



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

Laurence Després, Clément Henniaux, Delphine Rioux, Thibaut Capblancq, Sara Zupan, Tatjana Čelik, Marcin Sielezniew, Lucio Bonato, Gentile Francesco Ficetola

► To cite this version:

Laurence Després, Clément Henniaux, Delphine Rioux, Thibaut Capblancq, Sara Zupan, et al.. Inferring the biogeography and demographic history of an endangered butterfly in Europe from multilocus markers. *Biological Journal of the Linnean Society*, 2019, 126 (1), <10.1093/biolinnean/bly160>. <hal-04849988>

HAL Id: hal-04849988

<https://hal.science/hal-04849988v1>

Submitted on 19 Dec 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

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

3 Laurence Després^{1,*}, Clément Henniaux¹, Delphine Rioux¹, Thibaut Capblancq¹, Sara Zupan², Tatjana
4 Čelik³, Marcin Sielezniew⁴, Lucio Bonato⁵, Gentile Francesco Ficetola^{1,6}

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

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

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

10 ⁴ Laboratory of Insect Evolutionary Biology and Ecology, Institute of Biology, University of
11 Białystok, 15-245 Białystok, Poland

12 ⁵ Department of Biology, Università degli Studi di Padova, Padova, Italy

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

14
15 Date: May, 2018

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

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

19 Running title: False ringlet population genomics in Europe

Abstract

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

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

Introduction

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

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

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

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

Samples collection

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

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

DNA extraction

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

ddRADseq library preparation and SNP calling

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

same error occurring independently at the same nucleotide position in the same read) are very unlikely. Reads with depth coverage < 5 were excluded from further analyses. Genotyping errors (locus and allelic dropout) were estimated by comparing 6 replicate pairs (3 inter-libraries and 3 intra-library replicate pairs).

The ~68 million DNA reads obtained were used to call SNP genotypes with the *STACKS* pipeline (Catchen *et al.*, 2013) as follows: the *process_radtags* function was first run to demultiplex the data and filter the reads on their quality. We removed reads with length < 100 nucleotides and cut all reads to this value, resulting in more than 92% of total reads retained. On average, we retained 89% of total reads by individual after removing reads of low quality or with uncalled bases (options -q, -c and -r). Individuals with $< 100,000$ reads were discarded ($n=2$, Tables 1 and S1). Only *SbfI* reads were retained for *de novo* assembly on each individual using a maximum of 7 mismatches to merge two stacks into a polymorphic locus (-M; *ustacks* function). This threshold was chosen after inspecting the effect of increased values of M on the proportion of polymorphic loci (Figure S1). Highly-repetitive stacks and over merged tags were dropped using the “Removal” (-r) and the “Deleveraging” (-d) options. A catalog of the loci from all the individuals was built, with a maximum of 9 mismatches for merging two individual loci (-n; *cstacks* function). Loci within each individual were searched against the catalog (*sstacks* function) and a SNP dataset was produced with the genotype of each individual for every polymorphic position (*populations* function).

Genetic diversity and genetic structure estimation

For genetic diversity indices and analysis of population structure, only SNPs present in more than 60% of the whole sampling were retained to avoid an excess of missing data. SNPs with a minimum allele frequency lower than 5% were removed from the data set and only one polymorphic site was kept for each RAD-tag ('write_random_snp' option) in order to analyze only unlinked polymorphisms.

Genetic diversity by individual (individual heterozygosity, or observed heterozygosity H_o), within each sampled locality (population genetic diversity, or expected heterozygosity H_e), and between all population pairs was assessed using hierfstat R package. F_{IS} and F_{ST} estimates were calculated according to Weir & Cockerham (1984). Confidence intervals (95%) for H_o , H_e and F_{IS} were assessed by 1,000 bootstraps across loci.

Clustering of individuals into homogenous genetic clusters ranging from $K=1$ to $K=32$ was tested using Structure 2.3.4 (Falush, Stephens & Pritchard, 2003). For each run, a burn-in period of 5,000 steps was followed by 20,000 iterations under the admixture model and the assumption of correlated allele frequencies among populations. For each K, 10 runs were performed. Estimated log probabilities ($\ln P(D)$) were averaged across runs and compared to determine the posterior probability of each K using Clumpak (Kopelman *et al.*, 2015). The best K was selected using the ΔK method (Evanno, Regnaut & Goudet, 2005) in Structure Harvester (Earl, 2012). As one single K value only

provides an incomplete picture of overall population structure, we explored the pattern of population structure within the main clusters detected (Janes *et al.*, 2017).

In order to assess the respective roles of geographic distance and orographic barriers in population genetic differentiation of this species, which is depending on low-altitude habitats (see Introduction), we performed a multiple linear regression on distance matrix (MRM) in package R ‘ecodist’ where genetic distance ($F_{ST}/(1 - F_{ST})$) was treated as a response matrix. The straight-line geographic distances (square-root transformed) and presence of mountain ranges higher than 1100 m a.s.l. were set as the explanatory matrices. The available data on historic and present distribution of the species in Europe showed that 700 m a.s.l. is the highest altitude for the most localities of the species (Verovnik *et al.* 2012; Bonato *et al.*, 2014). The exceptions are scarce localities on southern foothills of the Alps where species was found on semi-open dry grasslands at 750–1100 m a.s.l. (e.g. Čelik & Rebeušek, 1996).

For phylogenetic inference, we used every locus present in at least 59 individuals of the whole sampling (N=96), including invariant positions. Heterozygote positions were coded with IUPAC code. We used full sequences rather than just SNPs because it was shown to be preferable from the perspectives of branch length and topological accuracy (Leaché *et al.*, 2015). The maximum likelihood phylogenetic tree was generated using RAxML (Stamatakis, 2014) with 100 rapid bootstrap inferences following search for the best ML tree using the GTR + G model for rate heterogeneity.

COI sequencing and phylogeographic analysis

We sequenced also a mitochondrial marker for 38 individuals representative of 24 localities from all the main regions studied (Table 1). COI was amplified using the primer pairs LCO–HCO and Jerry–Pat (Wahlberg & Freitas, 2007) (PCR protocol: 95°C for 10 min; 35 cycles of 95°C for 30 s, 50°C for 30 s and 72°C for 30 s; followed by a final extension period of 72°C for 7 min) and sequenced by Genewiz Company, UK. The resulting chromatograms were visualized in the software BIOEDIT ver. 7.2.5 (Hall, 1999) and aligned using ClustalW and by eye. Five sequences of *C. oedippus* available in BOLD were added to the multiple alignment. They originate from the following localities: Ruggell, Liechtenstein (BOLD accession code: PHLAF624-11); Munich, Oberbayern (GenBank code: GU707147); Romano d’Ezzelino, Vicenza province, Italy (BOLD code: PHLSA390-11); Obluchye, European Russia (GenBank code: EU920755); Tavalzhanka, Kazakhstan (BOLD code: LWA191-06). Additionally, sequences from eight outgroup species were chosen based on the most recent phylogeny of Coenonymphina butterflies published (Kodandaramaiah & Wahlberg, 2009), including six *Coenonympha* species (*C. tullia*, *C. hero*, *C. glycerion*, *C. nolckeni*, *C. phryne* (previously under *Triphysa*), *C. myops* (previously under *Lyela*); Genbank codes EU920762, EU920750, EU920749, EU920754, EU920739, EU920741), and two species of strictly related genera (*Heteronympha merope* and *Mydosama terminus*; EU92073, DQ338765). The maximum likelihood

phylogenetic tree was generated after selecting for the best model of molecular evolution using Mega7 version 7.0.14 (Kumar, Stecher & Tamura, 2016).

Demographic scenarios and population size inferences

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

Species distribution models

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

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

Results

Nuclear genetic diversity

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

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

Population structure

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

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

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

Mitochondrial diversity

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

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

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

Historical demographic scenarios and population size inference

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

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

Species distribution models

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

Discussion

Biogeographical history of *C. oedippus*

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

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

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

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

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

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

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

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

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

Contemporary gene flow across populations

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

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

Genetic erosion and drift

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

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

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

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

Our results also show that the fragmentation of *C. oedippus* populations in France started far before intensive agriculture and urbanization, as the split between Atlantic and Western Alps lineages was estimated at ~6 kya. This fragmentation into two lineages in France is unlikely to be only due to the ecological barrier of the Massif Central, although we found that mountain ranges are relevant 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. Bräu; 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

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

Figure caption

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

Figure 2. Boxplots for A) observed heterozygosity (H_o), B) expected heterozygosity (H_e), C) F_{is} , within localities sampled in each of the 5 European geographical regions: Atlantic, Western Alps, Southern Alps, Northern Alps and East European regions.

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

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

Figure 5. Results of the ABC demographic analysis. A) Analysis within Europe ($n=96$ individuals; 1,314 SNPs), for the splitting between the three main lineages, from West to East: Atlantic (N_1), Western Alps (N_2), and remaining regions corresponding to 'East lineage' (N_3). Only the best scenario is shown (posterior probability 0.997; 95% C.I.: 0.995–0.999). B) Analysis within France ($n=38$ individuals; 1,123 SNPs), for the splitting between the Atlantic region (N_1), Rhône valley (N_2) and Isère valley (N_3). Only the best scenario (with bottlenecks) is shown (posterior probability 0.99; 95% C.I.: 0.976–1.000). The posterior distribution (mean, median and 95% C.I.) for each parameter is indicated. N : effective population size; t : time since splitting or since bottleneck.

Figure 6. Results of species distribution models: suitability for *C. oedippus* under A) present day; B) mid-Holocene (6 kya); C) last glacial maximum (21 kya) climatic conditions. 0.0883 is the 10% 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 Pérrier	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	Fis	95% C.I. Fis
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, Michalczyk 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.