



# Pleistocene origins of chorusing diversity in Mediterranean bush-cricket populations (*Ephippiger diurnus*)

Yareli Esquer Garrigos, Rejane Streiff, Virginie Party, Sabine Nidelet, Miguel Navascués, Michael D. Greenfield

## ► To cite this version:

Yareli Esquer Garrigos, Rejane Streiff, Virginie Party, Sabine Nidelet, Miguel Navascués, et al.. Pleistocene origins of chorusing diversity in Mediterranean bush-cricket populations (*Ephippiger diurnus*). Biological Journal of the Linnean Society, 2019, 126 (3), pp.598-613. 10.1093/biolinnean/bly195 . hal-02154694

**HAL Id: hal-02154694**

**<https://hal.science/hal-02154694>**

Submitted on 18 Aug 2023

**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.

# **Pleistocene origins of chorusing diversity in Mediterranean bush-cricket populations (*Ephippiger diurnus*)**

Yareli Esquer-Garrigos, Réjane Streiff, Virginie Party, Sabine Nidelet, Miguel Navascués, Michael D Greenfield

## **Abstract**

We studied the Pleistocene diversification of a relatively endemic Mediterranean insect (*Ephippiger diurnus*; Orthoptera: Tettigoniidae) to understand how species with restricted range may nonetheless exhibit the complex phylogeography normally associated with broad distribution. A time-calibrated molecular phylogeny based on two mitochondrial genes showed that *E. diurnus* diverged into two major clades, distinguished largely by male song, before or early during the Pleistocene. Several subclades also diverged before the most recent glacial period. Data from 20 microsatellite loci indicated higher genetic diversity in populations along the Mediterranean coast in France, consistent with the hypothesis that glacial refuges were located there. ‘Isolation by distance’ accounts for much genetic differentiation between populations, but some adjacent populations are highly differentiated. A Bayesian approach defined genetically distinct clusters and assigned individuals to their most probable cluster. Clusters corresponded to clades in the phylogenetic tree, and we used cluster assignments to estimate interclade gene flow in areas of potential secondary contact. Gene flow is negligible in potential contact areas in the Pyrenees, but a narrow hybrid zone featuring a steep cline exists on the coast. This hybrid zone suggests that the major clades represent distinct species that diverged within a restricted area during the Pleistocene.

**Keywords:** acoustic insects, cryptic species, endemic species, hybrid zone, Pleistocene refuge, post-glacial expansion, secondary contact

## **INTRODUCTION**

It is generally acknowledged that the Pleistocene glaciations have had a major impact on biogeography in many parts of the world (Hewitt, 2000, 2004; Avise, 2009; Stewart et al., 2010). The impact was particularly strong in Europe, where current distributions and population genetic structures of all three faunal and floristic elements, Arctic-alpine species (Wallis et al., 2016), continental species and warm-adapted ‘Mediterranean species’, were influenced, each responding in a characteristic fashion (Schmitt, 2007). For example, it is inferred that Mediterranean species retreated far southward to climatic refuges on the Iberian, Italian and Balkan peninsulas, or even beyond, during glacial advances and then returned northward during interglacial warm periods (Hewitt, 1999). Passing the glacial periods in separated refuges led to genetic differentiation among the isolated populations, and a species’ current phylogeography typically reflects those genetic differences acquired during isolation. Populations in Western Europe may largely descend from Iberian refuges, whereas populations farther east are likely to descend from Balkan and, to a lesser extent, Italian ones (Taberlet et al., 1998). East–west trending mountain chains, namely the Alps and Pyrenees, acted as barriers that blocked gene flow during the northward return, sustaining the genetic differentiation in many species. In some cases, populations with separate ancestries from different refugia later achieved secondary contact. Depending on reproductive barriers acquired during isolation, such contact resulted in a narrow hybrid zone (e.g. Butlin & Hewitt, 1985; Bella et al., 2007) or in wider genetic mixing between the lineages (e.g. Kühne et al., 2017; Tritsch et al., 2018).

Most biogeographical studies of the Mediterranean element have focused on species that are widely distributed throughout Europe and that disperse well (Hewitt, 1999; Schmitt, 2007). But other species found in general habitats of this element have more restricted ranges and do not disperse

effectively over long distances. Does the model outlined above apply in a similar manner to these more endemic species, differing only in geographical scale? Did their populations also retreat to separate refuges within the species range? If so, did isolation in separate, but relatively close, refuges yield a comparable level of genetic differentiation in current populations? Was experiencing the Pleistocene glacial periods more or less likely to generate divergence among these species?

We studied the problem of biogeography of restricted-range species in *Ephippiger diurnus* (western saddle bush-cricket; Orthoptera: Tettigoniidae: Bradyporinae), a Mediterranean bush-cricket found mostly in southern France (Duijm, 1990). The insects are flightless and typically occur in open, shrub habitat (garrigue) from sea level to elevations as high as 2000 m in parts of the eastern Pyrenees. In most areas, *E. diurnus* are distributed in local populations that are separated from neighbouring ones by several kilometres. This discontinuous distribution might reflect current fragmentation of the garrigue habitat, reinforced by the requirement of insects for gravelly soil for oviposition and low mobility in both the nymph and adult stages. Relevant to our objectives, the assemblage of *E. diurnus* populations collectively present a striking phylogeographical pattern across their range.

Recent genetic studies using mitochondrial and microsatellite markers indicate that *E. diurnus* populations are structured in two main divisions (cf. Spooner & Ritchie, 2006; Party et al., 2015; Greenfield et al., 2016). Clade 1 includes populations along the Mediterranean littoral south of the city of Narbonne, France and inland to the west on the southern slopes of the Pyrenees within France and the northeast corner of Spain. Clade 2 comprises populations along the littoral north and east of Narbonne and inland to the west on the northern slopes of the Pyrenees and north along the Rhone Valley and into the Massif Central. Prominent subclades exist within clades 1 and 2 and represent populations grouped within certain valleys: subclades in the Cerdagne valley in clade 1 and along the Ariège and upper Garonne rivers in clade 2. The main phenotypic differences between clades 1 and 2 are in the male calling song and chorusing and the female preferences for song (Party et al., 2014; Barbosa et al., 2016; Marin-Cudraz & Greenfield, 2016; Greenfield et al., 2017). Males in clade 1 populations produce calls with three or more syllables, whereas clade 2 populations have only one or two-syllable calls (cf. Duijm, 1990; for earlier genetic studies, see Oudman et al., 1989, 1990). Females in clade 1 populations prefer a high number of call syllables, considerably greater than the mean in their population, whereas in clade 2 the females prefer a number of call syllables equivalent to or only slightly greater than the mean.

The *E. diurnus* phylogeny and genetic structure based on mitochondrial and microsatellite markers, respectively, have robust branching topologies that are largely congruent (cf. Party et al., 2015; Greenfield et al., 2016). Laboratory studies showed that males and females from separate subclades, or even from clades 1 and 2, show little or no aversion to crossing, despite song and chorusing differences (Ritchie, 1996; Mahamoud-Issa et al., 2017). In some cases, males from another group are even preferred by local females. Crosses produce hybrids intermediate in male song and female preference (Ritchie, 1996), but the full viability and fertility of these hybrids remain unknown. Low mobility and habitat preferences could explain, in part, how the strong phylogeographical structure of *E. diurnus* persists in the absence of mating barriers.

Here, we analyse molecular genetic data from a wide array of *E. diurnus* populations to estimate when the clades and major subclades diverged, the genetic diversity of these groups, whether population sizes have increased or decreased in the past, and the levels of genetic differentiation. We used this information to align the *E. diurnus* phylogeny with the Pleistocene glaciations, infer where the Pleistocene refuge(s) were, identify populations that might have experienced bottlenecks in the past, and evaluate the role of ‘isolation by distance’ in population differentiation. We applied Bayesian clustering techniques to assign individuals to genetically distinct groups (clusters) and used these assignments to evaluate the level of recent gene flow between populations, especially

where marked genetic discontinuities occur in the landscape. Our findings reveal a surprising level of biogeographical complexity at a reduced spatial scale and that some relatively endemic species also might have undergone high divergence, including speciation, during the Pleistocene.

## MATERIAL AND METHODS

### Sampling a highly variable insect

We sampled *E. diurnus* collected from 27 localities in southern France and northeastern Spain between 2011 and 2016 (Table 1; Fig. 1). The localities were chosen to cover the region where the insects are relatively abundant. *Ephippiger diurnus* are reported from most of France and border regions of Spain, Switzerland, Germany and Belgium (Bellmann & Luquet, 2009), whereas they are rare in western and northern France and adjacent countries to the north and east (IUCN Red List of Threatened Species, 2017), and we did not sample these areas of sparse population. All insects were adults and included comparable numbers of males and females at each locality. We extracted total DNA from a hind femur of each insect using the DNAeasy blood and tissue kit (Qiagen, Hilden, Germany) following the manufacturer's instructions.

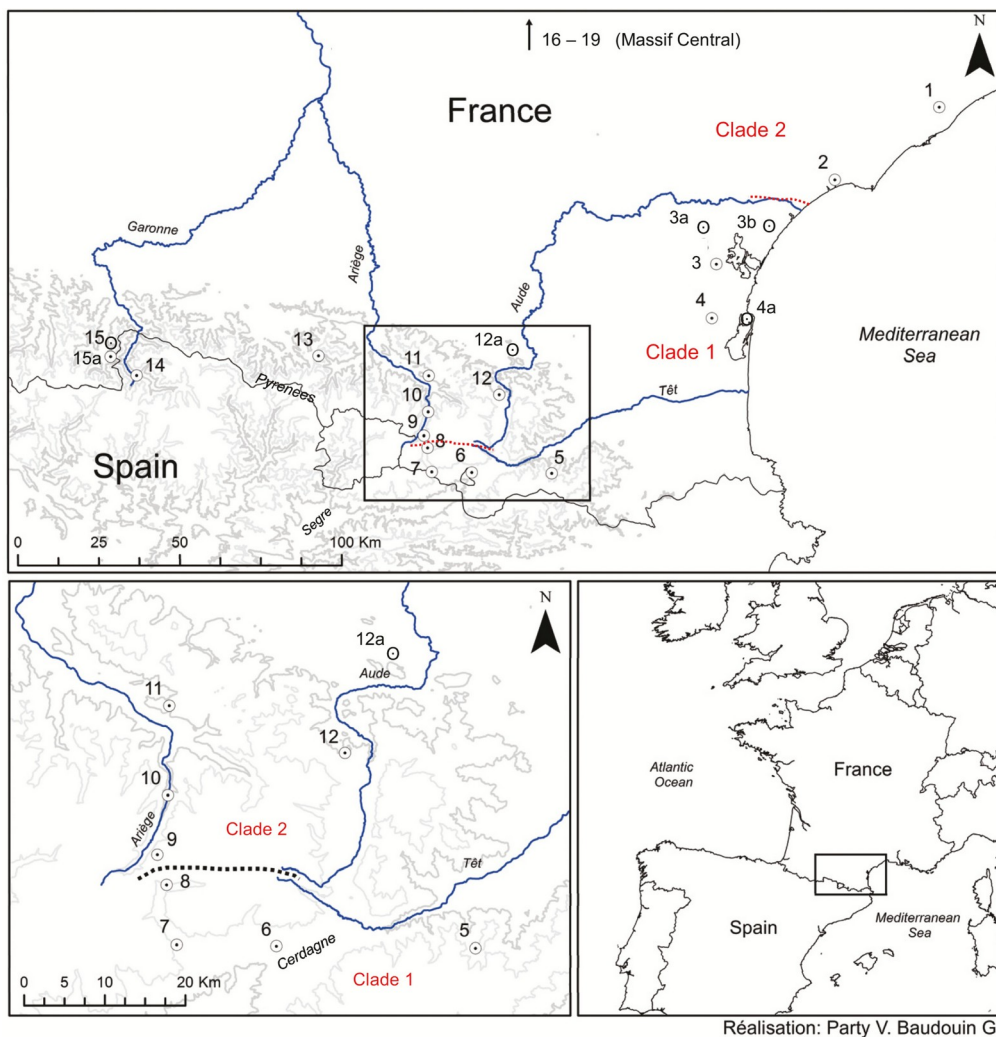
Table 1. *Ephippiger diurnus* localities, showing geographical coordinates, clade, male song type and sample sizes

Locality	Latitude (N)	Longitude (E)	Elevation (m)	Clade*	Song†	n‡
Aubrac	44°37'06"	2°57'56"	1270	2		1 0, 0, 23
Aurel	44°07'52"	5°25'44"	760	2		1, 1, 0
Carcanières	42°42'40"	2°06'26"	1220	2	2.78	6, 3, 7
Caumont	43°53'40"	4°56'41"	60	2		2, 2, 0
Cigalère	42°47'15"	0°38'12"	1560	2	1.5	4, 3, 9
Col de Chioula	42°45'16"	1°50'13"	1430	2	2.02	6, 2, 23
Col de la Croix Morand	45°35'37"	2°51'21"	1370	2		1 0, 0, 14
Col de Mantet	42°29'05"	2°18'58"	1700	1	4.28	6, 4, 23
Col de Puymorens	42°33'01"	1°50'32"	1750	1	3.64	4, 1, 23
Espezel	42°49'15"	2°01'54"	860	2		3, 1, 6
Feuilla	42°55'20"	2°54'19"	210	1	5.85	13, 13, 23
Font Romeu	42°29'24"	2°00'40"	1620	1	3.11	5, 2, 18
Hospitalet près l'Andorre	42°35'02"	1°47'58"	1580	2	2.04	5, 4, 7
Latour-de-Carol	42°29'05"	1°51'36"	1330	1	3.28	4, 2, 23
Le Lioran	45°05'29"	2°44'08"	1260	2		1 3, 3, 11
Leucate	42°55'45"	3°00'41"	10	1	7	5, 4, 21
Massif de la Clape	43°10'28"	3°06'30"	130	1		4 0, 0, 10
Mérens-les-Vals	42°38'55"	1°49'59"	1090	2	1.86	4, 1, 23
Mireval	43°30'12"	3°47'00"	20	2	1	5, 2, 20
Montségret	43°06'04"	2°47'17"	80	1		4 0, 0, 10
Peyriac de Mer	43°04'03"	2°55'41"	50	1	5.25	5, 4, 20
Plan d'Aups	43°19'58"	5°45'20"	670	2		2, 2, 0
Port de Lers	42°48'20"	1°25'26"	1360	2	1.15	4, 1, 23
Pouzol	45°12'00"	3°10'	785	2	1.02	4, 3, 23
Sode	42°48'40"	0°37'00"	930	2	1.5	8, 5, 23
Vias	43°18'28"	3°22'34"	27	2	2.08	6, 3, 22
Vilamós	42°44'43"	0°44'24"	1240	2	1.42	5, 3, 23

\* Clade as indicated by mtDNA (COI and COII) phylogenies or position in genetic structure constructed from microsatellite loci data.

† Mean number of syllables in male song (Greenfield et al., 2016).

‡ Sample sizes for COI and COII phylogenies and for genetic analyses based on microsatellite loci data, respectively. The COI sequences from 18 of the 23 sampled localities had been used in a previously published mitochondrial DNA phylogeny (Party et al., 2015). All COII sequences were obtained in the present study. Microsatellite data for 20 of the 24 sampled localities had been used in a previously published genetic analysis (Greenfield et al., 2016), but those data were based on only 15 of the 20 microsatellite loci analysed in the present study.



Réalisation: Party V. Baudouin G.

Figure 1. Locations of 24 *Ehippiger diurnus* populations in southern France and northeastern Spain sampled for mitochondrial DNA and microsatellite loci (see Table 1). Dotted lines indicate approximate boundaries between clades 1 and 2 where known (Party et al., 2015). Contour lines represent 1200 and 1800 m elevations. Localities: (1) Mireval; (2) Vias; (3) Peyriac de Mer; (3a) Montsérret; (3b) Massif de la Clape; (4) Feuilla; (4a) Leucate; (5) Col de Mantet; (6) Font Romeu; (7) Latour-de-Carol; (8) Col de Puymorens; (9) Hospitalet près l'Andorre; (10) Mérensles-Vals; (11) Col de Chioula; (12) Carcanières; (12a) Espezel; (13) Port de Lers; (14) Vilamós; (15) Sode; (15a) Cigalère; (16) Aubrac; (17) Le Lioran; (18) Pouzol; and (19) Col de la Croix Morand. The map was generated with ArcGIS v.10.0 for desktop, ESRI (<http://www.esri.com/software/arcgis/arcgis-for-desktop>) and adapted from Greenfield et al. (2016).

*Ehippiger diurnus* has had a complicated taxonomic history. Several colour forms were recognized early and accorded subspecies and species status in different reports (Bellmann & Luquet, 2009; see Duijm & Oudman, 1983). The species was also confused with *Ehippiger ehippiger* (eastern saddle bush-cricket), which is now recognized as occurring in the Balkan Peninsula and central and eastern Europe (IUCN Red List of Threatened Species, 2017). Recent studies using mitochondrial DNA (mtDNA) markers focused on the complex phylogeography of *E. diurnus* and showed that the several clades were not congruent with the early morphological and colour forms (Spooner & Ritchie, 2006). Our present study expands these recent efforts by greatly increasing the number of samples and using more precise phylogenetic analyses and microsatellite markers that we developed for *E. diurnus*. These samples, analyses and markers afforded us the opportunity to test specific hypotheses on the species' history, genetic differentiation and gene flow and to make informed inferences and generalizations on phylogeography.

## Divergence of clades and major subclades

We measured genetic divergence of *E. diurnus* clades by analysing a previously published mtDNA dataset (Party et al., 2015) that we completed with additional samples, sequences and localities (Table 1). For this complete dataset, we amplified a partial sequence of mitochondrial cytochrome c oxidase subunit I (COI; 511 bp) in 110 insects from 23 of the localities. In the insects where a COI sequence was recovered, we amplified a second fragment, cytochrome c oxidase subunit II (COII; 530 bp). This second amplification was accomplished with universal insect primers followed by COII primers that we designed specifically for *E. diurnus* (Supporting Information, Appendix S1).

A best-fit model of molecular evolution for our dataset was evaluated in jModeltest v.2.1.4 (Darriba et al., 2012) using Bayesian information criteria (Keane et al., 2006). Only unique haplotype sequences for COI and COII were used. We did not retain insects whose COI or COII sequences included missing data, ambiguities or gaps, or heteroplasmic sites; the latter may have represented nuclear mitochondrial DNA segments (numts), which are a noted problem in Orthoptera (Song et al., 2014). Independent phylogenetic reconstruction for COI and COII was done by Bayesian inference in BEAST v.2.2.1 (Bouckaert et al., 2014). We used a relaxed-clock lognormal model, a birth–death tree prior, and a Markov chain Monte Carlo (MCMC) chain length of 50 000 000, sampled every 5000 generations, for both COI and COII. Phylogenies were summarized as maximum-clade credibility trees in TreeAnnotator v.2.2.1 (<http://beast.bio.ed.ac.uk>; Rambaut & Drummond, 2010) and were visualized and edited in FigTree v.1.4.0 (<http://tree.bio.ed.ac.uk/software/figtree/>). One sequence of the closely related species *Sorapagus catalaunicus* (Tettigoniidae: Bradyporinae: Ephippigerini) was included as an outgroup.

Our initial phylogenies constructed with COI and COII data separately had similar branching topologies with robust major nodes (posterior probabilities > 0.80). We therefore proceeded to estimate phylogenetic relationships and divergence times between the clades and subclades using a concatenated approach in BEAST that linked the COI and COII datasets.

Lacking fossils or other external calibration points, we estimated dates for clade bifurcation in our concatenated phylogeny using a provisional molecular clock based on the COI mutation rate previously determined for the genus *Banza* (Tettigoniidae: Conocephalinae) from the Hawaiian archipelago (Shapiro et al., 2006). We relied on the *Banza* rate (0.04 substitutions per site Myr<sup>-1</sup>) because it represents Tettigoniidae, confirmed geological calibration points, and phylogenetic changes restricted to the last 3 Mya. Mitochondrial DNA mutation rates can vary considerably between different taxonomic groups (Papadopoulou et al., 2010), and short-term (1–2 Mya) rates are generally higher than those estimated over longer time intervals (Ho et al., 2005). We used a normal distribution of priors to infer the lower and upper clock rate estimates, setting the mean  $\pm$  SD = 0.04  $\pm$  0.0051 and the 97.5% highest posterior density (HPD) = 0.03–0.05. Given that COII clock rates were not available for Tettigoniidae, we assumed the same rate as for COI but specified a broader normal distribution of priors, setting the mean  $\pm$  SD = 0.04  $\pm$  0.235 and the 97.5% HPD = 0.0–0.501. We used lognormal relaxed-clock models to accommodate variation in mean branch rate among gene trees. The tree models for gene partitions were linked, and we used a birth–death tree prior. The MCMC chain length was 100 000 000, sampled every 10 000 generations. Convergence of parameters and tree summarization were evaluated in BEAST as above. Given the marked differences in male song and female preference, we predicted ancient divergence times, early during the Pleistocene or even before, between the major clades.

## Geographical patterns of genetic diversity

We evaluated genetic diversity with a set of microsatellite data from 428 insects collected at 24 of the localities (Table 1). Fifteen microsatellite loci were amplified in each insect using the primer

sequences and PCR conditions described by Esquer-Garrigos et al. (2016). One locus, Eeph\_8, from that previous study was not included here because its size overlap with other loci in some localities hindered multiplex PCR. In addition to the 15 loci noted above, we amplified five loci previously characterized by Hockham et al. (1999) and Hamill et al. (2006). After amplification, fragment analysis was performed on a 3730 xl DNA Analyzer (ABI), and genotypes were scored in GeneMapper v.5.0 (ABI) (Supporting Information, Appendix S2).

Using genotype data from the 20 microsatellite loci above, we calculated the number of alleles ( $N_a$ ), rarefied allelic richness (ARr) based on the smallest sample size ( $n = 6$ ), observed heterozygosity ( $H_o$ ) and expected heterozygosity ( $H_e$ ) for each locality using the ‘genepop’ and ‘hierstat’ packages in R (Goudet, 2005; Rousset, 2008). We assumed that Pleistocene refuge(s) were situated near the Mediterranean coast within the species’ current range and predicted higher genetic diversity indices at low-elevation coastal localities, where populations should generally have been stable over a very lengthy period (Hewitt, 1999). At a local scale, we also evaluated the pattern of genetic diversity along river valleys, which we assumed served as protected colonization routes to inland and montane areas as the climate warmed after the Pleistocene (cf. Spooner & Ritchie, 2006). Along valleys with multiple localities, we predicted lower genetic diversity in upstream populations, which should have undergone additional genetic bottlenecks. We inferred historical changes in demography using an approximate Bayesian computation framework visualized in skyline plots (Supporting Information, Appendix S3).

We checked for linkage disequilibrium between the 20 microsatellite loci, using genepop v.4.2 (Raymond & Rousset, 1995; Rousset, 2008) as implemented by the ‘genepop’ package in R (R Core Team, 2016), and concluded that the markers are largely independent. Among all 24 localities we found significant ( $P < 0.05$ ) linkage disequilibrium in only 180 out of a total 4560 pairwise comparisons between loci, and these associations were not consistent across pairs of loci or localities.

### **Genetic differentiation in the landscape**

We evaluated the level of genetic differentiation among sampled localities by calculating pairwise fixation index ( $F_{ST}$ ) values from our microsatellite data, using genepop v.4.2 as above. Given that basic analyses of our microsatellite data revealed many null alleles among loci and localities (Supporting Information, Appendix S2 and Table S1), we calculated  $F_{ST}$  using data that included null alleles and then repeated our calculations using the FreeNA correction (Chapuis & Estoup, 2007) that excluded them.

We examined the role of ‘isolation by distance’ (Wright, 1943) in genetic differentiation via a Mantel test that compared the matrices for pairwise  $F_{ST}$  values and pairwise geographical separations (Diniz-Filho et al., 2013). The Mantel tests were performed with 10 000 iterations in XLSTAT v.2018.1 and applied first to matrices compiled from all 24 localities and then to matrices compiled separately from clade 1 and clade 2 localities. We also examined the basic correlation between pairwise  $F_{ST}$  values and pairwise geographical separations. Given the low mobility of *E. diurnus*, we predicted that  $F_{ST}$  values would strongly reflect the geographical distance separating localities.

### **Genetic structure of *E. diurnus* populations**

We used our microsatellite dataset to construct a neighbour-joining (NJ) tree of populations (Saitou & Nei, 1987) based on Nei’s DA measure of genetic distance (Nei et al., 1983) and visualized in POPTREE2 (Takezaki et al., 2010). The robustness of the nodes was estimated using 1000 bootstrap replicates obtained by resampling loci. Our new NJ tree improved upon a recently

published tree (Greenfield et al., 2016) generated from fewer insects, localities and microsatellite markers, and it allowed us to describe the geographical ranges of clades and subclades from the mtDNA phylogenies, and gene flow between these divisions, with greater precision.

Interclade gene flow

We used a Bayesian approach (STRUCTURE v.2.3.3; Pritchard et al., 2000) to identify the most likely number (K) of distinct genetic clusters in our microsatellite data and assign individual insects to their most probable cluster (Supporting Information, Appendix S4). At all sampled localities, we evaluated the assignment probabilities  $q$  of every individual to each of the K clusters. Results were analysed and displayed with Clumpak v.beta (Kopelman et al., 2015). We used the cluster analysis to evaluate recent gene flow, i.e. introgression, between clades 1 and 2 and between subclades within clade 1. Two conditions distinguished recent gene flow from incomplete lineage sorting that occurred at the time of divergence (cf. Zhou et al., 2017): (1) X and Y are the majority clusters throughout clades A and B, respectively, and the representation of minority cluster Y in clade A increases in cline-like fashion across localities towards the geographical boundary with clade B; and (2) given condition 1, the representation of minority cluster Y in the clade A locality nearest the interclade boundary is highly variable among individuals, i.e. some individuals are dispersing parental forms, whereas others are recently formed hybrids. Our estimates of gene flow indicate the level of current separation in zones of potential secondary contact following differentiation during Pleistocene isolation.

We performed a discriminant analysis of principal components (DAPC; Jombart et al., 2010) on our microsatellite data to obtain an additional ordination of individuals among genetic groups (Supporting Information, Appendix S5). Unlike the Bayesian approach, DAPC does not rely on any population genetic model.

## **RESULTS**

### **Divergence of clades and major subclades**

Our molecular clock calculations indicated ancient divergence times for the two main *E. diurnus* clades and some subclades. Separation of clades 1 and 2 is estimated at 2.78 Mya, with 1.48 Mya as a younger limit (Fig. 2). Within clade 1, the subclade in the Cerdagne Valley is estimated as having diverged 1.23 Mya, with 0.35 Mya as a younger limit.



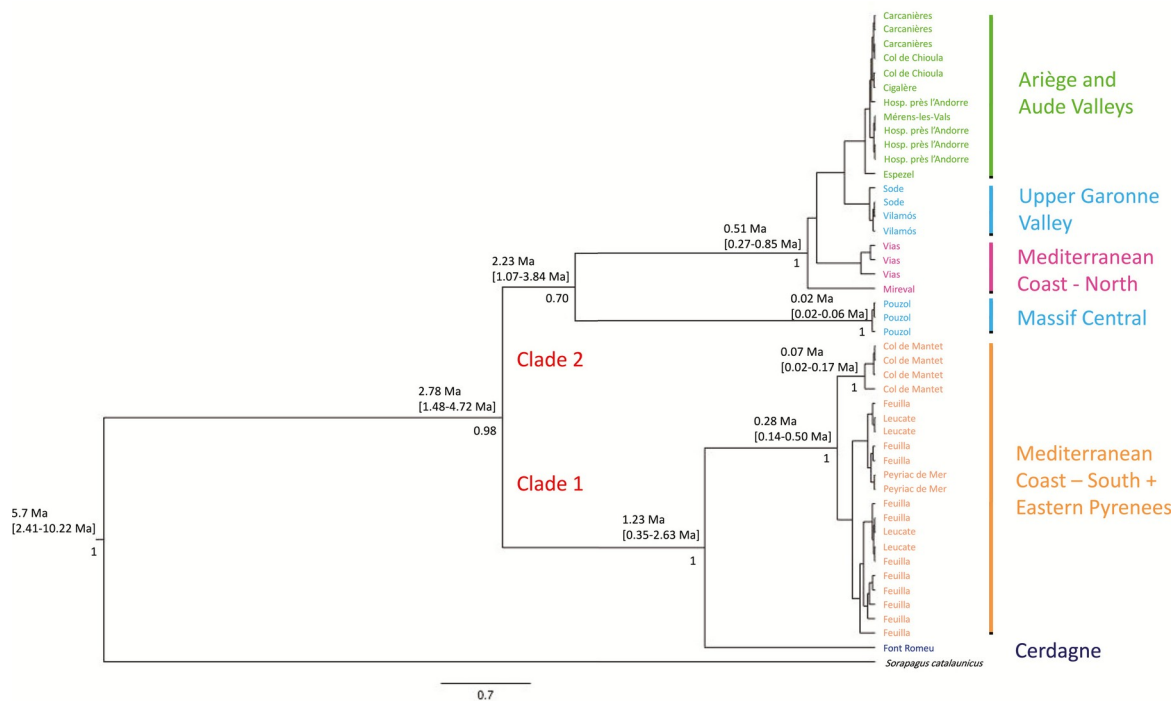


Figure 2. Phylogeny and estimated bifurcation dates of *Ehippiger diurnus* populations in southern France and northeastern Spain indicated by mitochondrial DNA data (cytochrome oxidase c subunits I and II; COI and COII). Data represent samples from 45 *E. diurnus* individuals whose COI or COII sequences lacked ambiguities, gaps and heteroplasmic sites (see Supporting Information, Appendix S1); the outgroup is *Sorapagus catalaunicus*. The phylogeny is a concatenation of COI and COII data. Values below nodes represent posterior probabilities; values above nodes are median divergence dates estimated by molecular clock rate, with the 95% highest posterior density range for divergence given in parentheses. Horizontal line at the bottom represents time scale in millions of years. Populations are colour-coded according to geographical region, listed at the far right; population localities are given in Table 1 and Figure 1.

## Geographical patterns of genetic diversity

Our 24 sampled localities exhibited markedly different levels of genetic diversity (Table 2), and two striking patterns emerged readily from geographical analysis. As predicted, three of the four measures of genetic diversity, ARr, Ho and He, decreased significantly with both linear distance from the Mediterranean and higher elevation (Pearson correlation,  $P < 0.01$ ; see Fig. 3 for ARr relationships). This pattern was also found within the nine clade 1 localities, but within the 15 clade 2 localities the decreases were found only with elevation.

Table 2. Genetic diversity parameters as determined from analysis of 20 microsatellite loci at 24 of the sampled localities

Locality	Na*	ARr*†	Heterozygosity*	
			Ho	He
Aubrac	4.2	2.61	0.42	0.51
Carcanières	2.45	2.11	0.37	0.4
Cigalère	3.8	2.68	0.45	0.5
Col de Chioula	5.5	2.85	0.43	0.56
Col de la Croix Morand	4.15	2.74	0.43	0.54
Col de Mantet	5.05	2.96	0.36	0.62
Col de Puymorens	3.15	2.45	0.33	0.51
Espezel	3.2	2.55	0.44	0.46
Feuilla	5.35	3.98	0.47	0.78
Font Romeu	4.65	2.95	0.4	0.6

Hospitalet près l'Andorre	5.4	2.44	0.43	0.45
Latour-de-Carol	9.85	2.91	0.41	0.57
Le Lioran	5.95	2.61	0.35	0.49
Leucate	4.05	3.61	0.49	0.7
Massif de la Clape	4.05	3.6	0.46	0.69
Mérens-les-Vals	9.4	2.4	0.34	0.47
Mireval	6.55	3.89	0.55	0.74
Montségret	6.5	3.82	0.46	0.73
Peyriac de Mer	5.95	3.42	0.54	0.7
Port de Lers	7.9	2.56	0.41	0.51
Pouzol	3.8	2.85	0.38	0.54
Sode	5.5	2.68	0.39	0.52
Vias	9	3.73	0.49	0.71
Vilamós	6.75	2.91	0.41	0.53

Abbreviations: ARr, rarefied allelic richness; He, expected heterozygosity; Ho, observed heterozygosity; Na, number of alleles.

\*Calculations according to Goudet (2005) and Rousset (2008).

†Rarefied allelic richness based on six randomly chosen insect samples.

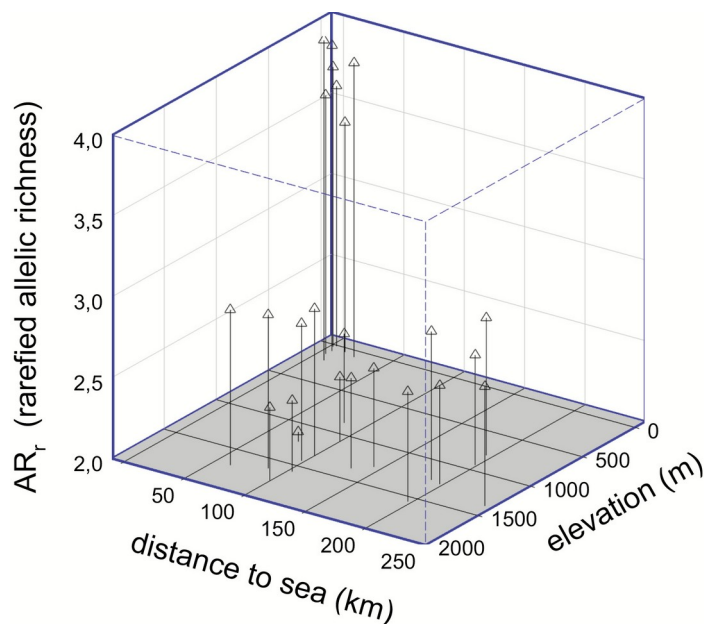


Figure 3. Genetic diversity estimates for 24 *Ephippiger diurnus* populations sampled in southern France and northeastern Spain for microsatellite loci (Table 1; Fig. 1). Graph indicates rarefied allelic richness (ARr) vs. the shortest straight-line distance to the Mediterranean Sea (Pearson  $r = -0.661$ ,  $P < 0.001$ ) and vs. elevation above sea level ( $r = -0.818$ ,  $P < 0.001$ ).

Over a smaller geographical scale, we found similar relationships within several of the major subclades associated with river valleys. Along the Cerdagne Valley (three localities) in clade 1 and the Alagnon (two localities), Ariège (three localities), Aude (two localities) and upper Garonne Valleys (three localities) in clade 2 we found lower ARr values in the populations situated farther upstream and/or at relatively higher elevation (Table 3). The ARr values for coastal populations were generally high (Fig. 3), and the highest values were found at Feuilla in clade 1 and at Mireval in clade 2 (Table 2). Our approximate Bayesian computation did not indicate major differences in demographic history between coastal and inland populations (Supporting Information, Appendix S3 and Figure S1). Instead, all populations showed a constant or increasing size over time.

Table 3. Rarefied allelic richness (ARr) in relationship to relative upstream position and relative elevation for localities situated along river valleys in the Pyrénées and the Massif Central

Valley	Locality*	Relative upstream	Relative	ARr
--------	-----------	-------------------	----------	-----

		position†	elevation‡	
Cerdagne	Latour-de-Carol (7)	1	1	2.91
	Font Romeu (6)	2	3	2.95
	Col de Puymorens (8)	2	4	2.45
Aude	Espezel (12a)	1	1	2.55
	Carcanières (12)	2	2	2.11
Ariège	Col de Chioula (11)	1	2	2.85
	Mérens-les-Vals (10)	2	1	2.4
	Hospitalet près de l'Andorre (9)	3	3	2.44
Upper Garonne	Sode (15a)	1	1	2.68
	Vilamós (15)	1	3	2.91
	Cigalère (14)	1	5	2.68
Alagnon	Pouzol	1	1	2.85
	Le Lioran	4	4	2.61

\*Parenthetical values are locality numbers on maps in Figures 1 and 7A.

†Values indicate 10 km increments in an upstream direction.

‡Values indicate 100 m increments in elevation.

## Genetic differentiation in the landscape

Analyses of pairwise  $F_{ST}$  values supported the involvement of isolation by distance in the genetic differentiation of *E. diurnus* populations. Our initial examination indicated that  $F_{ST}$  values determined from data excluding null alleles did not differ from values determined from data that included them (Pearson  $r = 0.98$ ), and we use the former in all calculations below. The matrices of pairwise  $F_{ST}$  values (A) and pairwise inter-locality distances (B), measured for all 24 sampled localities, were significantly correlated [Mantel test;  $R(AB) = 0.219$ ,  $P = 0.001$ ]. However, calculations using data restricted to the nine clade 1 [ $R(AB) = 0.764$ ,  $P < 0.0001$ ] or to the 15 clade 2 localities [ $R(AB) = 0.873$ ,  $P < 0.0001$ ] revealed much stronger correlations between  $F_{ST}$  and distance matrices, suggesting that processes in addition to isolation by distance were responsible for genetic differentiation. For clade 2, we repeated the calculations after removing the four Massif Central localities so that the matrices and geographical regions would be closer in size to those used for clade 1 (cf. Diniz-Filho et al., 2013). Again, we found a highly significant correlation between the matrices [ $R(AB) = 0.888$ ,  $P < 0.0001$ ].

Simple comparison between pairwise  $F_{ST}$  values and inter-locality distances corroborated our Mantel test results. Analysing data from all 24 sampled localities,  $F_{ST}$  values and inter-locality distances were significantly correlated (Pearson  $r = 0.204$ ,  $P = 0.001$ ; Fig. 4). However, Fig. 4 shows that although no widely separated locality pairs had low  $F_{ST}$  values, many neighbouring locality pairs had high values. We focus on these latter anomalies in the section on interclade gene flow below.

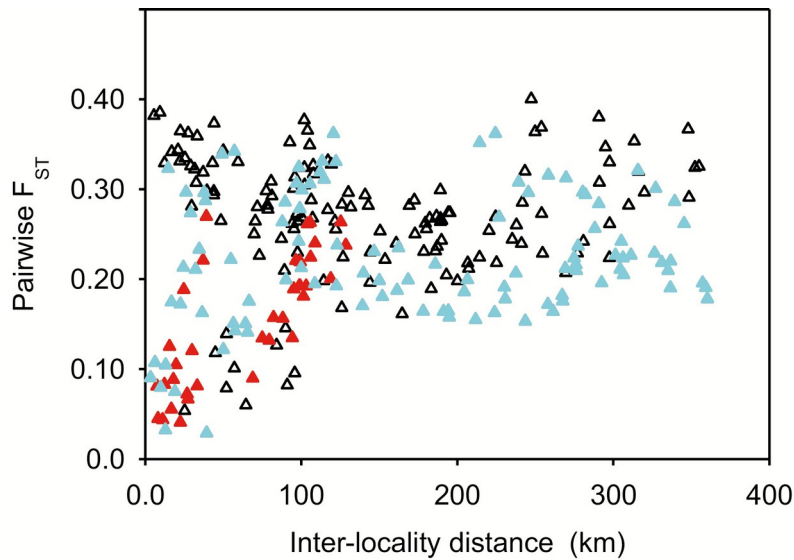


Figure 4. Pairwise fixation index ( $F_{ST}$ ) vs. inter-locality distance for 24 *Ehippiger diurnus* populations sampled in southern France and northeastern Spain (Table 1; Fig. 1) for 20 microsatellite loci (Pearson  $r = 0.204$ ,  $P = 0.001$ ). Symbols: red triangles, pairs of clade 1 populations; blue triangles, pairs of clade 2 populations; and open triangles, interclade population pairs.

### Genetic structure of *E.diurnus* populations

Our unrooted NJ tree (Fig. 5) is topologically comparable to the phylogenetic trees we constructed from concatenated COI and COII mtDNA datasets (Fig. 2) and to the NJ tree constructed from a smaller set of microsatellite data (Greenfield et al., 2016). This current tree, constructed from our expanded set of microsatellite data, presents a strongly supported structure (posterior probability > 0.80) corresponding to clades 1 and 2 and to the major subclades found in specific geographical regions (Fig. 2). These regions are as follows: Cerdagne Valley (subclade A); southern Mediterranean coast + eastern Pyrenees (subclade B); northern Mediterranean coast (subclade C); Ariège + Aude valleys (subclade D); and upper Garonne valley + Massif Central (subclade E). The correspondence between the unrooted NJ tree and the clades should not be construed to imply that the NJ tree reflects phylogeny. Rather, we use microsatellite data from the NJ tree only to obtain more precise genetic characteristics for the clades and subclades found in specific regions and to use these characteristics to estimate interclade and intersubclade gene flow.

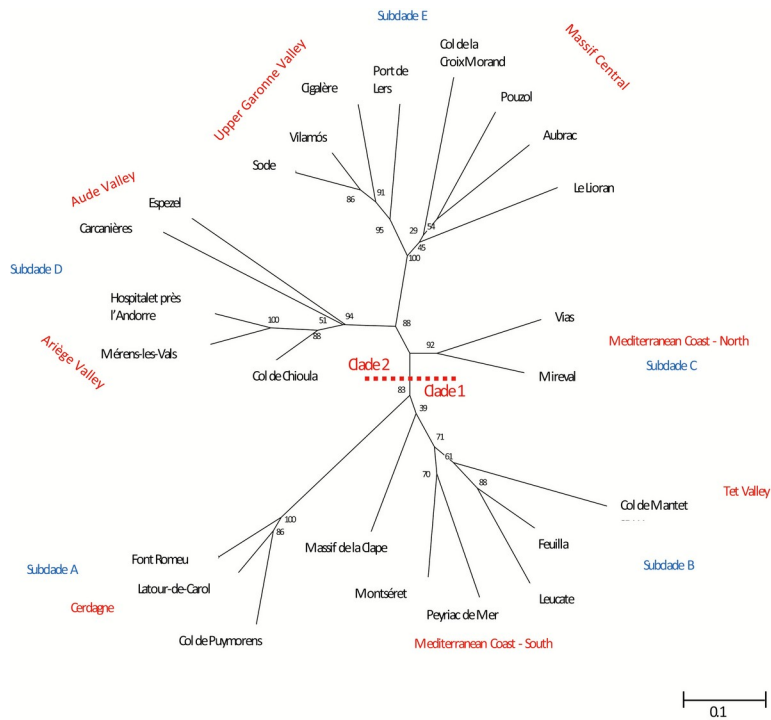


Figure 5. Unrooted neighbour-joining (NJ) tree showing genetic structure of 24 *Ehippiger diurnus* populations, sampled in southern France and northeastern Spain (Table 1; Fig. 1), as generated by data from 20 microsatellite loci. Scale at bottom right indicates 0.1 nucleotide substitutions per site; values at nodes represent posterior probabilities. Geographical regions and subclades (from mitochondrial DNA phylogeny; Fig. 2) are labelled in red and blue, respectively, outside the branches. Indication of subclades does not imply that the NJ tree reflects phylogeny; instead, these indications are given only to specify the microsatellite data associated with populations in the mitochondrial DNA phylogeny.

### Interclade gene flow

Designating  $K = 5$ , the most likely number of distinct genetic clusters in our data (Supporting Information, Appendix S4 and Figure S2), we analysed  $q$  values, the assignment probabilities of individuals to each of the five clusters, in localities at potential secondary contact zones between clades 1 and 2. The five clusters, designated A–E, are each the majority one in the corresponding subclade (A–E) and region shown in Fig. 5. Our analyses revealed substantial introgression between clades 1 and 2 where they approach each other on the Mediterranean coast near Narbonne. Within clade 1 localities, mean  $q$  values for (minority) cluster C increased markedly toward the boundary with clade 2 (Figs 6, 7), and in the two localities nearest the boundary, Montsérret and Massif de la Clape,  $q$  values for cluster C varied considerably among sampled individuals (coefficients of variation = 1.27 and 0.43, respectively; Fig. 6). These conditions remained when we repeated our cluster analysis with  $K = 6$  and  $K = 7$ . Within clade 2 localities, mean  $q$  values for (minority) cluster B remained uniformly low, and there was no clear increase towards the boundary with clade 2 for  $K = 5–7$  (Figs 6, 7). Thus, introgression appeared to be unidirectional, from clade 2 into clade 1.

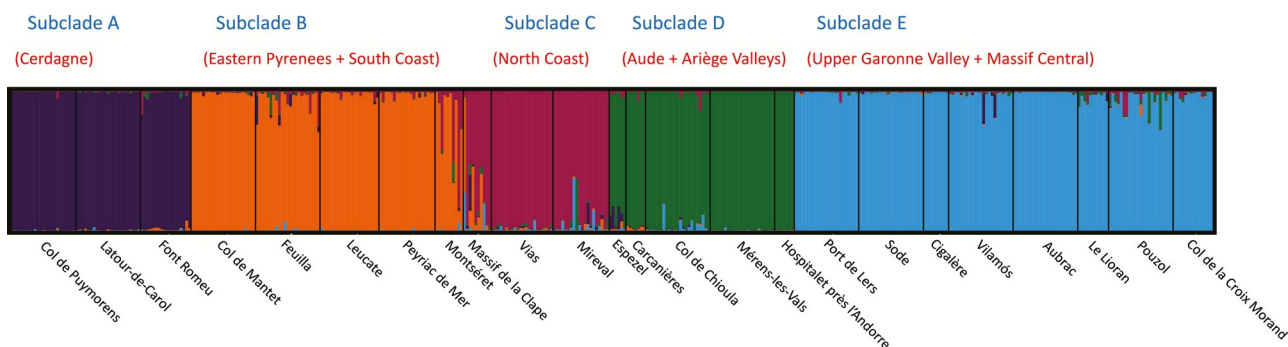


Figure 6. Bayesian STRUCTURE plot (Pritchard et al., 2000) showing the assignment of *Ehippiger diurnus* to five genetically distinct clusters, each coded by a different colour; see the Supporting Information (Appendix S4) for determination that five is the most likely number of distinct clusters. Clusters are based on data from 20 microsatellite loci sampled from 428 individuals collected at 24 localities in southern France and northeastern Spain. The 428 individuals are each represented by a narrow vertical line, and the probability ( $q = 0-1$ ) of assignment of an individual to a given cluster is indicated by the height of the line segment bearing the colour of that cluster. Thus, individuals represented by lines composed of two or more segments of different colours may belong to the several distinct clusters, whereas those represented by lines composed of a single colour are likely to be pure. Subclades (Figs 2 and 5) and geographical regions strongly associated with each cluster are indicated in blue and red, respectively, above the plot.

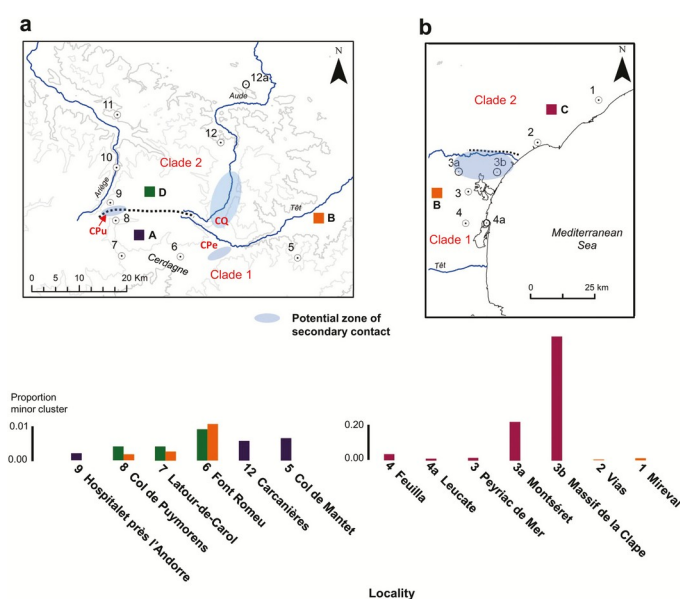


Figure 7. Zones of potential secondary contact between clades and levels of possible interclade introgression. Coloured squares designated by letters indicate the majority genetic cluster (see Fig. 6) on each side of a contact zone. Coloured vertical bars below the maps indicate the identity and proportion of the minority genetic cluster(s) in each locality adjacent to or within the contact zones. A, region of potential contacts surrounding the Cerdagne Valley in the Pyrenees (see Fig. 1); contour lines represent 1200 and 1800 m elevations. Abbreviations: CPe, Col de la Perche (1581 m); CPu, Col de Puymorens (1920 m); and CQ, Col de Quillane (1713 m). B, region of potential contact midway along the Mediterranean coast. The minority genetic cluster for locality 3b (Massif de la Clape) is designated as C, despite a proportion = 0.80, because overall this population is more similar to locality 3a (Montséret) than to localities 2 and 1 (Vias and Mireval) (Fig. 5). Minority cluster proportions are calculated as mean  $q$  values (see Fig. 6).

Clades 1 and 2, and two different subclades within clade 2, also approached one another around the Cerdagne Valley in the Pyrenees, but we did not observe high levels of introgression in this region (Fig. 7A). At the Col de Puymorens (elevation 1920 m), clades 1 and 2 are separated by only 5 km,

but mean  $q$  values for (minority) cluster D to the south (Col de Puymorens and Latour-de-Carol localities, clade 1) and for (minority) cluster A to the north (Hospitalet près l'Andorre locality, clade 2) are essentially zero for  $K = 5-7$ . At the Col de Quillane (elevation 1713 m), clades 1 and 2 are separated by 20 km, and mean  $q$  values for (minority) cluster D to the south (Font Romeu locality, clade 1) and for (minority) cluster A to the north (Carcanières locality, clade 2) are also very low but slightly higher than at the Col de Puymorens (Fig. 7A). We have repeatedly made extensive field observations at both cols and failed to find *E. diurnus* at the summits. Likewise, two different subclades of clade 1 approach one another at the Col de la Perche (elevation 1581 m) at the east end of the Cerdagne Valley, and mean  $q$  values for minority cluster B to the west (Font Romeu locality, Cerdagne Valley) and for minority cluster A to the east (Col de Mantet, Tet River valley) are very low (Fig. 7A). Among the three Cerdagne Valley localities, we did observe a small increase in the mean  $q$  values for (minority) cluster B towards the east and the boundary with the subclade representing the eastern Pyrenees + the southern Mediterranean coast (Fig. 7A). Moreover, at the Font Romeu locality the  $q$  values for both minority clusters (B and D) varied considerably among sampled individuals (coefficients of variation = 1.55 and 1.81, respectively; Fig. 6). Evidently, low levels of introgression occurred at the Col de Quillane and the Col de la Perche, but not at the Col de Puymorens.

Discriminant analysis (DAPC) of our microsatellite data identified more genetic groups than the Bayesian approach, but the nine DAPC groups were ordinated in a similar manner to the five Bayesian clusters (Supporting Information, Appendix S5 and Figure S3). The groups were distributed in three distinct sectors: a sector of one group corresponding to subclade A (Cerdagne Valley) in clade 1, a sector of two groups corresponding to subclade B (eastern Pyrenees and the southern Mediterranean coast) in clade 1, and a sector of six groups corresponding to the several subclades (C, D and E) in clade 2.

## DISCUSSION

### Pleistocene divergence, refuges and recolonization

Analysis of the phylogeny constructed from our concatenated mtDNA dataset confirmed the suggestion by Spooner & Ritchie (2006) that divergence within *E. diurnus* began early. Our molecular clock approach (Fig. 2) indicates the primary divergence, between clades 1 and 2, at 2.7 Mya, approximating the Pliocene–Pleistocene boundary. Within clade 1, we estimate divergence of subclade A in the Cerdagne valley from subclade B on the southern Mediterranean coast and in the Tet valley at 1.2 Mya. Even allowing for the wide confidence intervals surrounding these estimated dates, clades 1 and 2 and subclades A and B passed at least the most recent glacial period (Würm glaciation, 117 000–11 000 years ago) separated from one another. These estimated divergence dates fall at the beginning of the range reported for European biota (Hewitt, 2000) and might reflect strong adaptation to warm climate and low mobility.

The pattern of genetic diversity in *E. diurnus* populations is consistent with the model of a Mediterranean refuge during Pleistocene glacial periods (Hewitt, 1999; Schmitt, 2007; Stewart et al., 2010). Higher ARr values exist at all coastal localities, indicating populations that had been relatively stable over a lengthy period. No observations suggest that higher genetic diversity along the coast reflects higher population sizes or greater connectedness between these populations. On the contrary, only one coastal locality, Vias, had relatively high populations over the years of our sampling, whereas five of the montane localities, Col de Mantet, Font Romeu, Col de Puymorens, Mérens les Vals and Col de Chioula, regularly did. Moreover, we have no reason to expect higher effective population sizes at coastal localities. Owing to greater urbanization and agricultural land use in coastal areas, connectedness of *E. diurnus* populations along the coast would have been seriously disrupted for many years.

At a finer scale, the pattern of genetic diversity suggests recolonization up river valleys after the Last Glacial Maximum, 22 000 years ago, and the Younger Dryas, 13 000 years ago. The  $AR_r$  values decreased with distance upstream and/or elevation along all five river valleys for which multiple localities were sampled. *Ephippiger diurnus* probably advanced in stepping-stone fashion between habitable localities, and each advance would have represented a founder population that retained only a portion of the genetic variance of the lower, downstream site from which it originated. This inference is consistent with the expectation that post-glacial expansion in a warm-adapted Mediterranean species would have proceeded initially along climatically protected corridors.

Unlike many Mediterranean species, no information on *E. diurnus* points towards a Pleistocene refuge far to the south on a European peninsula, or beyond. Along the Mediterranean coast, *E. diurnus* is currently distributed from the Maritime Alps in southeastern France to the Ebro River in the northeast corner of Spain (IUCN Red List of Threatened Species, 2017). Two other *Ephippiger* species, *Ephippiger provincialis* and *Ephippiger terrestris*, are sympatric with *E. diurnus* in the southeast corner of France, yet other species replace *E. diurnus* in Italy, the Balkan Peninsula and central and eastern Europe, and the genus is not found southwest of the Ebro River (Barat, 2012). From our estimated divergence dates, we can infer at least three separate refuges within that coastal range during the most recent glaciation, one for clade 2 to the north and two for subclades A and B of clade 1 to the south. However, we cannot pinpoint the locations of these refuges with available information. Additional, ‘cryptic refuges’ might have existed for subclades D and E of clade 2, possibly in protected inland sites (Schmitt, 2007; Kühne et al., 2017). Isolation by distance, secondary contact and interclade gene flow

Mantel test results show that isolation by distance is a significant factor, but not the sole one, in the differentiation of *E. diurnus* populations. In many cases, localities separated by relatively short distances nonetheless feature high pairwise  $F_{ST}$  values, and we focused on these sharp discontinuities to understand gene flow and the level of genetic differentiation in *E. diurnus*. The most marked genetic discontinuities surround the Cerdagne Valley in the Pyrenees (cf. Oudman et al., 1989, 1990) and run perpendicular to the Mediterranean coast near the Aude River.

Three different subclades, A, B and D, are parapatric along the northern and eastern boundaries of the Cerdagne Valley, but little or no introgression is observed. Individuals from the three subclades are positioned in distinct, largely non-overlapping sectors in our discriminant analysis (Supporting Information, Appendix S5), and their estimated divergences may predate 1.2 Mya. Of three potential secondary contact zones bordering the Cerdagne Valley, separation of subclades A and D at the highest zone, Col de Puymorens, is most easily understood; low winter temperatures are likely to prevent egg development and thus preclude settlement (Kidd & Ritchie, 2000). However, neither the Col de Quillane nor the Col de la Perche exceeds elevation limits for *E. diurnus* in the area, and the negligible introgression between subclades A and D and between subclades A and B across these two zones is curious (cf. Spooner & Ritchie, 2006). Although low mobility and possible colonization of high elevations only in the last millennium (Mahamoud-Issa et al., 2017), when interglacial climate warmed to its present conditions (Terral & Mengual, 1999), are two explanations for the absence of mixture, analysis of the genetic discontinuity on the Mediterranean coast offers a very different and more complete perspective.

In contrast to potential contact in the Pyrenees, clade 2 (subclade C) does exhibit introgression into clade 1 (subclade B) on the coast. Mean  $q$  values for cluster C increase abruptly over a short distance, with no obvious physical barrier, between Peyriac de Mer and Vias. This steep cline has the characteristics of a narrow hybrid ‘tension zone’ (Barton & Hewitt, 1985), and it indicates some amount of genetic incompatibility between clades 1 and 2. That incompatibility might be expressed



as relative non-viability in descendants of hybrid crosses, which would not have been detected in previous laboratory studies. It is relevant that we observed *E. diurnus* populations at the two hybrid localities, Montséret and Massif de la Clape, to be very sparse in comparison with localities in the core coastal areas of clade 1 to the south and of clade 2 to the north. Unlike the situation in the Pyrenees, *E. diurnus* probably dispersed in coastal areas many millennia ago, and recent colonization is not a likely explanation for the narrow hybrid zone.

## **Speciation**

We infer that clades 1 and 2 are distinct species (taxa retaining their identities despite gene flow; Mallet, 2008; Harrison & Larson, 2014) based on their separation in both montane and coastal areas. Their mtDNA divergence greatly exceeds the general level distinguishing eastern and western subspecies of plants and animals in Europe (Taberlet et al., 1998), and it is typical of interspecific differences in insects (e.g. Funk, 1999). The process of reinforcement (Servedio & Noor, 2003; Pfennig, 2016), potentially effective where dispersal is low but not entirely absent (Nosil et al., 2003), could have played a role in behaviourally isolating the clades at the Col de Quillane in the Pyrenees to the extent that a hybrid population is no longer present. For example, differences in the daily timing and microhabitat of mating activity might have increased, and subtle differences in song and preference might have been linked with that activity. Such reinforcement might also have been operating in the coastal contact zone. But reinforcement can be a dynamic process, and we might currently be witnessing a phase where isolation, possibly afforded by behaviour expressed specifically in the field, has partly broken down along the coast. Subclades A and B within clade 1 also have features characteristic of distinct species. Genetic divergence is high, and introgression between Font Romeu (subclade A) in the Cerdagne Valley and Col de Mantet (subclade B) in the Tet Valley, localities separated by 25 km, is minimal. After the Last Glacial Maximum and Younger Dryas, ancestors of subclade A populations most probably colonized the Cerdagne Valley from the Mediterranean coast via the Segre River Valley, ~100 km south of the border between France and Spain. This route offered the lowest, therefore most protected, entrance to the Cerdagne Valley, and its length is consistent with the long branch of subclade A in the *E. diurnus* NJ tree and DAPC ordination. But we currently have no information pertaining to a potential contact zone between subclades A and B along the coast in the northeast corner of Spain, and species designation for these groups remains tentative. At this point, we refrain from linking our species designations to previous subspecies and species names, e.g. *crucifer*, *cunii*, *vitium*, from the *E. diurnus* ‘complex’. These previous names were largely associated with colour forms and morphology, and their geographical ranges might not coincide exactly with ranges of the clades and subclades from the present study.

## **Complex phylogeography within a small area**

Our study demonstrates that relatively endemic species, facing appropriate circumstances, can have biogeographical features at least as complex as those typically observed in more widespread taxa. In the case of the Mediterranean fauna of Europe, a species with limited range may exhibit strong phylogeographical structure originating from close, yet isolated, Pleistocene refugia. These refugia may not fit the expectation of location far onto a southern peninsula, a Mediterranean island or beyond. Low dispersal is likely to be coincident with limited range, and the lack of mobility would have restricted the areas capable of serving as climatic refuges during glacial periods. Strong genetic and phenotypic differentiation may distinguish clades and subclades that originated from the different refugia. In addition to but associated with low dispersal, differentiation was probably influenced by genetic bottlenecks created when small populations were isolated in limited areas. Moreover, the four successive Pleistocene glacial periods in Europe would have repeatedly exerted the above influences, each period adding to the differentiation acquired during previous episodes. This ‘accumulation effect’ (cf. Spooner & Ritchie, 2006) would have been facilitated by low dispersal that minimized the mixing of clades in glacial refugia and during interglacial expansion.

Thus, relatively endemic species, such as *E. diurnus*, can exhibit higher (not lower) variation and divergence than more widespread species.

Elsewhere, similar differentiation after the Pleistocene glaciations has been reported in Andean anostomatid crickets (Alfaro et al., 2018), which are also flightless Orthoptera. Such findings bolster the notion that low dispersal is conducive to divergence and cryptic speciation (Avice, 2009).

In *E. diurnus*, we observe differentiation most clearly in the male song and chorusing variants displayed by the subclades and even by individual populations. These variants range from songs that include a single syllable in Massif Central populations (Pouzol, Col de la Croix Morand) to those that include six or more syllables in populations along the southern Mediterranean coast (Leucate) (Table 1). We show that this variation is a legacy of the Pleistocene, and it represents an important element of present-day biodiversity. But our study also demonstrates that not all *E. diurnus* diversity is equal, because coastal populations harbour substantially more genetic diversity than inland localities. Extra effort is needed to conserve the Mediterranean populations, and similar considerations should be adopted in other species where geographical patterns of genetic diversity are found.

## **SHARED DATA**

Data deposited in the Dryad digital repository (Esquer-Garrigos et al., 2018).

## **SUPPORTING INFORMATION**

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

Appendix S1. Mitochondrial DNA amplification and analysis.

Appendix S2. Microsatellite loci and analysis.

Appendix S3. Historical demography.

Appendix S4. Bayesian clustering.

Appendix S5. Discriminant analysis of principal components.

Table S1. Population genetic parameters as determined from microsatellite analysis at 24 sampled localities.

Figure S1. Bayesian skyline plots of estimated effective population size (logarithmic scale) vs. time before present for five *Ephippiger diurnus* subclades (Fig. 5). Subclade B does not include Montséret and Massif de la Clape because of high levels of recent introgression at these localities (Figs 6, 7B). Black continuous lines represent median effective population size, and dotted lines are 95% highest posterior density intervals. BF is the Bayes factor for comparison of a model assuming population growth and a model assuming stable population size.

Figure S2. Bayesian clustering, by microsatellite loci data, of 428 *Ephippiger diurnus* sampled among 24 populations in southern France and northeastern Spain. A, mean and standard deviation of  $\ln(\text{probability of data})$  for each inferred number of clusters ( $K = 1\text{--}15$ ) as estimated from

STRUCTURE analysis. Values are computed over 20 runs per K. B, delta (K), defined as  $[\ln(\text{probability of data})]K - [\ln(\text{probability of data})]K - 1$ , vs. K for K = 2–14.

Figure S3. Discriminant analysis of principal components (DAPC; Jombart et al., 2010) applied to microsatellite data from 428 *Ephippiger diurnus* individuals collected at 24 localities in southern France and northeastern Spain (Fig. 6). A, DAPC showing the first two principal components. Nine groups are indicated by different colours and inertia ellipses; dots in the scatterplot represent the 428 individuals. See the Supporting Information (Appendix S5) for determination that nine is the most likely number of distinct groups. The nine groups are distributed in three sectors that largely correspond to subclade A (group 1), subclade B (groups 8 and 9) and subclades C, D and E (groups 2–7) (cf. Fig. 5). B, apportionment of the 24 populations, listed to the right, among nine groups, listed above. The size of the filled rectangle indicates the number of individuals from a population that are assigned to a given group.

## ACKNOWLEDGEMENTS

We thank F. Barbosa, G. Bourdais, O. Brunel-Pons, M. Deluen, S. Devers, J. Gibson, M. Goubault, C. Hebert, A. Kerbrat, M. Mahamoud-Issa, T. Marin-Cudraz, J. Penin, F. Plault, D. Rebar and V. Terwilliger for technical assistance, K. Berthier and S. Piry for help with analyses, and the Agence Nationale de la Recherche de France (contract ANR-11-BSV7-025-01; EVOLCHOR), the Centre National de la Recherche Scientifique (CNRS) and the Université de Tours for their financial support. Genotyping of microsatellite markers was accomplished at the Centre Méditerranéen Environnement Biodiversité (LabexCeMEB). G. Baudouin created the map of *E. diurnus* populations, and K. Berthier, M.-P. Chapuis, D. Rebar and several anonymous referees reviewed earlier versions of this manuscript.

## REFERENCES

- Alfaro FM, Muñoz-Ramírez CP, Zuñiga-Reinosos A, Trewick SA, Méndez MA. 2018. Phylogeography of the Chilean red cricket *Cratomelus armatus* (Orthoptera: Anostomatidae) reveals high cryptic diversity in central Chile. *Biological Journal of the Linnean Society* 123: 712–727.
- Avise J. 2009. Phylogeography: retrospect and prospect. *Journal of Biogeography* 36: 3–15.
- Barat J. 2012. Revisión preliminar de los géneros de *Ephippigerini* Brunner von Wattenwyl, 1878 (Orthoptera: Tettigoniidae: Bradyporinae). *Boletín de la Sociedad Entomológica Aragonesa (S.E.A.)* 50: 1–71.
- Barbosa F, Rebar D, Greenfield MD. 2016. Female preference functions drive inter-population divergence in male signaling: call diversity in the bushcricket *Ephippiger diurnus*. *Journal of Evolutionary Biology* 29: 2219–2228.
- Barton NH, Hewitt GM. 1985. Analysis of hybrid zones. *Annual Review of Ecology and Systematics* 16: 113–148.
- Bella JL, Serrano L, Orellana J, Mason PL. 2007. The origin of the *Chorthippus parallelus* hybrid zone: chromosomal evidence of multiple refugia for Iberian populations. *Journal of Evolutionary Biology* 20: 568–576.
- Bellmann H, Luquet G. 2009. *Sauterelles, grillons et criquets d'Europe occidentale*. Paris: Delachaux & Niestlé.
- Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu C-H, Xie D. 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Computational Biology* 10: e1003537.
- Butlin RK, Hewitt GM. 1985. A hybrid zone between *Chorthippus parallelus parallelus* and *Chorthippus parallelus erythropus* (Orthoptera, Acrididae): behavioural characters. *Biological Journal of the Linnean Society* 26: 287–299.

- Chapuis M-P, Estoup A. 2007. Microsatellite null alleles and estimation of population differentiation. *Molecular Biology and Evolution* 24: 621–631.
- Darriba D, Taboada GL, Doallo R, Posada D. 2012. JModelTest 2: more models. new heuristics and parallel computing. *Nature Methods* 9: 772.
- Diniz-Filho JAF, Soares TN, Lima JS, Dobrovolski R, Landeiro VL, de Campos Telles MP, Rangel TF, Bini LM. 2013. Mantel test in population genetics. *Genetics and Molecular Biology* 36: 475–85.
- Duijm M. 1990. On some song characteristics in *Ephippiger* (Orthoptera, Tettigoniidea) and their geographic variation. *Netherlands Journal of Zoology* 40: 428–453.
- Duijm M, Oudman L. 1983. Interspecific mating in *Ephippiger* (Orthoptera, Tettigoniidea). *Tijdschrift voor Entomologie* 126: 97–108.
- Esquer-Garrigos Y, Greenfield MD, Party V, Streiff R. 2016. Characterization of 16 novel microsatellite loci for *Ephippiger diurnus* (Orthoptera: Tettigoniidae) using pyrosequencing technology and cross-species amplification. *European Journal of Entomology* 113: 302–306.
- Esquer-Garrigos Y, Streiff R, Party V, Nidelet S, Navascués M, Greenfield M. 2018. Data from: Pleistocene origins of chorusing diversity in Mediterranean bush-cricket populations (*Ephippiger diurnus*). Dryad Digital Repository. doi:10.5061/dryad.m04r4b5
- Funk DJ. 1999. Molecular systematics of cytochrome oxidase I and 16S from *Neochlamisus* leaf beetles and the importance of sampling. *Molecular Biology and Evolution* 16: 67–82.
- Goudet J. 2005. HIERFSTAT: a package for R to compute and test hierarchical F-statistics. *Molecular Ecology Resources* 5: 184–186.
- Greenfield MD, Esquer-Garrigos Y, Streiff R, Party V. 2016. Animal choruses emerge from receiver psychology. *Scientific Reports* 6: 34369.
- Greenfield MD, Marin-Cudraz T, Party V. 2017. Evolution of synchronies in insect choruses. *Biological Journal of the Linnean Society* 122: 487–504.
- Hamill RM, Noor MAF, Watson ET, Ritchie MG. 2006. New microsatellite loci for the European bushcricket, *Ephippiger ephippiger* (Orthoptera: Tettigoniidae). *Molecular Ecology Notes* 6: 340–342.
- Harrison RG, Larson EL. 2014. Hybridization, introgression and the nature of species boundaries. *Journal of Heredity* 105: 795–809.
- Hewitt G. 2000. The genetic legacy of the Quaternary ice ages. *Nature* 405: 907–913.
- Hewitt GM. 1999. Post-glacial re-colonization of European biota. *Biological Journal of the Linnean Society* 68: 87–112.
- Hewitt GM. 2004. Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society B: Biological Sciences* 359: 183–195.
- Ho SYW, Phillips MJ, Cooper A, Drummond AJ. 2005. Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. *Molecular Biology and Evolution* 22: 1561–1568.
- Hockham LR, Graves JA, Ritchie MG. 1999. Isolation and characterization of microsatellite loci in the bushcricket *Ephippiger ephippiger* (Orthoptera: Tettigoniidae). *Molecular Ecology* 8: 905–906.
- IUCN Red List of Threatened Species. 2017. Available at: <http://www.iucnredlist.org/>
- Jombart T, Devillard S, Balloux F. 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics* 11: 94.
- Keane TM, Creevey CJ, Pentony MM, Naughton TJM, McInerney JO. 2006. Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. *BMC Evolutionary Biology* 6: 29.
- Kidd DM, Ritchie MG. 2000. Inferring the patterns and causes of geographic variation in *Ephippiger ephippiger* (Orthoptera, Tettigoniidae) using geographical information systems (GIS). *Biological Journal of the Linnean Society* 71: 269–295.
- 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.

- 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.
- Mahamoud-Issa M, Marin-Cudraz T, Party V, Greenfield MD. 2017. Phylogeographic structure without pre-mating barriers: do habitat fragmentation and low mobility preserve song and chorus diversity in a European bushcricket? *Evolutionary Ecology* 31: 865–884.
- Mallet J. 2008. Hybridization, ecological races and the nature of species: empirical evidence for the ease of speciation. *Philosophical Transactions of the Royal Society B: Biological Sciences* 363:2971–2986.
- Marin-Cudraz T, Greenfield MD. 2016. Finely tuned choruses: bushcrickets adjust attention to neighboring singers in relation to the acoustic environment they create. *Behavioral Ecology and Sociobiology* 70: 1581–1589.
- Nei M, Tajima F, Tateno Y. 1983. Accuracy of estimated phylogenetic trees from molecular data . *Journal of Molecular Evolution* 19: 153–170.
- Nosil P, Crespi BJ, Sandoval CP. 2003. Reproductive isolation driven by the combined effects of ecological adaptation and reinforcement. *Proceedings of the Royal Society B: Biological Sciences* 270: 1911–1918.
- Oudman L, Duijm M, Landman W. 1990. Morphological and allozyme variation in *Ephippiger ephippiger* (Orthoptera, Tettigoniidae). *Netherlands Journal of Zoology* 40: 454–483.
- Oudman L, Landman W, Duijm M. 1989. Genetic distance in the genus *Ephippiger* – a reconnaissance. *Tijdschrift voor Entomologie* 132: 177–181.
- Papadopoulou A, Anastasiou I, Vogler AP. 2010. Revisiting the insect mitochondrial molecular clock: the mid-Aegean Trench calibration. *Molecular Biology and Evolution* 27: 1659–1672.
- Party V, Brunel-Pons O, Greenfield MD. 2014. Priority of precedence: receiver psychology, female preference for leading calls and sexual selection in insect choruses. *Animal Behaviour* 87: 175–185.
- Party V, Streiff R, Marin-Cudraz T, Greenfield MD. 2015. Group synchrony and alternation as an emergent property: elaborate chorus structure in a bushcricket is an incidental byproduct of female preference for leading calls. *Behavioral Ecology and Sociobiology* 69: 1957–1973.
- Pfennig KS. 2016. Reinforcement as an initiator of population divergence and speciation. *Current Zoology* 62: 145–154.
- Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155: 945–959.
- R Core Team. 2016. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. Available at: <http://www.R-project.org/>
- Rambaut A, Drummond AJ. 2010. TreeAnnotator version 1.6.1. Available at: <http://beast.bio.ed.ac.uk>.
- Raymond M, Rousset F. 1995. GENEPOP (Version 1.2): population genetics software for exact tests and ecumenism. *Journal of Heredity* 86: 248–249.
- Ritchie MG. 1996. The shape of female mating preferences. *Proceedings of the National Academy of Sciences of the United States of America* 93: 14628–14631.
- Rousset F. 2008. genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Molecular Ecology Resources* 8: 103–106.
- Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* 4: 406–425.
- Schmitt T. 2007. Molecular biogeography of Europe: Pleistocene cycles and postglacial trends. *Frontiers in Zoology* 4: 11.
- Servedio MR, Noor MAF. 2003. The role of reinforcement in speciation: theory and data. *Annual Review of Ecology, Evolution and Systematics* 34: 339–364.
- Shapiro LH, Strazanac JS, Roderick GK. 2006. Molecular phylogeny of *Banza* (Orthoptera: Tettigoniidae), the endemic katydid of the Hawaiian Archipelago. *Molecular Phylogenetics and Evolution* 41: 53–63.

- Song HJ, Moulton MJ, Whiting MF. 2014. Rampant nuclear insertion of mtDNA across diverse lineages within Orthoptera (Insecta). *PLoS ONE* 9: e110508.
- Spooner LJ, Ritchie MG. 2006. An unusual phylogeography in the bushcricket *Ephippiger ephippiger* from southern France. *Heredity* 97: 398–408.
- 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.
- Taberlet P, Fumagalli L, Wust-Saucy A-G, Cosson J-F. 1998. Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology* 7: 453–464.
- Takezaki N, Nei M, Tamura K. 2010. POPTREE2: software for constructing population trees from allele frequency data and computing other population statistics with windows interface. *Molecular Biology and Evolution* 27: 747–752.
- Terral J-F, Mengual X. 1999. Reconstruction of Holocene climate in southern France and eastern Spain using quantitative anatomy of olive wood and archaeological charcoal. *Palaeogeography, Palaeoclimatology, Palaeoecology* 153: 71–92.
- Tritsch C, Stuckas H, Martens J, Pentzold S, Kvist L, Lo Valvo M, Giacalon G, Tietze DT, Nazarenko AA, Packert M. 2018. Gene flow in the European coal tit, *Periparus ater* (Aves: Passeriformes): low among Mediterranean populations but high in a continental contact zone. *Biological Journal of the Linnean Society* 124: 319–338.
- Wallis GP, Waters JM, Upton P, Craw D. 2016. Transverse alpine speciation driven by glaciation. *Trends in Ecology & Evolution* 31: 916–926.
- Wright S. 1943. Isolation by distance. *Genetics* 28: 114–138.
- Zhou Y, Duvaux L, Ren G, Zhang L, Savolainen O, Liu J. 2017. Importance of incomplete lineage sorting and introgression in the origin of shared genetic variation between two closely related pines with overlapping distributions. *Heredity* 118: 211–220.