Biological Journal of the Linnean Society, 2022, **137**, 359–373. With 5 figures.

Phylogeography reveals the origin of the two phenological forms of large blue, *Phengaris arion* (Lepidoptera: Lycaenidae)

JUDIT BERECZKI[1,*], MARCIN SIELEZNIEW[2], RUDI VEROVNIK[3], STOYAN BESHKOV[4], GENNADIJ KUZNETSOV[5], SIMONA BONELLI[6] and JÁNOS PÁL TÓTH[7]

[1]*Molecular Taxonomy Laboratory, Hungarian Natural History Museum, Budapest H-1083, Hungary*
[2]*Laboratory of Insect Evolutionary Biology and Ecology, Institute of Biology, University of Bialystok, Białystok 15-245, Poland*
[3]*Department of Biology, Biotechnical Faculty, University of Ljubljana, Ljubljana 1000, Slovenia*
[4]*National Museum of Natural History, Bulgarian Academy of Sciences, Sofia 1000, Bulgaria*
[5]*Independent researcher, Volgograd, Russia gen-mash@mail.ru, http://babochki-kavkaza.ru*
[6]*Department of Life Sciences and Systems Biology, University of Turin, Turin 10123, Italy*
[7]*Independent researcher, Dabas, Hungary acutiformis@yahoo.com*

Received 18 February 2022; revised 1 June 2022; accepted for publication 10 June 2022

Our main goal was to investigate the phylogeography of the butterfly *Phengaris arion* to reveal the evolutionary origin of its 'spring' and 'summer' forms. Molecular analyses based on highly variable microsatellites, together with *Wolbachia* screening, were carried out on 34 populations in Europe. We found three well-defined genetic lineages of different origins: the Apennine, the central and the eastern. The highly distinct Apennine lineage is limited by the Alps and evaluated as an Evolutionary Significant Unit (ESU). Therefore, the taxon name *ligurica*, described from the Ligurian coast (Italy), should not be applied to denote the 'summer form' of the Pannonian region. The central lineage is limited by the Carpathians and the most eastern ranges of the Alps, and lacks major range fluctuations related to glaciations, although there is evidence for extra-Mediterranean refugia in the Carpathian Basin. The eastern clade could have had refugia in central Asia. Our results exclude the potential allopatric origin of the 'spring' and 'summer' *arion*, and support the hypothesis that the existence of the two forms could be a result of local adaptation to the distinctive phenology of host plant flowering which is manifested in the genetic differences between them. *Wolbachia* infection has been ruled out as a driver of sympatric speciation in *P. arion*.

ADDITIONAL KEYWORDS: extra-Mediterranean refugia – genetic differentiation – microsatellites – 'spring' and 'summer' *arion* form – *Wolbachia*.

INTRODUCTION

The study of speciation is one of the most active areas of evolutionary biology (Turelli *et al.*, 2001). In past decades, substantial progress has been made in documenting and understanding species formation which has been greatly facilitated by the explosive development of molecular methods (Hewitt, 2001). Breakthroughs in DNA-based technology revolutionized evolutionary biology, and out of this revolution emerged a highly influential discipline known as phylogeography which concerns the geographical distribution of genealogical lineages (Avise, 1998). Knowledge on phylogeographical patterns is essential to understand the evolutionary history of species.

Geography is a widely recognized key factor in the process of speciation. The primary classification of speciation, into so-called geographic modes, is based on the pattern of geographic ranges observed among daughter species (Mayr, 1963; Bush, 1975; Templeton, 1981). Accordingly, there are two extremes, namely the allopatric and sympatric speciation models.

*Corresponding author. E-mail: bereczki.judit@gmail.com

Traditionally, allopatric speciation is regarded as a null model of speciation which only requires geographical disjunction and a long enough time of isolation (Futuyma & Mayer, 1980). Climatic oscillations, which had severe impacts on the distributions of many organisms, offered ample opportunity for such allopatric species formation (Avise, 2000; Hewitt, 2004; Stewart *et al.*, 2010). In temperate regions of Europe, species experienced contractions of ranges in glacial periods and expanded their distribution during inter- and postglacial cycles (Hewitt, 1996). Consequently, they survived the unfavourable phases in southern Mediterranean regions. Their disjunct glacial distribution patterns regularly resulted in the separation of different genetic lineages in the three major Mediterranean peninsulas (i.e. Iberian, Apennine and Balkan) combined with different basic patterns of postglacial expansion (Taberlet *et al.*, 1998; Hewitt, 1999, 2000; Habel *et al.*, 2005). In addition, a postglacial colonization route was postulated from Asia to Europe (Schmitt & Varga, 2012). The significance of refugial isolation in speciation is debatable and it also strongly depends on the individualistic responses of the species influenced by their life history traits and habitat preferences (Bhagwat & Willis, 2008; Stewart *et al.*, 2010). Based on different studies it seems that refugial isolation during one glacial cycle would often be insufficient for speciation to take place (Johnson *et al.*, 1996; Coope, 2004; Lister, 2004). However, as reviewed by Stewart *et al.* (2010), cryptic refugia could occasionally lead to rapid evolution, as populations fulfil several of the requirements for allopatric speciation (Mayr, 1954; Eldredge & Gould, 1972), particularly in the case of 'ecological' speciation under strong adaptive selection (Hendry *et al.*, 2007; Nosil *et al.*, 2009).

The opposite extreme, sympatric speciation, is divergence within a single geographical region such that the range of one nascent species completely overlaps with another (Fitzpatrick *et al.*, 2008). Sympatric speciation was once thought by many to be improbable, many examples and models have been published as several genetic analyses proposed as retrospective tests of sympatric speciation (Via, 2001; Berlocher & Feder, 2002, Foote, 2018; Richards *et al.*, 2019; Inskeep *et al.*, 2021). Reconstruction of the phylogenetic history of divergent taxa can provide an important line of empirical evidence for sympatric speciation, but within-species phylogeography might also be very useful. One key approach has been to evaluate whether phenotypically or ecologically divergent races found in sympatry are more closely related to one another than phenotypically equivalent races are in allopatry to each other.

Currently, substantial indications exist that sympatric speciation may have a significant role in the evolution of insects (Via, 2001). Host-associated biotypes, including host races in plant-feeding insects, have often been used as evidence of the initial stage of sympatric speciation representing the incipient stage of it (Bush, 1969; Tauber & Tauber, 1989; Feder, 1998; Filchak *et al.*, 2000). At the same time, certain intracellular bacteria may also act as speciation agents in insects. For example, members of the genus *Wolbachia* can generate reproductive isolation even within a single population of their insect host by cytoplasmic incompatibility between different bacterial strains (Hoffmann & Turelli, 1997; Werren *et al.*, 2008). Thereafter, the individuals harbouring a different type of *Wolbachia* infection potentially diverge genetically and evolve into new species (Breeuwer & Werren, 1993; Werren, 1997).

The existence of different host races, together with the possibility of sympatric speciation, has also emerged in the case of the obligatorily myrmecophilous butterfly *Phengaris arion* (Linnaeus, 1758) which has two phenological forms. The fast-flying, smaller-sized and dark violet-blue form (referred to as the 'spring *arion*' hereafter) usually flies from mid-May to mid-June and is linked to the early flowering host plant species of the genus *Thymus*. The slower, larger and light silvery blue form (referred to as the 'summer *arion*') is on the wing from the end of June to mid-August, and oviposits among flower buds of late-flowering *Thymus* species and/or *Origanum vulgare*. In addition, *Wolbachia* infestation has been already identified as a potential speciation trigger in *P. arion* (Bereczki *et al.*, 2011, 2014, 2015, 2020).

Although previous studies have found significant morphological differences between the two *arion* forms both in wing characteristics and genitalia, they could not reveal any genetic isolation between them based on either the investigated mitochondrial and nuclear gene regions or allozyme loci (Bereczki *et al.*, 2011, 2014, 2015). However, these authors raised the possibility that the markers analysed were not suitable for the detection of the divergence between those forms because of their low variability. Additionally, molecular studies based on the mitochondrial barcoding gene and the nuclear elongation factor 1α did not reveal any sign of the existence of the two *arion* forms in other European regions (Patricelli *et al.*, 2013). At the same time, significant differences were detected between the 'spring' and the 'summer' *arion* based on highly variable microsatellites (Bereczki *et al.*, 2020). Moreover, these markers have been effectively used to explore the population history of *P. arion* in Sweden, Denmark, Poland, Italy and the UK (Ugelvig *et al.*, 2011, 2012; Sielezniew & Rutkowski, 2012; Andersen *et al.*, 2014; Sielezniew *et al.*, 2015). Consequently, so far only microsatellites proved to be suitable for reconstructing the phylogeography of the target species and its forms.

Microsatellite based studies (Bereczki *et al.*, 2020) have revealed that the two phenological forms may meet all criteria of host plant races according to Drès & Mallet (2002). Therefore, they are good candidates as subjects of sympatric speciation. However, the authors also emphasized that spatial replicability should be tested on a larger geographical scale since the evolutionary processes in *P. arion* cannot be fully understood without more thorough knowledge on its phylogeography.

Here we investigate the phylogeography of *P. arion* based on highly variable microsatellites. Our main goals are (i) to reveal the existence and the origin of different genetic lineages, (ii) to explore whether the two phenological forms originate from separate refugia, that is whether they have an allopatric or sympatric origin, (iii) to test the possible role of *Wolbachia* in the evolution of *P. arion* forms on a large geographical scale, i.e. whether we can detect different bacterial strains in line with the genetic and/or phenotypic differentiation.

MATERIAL AND METHODS

Study species

Phengaris arion has a very special socially parasitic life cycle depending on the dual presence of specific initial host plant and host ant species. Females lay their eggs on flower buds of a specific initial host plant. Young larvae feed on developing seeds, quickly growing through three instars but gaining only a few percent of their final weight. After 2–3 weeks, larvae drop on the ground and wait for foraging *Myrmica* ant workers which adopt them. In the ant nest, larvae follow a 'predatory' strategy preying on an ant brood for 10–11 months, or continue development for an additional year (Thomas *et al.*, 1998; Schönrogge *et al.*, 2000). Relationships of *P. arion* with ants used to be considered as highly specific (Thomas *et al.*, 1989); however, more recent studies suggest multiple host ant use at least in some parts of the range (Tartally *et al.*, 2019). The average life span of the imagos is only few days (Nowicki *et al.*, 2005; Osváth-Ferencz *et al.*, 2017).

Phengaris arion has a Palaearctic distribution from France and Spain to China; however, it faces a serious conservation risk as in past decades its habitats have suffered a severe decrease and fragmentation especially in Europe. It became extinct in the Netherlands in 1964 (Tax, 1989), in the UK in 1979 (Thomas, 1995) and in Belgium in 1996 (Goffart, 1997). However, the species was successfully re-introduced into the UK (Thomas *et al.*, 2009) and it also recolonized Belgium (Goffart, 1997). Otherwise, it shows a serious retreat all over Europe, especially at the northern border of its distribution (Wynhoff, 1998). Therefore, *P. arion* is considered endangered on the European scale. The species is included in Annex IV of the European Habitats' Directive, and is listed in the IUCN Red List of Threatened Species as 'near threatened' and considered as 'endangered' in the European Red List of Butterflies (Munguira & Martin, 1997; Van Swaay *et al.*, 1998, 2010). Together with the other European *Phengaris* species, they are among the few insects for which specific conservation actions have been undertaken, and are regarded as 'flagship' species by many conservationists (Thomas, 1995).

Sampling

Altogether 313 specimens were analysed from 34 populations in seven European countries (Supporting Information, Table S1) of which ten were 'spring *arion*', 20 were 'summer *arion*' and four were of uncertain classification. The identification was carried out based on the collection time—considering the altitude and the latitude of the sample site—and the host plant. Populations exploiting *O. vulgare* were referred as 'summer *arion*' in every case while those using *Thymus* species could not be classified without knowledge on the local population dynamics since certain *Thymus* species flower at the same time as *O. vulgare*. Four syntopic sample pairs of 'spring *arion*' and 'summer *arion*' were available from north-eastern Hungary (labelled as KORa-KORb, ZABa-ZABb, SUSa-SUSb) and Russia (OSTa-OSTb, more details in Supporting Information, Table S1). Imagos were caught with a butterfly net and stored as dried material until molecular analyses.

Microsatellite studies

DNA was extracted from different kind of tissues (see Supporting Information, Table S1) following the protocol in Bereczki *et al.* (2014). Microsatellite polymorphism was studied at 12 loci, namely *Macu8*, *Macu11*, *Macu15*, *Macu44*, *Macu45*, *Macari02*, *Macari05*, *Macari08*, *Macari16*, *Macari19*, *Macari22* and *Macari23* characterized by Zeisset *et al.* (2005) and Ugelvig *et al.* (2011, 2012). During amplification we used fluorescent dye-labelled primers described by the authors mentioned above, and PCR reagents and conditions described in Rácz *et al.* (2015). After amplification, microsatellite products were multiplexed in three reactions (Multiplex 1 with the loci *Macu8*, *11*, *45*, *Macari02*, *16*, Multiplex 2 with *Macu44*, *Macari05*, *19*, *23* and Multiplex 3 with Macu15, Macari08, 22) and fragment analysis was carried out on an ABI 3130 Genetic Analyser in the Molecular Taxonomy Laboratory of the Hungarian Natural History

Museum (Budapest, Hungary). Allele sizes were estimated using Peak Scanner software (Thermo Fisher Scientific, Waltham, MA, USA). Micro-Checker 2.2.3 (Van Oosterhout *et al.*, 2004) was used for calculating null allele frequency by Monte Carlo simulation of expected homozygote frequencies and heterozygote allele size differences.

The genetic structure of the populations was analysed by two different methods

Firstly, the Geneland 4.9.2 package (Guillot *et al.*, 2005) was used to detect spatial discontinuities among populations based on geo-referenced multilocus genotypes. An uncorrelated allele frequency model was used to estimate the most probable number of clusters (K). Ten replicates were carried out to verify the consistency of the most probable K value, which was allowed to vary from 1 to 10. MCMC iterations were set to 1 000 000 and a thinning of 1000. We discarded the first 20% of iterations as 'burn-in'. The best analysis based on the 'mean logarithm of posterior probability' was visualized using QGIS 3.1 (QGIS, 2022).

Secondly, Structure 2.3.4 using a Bayesian-clustering algorithm (Pritchard *et al.*, 2000) was run to estimate the most probable number of genetically differentiated groups (K) in our populations and to assign the individuals to these groups without geographic information. These analyses were carried out with an initial burn-in of 100 000 and a running length of 500 000. In the evaluation of the results ΔK was computed which indicates the change in log probability between successive K values (Evanno *et al.*, 2005). Structure Harvester Web 0.6.93 (Earl & von Holdt, 2012) was used to compute the ΔK values. The package 'pophelper' in R (Francis, 2017) was applied to average the ten runs of the most probable K value given by Structure and correct for label switching. Finally, we constructed geo-referenced pie charts based on the most probable assignments using QGIS 3.1 (QGIS, 2022).

Non-metric multidimensional scaling (nMDS) was used to represent genetic relationships among *P. arion* populations based on Cavalli-Sforza's chord distance (Cavalli-Sforza & Edwards, 1967) calculated from allele frequency data using PAST 4.05 (Hammer *et al.*, 2001).

Wolbachia studies

The same DNA extracts were used for *Wolbachia* screening as were used in microsatellite studies. Each specimen was screened by the amplification of the highly conservative 16S ribosomal RNA gene with the *Wolbachia* specific W-Spec primers of Werren & Windsor (2000). The amplification procedure described in Rácz *et al.* (2015) was followed. We used positive (confirmed infected samples) and negative controls (master mix without any DNA sample) in each reaction. The success of the PCRs, i.e. *Wolbachia* presence, was checked by running 2 µL of product on 1% agarose gels stained with GelRed Nucleic Acid Stain (Biotium Inc., Fremont, CA, USA).

Wolbachia strain identification was carried out by the amplification of *Wolbachia* surface protein (WSP) following the PCR protocol in Bereczki *et al.* (2015). After sequencing, we defined the strains in the *Wolbachia* Multilocus Sequence Typing (MLST) database (http://pubmlst.org/wolbachia/).

Species distribution modelling

We used MaxEnt 3.4.4 (Phillips *et al.*, 2006) to predict the potential distribution of *P. arion* using BIOCLIM (Busby, 1991) and ENVIREM (Title & Bemmels, 2018) variables. MaxEnt is a widely used method for predicting species distributions using presence-only data (Phillips *et al.*, 2004; Warren & Seifert, 2011). MaxEnt's predictive performance is consistently competitive with the highest performing methods (Elith *et al.*, 2011).

Presence data for *P. arion* were used from the authors' own database, occurrence data from Filz & Schmitt (2015) and GBIF. Data from GBIF were filtered based on accuracy and only those with coordinate uncertainty of less than 1 km were used. Since presence locations were highly biased, we spatially thinned the data set using the 'spThin' package (Aiello-Lammens *et al.*, 2015) in R. Spatial thinning helps to reduce the effect of uneven, or biased, species occurrence collections on spatial model outcomes. The MaxEnt runs were performed with 100 presence points (Supporting Information, Fig. S1).

The climate variables were downloaded from the WorldClim (www.worldclim.com) and Environmental Rasters for Ecological Modeling (envirem.github.io) databases. The variables for the Present are the average for the years 1970–2000 while those for the Last Glacial Maximum (LGM) are ~21 000 years before present. Although MaxEnt is more robust in controlling for correlations between variables than stepwise regression (Elith *et al.*, 2011), strongly correlated variables ($r > 0.75$) are recommended to be excluded from the analysis (see Elith *et al.*, 2010; Stohlgren *et al.*, 2010).

To identify the most important set of uncorrelated variables and to fine-tune Maxent's regularization multiplier the 'MaxentVariableSelection' package (Jueterbock *et al.*, 2016) was used in R. A jack-knife test was applied using MaxEnt and results of 'with only variable' measurements were also considered during variable selection.

The discrimination ability of the model was evaluated by the area under the curve (AUC) metric. The value of AUC varies between 0.0 and 1.0, where 1.0 is considered a perfect prediction and 0.5 or less is considered no better than random (Fielding & Bell, 1997; Franklin & Miller, 2009).

The distribution model was projected back to the LGM, that is ~21 000 years before present. For the projections we used the predictions of two different global circulation models (CCSM4 and MIROC-ESM). The results were visualized on a binary presence (1) absence (0) raster using the ten-percentile training presence threshold rule. To evaluate the area dynamics of the studied species, we used these binary rasters for current climate and LGM scenarios. The presence values for the LGM have been changed from 1 to 2 followed by grid overlaying which resulted in four possible values for each cell: (1) where the species potentially occurred during the LGM but currently does not; (2) areas where the species does not occur: areas that are neither suitable under current conditions nor under LGM conditions; (3) areas where the species could potentially occur in both Present and LGM climates; and (4) areas where the species potentially occurs currently, but which were not suitable during the LGM [for more details on the methodology see Scheldeman & van Zonneveld, (2010)].

RESULTS

Microsatellite studies

The Micro-Checker analysis did not detect systematic evidence for null alleles at any of the studied microsatellite loci, thus the whole data set was used for the further analyses.

Geneland analysis identified three genetic lineages (Fig. 1). The so-called Apennine clade included populations from Italy and western Slovenia, the central lineage comprised samples from eastern Slovenia through the Carpathian Basin to southern Poland and western Romania, and the eastern group contained the majority of Polish populations and the samples from eastern Romania and Russia.

This tripartite division of the pattern was confirmed by the Structure analysis (see K = 3 in Fig. 2). However, as indicated by ΔK (Supporting Information, Fig. S2), the genetic pattern is further subdivided (see K = 8 in Fig. 2).

The isolation of the samples from Italy and western Slovenia can be also clearly seen from the Structure analysis. A separate analysis of this Apennine lineage revealed that the population from Cuneo (CUN, Italy) is highly distinct from the other Italian populations together with the analysed western Slovenian sample (see K = 2 in Fig. 3). At the same time, Structure analysis assigned the vast majority of the individuals according to their geographic origin at K = 5 which denotes the genetic uniqueness of each studied population within the Apennine clade (see K = 5 in Fig. 3).

Another very distinct genetic entity, the so-called eastern clade, originated from Russia from where westward gene flow took place toward Poland and eastern Romania. The third (central) genetic group involved the rest of the samples including specimens mostly from Bulgaria, Slovenia, Hungary and Romania (Fig. 1B). Interestingly, one of the south Polish populations (SRO) clustered with these samples as well. Populations from the Carpathian Basin and the Balkans shared common clusters. It is notable that the population from Kaszonyi Hill (Hungary) greatly differed from all other samples forming a separate genetic cluster (indicated in light blue in Fig. 2) from where some genetic material filtered into the surrounding populations.

The genetic distinctness of 'spring *arion*' and 'summer *arion*' originating from the same locality in Hungary was significant (SUSa-SUSb: $F_{ST} = 0.021$, $P = 0.013$; KORa-KORb: $F_{ST} = 0.065$, $P = 0.0001$; ZABa-ZABb: $F_{ST} = 0.042$, $P = 0.0001$). However, syntopic samples from Russia did not differ from each other (OSTa-OSTb: $F_{ST} = 0.069$, $P = 0.056$). Although we did not have the opportunity to compare syntopic sample pairs in Poland, the two *arion* forms differed remarkably here, more than would be justified by geographical distance (Fig. 2).

Non-metric multidimensional scaling (nMDS) (Fig. 4) also confirmed the distinctness of the Apennine clade which was composed of Italian and western Slovenian samples. At the same time, the central and eastern groups were in close proximity to each other. Only the populations from Russia displayed a higher level of differentiation from the rest of the samples.

Wolbachia studies

Altogether the prevalence of *Wolbachia* was 97.8%. However, *Wolbachia*-free specimens occurred only in samples from Russia (see Supporting Information, Table S1) where the infection level was 53.3%. All individuals on all the other sites were infested irrespective of phenology, differential host plant usage or the geographic origin.

WSP allele no. 685, which has been previously described from *P. arion*, was first identified in the Balkan Peninsula. Unfortunately, we could not obtain any *Wolbachia* sequence from Russia.

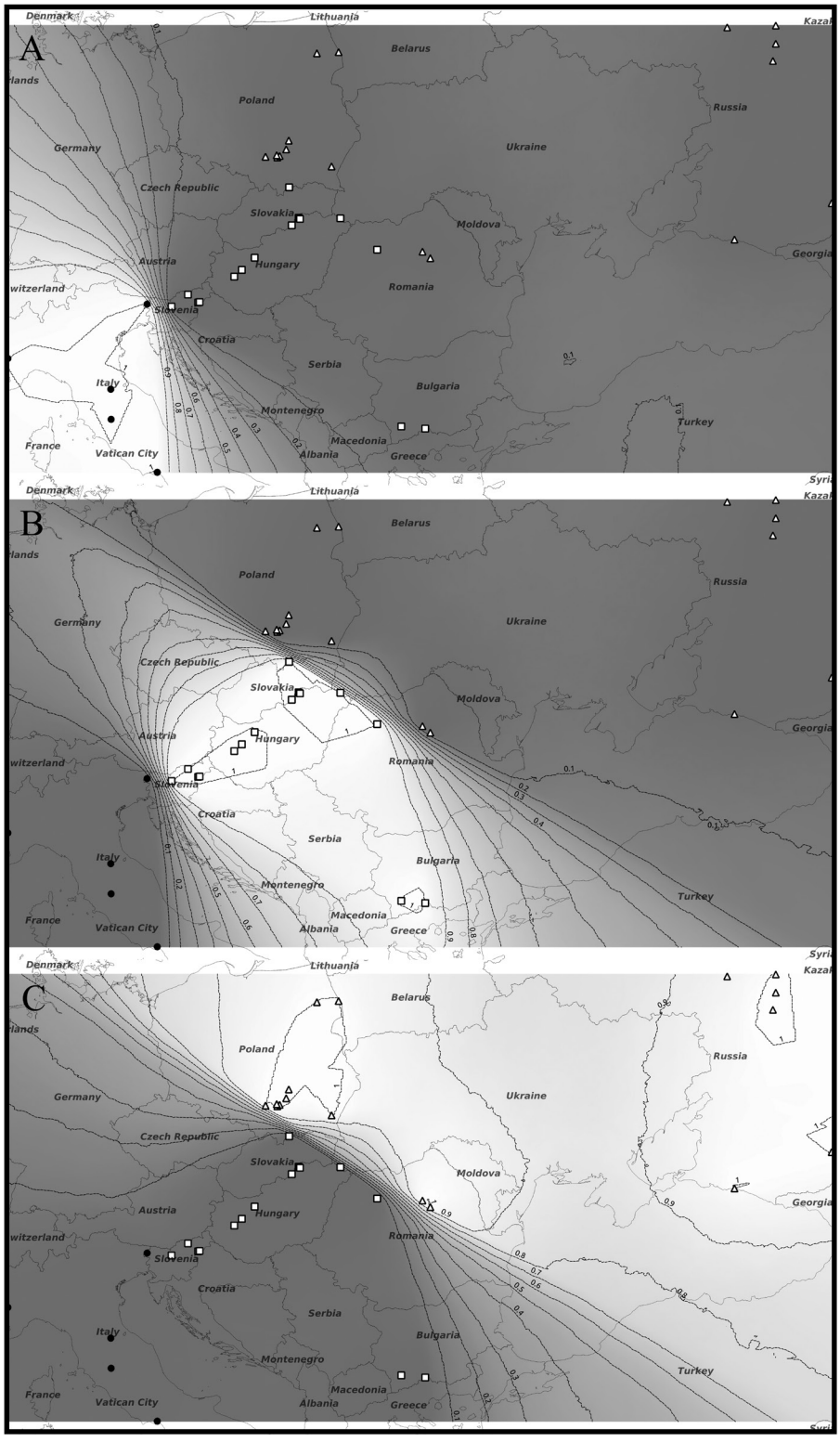

Figure 1. Geographical locations of *P. arion* populations sampled. Different colours indicate posterior probability of belonging to subsamples 1–3 detected in the Geneland analysis (colours are arbitrary to differentiate between population groupings). A, Apennine lineage is indicated by circles. B, central (Carpathian-Balkan) lineage is indicated by squares. C, eastern lineage is indicated by triangles.

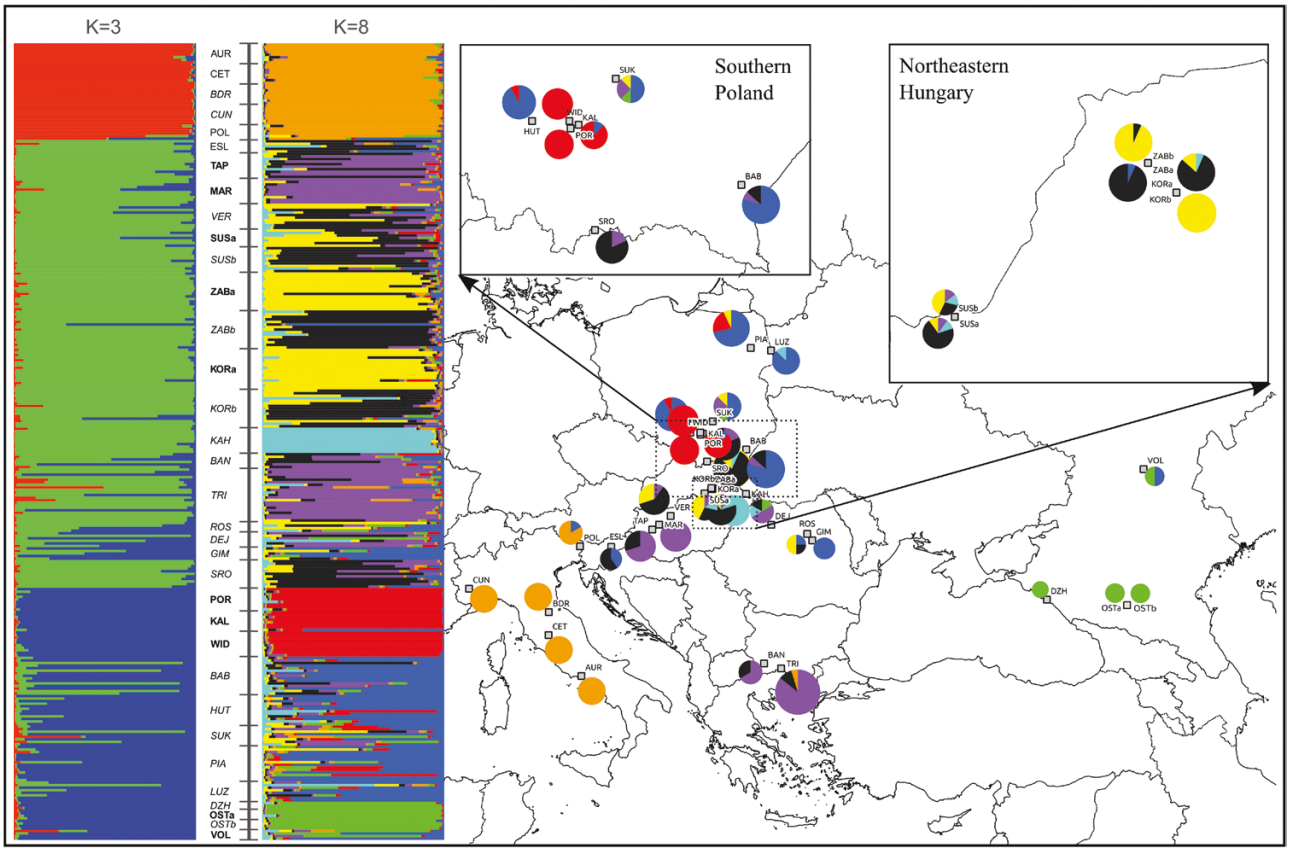

Figure 2. Bayesian assignment of individuals based on 12 microsatellite loci. Pie charts for each site represent the proportion of individuals assigned to each of the eight clusters. Areas of circles are proportional to the number of individuals analysed. On bar plots 'spring *arion*' samples are indicated in bold, 'summer *arion*' are in italics and samples with uncertain origin indicated are in normal letters. For full site names and other details, see Supporting Information, Table S1.

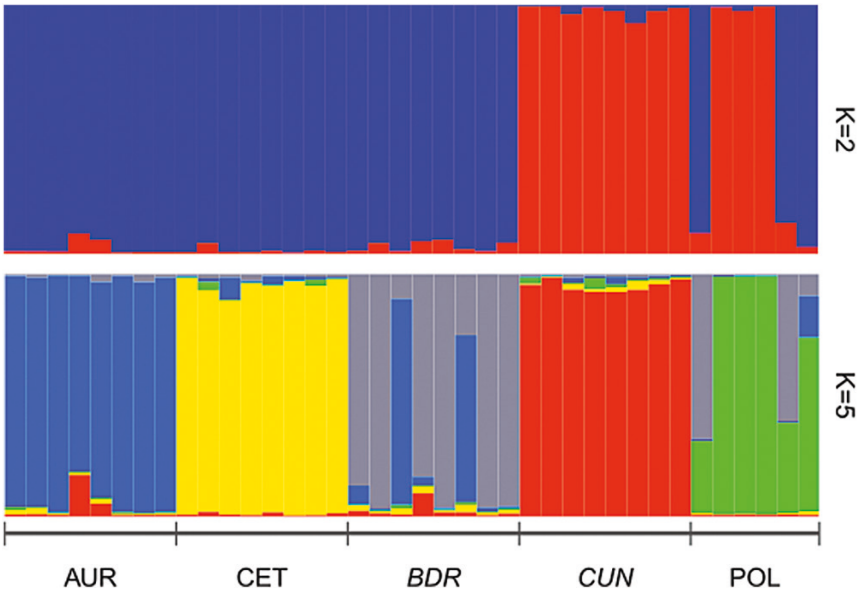

Figure 3. Bayesian assignment of individuals in the subsample of the Appenine genetic lineage. 'Summer *arion*' samples are in italics and samples with uncertain origin indicated with normal letters. For full site names and other details, see Supporting Information, Table S1.

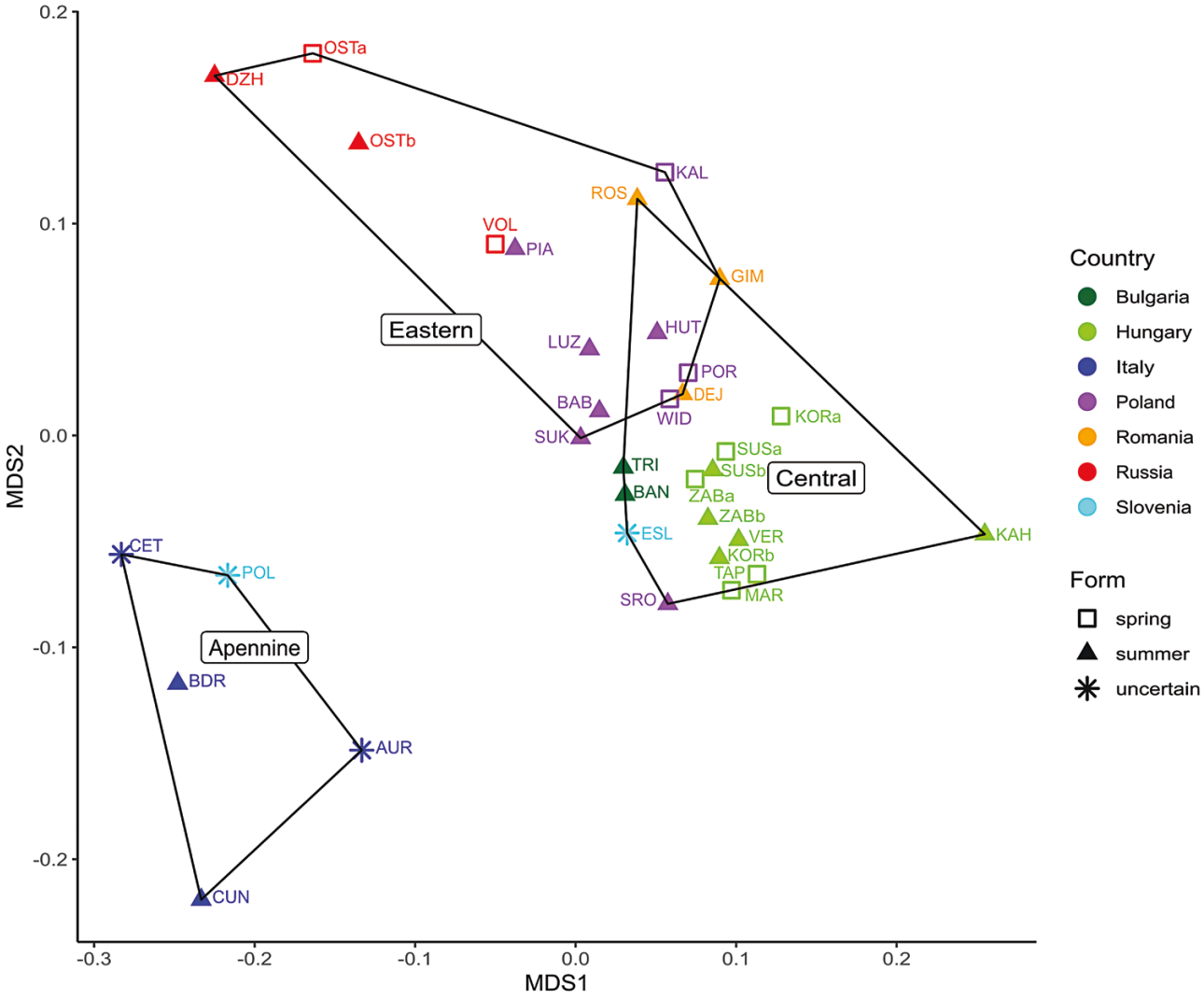

Figure 4. Non-metric multidimensional scaling (nMDS) ordination plot based on Cavalli-Sforza's chord distance calculated from microsatellite allele frequency data at 12 loci. (3D stress value = 0.1071). Each symbol represents one population. For abbreviations, see Supporting Information, Table S1.

Species distribution modelling (SDM)

The MaxEnt models yielded a good fit for the known distribution of *P. arion* (AUC = 0.895, SD = 0.023). It is remarkable that all variables which proved to be significant to predict the species distribution are related to the humidity of the climate while the temperature related parameters seem to be of secondary importance.

According to SDM predictions, the distribution of *P. arion* showed less fluctuation during the glaciation than is known for other temperate species as large areas suitable for the species were available during the LGM in Europe (Fig. 5). Namely, *P. arion* may have persisted in several extra-Mediterranean localities including even southern England (Fig. 5) during the LGM. It is also clear that *P. arion* could not survive north of the Carpathians and eastern Alps under continental climatic conditions (i.e. eastwards from Germany). Furthermore, SDM predicted potential refugial areas to the east in central Asia during the LGM.

DISCUSSION

This study is the first attempt to interpret the nature of the 'spring' and 'summer' form of *P. arion* in a biogeographical context despite limited sampling that does not cover the whole distribution of the species. However, this study is still based on the most complete geographical sampling of the target species using microsatellites compared to previous studies

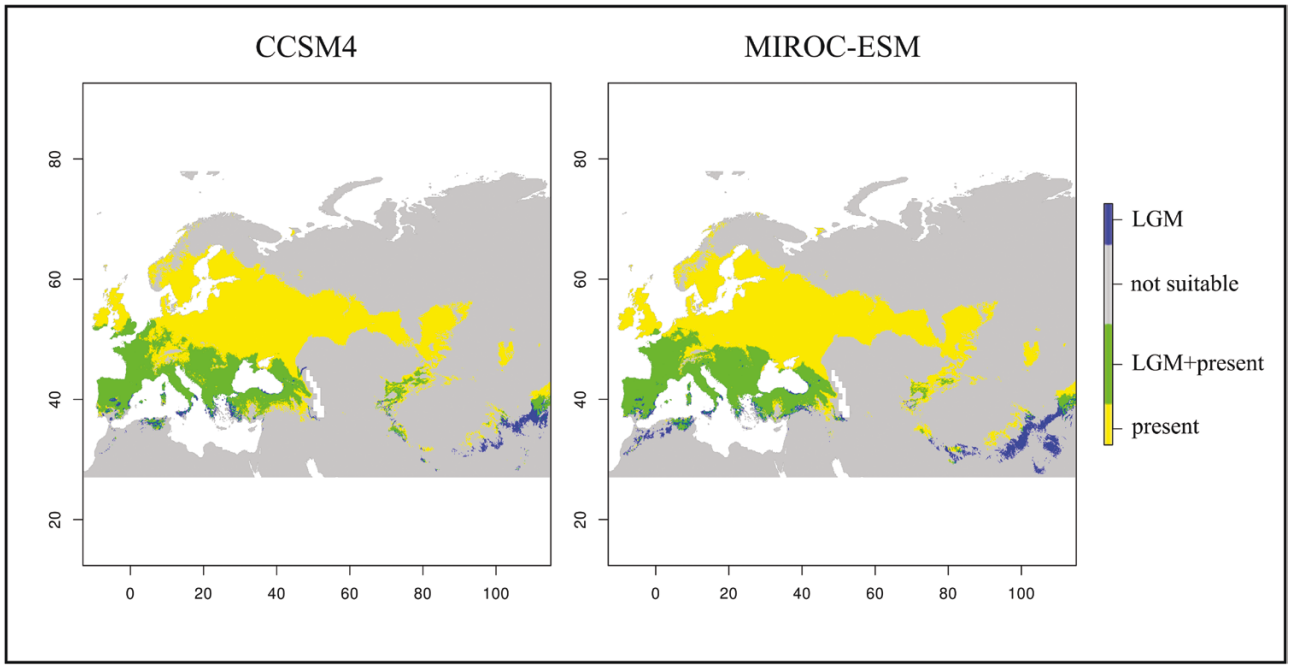

Figure 5. Predicted area dynamic for *P. arion* considering two timescales (Present and the Last Glacial Maximum) using two climate models (CCSM4 and MIROC-ESM).

(Ugelvig *et al.*, 2011, 2012; Sielezniew & Rutkowski, 2012; Patricelli *et al.*, 2013; Andersen *et al.*, 2014; Sielezniew *et al.*, 2015; Bereczki *et al.*, 2020).

Our results clearly indicate the presence of three genetic lineages in the studied samples. The Appenine clade includes Italian and western Slovenian samples. The eastern clade comprises populations from Poland, Russia and Romania. The third (central) genetic group involves the rest of the samples.

A previous study on *P. arion* using microsatellites (Sielezniew *et al.*, 2015) already showed the pronounced genetic structure of the non-Alpine Italian populations which differ from Polish populations as well. The authors hypothesized that the large genetic difference among these Italian populations may be interlinked with the history of the species during Plio-Pleistocene glaciations as many species underwent climate-linked cycles of fragmentation and allopatric divergence within the Apennine refugium (a multiple-refugia scenario, e.g. Canestrelli *et al.*, 2008; Canestrelli & Nascetti, 2008). However, they also emphasized that biogeographical history is not the only factor shaping population structure. It could also be more recently influenced by the distinctive demographic features of populations. Reduced connectivity of habitats resulting from, e.g. more forest cover at high altitudes, may have led to a greater genetic differentiation among populations in isolated mountainous areas (Keyghobadi *et al.*, 2005). According to this hypothesis, Italian populations could be more differentiated simply due to landscape heterogeneity.

Our study on a wider geographical scale clearly confirms the results concerning the divergence of the non-Alpine Italian populations and their pronounced genetic structure, and supports the combination of the explanations raised by the cited authors (Sielezniew *et al.*, 2015). Based on SDM, the Apennine Peninsula is a potential refugial area for *P. arion* where the long-term survival of the species has been assured during the LGM. The populations with presumably prolonged demographic stability occurring in these areas may have evolved in allopatric conditions in separate mountainous range and the pronounced genetic structuring was maintained by the landscape heterogeneity. It is also clear that the Alps could have acted as a barrier to the possible expansion of this Appenine genetic pool of the species towards central and eastern Europe as is known for several other species (Bilton *et al.*, 1998; Hewitt, 1999; Dapporto, 2010; Zinetti *et al.*, 2013). Because of its high genetic distinctness, we evaluate the Appenine genetic lineage as an Evolutionary Significant Unit (Casacci *et al.*, 2014).

Regarding the separateness of the eastern genetic lineage, a different scenario is likely since it is mainly distributed in areas unsuitable for the species during the last glaciation, thus they had to be re-colonized. Based on our findings, it is highly probable that the source populations of this re-colonization originated from central Asia where SDM predicted potential refugial area. This colonization route is also supported

by the *P. arion* study by Ugelvig (2010) using mitochondrial sequences. Furthermore, the results in Patricelli *et al.* (2013) using the barcode gene and the nuclear elongation factor 1α also do not exclude such a scenario.

Within the eastern genetic lineage, the higher level of differentiation of the samples originating from Russia shown in our study could simply be the result of isolation by distance. In contrast, the high genetic distinctness of the Polish 'spring *arion*' samples from the geographically close 'summer *arion*' populations can be explained more reasonably by a phenological shift between the two *arion* forms which greatly reduces the gene flow among populations flying in different time periods.

The rest of the samples belonging to the central genetic cluster originate from two geographic regions, i.e. the Carpathian Basin and the Balkans. Interestingly, one of the south Polish populations (SRO) clustered with these samples as well, and was therefore highly distinct from the other nearby southern Polish samples, as was already detected by Sielezniew *et al.* (2015). It is worth emphasizing that this is the only sampled Polish population inhabiting the southern slopes of the Carpathians. Therefore, the observed pattern indicates that the expansion of the central genetic group towards north and east was limited by this mountain range. The populations from the Carpathian Basin and the Balkans show genetic similarity, which could mean that they have a common evolutionary history.

Although Ugelvig (2010) suggested a refugium for *P. arion* on the Balkan Peninsula based on mitochondrial studies, SDM predicted extensive area suitable for the species during the LGM in the Carpathian Basin and elsewhere in central Europe, so it is very likely that the distribution area did not fluctuate considerably there. The revealed pattern at microsatellite loci also supports this finding as it indicates that gene flow appears to have been intense within this entire region. At the same time, the high distinctness of the sample from the Kaszonyi Hill (Hungary) within this genetic cluster suggests that there may have been separate extra-Mediterranean refugia (Schmitt & Varga, 2012) in the Carpathian Basin. This distinctiveness was already indicated by several private mutations even in the mitochondrial sequences, which is much less variable and largely homogeneous in the majority of the distribution (Bereczki *et al.*, 2014).

The results of our study exclude the potential allopatric origin of the 'spring *arion*' and 'summer *arion*' forms as they are both part of the central (Carpathian-Balkan) and the eastern lineages. The 'spring *arion*' and 'summer *arion*' sample pairs originating from the same locality in the Carpathian Basin exhibit significant genetic differentiation. This is consistent with the findings of a previous study, which investigated the differentiation of syntopic sample pairs by a multilevel approach in a small area, namely the Aggtelek Karst region of Hungary (Bereczki *et al.*, 2020). The authors concluded that the two phenological forms of *P. arion* may represent the incipient stage of sympatric speciation which is facilitated by the adaptation to the distinct phenology of flowering of the host plants. Negative selection acts against the intermediate individuals, which are on the wing in an inappropriate time frame for egg laying. Thus, disruptive selection affects and produces bimodal distributions of phenotypes. However, the phenology of host plants flowering is not entirely distinct and fluctuates from year to year. Therefore, the two *P. arion* phenological forms can occasionally exchange genes depending on the length of the overlap of the flight periods. Such gene flow prevents the completion of the speciation process.

This kind of adaptation to host plant phenology is well exemplified also in the two forms of *Phengaris alcon* ([Denis & Schiffermüller], 1775) exploiting different initial host plants (*Gentiana cruciata* L. vs. *Gentiana pneumonanthe* L.) and host ant species which usually results in differentiation of habitat use and phenology. Generally, the two main host plants of *Phengaris alcon* grow in different habitats, however, they co-occur at a few sites usually with a temporal separation of the flowering periods between them. As in *P. arion*, the flowering varies from year to year and overlapping phenology occasionally occurs. In such cases, butterflies lay their eggs on both host plants. This can lead directly to fluctuations in genetic differentiation among forms as has been shown for sympatric occurrence (Răscruci, Transylvania) with F_{ST} values decreasing more than by a half from 2007 to 2011 (Bereczki *et al.*, 2018). Such dynamics could be the result of varying phenology driven by changes in yearly weather conditions. A larger overlap in the flight period of the two forms can result in a higher level of gene flow between forms.

Thus, the lack of differentiation among *P. arion* phenological forms during a single vegetation period in Ostrogorka (OSTa and OSTb in Russia) could be a result of such an extensive phenological overlap in flowering of the host plants in some of the previous years. Namely, it is important to emphasize that the differentiation between the 'spring' and 'summer' type of *P. arion* is a result of local adaptation that strongly depends on the local phenological conditions. However, it is important to note that the genetic difference between the two forms only marginally failed to be significant, thus it is also possible that adding more samples would have produced different results.

It is also important to note that *Wolbachia* infection has been rejected as a driver of sympatric speciation in

P. arion on a larger geographical scale as only a single strain was detected in all geographic regions (Patricelli *et al.*, 2013; Bereczki *et al.*, 2015). More importantly, the same strain was found in syntopic sample pairs of 'spring *arion*' and 'summer *arion*' (Bereczki *et al.*, 2020) which suggests that *Wolbachia* does not have any effect on the evolutionary processes in *P. arion*.

Our results also clarify a nomenclatural question: namely, whether the name *ligurica* can be used to specify the 'summer form' of *P. arion* in the Pannonian region. This name was introduced by Fritz Wagner originally as a *varietas* to characterize the form which occurs at the Ligurian coast (in Italy and France) (see *Lycaena arion ligurica* Wagner, 1904) and became established in the Hungarian literature through the work of Bálint (1990, 1994, 1996) and Varga (Gyulai *et al.*, 2010). The 'summer form' has been identified as *ligurica* based on external morphological characters of wings. However, Bálint (2015) has already rejected the use of this name for the populations of the Carpathian Basin based on the designation and morphological examination of the neotype of *Lycaena arion ligurica* from San Remo. Our genetic results also support that the name *ligurica* has been misused to denote the 'summer' form of *P. arion*, given that the Italian and Carpathian Basin populations belong to two separate genetic lineages.

In summary, the results of the present study exclude the potential allopatric origin of the 'spring *arion*' and 'summer *arion*', and support the hypothesis that the existence of the two forms could be a result of local adaptation to the distinctive phenology of host plant flowering which is manifested in the genetic differences between them. Our data indicate that ecological specialization of *P. arion* is not associated with lineage sorting and could be an example of convergent evolution as it is hypothesized for congeneric *Phengaris alcon* (Koubínová *et al.*, 2017). However, the differentiation between the 'spring *arion*' and 'summer *arion*' depends highly on the local phenological conditions (fluctuations). *Wolbachia* infection has been ruled out as a driver of sympatric speciation in *P. arion* on a larger geographical scale as well.

The three well-defined genetic lineages in *P. arion* recognized by our study seem to have different origin. The highly distinct Apennine lineage is limited by the Alps and may have survived long time under allopatric conditions in the mountainous ranges of the Apennines where the pronounced genetic structuring is maintained by the orography. Although the Apennine lineage could deserve a separate taxonomic status because of high genetic distinctness, we prefer to evaluate it only as an ESU for the conservation purposes. It is also clear that the taxon name *ligurica* should not be applied to denote the 'summer form'. The central lineage is limited in the North and East by the Carpathians and lacks major range fluctuations related to glaciations, although there is evidence for extra-Mediterranean refugia in the Carpathian Basin. The eastern clade could have originated from the refugia in central Asia.

ACKNOWLEDGEMENTS

The study was supported by the Hungarian Scientific Research Fund (OTKA K109223). J.B. was supported by a János Bolyai Scholarship of the Hungarian Academy of Sciences. The authors gratefully acknowledge A. Váradi, Z. Varga, A. Huber, V. Zakšek and N. Spring for the help in sample collection. The technical assistance of V. Mester in the molecular work is very much appreciated. We thank M. Tuschek for her work in the Molecular Taxonomy Laboratory of the Hungarian Natural History Museum (Budapest, Hungary). The support of the Nature Conservation Authorities is also greatly appreciated. Sampling was carried out in accordance with the following permits: OKTF-KP/4153-39/2016, ANP 730/2015, MOP 35601-34/2013, prot. DPN-20-0023821, DFKOPogiz-4200/I-11/431/08/ls, DZP-WG.6401.01.26.2015.JRO, DIR-59318-1-2. The authors have no conflicts of interests to declare.

DATA AVAILABILITY

Microsatellite genotypes have been uploaded as part of the Supporting Information. *Wolbachia* WSP sequence information has been uploaded into the *Wolbachia* MLST database.

REFERENCES

Aiello-Lammens ME, **Boria RA**, **Radosavljevic A**, **Vilela B**, **Anderson RP. 2015.** spThin: an R package for spatial thinning of species occurrence records for use in ecological niche models. *Ecography* **38:** 541–545.

Andersen A, **Simcox DJ**, **Thomas JA**, **Nash DR. 2014.** Assessing reintroduction schemes by comparing genetic diversity of reintroduced and source populations: a case study of the globally threatened large blue butterfly (*Maculinea arion*). *Biological Conservation* **175:** 34–41.

Avise JC. 1998. The history and purview of phylogeography: a personal reflection. *Molecular Ecology* **7:** 371–379.

Avise JC. 2000. *Phylogeography: the history and formation of species.* Cambridge: Harvard University Press.

Bálint Z. 1990. A kárpát-medence boglárkalepkéinek revíziója. *Janus Pannonius Múzeum Évkönyve* **34:** 47–62.

Bálint Z. 1994. Magyarország nappali lepkéi a természetvédelem tükrében. *Somogy Megyei Múzeumok Közleményei* **10:** 183–206.

Bálint Z. 1996. *A kárpát-medencei nappali lepkéi, 1 rész: pillangófélék, fehérlepkefélék, boglárkalepke-félék, mozaiklepke-félék*. Budapest: Magyar Madártani és Természetvédelmi Egyesület.

Bálint Z. 2015. Notes on the Pannonian species-group names proposed for the large blue butterfly (*Maculinea arion* (Linnaeus, 1758), Lycaenidae, Lepidoptera). *Folio Musei Historico-Naturalis Bakonyiensis* **32:** 103–120.

Bereczki J, **Rácz R**, **Varga Z**, **Tóth JP. 2015.** Controversial patterns of *Wolbachia* infestation in the social parasitic *Maculinea* butterflies (Lepidoptera: Lycaenidae). *Organisms Diversity & Evolution* **15:** 591–607.

Bereczki J, **Pecsenye K**, **Varga Z**, **Tartally A**, **Tóth JP. 2018.** *Maculinea rebeli* (Hirschke) - a phantom or reality? Novel contribution to a long-standing debate over the taxonomic status of an enigmatic Lycaenidae butterfly. *Systematic Entomology* **43**: doi:10.1111/syen.12259.

Bereczki J, **Póliska S**, **Váradi A**, **Tóth JP. 2020.** Incipient sympatric speciation via host race formation in *Phengaris arion* (Lepidoptera: Lycaenidae). *Organisms Diversity & Evolution* **20:** 63–76.

Bereczki J, **Tóth JP**, **Tóth A**, **Bátori E**, **Pecsenye K**, **Varga Z. 2011.** The genetic structure of phenologically differentiated large blue (*Maculinea arion*) populations (Lepidoptera: Lycaenidae) in the Carpathian Basin. *European Journal of Entomology* **108:** 519–527.

Bereczki J, **Tóth JP**, **Sramkó G**, **Varga Z. 2014.** Multilevel studies on the two phenological forms of large blue (*Maculinea arion*) (Lepidoptera: Lycaenidae). *Journal of Zoological Systematics and Evolutionary Research* **52:** 32–43.

Berlocher SH, **Feder JL. 2002.** Sympatric speciation in phytophagous insects: moving beyond controversy? *Annual Review of Entomology* **47:** 773–815.

Bhagwat SA, **Willis KJ. 2008.** Species persistence in northerly glacial refugia of Europe: a matter of chance or biogeographical traits? *Journal of Biogeography* **35:** 464–482.

Bilton DT, **Mirol PM**, **Mascheretti S**, **Fredga K**, **Zima J**, **Searle JB. 1998.** Mediterranean Europe as an area of endemism for small mammals rather than a source for northwards postglacial colonization. *Proceedings of the Royal Society of London B: Biological Sciences* **265:** 1219–1226.

Breeuwer JA, **Werren JH. 1993.** Cytoplasmic incompatibility and bacterial density in *Nasonia vitripennis*. *Genetics* **135:** 565–574.

Busby JR. 1991. BIOCLIM: a bioclimate analysis and prediction system. *Plant Protection Quarterly* **6:** 8–9.

Bush GL. 1969. Sympatric host race formation and speciation in frugivorous flies of the genus *Rhagoletis* (Diptera, Tephritidae). *Evolution* **23:** 237–251.

Bush GL. 1975. Modes of animal speciation. *Annual Review of Ecology and Systematics* **6:** 339–364.

Canestrelli D, **Cimmaruta R**, **Nascetti G. 2008.** Population genetic structure and diversity of the Apennine endemic stream frog, *Rana italica* - insights on the Pleistocene evolutionary history of the Italian peninsular biota. *Molecular Ecology* **17:** 3856–3872.

Canestrelli D, **Nascetti G. 2008.** Phylogeography of the pool frog *Rana* (*Pelophylax*) *lessonae* in the Italian peninsula and Sicily: multiple refugia, glacial expansions and nuclear–mitochondrial discordance. *Journal of Biogeography* **35:** 1923–1936.

Casacci LP, **Barbero F**, **Balletto E. 2014.** The 'Evolutionarily Significant Unit' concept and its applicability in biological conservation. *Italian Journal of Zoology* **81:** 182–193.

Cavalli-Sforza LL, **Edwards AWF. 1967.** Phylogenetic analysis. Models and estimation procedures. *American Journal of Human Genetics* **19:** 233–257.

Coope GR. 2004. Several million years of stability among insect species because of, or in spite of, Ice Age climatic instability? *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences* **359:** 209–214.

Dapporto L. 2010. Speciation in Mediterranean refugia and post-glacial expansion of *Zerynthia polyxena* (Lepidoptera, Papilionidae). *Journal of Zoological Systematics and Evolutionary Research* **48:** 229–237.

Drès M, **Mallet J. 2002.** Host races in plant-feeding insects and their importance in sympatric speciation. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences* **357:** 471–492.

Earl DA, **vonHoldt BM. 2012.** STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* **4:** 359–361.

Eldredge N, **Gould SJ. 1972.** Punctuated equilibria: an alternative to phyletic gradualism. In: Schopf TJM, ed. *Models in paleobiology.* New York: Freeman, Cooper & Co., 82–115.

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.

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

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.

Feder JL. 1998. The apple maggot fly, *Rhagoletis pomonella*: flies in the face of conventional wisdom about speciation? In: Howard DJ, Berlocher SH, eds. *Endless forms: species and speciation.* Oxford: Oxford University Press, 130–144.

Fielding AH, **Bell JF. 1997.** A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation* **24:** 38–49.

Filchak KE, **Roethele JB**, **Feder JL. 2000.** Natural selection and sympatric divergence in the apple maggot *Rhagoletis pomonella*. *Nature* **407:** 739–742.

Filz KJ, **Schmitt T. 2015.** Niche overlap and host specificity in parasitic *Maculinea* butterflies (Lepidoptera: Lycaenidae) as a measure for potential extinction risks under climate change. *Organisms Diversity & Evolution* **15:** 555–565.

Fitzpatrick BM, **Fordyce JA**, **Gavrilets S. 2008.** What, if anything, is sympatric speciation? *Journal of Evolutionary Biology* **21:** 1452–1459.

Foote AD. 2018. Sympatric speciation in the genomic era. *Trends in Ecology & Evolution* **33:** 85–95.

Francis RM. 2017. pophelper: an R package and web app to analyse and visualize population structure. *Molecular Ecology Resources* **17:** 27–32.

Franklin J, **Miller JA. 2009.** *Mapping species distributions.* Cambridge: Cambridge University Press.

Futuyma DJ, **Mayer GC. 1980.** Non-allopatric speciation in animals. *Systematic Biology* **29:** 254–271.

Goffart P. 1997. *Libellules et papillons en Wallonie*. Louvain-la-Neuve: Université Catholique de Louvain-la-Neuve.

Guillot G, **Mortier F**, **Estoup A. 2005.** Geneland: a computer package for landscape genetics. *Molecular Ecology Notes* **5:** 712–715.

Gyulai P, **László MG**, **Pekarsky O**, **Peregovits L**, **Ronkay G**, **Ronkay L**, **Szabóky C**, **Varga Z**, **Witt TJ. 2010.** *Magyarország nagylepkéi (Macrolepidoptera of Hungary)* (Z Varga, ed.). Budapest: Heterocera Press.

Habel JC, **Schmitt T**, **Müller P. 2005.** The fourth paradigm pattern of post-glacial range expansion of European terrestrial species: the phylogeography of the marbled white butterfly (Satyrinae, Lepidoptera). *Journal of Biogeography* **32:** 1489–1497.

Hammer O, **Harper DAT**, **Ryan PD. 2001.** PAST: paleontological statistics software package for education and data analysis. *Paleontologica Electronica* **4:** 9.

Hendry AP, **Nosil P**, **Rieseberg LH. 2007.** The speed of ecological speciation. *Functional Ecology* **21:** 455–464.

Hewitt GM. 1996. Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society* **58:** 247–276.

Hewitt GM. 1999. Post-glacial re-colonization of European biota. *Biological Journal of the Linnean Society* **68:** 87–112.

Hewitt G. 2000. The genetic legacy of the Quaternary ice ages. *Nature* **405:** 907–913.

Hewitt GM. 2001. Speciation, hybrid zones and phylogeography - or seeing genes in space and time. *Molecular Ecology* **10:** 537–549.

Hewitt GM. 2004. Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences* **359:** 183–195.

Hoffmann AA, **Turelli M. 1997.** *Cytoplasmic incompatibility in insects.* Oxford: Oxford University Press.

Inskeep KA, **Doellman MM**, **Powell THQ**, **Berlocher SH**, **Seifert NR**, **Hood GR**, **Ragland GJ**, **Meyers PJ**, **Feder JL. 2021.** Divergent diapause life history timing drives both allochronic speciation and reticulate hybridization in an adaptive radiation of *Rhagoletis* flies. *Molecular Ecology*. doi: 10.1111/mec.15908.

Johnson TC, **Scholz CA**, **Talbot MR**, **Kelts K**, **Ricketts RD**, **Ngobi G**, **Beuning K**, **Ssemmanda I**, **McGill JW. 1996.** Late Pleistocene desiccation of Lake Victoria and rapid evolution of cichlid fishes. *Science* **273:** 1091–1093.

Jueterbock A, **Smolina I**, **Coyer JA**, **Hoarau G. 2016.** The fate of the Arctic seaweed *Fucus distichus* under climate change: an ecological niche modeling approach. *Ecology and Evolution* **6:** 1712–1724.

Keyghobadi N, **Roland J**, **Strobeck C. 2005.** Genetic differentiation and gene flow among populations of the alpine butterfly, *Parnassius smintheus*, vary with landscape connectivity. *Molecular Ecology* **14:** 1897–1909.

Koubínová D, **Dincă V**, **Dapporto L**, **Vodă R**, **Suchan T**, **Vila R**, **Alvarez N. 2017.** Genomics of extreme ecological specialists: multiple convergent evolution but no genetic divergence between ecotypes of *Maculinea alcon* butterflies. *Scientific Reports* **7:** 13752.

Lister AM. 2004. The impact of Quaternary Ice Ages on mammalian evolution. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences* **359:** 221–241.

Mayr E. 1954. Change of genetic environment and evolution. In: Huxley J, Hardy AC, Ford EB, eds. *Evolution as a process.* London: Allen & Unwin, 157–180.

Mayr E. 1963. *Animal species and evolution*. Cambridge: Harvard University Press.

Munguira M, **Martin J. 1997.** *Action plan for the* Maculinea *butterflies in Europe.* Strasbourg: Council of Europe.

Nosil P, **Harmon LJ**, **Seehausen O. 2009.** Ecological explanations for (incomplete) speciation. *Trends in Ecology & Evolution* **24:** 145–156.

Nowicki P, **Settele J**, **Thomas JA**, **Woyciechowski M. 2005.** A review of population structure of *Maculinea* butterflies. In: Settele J, Kühn E, Thomas JA, eds. *Studies on the ecology and conservation of butterflies in Europe.* Sofia: Pensoft Publishers, 144–149.

Van Oosterhout C, **Hutchinson WF**, **Wills DPM**, **Shipley P. 2004.** Micro-checker: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes* **4:** 535–538.

Osváth-Ferencz M, **Bonelli S**, **Nowicki P**, **Peregovits L**, **Rákosy L**, **Sielezniew M**, **Kostro-Ambroziak A**, **Dziekańska I**, **Kőrösi A. 2017.** Population demography of the endangered large blue butterfly *Maculinea arion* in Europe. *Journal of Insect Conservation* **21:** 411–422.

Patricelli D, **Sielezniew M**, **Ponikwicka-Tyszko D**, **Ratkiewicz M**, **Bonelli S**, **Barbero F**, **Witek M**, **Buš 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.

Phillips SJ, **Anderson RP**, **Schapire RE. 2006.** Maximum entropy modeling of species geographic distributions. *Ecological Modelling* **190:** 231–259.

Phillips SJ, **Dudík M**, **Schapire RE. 2004.** A maximum entropy approach to species distribution modeling. In: Proceedings of the Twenty-First International Conference on Machine Learning. Banff, AB, Canada. 4 -8 July **2004**; New York, NY, USA: Association for Computing Machinery; **2004**. pp. 655–662.

Pritchard JK, **Stephens M**, **Donnelly P. 2000.** Inference of population structure using multilocus genotype data. *Genetics* **155:** 945–959.

QGIS. 2022. *QGIS Geographic Information System*. QGIS Association. https://qgis.org/en/site/getinvolved/faq/index.html.

Rácz R, **Bereczki J**, **Sramkó G**, **Kosztolányi A**, **Tóth JP**, **Póliska S**, **Horváth A**, **Barta E**, **Barta Z. 2015.** Isolation and characterisation of 15 microsatellite loci from *Lethrus apterus* (Coleoptera: Geotrupidae). *Annales Zoologici Fennici* **52:** 45–50.

Richards EJ, **Servedio MR**, **Martin CH. 2019.** Searching for sympatric speciation in the genomic era. *BioEssays* **41:** 1900047.

Scheldeman X, **van Zonneveld M. 2010.** *Training manual on spatial analysis of plant diversity and distribution*. Rome: Bioversity International.

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

Schönrogge K, **Wardlaw JC**, **Thomas JA**, **Elmes GW. 2000.** Polymorphic growth rates in myrmecophilous insects. *Proceedings of the Royal Society B: Biological Sciences* **267:** 771–777.

Sielezniew M, **Patricelli D**, **Rutkowski R**, **Witek M**, **Bonelli S**, **Buś MM. 2015.** Population genetics of the endangered obligatorily myrmecophilous butterfly *Phengaris* (=*Maculinea*) *arion* in two areas of its European range. *Insect Conservation and Diversity* **8:** 505–516.

Sielezniew M, **Rutkowski R. 2012.** Population isolation rather than ecological variation explains the genetic structure of endangered myrmecophilous butterfly *Phengaris* (=*Maculinea*) *arion*. *Journal of Insect Conservation* **16:** 39–50.

Stewart JR, **Lister AM**, **Barnes I**, **Dalén L. 2010.** Refugia revisited: individualistic responses of species in space and time. *Proceedings of the Royal Society B: Biological Sciences* **277:** 661–671.

Stohlgren TJ, **Ma P**, **Kumar S**, **Rocca M**, **Morisette JT**, **Jarnevich CS**, **Benson N. 2010.** Ensemble habitat mapping of invasive plant species. *Risk Analysis* **30:** 224–235.

Van Swaay C, **Cuttelod A**, **Collins S**, **Maes D**, **Munguira L**, **Šašić M**, **Settele J**, **Verovnik R**, **Verstrael T**, **Warren M**, **Wiemers M**, **Wynhoff I. 2010.** *European Red List of butterflies*. Luxembourg: Publications Office of the European Union.

Van Swaay C, **Warren M**, **Grill A. 1998.** *Convention on the conservation of European wildlife and natural habitats: threatened butterflies in Europe*. Strasbourg: Council of Europe.

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

Tartally A, **Thomas JA**, **Anton C**, **Balletto E**, **Barbero F**, **Bonelli S**, **Bräu M**, **Pietro CL**, **Csősz S**, **Czekes Z**, **Dolek M**, **Dziekańska I**, **Elmes G**, **Fürst MA**, **Glinka U**, **Hochberg ME**, **Höttinger H**, **Hula V**, **Maes D**, **Munguira ML**, **Musche M**, **Nielsen PS**, **Nowicki P**, **Oliveira PS**, **Peregovits L**, **Ritter S**, **Schlick-Steiner BC**, **Settele J**, **Sielezniew M**, **Simcox DJ**, **Stankiewicz AM**, **Steiner FM**, **Švitra G**, **Ugelvig LV**, **Van Dyck H**, **Varga Z**, **Witek M**, **Woyciechowski M**, **Wynhoff I**, **Nash DR. 2019.** Patterns of host use by brood parasitic *Maculinea* butterflies across Europe. *Philosophical Transactions of the Royal Society B: Biological Sciences* **374:** 20180202.

Tauber CA, **Tauber MJ. 1989.** Symaptric speciation in insects: perception and perspective. In: Otte D, Endler J, eds. *Speciation and its consequences*. Sunderland: Sinauer Press, 307–344.

Tax MH. 1989. *Atlas van de Nederlandse dagvlinders*. Nederland: s'-Graveland: De Vlinderstichting, Wageningen and Stichting tot Behoud van Natuurmonumenten.

Templeton AR. 1981. Mechanisms of speciation - a population genetic approach. *Annual Review of Ecology and Systematics* **12:** 23–48.

Thomas JA, **Elmes GW**, **Wardlaw JC**, **Woyciechowski M. 1989.** Host specificity among *Maculinea* butterflies in *Myrmica* ant nests. *Oecologia* **79:** 452–457.

Thomas J. 1995. The ecology and conservation of *Maculinea arion* and other European species of large blue butterfly. In: Pullin A, ed. *Ecology and conservation of butterflies*. London: Chapman & Hall, 180–196.

Thomas JA, **Elmes GW**, **Wardlaw JC. 1998.** Polymorphic growth in larvae of the butterfly *Maculinea rebeli*, a social parasite of *Myrmica* ant colonies. *Proceedings of the Royal Society B: Biological Sciences* **265:** 1895–1901.

Thomas JA, **Simcox DJ**, **Clarke RT. 2009.** Successful conservation of a threatened *Maculinea* butterfly. *Science* **325:** 80–83.

Title PO, **Bemmels JB. 2018.** ENVIREM: an expanded set of bioclimatic and topographic variables increases flexibility and improves performance of ecological niche modeling. *Ecography* **41:** 291–307.

Turelli M, **Barton NH**, **Coyne JA. 2001.** Theory and speciation. *Trends in Ecology & Evolution* **16:** 330–343.

Ugelvig LV. 2010. *Ecological genetics and evolution of the large blue butterfly - consequences of an extraordinary lifecycle*. Museum Tusculanum, Biologisk Institut.

Ugelvig LV, **Andersen A**, **Boomsma JJ**, **Nash DR. 2012.** Dispersal and gene flow in the rare, parasitic large blue butterfly *Maculinea arion*. *Molecular Ecology* **21:** 3224–3236.

Ugelvig LV, **Nielsen PS**, **Boomsma JJ**, **Nash DR. 2011.** Reconstructing eight decades of genetic variation in an isolated Danish population of the large blue butterfly *Maculinea arion*. *BMC Evolutionary Biology* **11:** 201.

Via S. 2001. Sympatric speciation in animals: the ugly duckling grows up. *Trends in Ecology & Evolution* **16:** 381–390.

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.

Werren JH. 1997. Biology of *Wolbachia*. *Annual Review of Entomology* **42:** 587–609.

Werren JH, **Baldo L**, **Clark ME. 2008.** *Wolbachia*: master manipulators of invertebrate biology. *Nature Reviews Microbiology* **6:** 741–751.

Werren JH, **Windsor DM. 2000.** *Wolbachia* infection frequencies in insects: evidence of a global equilibrium? *Proceedings Biological Sciences* **267:** 1277–1285.

Wynhoff I. 1998. REVIEW: The recent distribution of the European *Maculinea* species. *Journal of Insect Conservation* **2:** 15–27.

Zeisset I, **Damm Als T**, **Settele J**, **Boomsma JJ. 2005.** Microsatellite markers for the large blue butterflies *Maculinea nausithous* and *Maculinea alcon* (Lepidoptera: Lycaenidae) and their amplification in other *Maculinea* species. *Molecular Ecology Notes* **5:** 165–168.

Zinetti F, **Dapporto L**, **Vovlas A**, **Chelazzi G**, **Bonelli S**, **Balletto E**, **Ciofi C. 2013.** When the rule becomes the exception. No evidence of gene flow between two *Zerynthia* cryptic butterflies suggests the emergence of a new model group. *PLoS One* **8:** e65746.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Table S1. Information on all specimens sampled with individual microsatellite genotypes at all 12 loci and the details of the sampled populations (location data, sampling time, phenology and food plant where they are available as well as the source of DNA extraction and *Wolbachia* infection status).
Figure S1. Presence points used for species distribution modelling (SDM). The original data set (grey dots) has been balanced using spatial thinning. The analyses were performed with 100 presence points (black dots).
Figure S2. ΔK obtained by Structure Harvester which indicates the change in log probability between successive K values.