254 Competing hypotheses regarding population divergence at the European scale based on the nuclear
255 data were compared using Approximate Bayesian Computation (ABC) as implemented in DIYABC
256 v2.1 (Cornuet *et al.*, 2014). Based on the results from STRUCTURE, which identified three main
257 genetic clusters (from West to East), we tested whether the geographically intermediate lineage
258 (Western Alps region) was more related to the western (Atlantic region) or to the eastern lineage
259 (remaining regions). For each scenario we allowed population size changes after each split time. The
260 competing scenarios were set using uniformly broadly distributed priors (10^2-10^7 individuals for
261 population sizes and 10,000-700,000 years for divergence times). As *C. oedippus* is a univoltine
262 species (Bonato *et al.*, 2014; Čelik, Vreš & Seliškar, 2009; Verovnik *et al.*, 2012), divergence times
263 were directly estimated in years. For each scenario, 100,000 data sets were simulated and the posterior
264 probability was computed by performing a logistic regression on the 1% of simulated data closest to
265 the observed data set (Cornuet *et al.*, 2014). Summary statistics of observed/simulated dataset
266 comparisons were mean genetic diversity within populations, and F_{ST} and Nei's distances among
267 populations, using only SNPs with a minimum allele frequency >5%. We further estimated divergence
268 time and tested for recent bottlenecks within the western and intermediate lineages, and we tested
269 alternative splitting hypotheses within the eastern lineage.

270

271 *Species distribution models*

272 We used Maximum Entropy Modelling (MaxEnt) to build species distribution models (SDMs) relating
273 the distribution of *C. oedippus* to climatic variables, and to assess potential distribution changes since
274 the LGM. MaxEnt is a presence-background modelling tool; comparative analyses showed that
275 MaxEnt is among the SDMs with best predictive performance (Elith *et al.*, 2006, 2011). Models were
276 calibrated on the basis of 463 presence records, obtained from the literature, from the Global
277 Biodiversity Information Facility (GBIF, 2016) and from our own surveys (Table S2). As climatic
278 variables, we considered a set of variables that represent the climatic conditions experienced by the
279 species through the year: mean summer temperature, mean winter temperature, temperature
280 seasonality, summed precipitation during the summer, and summed precipitation during winter.
281 Variables were extracted from the Worldclim dataset at the 10 arc-primes resolution (approx. 15 km
282 within the study area) (Hijmans *et al.*, 2005); for analyses, we only retained one presence record per
283 each cell. We built models with linear, quadratic and hinge features; we run preliminary models with a
284 range of different regularization multipliers (1, 2, 3, 4, 5, 7 and 10) and selected the best regularization

285 multiplier on the basis of corrected Akaike's Information Criterion (AICc) (Warren & Seifert, 2011).
286 We used a 10-fold cross-validation to assess the predictive performance of the best-AICc model
287 (Nogués-Bravo, 2009). Predictive performance was evaluated on the basis of the area under the curve
288 of the receiver operator plot of the test data (AUC), averaged over the ten replicated runs (Manel,
289 Williams & Ormerod, 2001). We assumed that a cell is suitable if its suitability value was higher than
290 the 10% presence threshold (averaged over the cross-validated runs); we assumed a high suitability if
291 suitability was higher than 0.5 (Pearson *et al.*, 2007; Elith *et al.*, 2011). Models were then projected to
292 the mid-Holocene (6 kya) and LGM (21 kya) conditions, using MPI-ESM model. When projecting to
293 past climates, we assessed whether models were projected into climatic conditions different from the
294 ones found in the calibration climate using clumping and evaluating if climatic variables are outside
295 the training range (Elith, Kearney & Phillips, 2010).

296

297 ## Results

298 ### *Nuclear genetic diversity*

299 More than 60 million high quality reads were obtained with an average 600,000 reads/sample. A total
300 of 102 samples (96 individuals and 6 replicates), with an average of 7,500 loci per individual (mean
301 coverage/locus: 60) were kept for genetic analysis (Table S1). The three samples from Ger (Atlantic
302 region) passed this filter, indicating that DNA extracted from two legs can be enough to successfully
303 achieve the ddRADseq experiment. A total of 1,594 loci (100 bp each, including 126 monomorphic
304 loci) present in >60% of the whole sampling (i.e. ≥59 individuals) were considered. A total of 1,314
305 independent SNPs were retained by selecting one random SNP per locus with minimum allele
306 frequency >5%. Genotyping errors ranged from 1 to 10% for locus dropout (absence of a locus in the
307 replicate) but allelic dropout (heterozygous position genotyped as homozygous in the replicate) was
308 always ≤1.5%.

309 Observed heterozygosity of individuals (Ho) ranged from 0.109 (Western Alps region) to
310 0.169 (Atlantic region) (Table 2; Figure 2A), and was significantly lower in the populations of the
311 Western Alps region compared to other regions ($F_{4,27}$=7.01, $P < 0.01$, adjusted R^2=43.7%); population
312 diversity (He) ranged from 0.152 (in the population MTF, Western Alps region) up to 0.266 (in one
313 population from the Southern Alps region) (Table 2; Figure 2B), and was again significantly lower in
314 Western Alps region than other regions ($F_{4,27}$=13.22, $P < 0.01$, adjusted R^2=61%). Inbreeding
315 coefficients (*F*is) ranged between 0.291 and 0.442 and did not significantly differ between regions
316 ($F_{4,27}$=2.146, $P > 0.05$; Figure 2C).

317

318 ### *Population structure*

319 The STRUCTURE Bayesian assignment approach showed that *C. oedippus* populations are
320 genetically differentiated across Europe. The highest likelihood was for K=7 and ΔK was maximum
321 for K=3 (Figure 3 and Figure S2). At K=3, a primary separation was found between the following
322 three groups of populations, from West to East: (i) all populations in the Atlantic region; (ii) all
323 populations in the Western Alps region, with the possible exception of MTF (Isère valley); (iii) all
324 populations in the remaining regions, i.e. Northern Alps, Southern Alps and East European region.
325 The population MTF remained uncertainly assigned to one or the other of the two latter groups. At
326 K=4 all populations from the Southern Alps region formed a distinct group with the exception of the
327 easternmost one (LB, Ljubljansko barje, Slovenia), which remained uncertainly assigned. At K=5 the
328 population from the Isère valley (MTF) was separated from all the others. At K=6 the populations
329 from the Northern Alps region separated from those in the East European region. At K=7, some
330 evidence of admixture was retrieved between the latter groups. At K=8, the two populations from SW
331 Slovenia (CD, COE), which are the only sampled populations from dry ecotype, separated clearly
332 from the remaining populations of the Southern Alps region, which showed some differentiation
333 between a western group and an eastern group.

334 Pairwise *F*st ranged from 0 to 0.36 (Table S3). At the entire European scale, there was no
335 significant correlation between geographical and genetic distances (P=0.42) of sampled populations.
336 Genetic differentiation between populations separated by mountain ranges higher than 1100 m was
337 significantly higher ($P < 0.01$) compared to other populations. At the regional scale, a strong and
338 significant pattern of isolation by distance (ibd) was found across the populations from the Western
339 Alps region (R^2=0.98). Moderate and significant ibd was found across the populations within Atlantic
340 region (R^2=0.55), and low but significant ibd was found across those from the Southern Alps region
341 (R^2=0.11). Instead, no ibd was observed across populations in the East European region (Figure S3).

342 In the maximum likelihood tree based on the nuclear dataset (Figure S4), the two replicates for
343 each replicated individual (n=6) grouped together with 100% bootstrap support (BS), whereas
344 individuals from a single locality grouped together only for some populations, especially those from
345 the Northern Alps region and East European region. Relationships between populations were overall
346 poorly supported, with only a few well supported groups, from West to East: (i) all populations of the
347 Atlantic region (99% BS), within which the population from the Pyrenees was well separated (Co;
348 100% BS); (ii) all populations of the Western Alps region (87% BS), with two subgroups, i.e. the
349 population from Isère valley (MTF; 100% BS) and all others from Rhône valley (100% BS); (iii) the
350 two populations from Liechtenstein (RUG, SCH; 99% BS); (iv) both populations of dry ecotype from
351 the Southern Alps region (CD, COE; 100% BS); (v) two populations in the East European region
352 (KAM, UHO; 80% BS).
353

354 *Mitochondrial diversity*

355 A total of 17 haplotypes were found for the CO1 fragment sequenced from 43 individuals: 38
356 individuals in our sample (Table 1) and 5 specimens from BOLD database (Figure 4A). The most
357 common haplotype (Hap_1) was shared by 16 individuals out of a total of 43 and was found across
358 central and eastern Europe but not in the Atlantic populations. It differed by only one mutation from
359 the second-most represented haplotype (Hap_3), which was found in the Atlantic and in the Southern
360 Alps regions. Most other haplotypes were very similar (1-5 mutations from either Hap_1 or Hap_3),
361 including the haplotype of a previously sequenced individual from European Russia (Hap_17), but
362 with the remarkable exception of the haplotype of the single individual sampled from central Asia
363 (Hap_16). The latter had 3.5–3.7% divergence from all the other haplotypes, while pairwise
364 divergence within Europe did not exceed 0.44% (Figure 4B). Many haplotypes were found in the
365 Southern Alps region (10 haplotypes for 16 sampled specimens) while only the most common
366 haplotype was found in the Western Alps region.

367 Mitochondrial nucleotide diversities (π and θ; Figure 4) were highest in the populations from
368 the Southern and the Northern Alps regions (>0.0035), moderate in Atlantic region and East European
369 region (0.0015–0.0020), and null in Western Alps region.

370 A phylogenetic analysis of the CO1 sequences (Figure S5) did not recover well-supported
371 relationships between different populations of *C. oedippus*.

372

373 *Historical demographic scenarios and population size inference*

374 The most likely scenario (Figure 5A) was a first divergence between the western groups of
375 populations (Atlantic region and Western Alps region) and the other European populations, around 66
376 kya (95% C.I.: 30-95 kya), followed by a much more recent divergence between the Atlantic group of
377 populations and the Western Alps group, around 6 kya (95% C.I.: 1-10 kya). These divergence events
378 were associated to moderate population size changes, and the inferred population sizes for the three
379 lineages ranged between 10^5 and 10^6 individuals at splitting times; a strong population decline was
380 observed in the Western Alps lineage during the last 1,000 years with estimated current median
381 effective population size only around 8,000 individuals (Figure 5A).

382 When focusing on the western populations (Atlantic and Western Alps regions), the scenario
383 involving population bottleneck was much more likely than a scenario involving only population
384 divergence without population size change (Figure 5B). An impressive decline was detected in all
385 populations during the last 2,000 years, with those in the Atlantic region declining from 10^7 to 10^3,
386 those in the Rhône valley from 10^6 to 10^3 individuals and the population of the Isère valley from 10^6 to
387 ~700 individuals. When focusing on the remaining European populations, the analyses were not able
388 to distinguish between alternative scenarios for the splitting and/or admixture among the populations
389 of the East European region, those from the Northern Alps region and those in the Southern Alps
390 region (results not shown).

391

392 *Species distribution models*

393 MaxEnt models showed excellent performance in describing present-day distribution in
394 central and western Europe (Figure 6A); the average AUC across the cross-validated runs was 0.93
395 (SD=0.028). Summer temperature and summer precipitation were the variables with the strongest
396 contribution to the model (35% and 29% respectively). Suitability in the mid-Holocene (6 kya) was
397 similar to the present-day situation, with broader highly suitable areas north and east of the present
398 distribution (Figure 6B). The situation was very different in the LGM (21 kya). In this period, the
399 model suggested three suitable areas, all limited to coastal regions. Two small suitable areas were at
400 opposite ends of the Pyreneean chain (Figure 6C). Furthermore, a broader suitable area was present in
401 the Italian peninsula and in the Adriatic region, partially fragmented along the East-West axis. Both in
402 the mid-Holocene and in the LGM, suitable areas showed very low clumping and within suitable areas
403 no climatic variable was outside the range of calibration conditions.

404

405 Discussion

406 *Biogeographical history of* C. oedippus

407 The analysis of nuclear and mitochondrial variation of *C. oedippus* specimens collected throughout
408 most of the European range of the species, together with species distribution modelling, suggests that
409 the ancestors of all current European populations survived the last Pleistocene glacial period in at least
410 two refugia, most probably separated by the Alps.

411 The absence of geographical structure in the variation of the mitochondrial CO1 marker, with
412 similar haplotypes present from Russia to W France, suggests rapid expansion of the species
413 throughout central Europe after the last glacial period. The current populations in the Southern Alps
414 region account for most of the mt haplotype diversity, but comparable diversity persists in the small
415 populations surviving in the Northern Alps region, suggesting that the two regions were
416 interconnected at the beginning of the current interglacial period, without strong population
417 bottlenecks but rather a continuous northwards expansion wave during warming. The star-like patterns
418 in the haplotype network suggest two distinct expansion events, presumably from Southern Alps
419 region for Hap_1 (with unique derived haplotypes in Southern Alps, Northern Alps and East European
420 regions) and from Atlantic region for Hap_3 (with some derived haplotypes in this region only).
421 Interestingly two allopatric centers of differentiation during the last glacial period (Atlantic-
422 Mediterranean and Adriatic-Mediterranean) are also the most likely origin for two other satyrine

423 butterflies i.e. *Maniola jurtina* (Schmitt, Röber & Seitz, 2005) and *Conenonympha arcania* (Besold *et*
424 *al.*, 2008).

425 The distinct, highly divergent haplotype found in Kazakhstan, suggests that there was at least
426 one other more eastern refugium for the species during the Pleistocene glaciations, but this refugium
427 did not contribute to the recolonization of Europe after the LGM. More samples from central and
428 eastern Europe (e.g., Austria, Hungary, Belarus, Ukraine) and from Asia would be necessary to
429 reconstruct *C. oedippus* postglacial biogeographical history throughout its whole distribution range.

430 In contrast with the lack of geographical structure observed for the mt marker, both the genetic
431 structure analysis and the coalescence ABC simulations based on a large ddRADseq SNP dataset
432 support three main genetic lineages in Europe. An eastern lineage (comprising the populations of Italy,
433 Slovenia, Liechtenstein, Germany and Poland) separated from a western lineage (France) around 66
434 kya (before the LGM), while – within the latter – the populations of the Atlantic region separated from
435 those in the Western Alps region after the LGM (~6 kya).

436 This demographic scenario is supported by species distribution modelling based on current
437 occurrence of the species. The MaxEnt result suggests that the species distribution is constrained
438 mainly by the annual mean temperature (with an optimum between 12 and 13°C) and summer
439 precipitations (with an optimum around 350 mm). Only four small southern areas were potentially
440 suitable for the species during the LGM, but with a rapid increase in suitable area with climate
441 warming.

442 During mid-Holocene warming, the climatically suitable area increased towards north,
443 allowing gene flow between populations in different French regions. The subsequent separation
444 between the populations of the Atlantic regions and those of the Western Alps region could be
445 determined by habitat loss due to forest expansion during the rapid warming that followed LGM.
446 Indeed, simulations of the potential land cover after LGM in Europe consistently suggest that
447 extensive forests occupied large areas of Europe, particularly north and west of the Alps (Strandberg *et*
448 *al.*, 2011).

449 The current estimated effective population size is far higher for the lineage distributed in
450 central and eastern Europe (10^6 individuals) and that surviving in the Atlantic region (10^5) compared
451 to the populations of the Western Alps (10^3). Furthermore, the strong decline observed in the latter
452 region is recent, with dramatic population decline estimated from 10^6 down to 10^3 during the last
453 centuries. This scenario based on nuclear markers is also supported by the highest mtDNA haplotype
454 diversity and divergence found in Southern Alps and Northern Alps regions, suggesting that different
455 haplotypes were randomly lost during/following fragmentation in the Atlantic and Western Alps
456 regions. Indeed, although in both latter regions the haplotypes found were common haplotypes in

457 Europe, none was shared between the two regions. The lack of mtDNA variability across all
458 populations from Western Alps region, where only one haplotype was found, supports a dramatic
459 population decline in this region.

460

461 *Contemporary gene flow across populations*

462 In contrast to the mitochondrial marker, the ddRADseq multilocus analysis allowed to differentiate the
463 samples according to their geographical location, with a clear E-W and N-S population genetic
464 differentiation. However, in accordance with the analysis of mtDNA haplotypes, we found little
465 support for highly diverging lineages in Europe: there were only few informative sites (i.e., differently
466 fixed nucleotides across populations), and the relationships between populations were overall poorly
467 resolved. Of the main genetic groups well supported both in phylogenetic (ML tree) and population
468 genetic (STRUCTURE) analyses, three correspond to subspecies previously described based on wing
469 coloration pattern variation, e.g. *aquitanica* Varin, 1952 in Atlantic (including Charente-Maritime,
470 Landes and Pyrenees), *rhodanica* Varin, 1964 in the Rhône valley and *herbuloti* Varin, 1952 in the
471 Isère valley. Further morphological analysis would be required to test whether these distinct genetic
472 groups can indeed be distinguished based on phenotypic traits.

473 At the European scale, the pairwise genetic differentiation (F_{ST}) was moderate (0.04–0.15 on
474 average) given the wide geographical range sampled. The weak genetic structure and isolation-by-
475 distance patterns observed within geographical regions suggest that populations were presumably
476 more connected in the recent past. Indeed, historical records from the beginning of the 20th century
477 suggest a much larger distribution throughout France, Switzerland and Germany. Low altitude
478 wetlands and oligotrophic grasslands are the habitats that suffered the most from intensive agriculture
479 development, land draining and urbanization since the early 20th century throughout Europe, especially
480 in western Europe (Levers *et al.*, 2016). For some other butterfly species it is also suggested that their
481 current genetic structure may be explained better with past than present distribution (Orsini *et al.*,
482 2008; Sielezniew *et al.*, 2012).

483 *Genetic erosion and drift*

484 The lowest genetic diversity was found in Rhône and Isère valleys (Western Alps region), with He <
485 0.20. The population MTF (Isère valley) was significantly less diversified than any other population,
486 while the highest diversity was observed in populations from Atlantic, Southern Alps and East
487 European regions, with He > 0.25. The low genetic diversity observed in some populations, especially
488 MTF (0.15), suggests allele loss through genetic drift in isolated population with low effective size.

489 In the Western Alps region, the population size has been estimated from a few hundred in
490 Isère valley (MTF) to several thousand individuals in Rhône valley (PCC) by capture-mark-recapture
491 (unpublished data). These direct estimates from the field fit well population size estimates from the
492 gene coalescence simulations, suggesting that our prior distributions and model selection through
493 ABC procedure are realistic. The available habitat is several hundred ha in Rhône valley, while it is
494 restricted to 6 ha of protected area in Isère valley. Despite the nearest populations being about 60 km
495 apart, the high F_{ST} values (around 0.33) between this population and the neighboring ones suggests
496 that gene flow has been interrupted since a long time between MTF and other Western Alps
497 populations (LV, PCC, CNC, CSC). The latter populations have similar levels of genetic diversity
498 (around 0.20, all 95% CIs overlapping): LV is a protected site, only a few ha in size, but
499 geographically close (~5 km) to a larger habitat in Savoie (several hundred ha, 3 populations sampled:
500 PCC, CNC, CSC). Genetic differentiation between these two areas is low, suggesting that ongoing
501 gene flow has likely helped to maintain a relatively high genetic diversity, which could reflect the
502 legacy of formerly large and interconnected populations.

503 In comparison to Western Alps region, the populations of Atlantic region are much more
504 diversified and connected, with pairwise Fst usually not exceeding 0.10, except for the southernmost
505 population from the Pyrenees (Co), which is more than 100 km from the closest sampled population
506 (Table S3), and is also the less diverse population within the region (He=0.217, Table 2). This
507 suggests that populations in Atlantic region are still genetically connected or were connected in the
508 recent past, in accordance with a large climatically suitable area in this region.

509 The same pattern of isolation by distance is observed among populations throughout northern
510 Italy to central Slovenia (Southern Alps region). However, the population LB from central Slovenia,
511 was found admixed with populations of both East European and Southern Alps regions, but not with
512 the nearby populations CD and COE, which formed a distinct genetic cluster (Figure 3). While LB
513 inhabits wet grasslands, CD and COE live in a distinct karstic habitat, sub-mediterranean dry
514 grasslands in different successional stages up to light woods (Čelik & Verovnik, 2010), which are
515 drier and from phytosociological aspect different from the typical wet grasslands where *C. oedippus* is
516 mostly found in Europe. The distinctiveness of this habitat might have limited gene flow and / or
517 promoted local adaptations, but a larger sampling (both in terms of individuals from the two contrasted
518 habitats and of SNPs across the genome) would be necessary to test these hypotheses.

519 Our results also show that the fragmentation of *C. oedippus* populations in France started far
520 before intensive agriculture and urbanization, as the split between Atlantic and Western Alps lineages
521 was estimated at ~6 kya. This fragmentation into two lineages in France is unlikely to be only due to
522 the ecological barrier of the Massif Central, although we found that mountain ranges are relevant
523 barriers to gene flow in this species. In Europe, human populations strongly expanded as early as 11

kya, in link with the Neolithic agricultural revolution that sustained substantial population growth (Barker, 2009), and the agro-ecosystems developed by Gallic people and during the Middle Age might already had a negative impact on natural grassland ecosystems. On another hand, natural re-forestation could also explain the decline of this open-land butterfly. By maintaining semi-natural open-habitats human activities might mitigate its negative impact on *C. oedippus* in terms of land monopolization for agriculture and urbanization. Most of the current European populations of the species are found in protected areas that are managed in order to maintain the environment open. In addition to openness of the habitat, oligotrophic soil favoring grasses and sedges (i.e. larval hostplants with erect leaf orientation) over other herbs (with plane leaf orientation), appear to be key factor for pre-adult stages (Čelik *et al.*, 2015). It creates microhabitats with herb vegetation structure providing suitable microclimatic conditions and micro-spatial connectivity between hostplants.

Therefore, extension of forests after LGM is probably a more realistic factor than early agriculture to explain the fragmentation of *C. oedippus* in the Neolithic. In Poland and Belarus the butterfly is restricted almost exclusively to some fen communities (Sielezniew, 2012; Kulak & Yakovlev, 2018) which could be relatively stable open ecosystems before recent human-induced drainage and eutrophication (Jabłońska *et al.*, 2014). At the moment they have to be managed to prevent ecological succession and one population went extinct before that need was realized (Sielezniew *et al.* 2010). However, many historical population extinctions were recorded in France (Lhonore & Lagarde, 1999), Switzerland (Dušej *et al.*, 2010), Germany (Bräu *et al.*, 2010), Italy (Bonelli *et al.*, 2010; Bonato *et al.*, 2014), Slovenia (Čelik *et al.*, 2015), Slovakia (Pastoralis & Reiprich, 1995), and Bulgaria (Staub & Aistleitner, 2006) during the last century, indicating that population decline of this butterfly species is ongoing nowadays.

Conclusion: Despite a highly fragmented distribution in Europe, populations of *C. oedippus* still conserve a high level of genetic diversity, except in few locations (e.g., MTF) where there is evidence for genetic erosion and lack of connectivity. This high genetic diversity appears to be a legacy from previously large and interconnected populations that expanded after the LGM from at least two distinct refugia probably located west and south of the Alps respectively.

Acknowledgments: We thank all the collectors and their respective institutions/administrations for giving permission to collect specimens of *C. oedippus* in protected areas for this study: in France, T. Le Moal, N. Déjean and V. Labourel (CEN Aquitaine), M. Holthoff (CEN Poitou-Charentes), P. Freydier (CEN Savoie), CEN Isère, and C. Guérin (Réserve Naturelle du Marais de Lavours, Ain); in Germany, M. Braü; in Liechtenstein, U. Hiermann and the Amt fuer Umwelt in Vaduz; in Italy, the Ministero dell'Ambiente e della Tutela del Territorio e del Mare, and more particularly D. Baratelli and G. Forni (Provincia de Varese), J. Guidici, R. Filipello (Parco La Mandria), P. Glerean (Sezione

559 Entomologica, Museo Friulano di Storia Naturale) and S. Bonelli (University of Torino); in Poland, K.
560 Pałka (Maria Curie-Skłodowska University, Lublin); G. Nève (University Marseille, France) for
561 giving access to preserved biological material; and P. Dupont (Museum National d'Histoire Naturelle,
562 France) for sharing information on past and present species distribution in France. This work was
563 supported by the Conseil Départemental de l'Isère (Pôle biodiversité).

564

565 Figure caption

566 Figure 1. Map of Europe with the 32 sampled localities (colored dots often overlapping), assigned to
567 five geographical regions: Atlantic (orange), Western Alps (dark purple), Southern Alps (blue),
568 Northern Alps (pink) and East European (green) regions (see Table 1 for details).

569 Figure 2. Boxplots for A) observed heterozygosity (Ho), B) expected heterozygosity (He), C) *F*is,
570 within localities sampled in each of the 5 European geographical regions: Atlantic, Western Alps,
571 Southern Alps, Northern Alps and East European regions.

572 Figure 3. Results of the Bayesian genetic clustering (STRUCTURE) based on 1,314 unlinked SNPs,
573 for the two most likely numbers of cluster K=3 and K=8 (see also Figure S2 where the probability of
574 assignment to a given cluster is indicated for each individual).

575 Figure 4. Mitochondrial variability (partial CO1 gene; 630 bp). Panel A: Minimum spanning network
576 for the 17 haplotypes found in 43 specimens of *C. oedippus*. Each color represents a different
577 geographical region, and the size of each pie represents the number of specimens sharing the same
578 haplotype. Panel B: Within-region genetic diversity expressed as π (pairwise nucleotide divergence), θ
579 and hdiv (haplotype diversity). N is the number of sequenced samples, and h the number of
580 haplotypes.

581 Figure 5. Results of the ABC demographic analysis. A) Analysis within Europe (n=96 individuals;
582 1,314 SNPs), for the splitting between the three main lineages, from West to East: Atlantic (N1),
583 Western Alps (N2), and remaining regions corresponding to 'East lineage' (N3). Only the best
584 scenario is shown (posterior probability 0.997; 95% C.I.: 0.995–0.999). B) Analysis within France
585 (n=38 individuals; 1,123 SNPs), for the splitting between the Atlantic region (N1), Rhône valley (N2)
586 and Isère valley (N3). Only the best scenario (with bottlenecks) is shown (posterior probability 0.99;
587 95% C.I.: 0.976–1.000). The posterior distribution (mean, median and 95% C.I.) for each parameter is
588 indicated. N: effective population size; t: time since splitting or since bottleneck.

589 Figure 6. Results of species distribution models: suitability for *C. oedippus* under A) present day; B)
590 mid-Holocene (6 kya); C) last glacial maximum (21 kya) climatic conditions. 0.0883 is the 10%
591 training presence threshold; 0.342 is the maximum test sensitivity plus specificity threshold; values >

592 0.5 indicate very high suitability, as 0.5 is the typical suitability of presence points used by MaxEnt for
593 calibration (Elith *et al.*, 2011).

594

Table 1. Sampled populations of *C. oedippus* and number of individuals employed for nuclear (ddRAD-Seq) and mitochondrial (CO1) sequencing (see also Figure 1).

Code	Locality	Country	Administrative region	Geographic region	N individuals	N ddRADSeq	N CO1	Collector	Institution
MEES	Mees	France	Pays Basque/Landes	Atlantic	3	3	1	R. Dupéré	CEN Aquitaine
Co	Ger	France	Pyrénées-Atlantiques	Atlantic	3	3	2	T. Le Moal	CEN Aquitaine
BBL	Bélin-Béliet	France	Gironde	Atlantic	3	3	1	N. Déjean	CEN Aquitaine
LOU	Louchats	France	Gironde	Atlantic	3	3	0	N. Déjean	CEN Aquitaine
PUY	Les Ardillasses	France	Vienne	Atlantic	3	3	1	M. Holthoff	CEN Poitou-Charente
HOL	Les Ragouillis	France	Vienne	Atlantic	2	2	0	M. Holthoff	CEN Poitou-Charente
ECH	Echourgnac	France	Dordogne	Atlantic	3	3	1	V. Labourel	CEN Aquitaine
PES	Le Périer	France	Dordogne	Atlantic	3	3	0	V. Labourel	CEN Aquitaine
LV	Lavours	France	Ain	Western Alps	3	3	3	C. Guérin	Réserve Naturelle du Marais de Lavours
CNC	Chindrieux Nord	France	Savoie	Western Alps	3	3	1	P. Freydier	CEN Savoie
CSC	Chindrieux Sud	France	Savoie	Western Alps	3	3	0	P. Freydier	CEN Savoie
PCC	Prés-Crottis	France	Savoie	Western Alps	3	3	1	P. Freydier	CEN Savoie
MTF	Montfort	France	Isère	Western Alps	3	3	3	L. Després	LECA
CSB	Caselette	Italy	Torino	Southern Alps	3	3	2	L. Després	LECA
LMD	Mandria	Italy	Torino	Southern Alps	3	3	1	L. Després	LECA
MAS	Massazza	Italy	Biella	Southern Alps	3	3	1	S. Bonelli	University of Torino
BIA	Biandronno	Italy	Varese	Southern Alps	5	5	1	G. Forni	Servizi Agricoltura e Foreste, Province of Varese
VIL	Villadosia	Italy	Varese	Southern Alps	4	4	2	D. Baratelli	Servizi Agricoltura e Foreste, Provincia di Varese
COR	Cornuda	Italy	Treviso	Southern Alps	3	3	2	F. Ficetola	LECA

TRBS	Castions di Strada	Italy	Udine	Southern Alps	3	3	0	P. Glerean	Sezione Entomologica, Museo Friulano di Storia naturale
COE	Opatje Selo	Slovenia	Nova Gorica	Southern Alps	3	3	2	T. Čelik	SRC SASA, Jovan Hadži Institute of Biology
CD	Gorjansko	Slovenia	Sežana	Southern Alps	3	3	2	T. Čelik	SRC SASA, Jovan Hadži Institute of Biology
LB	Ljubljansko barje	Slovenia	Ljubljana	Southern Alps	3	3	3	T. Čelik	SRC SASA, Jovan Hadži Institute of Biology
SCH	Schaan	Liechtenstein	Liechtenstein	Northern Alps	3	3	1	U. Hiermann	Amt fuer Umwelt, Vaduz
MUN	Munich	Germany	Oberbayern	Northern Alps	3	3	2	M. Braü	Bayerische Akademie für Naturschutz und Landschaftspflege
RUG	Ruggell	Liechtenstein	Liechtenstein	Northern Alps	3	3	1	U. Hiermann	Amt fuer Umwelt, Vaduz
SZO	Szorce	Poland	Podlasie	East European	3	3	0	M. Sielezniew	University of Bialystok
UHO	Uhowo	Poland	Podlasie	East European	3	2	0	M. Sielezniew	University of Bialystok
ZAW	Zawadowka	Poland	Lublin	East European	3	2	1	K. Palka	Maria Curie-Skłodowska University (UMCS), Lublin
KAM	Kamien	Poland	Lublin	East European	3	3	0	K. Palka	Maria Curie-Skłodowska University (UMCS), Lublin
ANT	Antoniowka	Poland	Lublin	East European	3	3	2	K. Palka	Maria Curie-Skłodowska University (UMCS), Lublin
SWA	Swaryczow	Poland	Lublin	East European	3	3	1	K. Palka	Maria Curie-Skłodowska University (UMCS), Lublin

Table 2: Genetic diversity indices per sampled locality, based on the ddRADSeq dataset.

Geographic region	Code	Ho	95% C.I. Ho	He	95% C.I. He	*F*is	95% C.I. *F*is
Atlantic	MEES	0.152	0.1377-0.1671	0.2448	0.2278-0.2617	0.3958	0.3277-0.4306
Atlantic	Co	0.1385	0.1244-0.1522	0.217	0.2005-0.2349	0.33	0.2999-0.4248
Atlantic	BBL	0.1512	0.1367-0.1661	0.2467	0.2298-0.2654	0.4171	0.3326-0.4471
Atlantic	LOU	0.1658	0.1512-0.1802	0.248	0.2310-0.2656	0.3663	0.2758-0.3887
Atlantic	PUY	0.1606	0.1475-0.1755	0.2573	0.2407-0.2739	0.4277	0.3236-0.4286
Atlantic	HOL	0.145	0.1289-0.1611	0.2336	0.2124-0.2559	0.425	0.3068-0.4500
Atlantic	ECH	0.1623	0.1463-0.1781	0.2437	0.2254-0.2629	0.3215	0.2715-0.3905
Atlantic	PES	0.1691	0.1549-0.1835	0.237	0.2201-0.2531	0.2983	0.2267-0.3426
Western Alps	LV	0.1183	0.1052-0.1321	0.2029	0.1838-0.2216	0.351	0.3482-0.4834
Western Alps	CNC	0.1408	0.1275-0.1542	0.2023	0.1853-0.2203	0.3081	0.2422-0.3645
Western Alps	CSC	0.1293	0.1160-0.1449	0.1924	0.1763-0.2095	0.2922	0.2612-0.3923
Western Alps	PCC	0.1335	0.1200-0.1470	0.195	0.1788-0.2113	0.3195	0.2541-0.3823
Western Alps	MTF	0.1094	0.0959-0.1228	0.1525	0.1380-0.1667	0.2915	0.2040-0.3544
Southern Alps	CSB	0.1538	0.1400-0.1678	0.2386	0.2198-0.2561	0.339	0.2937-0.4119
Southern Alps	LMD	0.1355	0.1220-0.1505	0.2325	0.2132-0.2504	0.4419	0.3599-0.4753
Southern Alps	MAS	0.1574	0.1443-0.1721	0.252	0.2349-0.2693	0.3701	0.3188-0.4289
Southern Alps	BIA	0.147	0.1342-0.1595	0.2274	0.2123-0.2432	0.338	0.2988-0.4021
Southern Alps	VIL	0.1626	0.1498-0.1758	0.266	0.2503-0.2826	0.3773	0.3376-0.4368
Southern Alps	COR	0.1314	0.1169-0.1462	0.2263	0.2069-0.2443	0.409	0.3545-0.4772
Southern Alps	TRBS	0.142	0.1189-0.1663	0.2473	0.2162-0.2788	0.4372	0.3332-0.5147
Southern Alps	COE	0.1572	0.1417-0.1726	0.2333	0.2152-0.2504	0.349	0.2657-0.3835
Southern Alps	CD	0.1525	0.1382-0.1673	0.2425	0.2241-0.2603	0.3747	0.3135-0.4317
Southern Alps	LB	0.1534	0.1394-0.1676	0.2251	0.2079-0.2419	0.33	0.2553-0.3782
Northern Alps	SCH	0.1488	0.1360-0.1633	0.2484	0.2307-0.2654	0.3707	0.3446-0.4553
Northern Alps	MUN	0.14	0.1264-0.1545	0.2282	0.2113-0.2461	0.4173	0.3280-0.4442
Northern Alps	RUG	0.1526	0.1381-0.1672	0.2266	0.2095-0.2428	0.3015	0.2636-0.3854
East European	SZO	0.1383	0.1222-0.1543	0.2157	0.1954-0.2353	0.4038	0.2857-0.4333
East European	UHO	0.1631	0.1463-0.1805	0.2404	0.2166-0.2647	0.2979	0.2367-0.4017
East European	ZAW	0.1651	0.1481-0.1823	0.235	0.2149-0.2562	0.3918	0.2214-0.3683
East European	KAM	0.1517	0.1377-0.1655	0.2374	0.2181-0.2539	0.3656	0.3050-0.4154
East European	ANT	0.1518	0.1358-0.1659	0.2536	0.2326-0.2737	0.406	0.3363-0.4579
East European	SWA	0.15	0.1360-0.1634	0.2315	0.2149-0.2487	0.3766	0.2911-0.4080

Supporting information

Figure S1: Variation of the proportion of polymorphic loci *de novo* reconstructed when increasing the number of mismatch M allowed between two reads to be merged. The threshold M=7 within an individual, and 9 between individuals, was chosen.

Figure S2: Results of the Bayesian genetic clustering (STRUCTURE) based on 1,314 unlinked SNPs, with number of clusters ranging from K=3 to K=8. For each K the mean log likelihood is indicated (10 replicates), and Evanno's ΔK is shown. Individuals (n=96) are represented by vertical bars and grouped by localities (n=32).

Figure S3: Relation between genetic (*F*st/(1 - *F*st)) and geographical distance (square root transformed) for the main groups of populations obtained with the STRUCTURE analysis, except for the three populations of the Northern Alps region (sample size too small for regression analysis).

Figure S4: Maximum likelihood tree based on 1,594 concatenated 100 bp ddRADseq fragments. Bootstrap values are shown next to the branches (100 replicates). The tree is drawn to scale, with branch lengths representing the number of substitutions per site, and rooted at midpoint as outgroup is unknown. The analysis involved 96 individuals and 6 replicates.

Figure S5: Maximum Likelihood tree under the General Time Reversible model allowing for invariable sites ([GTR+*I*], 62% sites) based on mitochondrial sequences (partial CO1). Bootstrap support is indicated on each node (500 replicates). Branch lengths at scale (number of substitutions per site). The analysis involved 51 sequences, including 43 sequences from *C. oedippus* and 8 outgroups.

Table S1: Results of the ddRADseq experiment on 104 samples (98 individuals, 3 intra-library replicates and 3 inter-libraries replicates); six samples with less than 100,000 reads were excluded from analysis (in italic). After quality filtering, the mean number of loci per individual was 7,552 and the mean coverage per locus was 60.

Table S2: Presence records of *C. oedippus* in Europe used for MaxEnt inferences

Table S3: Pairwise genetic distances (Weir-Cokerham *F*st, below diagonal) and geographical distances (in km, above diagonal) between sampled populations. Distances between populations from the same geographical region are indicated in italic.

References

Barker G. 2009. *The agricultural revolution in prehistory: why did foragers become farmers?* Oxford University Press on Demand.

Besold J, Schmitt T, Tammaru T, Cassel-Lundhagen A. 2008. Strong genetic impoverishment from the centre of distribution in southern Europe to peripheral Baltic and isolated Scandinavian populations of the pearly heath butterfly. *Journal of Biogeography* **35:** 2090-2101.

Bonato L, Uliana M, Beretta S. 2014. *Farfalle del Veneto : atlante distributivo.* Marsilio.

Bonelli S, Canterino S, Balletto E. 2010. Ecology of Coenonympha oedippus (Fabricius, 1787)(Lepidoptera: Nymphalidae) in Italy. *Oedippus* **26:** 25-30.

Bozano GC. 2012. *Guide to the Butterflies of the Palearctic Region. Satyrinae, part III*. Omnes Artes

Bräu M, Dolek M, Stettmer C. 2010. Habitat requirements, larval development and food preferences of the German population of the False Ringlet Coenonympha oedippus (Fabricius, 1787)(Lepidoptera: Nymphalidae)-research on the ecological needs to develop management tools. *Oedippus* **26:** 41-51.

Capblancq T, Despres L, Rioux D, Mavarez J. 2015. Hybridization promotes speciation in Coenonympha butterflies. *Molecular Ecology* **24:** 6209-6222.

Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA. 2013. Stacks: an analysis tool set for population genomics. *Molecular Ecology* **22:** 3124-3140.

Čelik T, Brau M, Bonelli S, Cerrato C, Vreš B, Balletto E, Stettmer C, Dolek M. 2015. Winter-green host-plants, litter quantity and vegetation structure are key determinants of habitat quality for Coenonympha oedippus in Europe. *Journal of Insect Conservation* **19:** 359-375.

Čelik T, Rebeušek F. 1996. *Atlas of threatened butterflies of Slovenia*. Slovensko entomološko društvo Štefana Michielija, Ljubljana, 102 p. (in Slovene with English Summary)

Čelik T, Verovnik R. 2010. Distribution, habitat preferences and population ecology of the False Ringlet Coenonympha oedippus (Fabricius, 1787)(Lepidoptera: Nymphalidae) in Slovenia. *Oedippus* **26:** 7-15.

Čelik T, Vreš B, Seliškar A. 2009. Determinants of within-patch microdistribution and movements of endangered butterfly Coenonympha oedippus (Fabricius, 1787)(Nymphalidae: Satyrinae). *Hacquetia* **8(2):** 115-128.

Cornuet J-M, Pudlo P, Veyssier J, Dehne-Garcia A, Gautier M, Leblois R, Marin J-M, Estoup A. 2014. DIYABC v2.0: a software to make approximate Bayesian computation inferences about population history using single nucleotide polymorphism, DNA sequence and microsatellite data. *Bioinformatics* **30:** 1187-1189.

Dincă V, Lukhtanov VA, Talavera G, Vila R. 2011. Unexpected layers of cryptic diversity in wood white Leptidea butterflies. *Nature communications* **2:** 324.

Doyle J, Doyle J. 1987. Genomic plant DNA preparation from fresh tissue-CTAB method. *Phytochem Bull* **19:** 11-15.

Dušej G, Wermeille E, Carron G, Ziegler H. 2010. Concerning the situation of the False Ringlet Coenonympha oedippus (Fabricius, 1787)(Lepidoptera: Nymphalidae) in Switzerland. *Oedippus* **26:** 38-40.

Earl DA. 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation genetics resources* **4:** 359-361.

Elith J, Graham CH, Anderson RP, Dudík M, Ferrier S, Guisan A, Hijmans RJ, Huettmann F, Leathwick JR, Lehmann A. 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography***:** 129-151.

Elith J, Kearney M, Phillips S. 2010. The art of modelling range-shifting species. *Methods in Ecology and Evolution* **1:** 330-342.

Elith J, Phillips SJ, Hastie T, Dudík M, Chee YE, Yates CJ. 2011. A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions* **17:** 43-57.

Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* **14:** 2611-2620.

Falush D, Stephens M, Pritchard JK. 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* **164:** 1567-1587.

GBIF. 2016. GBIF Occurrence Download https://doi.org/10.15468/dl.ywhpmz.

Grassi F, De Mattia F, Zecca G, Sala F, Labra M. 2008. Historical isolation and Quaternary range expansion of divergent lineages in wild grapevine. *Biological Journal of the Linnean Society* **95:** 611-619.

Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic acids symposium series: [London]: Information Retrieval Ltd., c1979-c2000., 95-98.

Hallmann CA, Sorg M, Jongejans E, Siepel H, Hofland N, Schwan H, Stenmans W, Muller A, Sumser H, Horren T, Goulson D, de Kroon H. 2017. More than 75 percent decline over 27 years in total flying insect biomass in protected areas. *Plos One* **12**.

Hebert PD, Penton EH, Burns JM, Janzen DH, Hallwachs W. 2004. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly Astraptes fulgerator. *Proceedings of the National Academy of Sciences of the United States of America* **101:** 14812-14817.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. 2005. Very high resolution interpolated climate surfaces for global land areas. *International journal of climatology* **25:** 1965-1978.

Jabłońska E, Falkowski T, Chormański J, Jarzombkowski F, Kłosowski S, Okruszko T, Pawlikowski P, Theuerkauf M, Wassen MJ, Kotowski W. 2014. Understanding the long term ecosystem stability of a fen mire by analyzing subsurface geology, eco-hydrology and nutrient stoichiometry–case study of the Rospuda Valley (NE Poland). *Wetlands* **34:** 815-828.

Janes JK, Miller JM, Dupuis JR, Malenfant RM, Gorrell JC, Cullingham CI, Andrew RL. 2017. The K= 2 conundrum. *Molecular Ecology*.

Keyghobadi N. 2007. The genetic implications of habitat fragmentation for animals. *Canadian Journal of Zoology* **85**: 1049–1064.

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

Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, Mayrose I. 2015. Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Molecular ecology resources* **15:** 1179-1191.

Kudrna O, Harpke A, Lux K, Pennerstorfer J, Schweiger O, Settele J, Wiemers M. 2011. *Distribution atlas of butterflies in Europe*. Gesellschaft für Schmetterlingsschutz eV.

Kühne G, Kosuch J, Hochkirch A, Schmitt T. 2017. Extra-Mediterranean glacial refugia in a Mediterranean faunal element: the phylogeography of the chalk-hill blue Polyommatus coridon (Lepidoptera, Lycaenidae). *Scientific Reports* **7:** 43533.

Kulak A, Yakovlev R. 2018. Peculiarities of biology and the current state of the populations of the False Ringlet Coenonympha oedippus (Fabricius, 1787) on the territory of Belarus. *Ukrainian Journal of Ecology* **8:** 342-349.

Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution* **33:** 1870-1874.

Leaché AD, Banbury BL, Felsenstein J, de Oca AN-M, Stamatakis A. 2015. Short tree, long tree, right tree, wrong tree: new acquisition bias corrections for inferring SNP phylogenies. *Systematic biology* **64:** 1032-1047.

Levers C, Butsic V, Verburg PH, Müller D, Kuemmerle T. 2016. Drivers of changes in agricultural intensity in Europe. *Land Use Policy* **58:** 380-393.

Lhonore J, Lagarde M. 1999. Biogeography, ecology and conservation of Coenonympha oedippus (Fab., 1787) (Lepidoptera : Nymphalidae : Satyrinae). *Annales De La Societe Entomologique De France* **35:** 299-307.

Louy D, Habel JC, Abadjiev S, Schmitt T. 2013. Genetic legacy from past panmixia: high genetic variability and low differentiation in disjunct populations of the Eastern Large Heath butterfly. *Biological Journal of the Linnean Society* **110:** 281-290.

Manel S, Williams HC, Ormerod SJ. 2001. Evaluating presence–absence models in ecology: the need to account for prevalence. *Journal of Applied Ecology* **38:** 921-931.

Nazareno AG, Bemmels JB, Dick CW, Lohmann LG. 2017. Minimum sample sizes for population genomics: An empirical study from an Amazonian plant species. *Molecular ecology resources* **17:** 1136-1147.

Nogués-Bravo D. 2009. Predicting the past distribution of species climatic niches. *Global Ecology and Biogeography* **18:** 521-531.

Orsini L, Corander J, Alasentie A, Hanski I. 2008. Genetic spatial structure in a butterfly metapopulation correlates better with past than present demographic structure. *Molecular Ecology* **17:** 2629-2642.

Pastoralis G, Reiprich A. 1995. Zoznam Motylov vyskytujucich sa na uzemi Slovenska. *Spišska Nova Ves, Komarno*.

Patricelli D, Sielezniew M, Ponikwicka-Tyszko D, Ratkiewicz M, Bonelli S, Barbero F, Witek M, Buś MM, Rutkowski R & Balletto E. 2013. Contrasting genetic structure of rear edge and continuous range populations of a parasitic butterfly infected by *Wolbachia*. *BMC Evolutionary Biology* **13**: 14.

Pearson RG, Raxworthy CJ, Nakamura M, Townsend Peterson A. 2007. Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. *Journal of Biogeography* **34:** 102-117.

Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE. 2012. Double Digest RADseq: An Inexpensive Method for De Novo SNP Discovery and Genotyping in Model and Non-Model Species. *Plos One* **7**.

Ritter S, Michalski SG, Settele J, Wiemers M, Fric ZF, Sielezniew M, Šašić M, Rozier Y, Durka W. 2013. Wolbachia infections mimic cryptic speciation in two parasitic butterfly species, Phengaris teleius and P. nausithous (Lepidoptera: Lycaenidae). *Plos One* **8:** e78107.

Schmitt T, Röber S, Seitz A. 2005. Is the last glaciation the only relevant event for the present genetic population structure of the Meadow Brown butterfly Maniola jurtina (Lepidoptera: Nymphalidae)? *Biological Journal of the Linnean Society* **85:** 419-431.

Schmitt T, Varga Z. 2012. Extra-Mediterranean refugia: The rule and not the exception? *Frontiers in Zoology* **9:** 22.

Sielezniew M. 2012. Strzępotek edypus Coenonympha oedippus. In: Makomaska-Juchiewicz M. BP, ed. *Monitoring gatunków zwierząt. Przewodnik metodyczny. Part II.* Warszawa. 258-273.

Sielezniew M, Palka K, Michalczuk W, Bystrowski C, Holowinski M, Czerwinski M. 2010. False Ringlet Coenonympha oedippus (FABRICIUS, 1787)(Lepidoptera: Nymphalidae) in Poland: state of knowledge and conservation prospects. *Oedippus* **26:** 20-24.

Sielezniew M, Rutkowski R, Ponikwicka-Tyszko D, Ratkiewicz M, Dziekańska I, Švitra G. 2012. Differences in genetic variability between two ecotypes of the endangered myrmecophilous butterfly Phengaris (= Maculinea) alcon–the setting of conservation priorities. *Insect Conservation and Diversity* **5:** 223-236.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30:** 1312-1313.

Staub R, Aistleitner U. 2006. Das Moor-Wiesenvögelschen (Coenonympha oedippus)—oder worauf es im grenzüberschreitenden Artenschutz ankommt. *Alpenrheintal—eine Region im Umbau. Analysen und Perspektiven der räumlichen Entwicklung, Verlag der Liechtensteinischen Akademischen Gesellschaft, Schaan*: 245-254.

Strandberg G, Brandefelt J, Kjellström E, Smith B. 2011. High-resolution regional simulation of last glacial maximum climate in Europe. *Tellus A* **63:** 107-125.

Taberlet P, Fumagalli L, Wust-Saucy AG, Cosson JF. 1998. Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology* **7:** 453-464.

van Halder I, Barbaro L, Corcket E, Jactel H. 2008. Importance of semi-natural habitats for the conservation of butterfly communities in landscapes dominated by pine plantations. *Biodiversity and Conservation* **17:** 1149-1169.

Van Swaay C, Cuttelod A, Collins S, Munguira ML, Maes D, Šašić M, Settele J, Verovnik R, Verstrael T, Warren M, Wiemers M, I. W. 2010. *Red data book of European Red List of Butterflies (Rhopalocera).* . Publications Office of the European Union: Luxembourg,.

Van Swaay C, Warren M. 1999. *Red data book of European butterflies (Rhopalocera)*. Council of Europe.

Varin G. 1964. Contribution à l'étude des Satyridae (Lepidoptères) Coenonympha oedippus Fabricius. Sous-espèce rhodanica Varin nova en Savoie. *Bulletin mensuel de la Société Linéenne de Lyon* **2:** 92-93.

Verovnik R, Rebeušek F, Jež M. 2012. Atlas of butterflies (Lepidoptera: Rhopalocera) of Slovenia. Center za kartografijo favne in flore, Miklavž na Dravskem polju.

Wahlberg N, Freitas AV. 2007. Colonization of and radiation in South America by butterflies in the subtribe Phyciodina (Lepidoptera: Nymphalidae). *Molecular phylogenetics and evolution* **44:** 1257-1272.

Warren DL, Seifert SN. 2011. Ecological niche modeling in Maxent: the importance of model complexity and the performance of model selection criteria. *Ecological Applications* **21:** 335-342.

Weir BS, Cockerham CC. 1984. Estimating F-statistics for the analysis of population structure. *Evolution* **38:** 1358-1370.