

1 Two ways to be endemic. Alps and Apennines are different functional refugia during climatic  
2 cycles

3

4 Mattia Menchetti<sup>1,2,†</sup>, Gerard Talavera<sup>3,†</sup>, Alessandro Cini<sup>1,4</sup>, Vania Salvati<sup>1</sup>, Vlad Dincă<sup>5</sup>,  
5 Leonardo Platania<sup>2</sup>, Simona Bonelli<sup>6</sup>, Emilio Balletto<sup>6</sup>, Roger Vila<sup>2</sup>, Leonardo Dapporto<sup>1\*</sup>

6

7 <sup>1</sup> ZEN Lab, Dipartimento di Biologia dell'Università di Firenze, via Madonna del Piano 6  
8 50019 Sesto Fiorentino, Italy.

9 <sup>2</sup> Institut de Biologia Evolutiva (CSIC - Universitat Pompeu Fabra), Barcelona, Catalonia,  
10 Spain.

11 <sup>3</sup> Institut Botànic de Barcelona (IBB, CSIC-Ajuntament de Barcelona), Passeig del Migdia  
12 s/n, 08038 Barcelona, Catalonia, Spain.

13 <sup>4</sup> Centre for Biodiversity & Environment Research, University College London, Gower Street,  
14 London, WC1E 6BT.

15 <sup>5</sup> Ecology and Genetics Research Unit, University of Oulu, PO Box 3000, Oulu, 90014,  
16 Finland.

17 <sup>6</sup> ZOOLAB, Dipartimento di Scienze della Vita e Biologia dei Sistemi, Università degli Studi  
18 di Torino, Via Accademia Albertina 13, 10123 Turin, Italy.

19 † Mattia Menchetti and Gerard Talavera contributed equally to this study.

20 \* corresponding author Leonardo Dapporto, leonardo.dapporto@unifi.it

21

22

23

24 Running title: Endemicity centres on Alps and Apennines

25

26

27 ABSTRACT

28 Endemics co-occur because they evolved *in situ* and persist regionally or because they  
29 evolved *ex situ* and later dispersed to shared habitats, generating evolutionary or ecological  
30 endemicity centres, respectively. We investigate whether different endemicity centres can  
31 intertwine in the region ranging from Alps to Sicily, by studying their butterfly fauna. We  
32 gathered an extensive occurrence dataset for butterflies of the study area (27,123 records, 269  
33 species, in cells of 0.5x0.5 degrees of latitude-longitude). We applied molecular-based  
34 delimitation methods (GMYC model) to 26,557 COI sequences of Western Palearctic  
35 butterflies. We identified entities based on molecular delimitations and the most recent  
36 checklist of European butterflies and objectively attributed occurrences to their most probable  
37 entity. We obtained a zoogeographic regionalisation based on the 69 endemics of the area.  
38 Using phylogenetic ANOVA we tested if endemics from different centres differ from each  
39 other and from non-endemics for key ecological traits and divergence time. Endemicity  
40 showed high incidence in the Alps and Southern Italy. The regionalisation separated the Alps  
41 from the Italian Peninsula and Sicily. The endemics of different centres showed a high  
42 turnover and differed in phenology and distribution traits. Endemics are on average younger  
43 than non-endemics and the Peninsula-Sicily endemics also have lower variance in divergence  
44 than those from the Alps. The observed variation identifies Alpine endemics as  
45 paleoendemics, now occupying an ecological centre, and the Peninsula-Sicily ones as  
46 neoendemics, that diverged in the region since the Pleistocene. The results challenge the  
47 common view of the Alpine-Apennine area as a single “Italian refugium”.

48

49 Keywords: butterflies, endemicity centres, GMYC, Italian Peninsula, regionalization, species  
50 traits

51

52

## 53 INTRODUCTION

54 Endemicity is a central concept in biogeography and conservation biology, denoting the  
55 condition of a taxon to be exclusively distributed in a given area (Anderson, 1994).  
56 Consequently, endemics cannot be identified without defining a region within vaster areas.  
57 The centres of endemism are regions where several endemics co-occur (Harrison & Noss,  
58 2017) given a combination of geographical, historical and ecological factors (Sandel et al.,  
59 2020; Zuloaga, Currie, & Kerr, 2019). They are typically limited by barriers and characterised  
60 by stable and often singular climatic conditions, so that genetic and faunistic divergence can  
61 accumulate (Ohlemüller et al., 2008; Sandel et al., 2020; Zuloaga et al., 2019). In this respect,  
62 the concept of centres of endemism largely overlaps with that of refugia (Keppel et al., 2012).  
63 A major goal of conservation biology is to identify and protect centres of endemism because,  
64 other than functioning as refuges from ongoing environmental changes, they host peculiar and  
65 irreplaceable biodiversity elements (Brooks et al., 2015; Harrison & Noss, 2017; Keppel et  
66 al., 2012).

67 Different mechanisms explain why endemics co-occur. They could have evolved *in situ* due  
68 to long-term environmental stability while barriers prevented gene-flow (Crother & Murray,  
69 2011). These areas, to which we refer as evolutionary endemicity centres (EVOcs, Figure 1),  
70 largely coincide with *in situ* refugia (Keppel et al., 2012) and evolutionary refugia (Davis,  
71 Pavlova, Thompson, & Sunnucks, 2013). A paradigm for EVOcs is the diversification  
72 occurred in warm refugia during Pleistocene cold periods (Brooks et al., 2015; Hewitt, 1999;  
73 Taberlet, Fumagalli, Wust-Saucy, & Cosson, 1998).

74 Another pathway generating clusters of endemics arise when, after major environmental  
75 changes, species widely distributed track their habitats to reduced “safe havens” (Crother &  
76 Murray, 2011). These taxa, evolved in different areas, end up co-occurring in habitat  
77 remnants to which we refer as ecological endemicity centres (ECOcs, Figure 1). ECOcs  
78 coincide with the *ex situ* refugia (Keppel et al., 2012) and ecological refuges (Davis et al.,  
79 2013). A paradigm for ECOcs is the co-occurrence of boreo-alpine species in areas formerly  
80 covered by ice-sheets (Mutanen et al., 2012).

81 These mechanisms are not mutually exclusive and, in heterogeneous environmental settings,  
82 composite endemicity centres assembled after evolutionary and ecological processes can  
83 emerge (Crother & Murray, 2011). The methods to identify centres of endemism are well  
84 established; conversely the mechanisms behind their emergence have often remained

85 unresolved. Their understanding requires the application of eclectic approaches combining  
86 palaeoecological reconstructions, high-resolution occurrence data, phylogeographic  
87 assessments and species functional traits for large homogenous taxa (Brooks et al., 2015;  
88 Davis et al., 2013; Keppel et al., 2012; Zuloaga et al., 2019). We hypothesized that endemic  
89 assemblages formed under different processes show different features. EVOc endemics  
90 should have a similar degree of genetic divergence from closest relatives since they have  
91 likely evolved under similar processes; this is not expected in ECOc representing secondary  
92 sympatry areas (Moritz et al., 2009). Moreover, ECOc endemics should mostly represent  
93 strict specialists for environmental settings rarefied after historical changes. Recognizing  
94 endemic entities is a challenging task. Indeed, macroecological studies usually employ entities  
95 recognized by taxonomists at the species level. On the other hand, phylogeography is rooted  
96 on the pervasive evidence that most species encompass a wide variation of spatially-  
97 structured diversity both as cryptic taxa and as genetic lineages. Genetic lineages are not  
98 recognised in taxonomic catalogues, are usually excluded by macroecological studies, are not  
99 protected and there are no protocols to include them in conservation plans (e.g. IUCN Red  
100 List) (Brooks et al., 2015). This exclusion results in a significant loss of the signal of the  
101 Quaternary processes they convey and discards their fundamental contribution to  
102 biogeography and conservation (Brooks et al., 2015; Vodă, Dapporto, Dincă, & Vila, 2015).  
103 Here, we identified entities in the entire butterfly fauna (269 species) occurring along the  
104 Alps, the Italian Peninsula and surrounding islands (Figure 2a) based on an “or-or” approach,  
105 where an entity is represented by a group of individuals recognised as a species by  
106 taxonomists or based on a phylogenetic-based species delimitation approach (GMYC) (Figure  
107 2b). GMYC is increasingly used in macroecology (Fujisawa, Vogler, & Barraclough, 2015)  
108 and sometimes it proved to be more effective than taxonomic assessments in documenting  
109 eco-evolutionary processes (Liu et al., 2018). Using this dataset, we investigate whether the  
110 two different kinds of centres of endemism can be identified. This region represents an ideal  
111 system to test this hypothesis because it is located in the centre of the Mediterranean, a major  
112 hotspot where particularly high biodiversity has emerged from the interplay between Africa  
113 and Eurasia and the possibility for many species to persist during the Pleistocene (Bonelli et  
114 al., 2018). The continuous S-shaped mountain-hill system comprising the Alps and the  
115 Apennines (37° to 48° of latitude), encompasses the glacial refugium of the Italian Peninsula  
116 (Dapporto et al., 2019; Hewitt, 1999, Taberlet et al., 1998) and mountain areas covered by ice  
117 caps during glacial maxima (Figure 2a). Typically, the region is considered as a single

118 biogeographic unit: the Italian refugium (Drovetski et al., 2018; Hewitt, 1999; Petit et al.,  
119 2003; Taberlet et al., 1998), but based on paleogeographic and paleoclimatic evidence (Figure  
120 2a), we hypothesise that distinct EVOcs and ECOcs occur in this region. To test this, we 1)  
121 combine a massive dataset of COI sequences and occurrence data to evaluate if ECOc(s) and  
122 EVOc(s) are recognisable by 2) verifying whether the endemics from the potential ECOc have  
123 stricter requirements in key ecological traits and by 3) verifying whether endemics from  
124 potential EVOc show similar and shorter divergence times, being supposedly mostly  
125 represented by intraspecific genetic lineages. Answering these questions can provide  
126 fundamental insights for understanding the evolution of endemism in biodiversity hotspots  
127 and informs towards more effective conservation strategies.

## 128 METHODS

### 129 **Sampling and datasets**

130 The study area includes the Alps ([www.alpconv.org](http://www.alpconv.org)), the Italian Peninsula, Sicily and the  
131 small Italian islands closer to this land than to any other (Figure 2a). We obtained 307,228  
132 records for butterfly species as recognised in Wiemers et al. (2018) within the study area for  
133 cells of 0.5x0.5 degrees of latitude and longitude, corresponding to 1277 km<sup>2</sup> in the centre of  
134 the study area (Rome) (sources described in Appendix S1). We generated occurrence maps  
135 for each species and compared them with the distribution of European butterflies (Kudrna,  
136 2019) with the goal to removed possible misplaced records. After filtering unique occurrences  
137 for each cell, we counted 27,123 records (available in DryadXXX). We gathered 26,557 COI  
138 (standard barcode, 658 bp) sequences from 519 species occurring in the Western Palearctic  
139 (DryadXXX). Among these, 23,563 COI sequences belong to the 269 species occurring in the  
140 study area (BIBSA and EUGENMAP BOLD datasets).

141

### 142 **Phylogeny and GMYC**

143 We collapsed the COI dataset to unique haplotypes using the “haplotype” function of the R  
144 package “pegas” (<https://cran.r-project.org/web/packages/pegas/index.html>). We used BEAST  
145 1.8 (Drummond, Rambau, & Suchard, 2013) to reconstruct five ultrametric phylogenetic  
146 trees, one for each butterfly family (the single European Riodinidae was merged with  
147 Lycaenidae) (available in Dryad). The number of haplotypes was 6459 (3232 Nymphalidae,  
148 644 Pieridae, 561 Hesperidae, 247 Papilionidae and 1775 Lycaenidae-Riodinidae). Each  
149 dataset included one outgroup for each of the other families. Two independent chains of 100

150 million generations were run in BEAST for each dataset. The substitution model was set to  
151 GTR+I+G with six gamma rate categories. A coalescent tree prior was set. Divergence times  
152 were estimated by applying a strict clock and a normal prior distribution centered on the mean  
153 between two widely used substitution rates of 1.5% uncorrected pairwise distance per million  
154 years (Quek, Davies, Itino, & Pierce, 2004), and 2.3 % (Brower, 1994). Values were sampled  
155 every 10% of the run length and convergence was inspected in Tracer v.1.6  
156 (<http://tree.bio.ed.ac.uk/software/tracer/>). We applied the general mixed Yule-coalescent  
157 model (GMYC, Fujisawa & Barraclough, 2013) for each family tree to identify evolutionary  
158 significant units (ESUs) using the R package “splits” ([https://cran.r-project.org/web/packages/  
159 SplitSoftening/index.html](https://cran.r-project.org/web/packages/SplitSoftening/index.html)) with default settings.  
160 We identified entities as taxa recognised by the taxonomic list of Wiemers et al. (2018) or as  
161 haplotypes belonging to different GMYC ESUs (Figure 2b). According to the GMYC results  
162 each species identified by Wiemers et al. (2018) could be 1) “single entity species (SE)”: all  
163 haplotypes of a species belong to a single GMYC ESU, 2) “multiple entity species (ME)”:  
164 haplotypes belong to two or more ESUs, 3) “lumped entities (LE)”: two or more species are  
165 recovered as a single ESU, and 4) “lumped + multiple entities (LME): species are split in  
166 multiple ESUs and lumped with other species (Figure 2b).  
167 For SE and LE all occurrences were attributed to the original species while for ME and LME,  
168 we attributed species occurrence to their most probable ESU by using “biodecrypt”  
169 (“recluster” R package, <https://rdrr.io/github/leondap/recluster/>). The function creates concave  
170 hulls based on the distribution of the sequences attributed to a given ESU and uses the relative  
171 hull geometries to attribute unknown occurrence data to a given species (Platania et al., 2020).  
172 The biodecrypt function also provides a measure for hull overlap as an evaluation of sympatry  
173 among cryptic entities.  
174 We identified as endemics those entities for which all COI sequences occurred exclusively  
175 within the study area.

176

### 177 *1) Which are the centres of endemism?*

178 To locate the centres of endemism we ran regionalisation analyses for the occurrence data of  
179 endemics in 0.5x0.5 cells. We used the “recluster.region” function in the R package  
180 “recluster” (<https://cran.r-project.org/web/packages/recluster/index.html>) specifically  
181 designed to retrieve biogeographic regions at the intracontinental scale. We obtained  
182 clustering solutions from 2 to 8 centres based on the Simpson turnover index, suited to

183 identify regions based on vicariant patterns of distribution. The “recluster.region” function  
184 also calculates the silhouette width and the explained dissimilarity, evaluating how cells  
185 resemble those of their own centre (cohesion) compared to other centres (separation). Once  
186 the centres were obtained, we identified their exclusive endemics using the “indval” function  
187 in the “labdsv” R package (<https://cran.r-project.org/web/packages/labdsv/labdsv.pdf>).

188

189 *2) Are endemics characterised by different ecological traits in different centres?*

190 The traits of species which entities belong to, were compared between endemics from  
191 different centres and between endemics and non-endemics from the same centre. We used a  
192 series of 10 ecological traits for European Butterflies (Middleton-Welling et al, in press;  
193 Platania et al., 2020). These traits were used to describe both the alpha niche (i.e. functional  
194 traits describing the primary functions of invertebrates, and the beta niche (features related to  
195 distributional and environmental preferences) (Table 1). Butterfly traits are highly  
196 intercorrelated and are usually reduced to factors by Principal Component Analyses (PCA).  
197 We applied PCA to life history and distribution traits using the function rda of the R package  
198 “vegan” (<https://cran.r-project.org/web/packages/vegan/index.html>). Those components  
199 showing eigenvalues higher than one were retained as variables.

200 To assess differences in traits we applied a phylogenetic ANOVA, using the “aov.phylo”  
201 function of the R package “geiger”

202 (<https://cran.r-project.org/web/packages/geiger/index.html>). As a reference phylogeny we  
203 used a time-calibrated phylogenetic tree for all 496 species of European butterflies, based on  
204 14 mitochondrial and nuclear genes (Wiemers, Chazot, Wheat, Schweiger, & Wahlberg,  
205 2020). We carried out pairwise comparisons through sequential Bonferroni corrections. We  
206 log transformed the number of host plants to improve its normality. To investigate if the traits  
207 show different variances among groups, we carried out tests of variance homogeneity  
208 (followed by pairwise comparisons with sequential Bonferroni correction) through the non-  
209 parametric Fligner-Killeen test, using the “check\_homogeneity” function of the R package  
210 “performance” (<https://cran.r-project.org/web/packages/performance/index.html>).

211

212 *3) Do the endemics from EVOc(s) show lower variance in genetic divergence?*

213 For each entity, we obtained genetic divergence from its closest entity in the five family trees  
214 using the function “distTips” of the R package “adephylo”  
215 (<https://cran.r-project.org/web/packages/adephylo/index.html>). We compared divergence

216 between endemics from different centres and between endemics and non-endemics of the  
217 same region using ANOVAs and variance comparisons as described above. We also  
218 compared divergence among types of entity using the same method. We did not apply a  
219 correction for phylogenetic autocorrelation because genetic distances are exactly the variable  
220 compared here. We compared the incidence of the SE, ME, LE and LME endemics among  
221 centres through a Chi Square test.

222

## 223 **RESULTS**

224 A comparison between the reference taxonomic list (Wiemers et al., 2018) and the GMYC  
225 ESUs resulted in 369 entities for the 269 species occurring in the study area. Overall, we  
226 recovered many endemics (represented by 69 entities). The percentage of endemics obtained  
227 by combining species and ESUs ( $69/369 = 18.7\%$ ), is higher than the percentage obtained for  
228 endemic species over taxonomic richness ( $27/269 = 10.0\%$ ). Among the 36 ME-LME  
229 representing endemic entities in the study region, 14 showed only two ESUs over the west  
230 Palaeartic and 10 showed three ESUs (Figure S3a). Occurrence data for each species were  
231 attributed to multiple entity taxa by biodecrypt, revealing that the lineages were mostly  
232 parapatric since on average they showed only  $4.99 \pm 9.98\text{sd} \%$  geographic overlap (Figure  
233 S3b).

234

### 235 *1) Which are the centres of endemism?*

236 The 69 endemics were not homogeneously distributed along the study area. The number of  
237 endemics showed a single peak of richness with about 20 endemics per cell over the Alps  
238 (Figure 3a,b, Figure S4-72 Appendix S1 for individual distributions). When the effect of local  
239 richness was removed by calculating percentages of endemics, Sicily emerged as a main  
240 endemism hotspot (Figure 3c), showing endemism percentages around 15%, similar to Alps  
241 (Figure 3d).

242 Regionalisation revealed the solution of  $k=2$  clusters, among cells containing at least two  
243 endemics, had a silhouette value of 0.556 and explained dissimilarity of 68.9%. This solution  
244 separates the Alps from the Italian Peninsula, Sicily and surrounding islands (hereafter  
245 Peninsula-Sicily centre) (Figure 4a). The “indval” function showed that 62 of 69 endemics  
246 exclusively occur in the Alps or in Peninsula-Sicily, suggesting a strong turnover (Simpson  
247 index = 0.79) (Figure 4a). A solution of  $k=3$ , splitting Peninsula-Sicily into Italian Peninsula  
248 and Sicily (red and yellow regions in Figure 4b) showed a higher silhouette (0.623) and an

249 increase of explained dissimilarity to 79.4%. This higher silhouette can be explained by a  
250 moderate distinction in endemics between the Italian Peninsula and Sicily, including 17  
251 species exclusive of a single region and 16 shared (Simpson index = 0.33, Figure 4b). A  
252 partition of  $k=4$  separated the Western and Eastern Alps (Figure 4b), showed a lower  
253 silhouette of 0.558 and explained 84.6% of dissimilarity, but the two Alpine regions showed a  
254 low turnover (16 exclusive vs of 26 shared with a low Simpson turnover index = 0.04).  
255 Partitions for higher  $k$  gradually lowered silhouette values (always  $<0.5$ ) and lost geographic  
256 coherence. For this reason, the solution with  $k=2$  (Alps, Peninsula-Sicily) is preferred for the  
257 highest turnover, followed by  $k=3$  (Alps, Peninsula, Sicily). The Peninsula-Sicily centre of  
258 endemism is about 270,000 km<sup>2</sup> while the Alpine centre is about 210,000 km<sup>2</sup>.

259

### 260 *2) Are endemics characterised by different traits in different centres?*

261 The PCA identified one function from the phenological traits showing eigenvalues higher  
262 than one (Figure S73-74, Appendix S1), mostly correlated with flight period and voltinism.

263 The PCA for distribution traits extracted a component positively correlating with distribution  
264 and altitudinal ranges, and a component positively correlated with minimum and maximum  
265 altitudes.

266 When comparing traits of endemics exclusive of centres obtained for  $k=2$ , Alps endemics  
267 showed a significantly shorter flight period compared to both Peninsula-Sicily endemics and  
268 Alpine non-endemics, occurred over smaller areas, at higher altitudes and with narrower  
269 altitude ranges. Notably, there were no significant differences between the traits of Peninsula-  
270 Sicily endemics and non-endemics (Table 2, Figure 5). In many cases the variances also  
271 differed (Table 2, Figure 5 b,f) and alpine endemics significantly showed lower variance for  
272 wingspan, host plant specialisation and phenology. No comparisons have been made using  
273 three centres (Alps, Peninsula, Sicily) because the entities would have been too few to obtain  
274 reliable results.

275

### 276 *3) Do the endemics from EVOc(s) show lower variance in genetic divergence?*

277 Phylogenetic ANOVA showed that different types of endemics showed different divergence  
278 ( $Df=3$ , Sum. Sq.=151.74,  $F=15.406$ ,  $P<0.001$ ) and post-hoc comparisons revealed that SE  
279 have the highest divergence, followed by ME and then by LE and LME (with similar  
280 divergence) (Figure S75 for p-values). Endemics showed a lower divergence from their  
281 closest relatives compared to non-endemics (Table 2, Figure 5f). Moreover, Peninsula-Sicily

282 endemics showed lower variance in divergence, compared to both Alpine endemics and non-  
283 endemics from their centre with a particularly high frequency around 2 MY (Figure 5f). SE,  
284 LE, LME showed higher frequencies in the Alps (13, 8 and 4, respectively) compared to  
285 Peninsula-Sicily (3, 5, 0, respectively), while ME were less frequent in the Alps compared to  
286 Peninsula-Sicily (17 vs 26). A Chi Square test of independence of the frequencies showed a  
287 highly significant effect ( $\chi^2= 12.118$ ,  $df = 3$ ,  $p=0.007$ ).

## 288 289 **DISCUSSION**

290 Along the region formed by the continuous mountain chain of the Alps and the Apennines, we  
291 identified two main centres of endemism (Alps and Peninsula-Sicily) based on the distribution  
292 of endemic species and GMYC ESUs. The endemics of the two centres showed differences in  
293 their ecological traits and in the variance of genetic divergences that align with the predictions  
294 for the existence of two different endemism centres: an ecological centre, originated as an *ex*  
295 *situ* refugium after the recent occupation of the formerly glaciated Alps and an evolutionary  
296 centre, originated as an *in situ* refugium in the Italian Peninsula and Sicily, which remained  
297 suitable for butterflies during Pleistocene glacial cycles.

### 298 299 **Two centres of endemism in the Alps-Apennines region**

300 Our result challenges the common perception of the Alpine-Apennine area as a single unit,  
301 known as the “Italian refugium” (Hewitt, 1999; Petit et al., 2003; Taberlet et al., 1998).  
302 Indeed, we objectively identified two centres of endemism: Alps and Peninsula-Sicily. The  
303 two centres shared only 7 endemics, versus 35 endemics exclusive from the Alps and 27 from  
304 Peninsula-Sicily. This occurred despite the continuity of the mountain-hill chain, the presence  
305 of high-altitude areas in the Apennines (a maximum of 2912m in mainland and of 3324m in  
306 Sicily) and the inclusion in the Alps centre of many low-altitude cells. Moreover, endemics  
307 from the Alps and Peninsula-Sicily are characterised by different traits and variance of genetic  
308 divergences. In the Peninsula-Sicily region the ecological spectrum of endemics is variegated  
309 since they belong to strictly Mediterranean species (*Hipparchia leighebi*, *Hipparchia*  
310 *blachieri*, *Zerynthia cassandra*, *Melanargia arge*, *Pyronia cecilia*) but also to typical  
311 mountain taxa (*Erebia pluto*, *Erebia montana*, *Parnassius mnemosyne*, *Melitaea varia*).  
312 Conversely, their genetic divergence is less variable: most taxa seem to have differentiated  
313 during the Pleistocene and are recognised as deeply diverging intraspecific lineages (ME).  
314 This pattern agrees with the current view of European Quaternary phylogeography, deeply  
315 impacted by long cold periods, when most central-northern Europe, Alps and Pyrenees were

316 covered by ice sheets surrounded by permafrost and tundra belts (Ehlers, Ehlers, Gibbard, &  
317 Hughes, 2011). During cold pulses, many temperate species persisted in separated glacial  
318 refugia (notably the peninsulas of Iberia, Italy and Balkans, the Mediterranean islands, and the  
319 Maghreb); during interglacials, they dispersed towards higher latitudes and altitudes (Hewitt,  
320 1999; Petit et al., 2003; Schmitt, 2007). Virtually all European taxa showed differentiation  
321 among these areas and signal of (repeated) post-glacial poleward expansion (forest plants,  
322 Petit et al., 2003; butterflies, Schmitt, 2007; Dapporto et al., 2019; mammals, Seddon,  
323 Santucci, Reeve, & Hewitt, 2001; springtails, Fiera, Habel, Kunz, & Ulrich, 2017). The high  
324 incidence of ME endemics dated to the onset of the Pleistocene and limited to Peninsula-  
325 Sicily fits with the definition of neoendemics, described as recently diverged species that  
326 failed to disperse out of their ancestral area (Flantua et al., 2020).

327 Sicily also has a moderate turnover of endemics with respect to the continental area and it  
328 might represent a distinctive EVOc. Sicily is a well-known endemism hotspot also for plants,  
329 with an endemism rate showing peaks higher than 20% (Medail & Quezel, 1997), very close  
330 to the values we retrieved for butterflies. The high incidence of endemic haplogroups in  
331 Sicilian butterflies has been recently documented together with the observation that i) species  
332 showing genetic differentiation have lower dispersal capability and stronger ecological  
333 impediments to dispersal and that ii) phenomena of *in situ* evolution and relictuality have  
334 generated the observed differentiation (Scalercio et al., 2020).

335 Many butterflies probably populated the European tundra during the long cold periods of the  
336 Pleistocene, like the genus *Erebia*, which shows a diversification centre in Europe (Peña,  
337 Witthauer, Klečková, Fric, & Wahlberg, 2015). Species richness in a given biome depends on  
338 the area the biome occupied along historical time (Jetz & Fine, 2012). Due to the large  
339 extension of tundra, steppic and subarctic biotas in Europe during most Pleistocene, the  
340 continent hosts many cold-adapted species that likely suffered range contractions during  
341 interglacials, including the current one, and now show disjunct distributions between southern  
342 mountains and northern subarctic-steppic environments (Mutanen et al., 2012; Ohlemüller et  
343 al., 2008). The high incidence of SE endemics showing high variance in divergence time in  
344 the Alps is in line with the hypothesis of an ecological refugium amassing species with  
345 different evolutionary histories that tracked their habitat during the recent interglacial. This  
346 mechanism is likely responsible for the distribution of most alpha diversity since formerly  
347 glaciated areas in the Alps, Pyrenees and Balkans are currently the richest areas for butterflies  
348 in Europe (Hawkins, 2010). Generally, boreo-alpine species show low intraspecific

349 differentiation between regions because of the recent geographic split (Mutanen et al., 2012),  
350 which explains the low incidence of ME we found in the Alps. The Alps also prove to be a  
351 more compact centre compared to the Peninsula-Sicily centre, as further divisions in sub-  
352 centres showed a low turnover. This is in line with the hypothesis of a recently formed fauna  
353 assembled after habitat tracking phenomena in the last 16 thousand years after the LGM. The  
354 large number of SE and LE endemics in the Alps showing higher (SE) and lower (LE)  
355 divergence than ME endemics, contributed to the higher genetic variance of Alpine compared  
356 to Peninsula-Sicily endemics. The high incidence of LE and LME fits with the mechanisms at  
357 the basis of EVOc formation since they generally represent diverged species that exchanged  
358 mitochondrial DNA by introgression following secondary sympatry in areas of contact. Their  
359 incidence in the Alps could have determined the absence of a significant higher genetic  
360 differentiation of Alpine compared to Peninsula-Sicily endemics. In general, the endemics of  
361 the Alps fit with the definition of palaeoendemics, described as relict species whose ranges  
362 became spatially restricted (Flantua et al., 2020).

363 Without doubt, the Apennines also functioned as an *ex situ* refugium, as indicated by the  
364 presence of several cold-adapted species (12 *Erebia* spp., 2 *Parnassius* spp., many  
365 Lycaenidae spp.) and because mountain areas host the richest butterfly communities of the  
366 Peninsula-Sicily centre. However, this phenomenon has involved different entities from the  
367 Alps since only three high-altitude endemics are shared between the two centres  
368 (*Polyommatus damon*, *Melitaea varia*, *Erebia montana*). Other shared ME endemics belong  
369 to altitude generalist taxa (*Lycaena alciphron*, *Melitaea aurelia* and *Melitaea cinxia*) and to a  
370 Mediterranean taxon (*Lycaena thersamon*). The extinction of several high-altitude species in  
371 the Apennines during the last interglacial could also account for the high turnover with the  
372 Alps. Indeed, Apennine high-altitude refugia could be too small, warm and isolated to allow  
373 the persistence of cold-adapted species (Marta, Lacasella, Cesaroni, & Sbordoni, 2019).  
374 Currently, several cold-adapted taxa show small, isolated populations in the Apennines that  
375 are considered on the brink of extinction (*Erebia pandrose* and *Erebia montana* in Northern-  
376 Central Apennines, *Erebia gorge* in Southern Apennines), locally declining (*P. mnemosyne*  
377 and *P. apollo*) or have gone recently extinct (*Erebia aethiops* and probably *Erebia gorge* in  
378 Northern Apennines) (E Balletto, Bonelli, & Cassulo, 2007; Cini et al., 2020; Piazzini &  
379 Favilli, 2020). This is also the reason why it has been suggested that distinctions should be  
380 applied in IUCN assessments for butterflies populations from Alps and Apennines (Bonelli et  
381 al., 2018).

382 Butterfly species with endemic genetic lineages and high intraspecific differentiation usually  
383 have a low mobility (lower wingspan and shorter flight period) and a low polyphagy (e.g.  
384 Dapporto et al., 2019; Scalerio et al., 2020). Conversely, we did not find differences in  
385 wingspan and hostplant generalism between endemics and non-endemics, while the shorter  
386 phenology of Alpine species is likely due to their adaptation to shorter summer seasons. It is  
387 then plausible that the possibility of divergence in the Peninsula-Sicily area was not facilitated  
388 by a lower mobility of the species because isolation during cold periods was likely too high  
389 and cancelled any potential effect of mobility.

390

### 391 **Signal for dispersal along and across Alps and Apennines: implications for conservation**

392 Given the effects of recent climate changes, which induced poleward shifts of kilometres per  
393 year for several butterfly species (Parmesan et al., 1999), we can assume that the distribution  
394 of many butterfly species has changed from the onset of the present interglacial. Many  
395 temperate taxa differentiated in the EVOc of Peninsula-Sicily could have dispersed through  
396 the Alps and occupied Central Europe (Hewitt, 1999), thus losing their status of endemics for  
397 this region. However, the high endemic rate, mostly in ME of the Peninsula-Sicily region  
398 indicates that northward shifts might be slowed down at least for many genetic lineages. The  
399 infrequent northward dispersal observed in Italian endemics is usually explained by the  
400 existence of the huge physical barrier of Alps (Drovetski et al., 2018; Hewitt, 1999). If so, we  
401 should find several endemics shared between the Apennines and the southern slopes of the  
402 Alps (pre-Alps), a phenomenon which does not happen. More likely, the barrier to dispersal is  
403 represented by the different climate occurring in the Alps and in Peninsula-Sicily. Alps and  
404 pre-Alps are characterised by *cold and not-dry season* climates in low-altitude areas (Dfb, Dfc  
405 climates in Köppen classification) and by *Polar Tundra* (ET) climate in high-altitude areas.  
406 Conversely, most Peninsula-Sicily shows a *Temperate with dry summer climate* (Csa, Csb),  
407 with a lower incidence of *Temperate not-dry season* and *Cold no-dry season with warm*  
408 *summer* areas (Csc, Dfb) (Beck et al., 2018).

409 We can thus reject the hypothesis that the Alps and the Apennines represent a corridor for  
410 most butterflies (L. Dapporto, Fattorini, Vodă, Dincă, & Vila, 2014) and presumably for other  
411 insect species, with two main breaks located close to the geographical boundary between the  
412 Alps and the Apennines (west Liguria) and on the Strait of Messina. The identification of  
413 these breaks has important consequences for the conservation of the populations living in the  
414 study area. Indeed, the strong turnover existing between the Alps, Peninsula and Sicily

415 demonstrates that each of these areas represents an independent management unit and needs  
416 specific protection. The endemics identified in this study occurred in two centres of similar  
417 size: about 200,000 km<sup>2</sup>. Due to the large size of these centres, future studies should focus on  
418 identifying smaller portions representing key biodiversity areas for potential conservation  
419 actions (Brooks et al., 2015). An analysis of butterfly richness weighted by their risk of  
420 extinction (IUCN assessments) ranked the Alps, the southern tip of the Italian Peninsula and  
421 eastern Sicily as the most important areas for butterfly conservation in Italy (Girardello,  
422 Griggio, Whittingham, & Rushton, 2009). Several National Parks protect important areas of  
423 the peninsula, while only four regional parks with more limited funding are located in Sicily.  
424 In particular, while the Calabrian side of the Strait of Messina is protected by the Aspromonte  
425 and National Park, on the Sicilian side, the Peloritani mounts are completely unprotected.

426

427 **The importance of integrating taxonomical and genetic approaches: methodological**  
428 **implications**

429 The method we used to identify endemic taxa may have strong implications for future studies.  
430 The introduction of phylogenetic diversity and endemism (Faith, 1992; Rosauer, Laffan,  
431 Crisp, Donnellan, & Cook, 2009) added the evolutionary dimension to the study of  
432 communities and to conservation biology (Laity et al., 2015). Such methods assume that  
433 ancient divergence or wider phylogenetic representation of communities have higher value in  
434 identifying areas of endemism and key areas for conservation (Laity et al., 2015).

435 Alternatively, we used a qualitative approach generating a list of endemics for regionalisation  
436 and comparison of species traits. Our evaluation, based on an “or-or” approach, allowed to  
437 include several species that were lumped in a COI-based GMYC analysis, likely due to events  
438 of post-speciation mitochondrial introgression (Dincă et al., 2015). In particular, 18 endemic  
439 entities, widely recognised as good species by butterfly specialists also based on nuclear  
440 markers (Wiemers et al., 2018), were recovered as LE and LME. On the other hand, without  
441 the ME highlighted by GMYC, the Peninsula-Sicily region could not have been identified due  
442 to the low incidence of SE and LE endemics. Accordingly, a study at species level on the  
443 Italian hotspots for Lepidoptera, Carabidae, amphibians and reptiles identified most  
444 irreplaceable areas in the Alps, while none was recovered in Apennines and in Sicily (Emilio  
445 Balletto et al., 2010).

446 Recent reviews indicate key areas of endemism for conservation should be recognised also  
447 based on the intraspecific genetic divergence they encompass (Brooks et al., 2015). If we

448 consider the GMYC entities as units of genetic divergence, most ME endemics of the study  
449 area only show two or three ESUs across the whole West Palaearctic; these fractions indicate  
450 that species defining the Alps and Peninsula-Sicily as centres of endemism encompass a  
451 considerable fraction (33-50%) of the whole genetic differentiation of the species they belong  
452 to (Figure S3).

## 453 **Conclusions**

454 We show that one of the best known European areas for butterfly endemism, genetic  
455 differentiation and richness is composed of two functionally different centres of endemism: an  
456 ecological endemicity centre in the Alps, and an evolutionary endemicity centre in the  
457 Peninsula-Sicily mostly determined by the occurrence of paleoendemics and neoendemics,  
458 respectively. Peninsula and Sicily can also be identified as two different sub-centres. This  
459 result challenges the established perception of the Alpine-Appennine area as a single unit,  
460 frequently termed the “Italian refugium”.

461 We show that, although ME and LE have on average a lower genetic divergence than SE  
462 species, they convey an equivalent and complementary biogeographic meaning, and that only  
463 an approach assuming an equal operational value allowed the recognition of two (or three)  
464 centres of endemism.

465 These findings can have profound implications for a precise identification of areas of  
466 endemism since similar mechanisms have likely acted on other diversity hotspots in the west  
467 Palaearctic (Iberia-Pyrenees-Sierra Nevada, Balkan Peninsula-Balkans-Carpathians, Turkey  
468 and mountain systems therein, Maghreb-Atlas, most areas of Iran). In a simplified scenario,  
469 the functionally different endemics are expected to show different responses to climatic  
470 changes, driving possible range contractions of Alpine endemics to higher altitude areas and  
471 poleward expansions of the Peninsula-Sicily endemics. The outcome of this scenario can be  
472 affected by the quality of the habitat taxa will track, which could be better for mountain  
473 species than for lowland ones, thus determining unpredictable trends (Hülber et al., 2020).

474 The possibility to discern functionally different endemic assemblages will facilitate predicting  
475 such changes and employing strategies oriented to their safeguarding.

476

477

478

479 **Acknowledgements**

480 Support for this research was provided by the Ramón y Cajal program of the Spanish Ministry  
481 of Science and Innovation to GT (RYC2018-025335-I), the Academy of Finland to VD  
482 (Academy Research Fellow, decision no. 328895), by project PID2019-107078GB-I00 / AEI /  
483 10.13039/501100011033 to RV and by the projects "Ricerca e conservazione sui lepidotteri  
484 diurni di sei Parchi Nazionali dell'Appennino centro-settentrionale" and "Ricerca e  
485 conservazione sugli Impollinatori dell'Arcipelago Toscano e divulgazione sui  
486 Lepidotteri del parco" to LD. We thank Francesca Barbero, Luca Pietro Casacci, Sylvain  
487 Cuvelier, Stefano Scalercio, Raluca Vodă, and other people providing help in data collection  
488 with a special reference to iNaturalist and Observado users.

489

490

491

492 REFERENCES

- 493 Anderson, S. (1994). Area and endemism. *The Quarterly Review of Biology*, 69(4), 451–471.
- 494 Balletto, E, Bonelli, S., & Cassulo, L. (2007). insecta Lepidoptera Papilionoidea  
495 (Rhopalocera). *Checklist and Distribution of the Italian Fauna. Ministero Dell'ambiente  
496 e Della Tutela Del Territorio, Direzione per La Protezione Della Natura. Cd-RoM.*
- 497 Balletto, Emilio, Bonelli, S., Borghesio, L., Casale, A., Brandmayr, P., & Vigna Taglianti, A.  
498 (2010). Hotspots of biodiversity and conservation priorities: A methodological approach.  
499 *Italian Journal of Zoology*, 77(1), 2–13.
- 500 Beck, H. E., Zimmermann, N. E., McVicar, T. R., Vergopolan, N., Berg, A., & Wood, E. F.  
501 (2018). Present and future Köppen-Geiger climate classification maps at 1-km  
502 resolution. *Scientific Data*, 5, 180214.
- 503 Bonelli, S., Casacci, L. P., Barbero, F., Cerrato, C., Dapporto, L., Sbordoni, V., ... Balletto, E.  
504 (2018). The first red list of Italian butterflies. *Insect Conservation and Diversity*, 11(5).  
505 doi: 10.1111/icad.12293
- 506 Brooks, T. M., Cuttelod, A., Faith, D. P., Garcia-Moreno, J., Langhammer, P., & Pérez-  
507 Espona, S. (2015). Why and how might genetic and phylogenetic diversity be reflected  
508 in the identification of key biodiversity areas? *Philosophical Transactions of the Royal  
509 Society B: Biological Sciences*, 370(1662), 20140019.
- 510 Brower, A. V. (1994). Rapid morphological radiation and convergence among races of the  
511 butterfly *Heliconius erato* inferred from patterns of mitochondrial DNA evolution.  
512 *Proceedings of the National Academy of Sciences*, 91(14), 6491–6495.
- 513 Cini, A., Barbero, F., Bonelli, S., Bruschini, C., Casacci, L. Pietro, Piazzini, S., ... Dapporto,  
514 L. (2020). The decline of the charismatic *Parnassius mnemosyne* (L.)(Lepidoptera:  
515 Papilionidae) in a Central Italy national park: a call for urgent actions. *Journal of Insect*

516 *Biodiversity*, 16(2), 47–54.

517 Crother, B. I., & Murray, C. M. (2011). Ontology of areas of endemism. *Journal of*  
518 *Biogeography*, 38(6), 1009–1015.

519 Dapporto, L., Fattorini, S., Vodă, R., Dincă, V., & Vila, R. (2014). Biogeography of western  
520 Mediterranean butterflies: Combining turnover and nestedness components of faunal  
521 dissimilarity. *Journal of Biogeography*, 41(9). doi: 10.1111/jbi.12315

522 Dapporto, Leonardo, Cini, A., Vodă, R., Dincă, V., Wiemers, M., Menchetti, M., ... Vila, R.  
523 (2019). Integrating three comprehensive data sets shows that mitochondrial DNA  
524 variation is linked to species traits and paleogeographic events in European butterflies.  
525 *Molecular Ecology Resources*, 19(6), 1623–1636. doi: 10.1111/1755-0998.13059

526 Davis, J., Pavlova, A., Thompson, R., & Sunnucks, P. (2013). Evolutionary refugia and  
527 ecological refuges: key concepts for conserving Australian arid zone freshwater  
528 biodiversity under climate change. *Global Change Biology*, 19(7), 1970–1984.

529 Dincă, V., Montagud, S., Talavera, G., Hernández-Roldán, J., Munguira, M. L., García-  
530 Barros, E., ... Vila, R. (2015). DNA barcode reference library for Iberian butterflies  
531 enables a continental-scale preview of potential cryptic diversity. *Scientific Reports*, 5,  
532 12395.

533 Drovetski, S. V, Fadeev, I. V, Raković, M., Lopes, R. J., Boano, G., Pavia, M., ... Aghayan,  
534 S. A. (2018). A test of the European Pleistocene refugial paradigm, using a Western  
535 Palaearctic endemic bird species. *Proceedings of the Royal Society B*, 285(1889),  
536 20181606.

537 Drummond, A. J., Rambau, A., & Suchard, M. (2013). *BEAST 1.8. 0*.

538 Ehlers, J., Ehlers, J., Gibbard, P. L., & Hughes, P. D. (2011). *Quaternary glaciations-extent*  
539 *and chronology: a closer look* (Vol. 15). Elsevier.

540 Faith, D. P. (1992). Conservation evaluation and phylogenetic diversity. *Biological*

541 *Conservation*, 61(1), 1–10.

542 Fiera, C., Habel, J. C., Kunz, M., & Ulrich, W. (2017). Large-scale phylogenetic structure of  
543 European springtails (Collembola) depends on species range size and postglacial  
544 colonization history. *Biological Journal of the Linnean Society*, 120(3), 664–674.

545 Flantua, S. G. A., Payne, D., Borregaard, M. K., Beierkuhnlein, C., Steinbauer, M. J.,  
546 Dullinger, S., ... Kreft, H. (2020). Snapshot isolation and isolation history challenge the  
547 analogy between mountains and islands used to understand endemism. *Global Ecology  
548 and Biogeography*.

549 Fujisawa, T., & Barraclough, T. G. (2013). Delimiting species using single-locus data and the  
550 Generalized Mixed Yule Coalescent approach: a revised method and evaluation on  
551 simulated data sets. *Systematic Biology*, 62(5), 707–724.

552 Fujisawa, T., Vogler, A. P., & Barraclough, T. G. (2015). Ecology has contrasting effects on  
553 genetic variation within species versus rates of molecular evolution across species in  
554 water beetles. *Proceedings of the Royal Society B: Biological Sciences*, 282(1799),  
555 20142476.

556 Girardello, M., Griggio, M., Whittingham, M. J., & Rushton, S. P. (2009). Identifying  
557 important areas for butterfly conservation in Italy. *Animal Conservation*, 12(1), 20–28.

558 Harrison, S., & Noss, R. (2017). Endemism hotspots are linked to stable climatic refugia.  
559 *Annals of Botany*, 119(2), 207–214.

560 Hawkins, B. A. (2010). Multiregional comparison of the ecological and phylogenetic  
561 structure of butterfly species richness gradients. *Journal of Biogeography*, 37(4), 647–  
562 656.

563 Hewitt, G. M. (1999). Post-glacial re-colonization of European biota. *Biological Journal of  
564 the Linnean Society*, 68(1–2), 87–112.

565 Hülber, K., Kuttner, M., Moser, D., Rabitsch, W., Schindler, S., Wessely, J., ... Dullinger, S.

566 (2020). Habitat availability disproportionately amplifies climate change risks for lowland  
567 compared to alpine species. *Global Ecology and Conservation*, e01113.

568 Jetz, W., & Fine, P. V. A. (2012). Global gradients in vertebrate diversity predicted by  
569 historical area-productivity dynamics and contemporary environment. *PLoS Biol*, 10(3),  
570 e1001292.

571 Keppel, G., Van Niel, K. P., Wardell-Johnson, G. W., Yates, C. J., Byrne, M., Mucina, L., ...  
572 Franklin, S. E. (2012). Refugia: identifying and understanding safe havens for  
573 biodiversity under climate change. *Global Ecology and Biogeography*, 21(4), 393–404.

574 Kudrna, O. (2019). *Distribution of Butterflies and Skippers in Europe* (SOM, Ed.). Prague.

575 Laity, T., Laffan, S. W., González-Orozco, C. E., Faith, D. P., Rosauer, D. F., Byrne, M., ...  
576 Moritz, C. C. (2015). Phylodiversity to inform conservation policy: An Australian  
577 example. *Science of the Total Environment*, 534, 131–143.

578 Liu, Z., Chen, G., Zhu, T., Zeng, Z., Lyu, Z., Wang, J., ... Yang, Z. (2018). Prevalence of  
579 cryptic species in morphologically uniform taxa—Fast speciation and evolutionary  
580 radiation in Asian frogs. *Molecular Phylogenetics and Evolution*, 127, 723–731.

581 Marta, S., Lacasella, F., Cesaroni, D., & Sbordoni, V. (2019). Effects of Holocene climate  
582 changes on alpine ecosystems: Nonequilibrium dynamics drive insect species richness on  
583 alpine islands. *Journal of Biogeography*, 46(10), 2248–2259.

584 Medail, F., & Quezel, P. (1997). Hot-spots analysis for conservation of plant biodiversity in  
585 the Mediterranean Basin. *Annals of the Missouri Botanical Garden*, 112–127.

586 Michaux, J. R., Libois, R., Paradis, E., & Filippucci, M.-G. (2004). Phylogeographic history  
587 of the yellow-necked fieldmouse (*Apodemus flavicollis*) in Europe and in the Near and  
588 Middle East. *Molecular Phylogenetics and Evolution*, 32(3), 788–798.

589 Moritz, C., Hoskin, C. J., MacKenzie, J. B., Phillips, B. L., Tonione, M., Silva, N., ...  
590 Graham, C. H. (2009). Identification and dynamics of a cryptic suture zone in tropical

591 rainforest. *Proceedings of the Royal Society B: Biological Sciences*, 276(1660), 1235–  
592 1244.

593 Mutanen, M., Hausmann, A., Hebert, P. D. N., Landry, J.-F., de Waard, J. R., & Huemer, P.  
594 (2012). Allopatry as a Gordian knot for taxonomists: patterns of DNA barcode  
595 divergence in Arctic-Alpine Lepidoptera. *PLoS One*, 7(10), e47214.

596 Ohlemüller, R., Anderson, B. J., Araújo, M. B., Butchart, S. H. M., Kudrna, O., Ridgely, R.  
597 S., & Thomas, C. D. (2008). The coincidence of climatic and species rarity: high risk to  
598 small-range species from climate change. *Biology Letters*, 4(5), 568–572.

599 Parmesan, C., Ryrholm, N., Stefanescu, C., Hill, J. K., Thomas, C. D., Descimon, H., ...  
600 Tammaru, T. (1999). Poleward shifts in geographical ranges of butterfly species  
601 associated with regional warming. *Nature*, 399(6736), 579–583.

602 Peña, C., Witthauer, H., Klečková, I., Fric, Z., & Wahlberg, N. (2015). Adaptive radiations in  
603 butterflies: evolutionary history of the genus *Erebia* (Nymphalidae: Satyrinae).  
604 *Biological Journal of the Linnean Society*, 116(2), 449–467.

605 Petit, R. J., Aguinagalde, I., de Beaulieu, J.-L., Bittkau, C., Brewer, S., Cheddadi, R., ...  
606 Lascoux, M. (2003). Glacial refugia: hotspots but not melting pots of genetic diversity.  
607 *Science*, 300(5625), 1563–1565.

608 Piazzini, S., & Favilli, L. (2020). Rediscovery of a species considered to be extinct in the  
609 Pollino massif (Calabria, Italy): *Erebia gorge* (Hübner,[1804])(Lepidoptera:  
610 Nymphalidae). *Fragmenta Entomologica*, 52(1), 43–46.

611 Platania, L., Menchetti, M., Dincă, V., Corbella, C., Kay-Lavelle, I., Vila, R., ... Dapporto, L.  
612 (2020). Assigning occurrence data to cryptic taxa improves climatic niche assessments:  
613 Biodecrypt, a new tool tested on European butterflies. *Global Ecology and*  
614 *Biogeography*, (April), 1–14. doi: 10.1111/geb.13154

615 Quek, S., Davies, S. J., Itino, T., & Pierce, N. E. (2004). Codiversification in an ant–plant

616 mutualism: stem texture and the evolution of host use in *Crematogaster* (Formicidae:  
617 Myrmicinae) inhabitants of *Macaranga* (Euphorbiaceae). *Evolution*, 58(3), 554–570.

618 Rosauer, D. A. N., Laffan, S. W., Crisp, M. D., Donnellan, S. C., & Cook, L. G. (2009).  
619 Phylogenetic endemism: a new approach for identifying geographical concentrations of  
620 evolutionary history. *Molecular Ecology*, 18(19), 4061–4072.

621 Sandel, B., Weigelt, P., Kreft, H., Keppel, G., van der Sande, M. T., Levin, S., ... Knight, T.  
622 M. (2020). Current climate, isolation and history drive global patterns of tree  
623 phylogenetic endemism. *Global Ecology and Biogeography*, 29(1), 4–15.

624 Scalercio, S., Cini, A., Menchetti, M., Vodă, R., Bonelli, S., Bordoni, A., ... Vila, R. (2020).  
625 How long is 3 km for a butterfly? Ecological constraints and functional traits explain  
626 high mitochondrial genetic diversity between Sicily and the Italian Peninsula. *Journal of*  
627 *Animal Ecology*.

628 Schmitt, T. (2007). Molecular biogeography of Europe: Pleistocene cycles and postglacial  
629 trends. *Frontiers in Zoology*, 4(1), 11.

630 Seddon, J. M., Santucci, F., Reeve, N. J., & Hewitt, G. M. (2001). DNA footprints of  
631 European hedgehogs, *Erinaceus europaeus* and *E. concolor*: Pleistocene refugia,  
632 postglacial expansion and colonization routes. *Molecular Ecology*, 10(9), 2187–2198.

633 Taberlet, P., Fumagalli, L., Wust-Saucy, A., & Cosson, J. (1998). Comparative  
634 phylogeography and postglacial colonization routes in Europe. *Molecular Ecology*, 7(4),  
635 453–464.

636 Vodă, R., Dapporto, L., Dincă, V., & Vila, R. (2015). Cryptic matters: overlooked species  
637 generate most butterfly beta-diversity. *Ecography*, 38(4), 405–409.

638 Wiemers, M., Balletto, E., Dincă, V., Fric, Z. F., Lamas, G., Lukhtanov, V., ... Verovnik, R.  
639 (2018). An updated checklist of the European Butterflies (Lepidoptera, Papilionoidea).  
640 *ZooKeys*, 811, 9–45. doi: 10.3897/zookeys.811.28712

641 Wiemers, M., Chazot, N., Wheat, C. W., Schweiger, O., & Wahlberg, N. (2020). A complete  
642 time-calibrated multi-gene phylogeny of the European butterflies. *ZooKeys*, 938, 97.  
643 Zuloaga, J., Currie, D. J., & Kerr, J. T. (2019). The origins and maintenance of global species  
644 endemism. *Global Ecology and Biogeography*, 28(2), 170–183.

645

646

647

648

649 **Data Accessibility Statement:**

650 R scripts and data to replicate the analysis are available in Dryad doi.XXXXXX. Previously  
651 published and newly generated COI data are available in the “BIBSA” and “EULEP”  
652 BOLD projects where the GenBank accession codes are also reported and in Dryad.

653

654 **Author Contributions:**

655 LD, MM, GT, AC, VS and RVi designed the research, LD, MM, AC, LP, VS validated  
656 records on iNaturalist, EB and SB collected literature records, all the authors collected field  
657 data on butterflies, VD, LD, MM, RVi and LP curated COI sequencing, MM, LD and AC  
658 collected trait data, LD and GT carried out the analyses, all the authors wrote the paper.

659

660

661 **Table 1.** Ecological traits describing alpha and beta niches.

662

Type	Trait	Measure description
Alpha traits	Trophic generalism (feeding trait)	The number of host plant genera
	Mobility (morphology trait)	The wingspan index (Middleton-Welling et al. in press) based on multiple bibliographic measurements of wingspan
Beta traits	Phenology (life history trait)	The number of months during which adults occur in Europe
		The first month when adults fly
		The last month when adults fly
		The maximum number of generations (voltinism) per year recorded in Europe
	Distribution and environmental preferences	The number of 30x30 km <sup>2</sup> occupied in Europe (range size)
		The maximum altitude reported
		The minimum altitude reported
		Altitudinal range

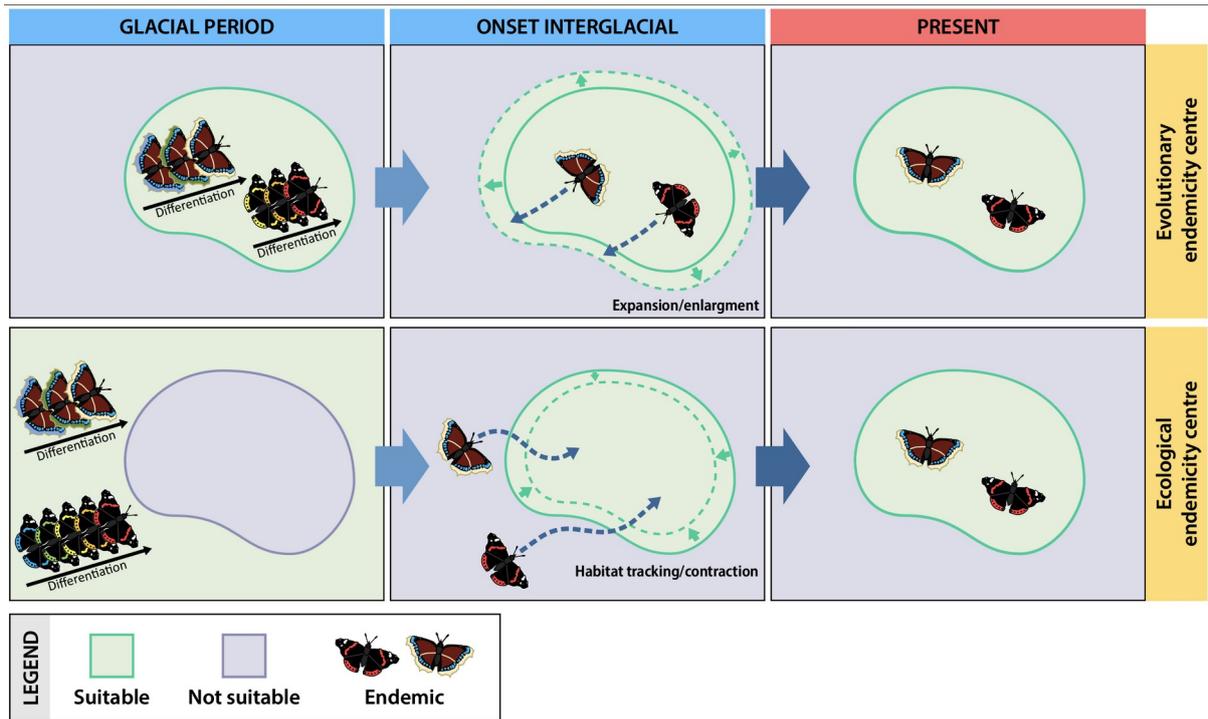
663

664

665 **Table 2.** Phylogenetic ANOVA and homogeneity of variance comparing species traits and  
666 divergence from the closest relative among endemics and non-endemics and between entities  
667 from the two centres: Sum-Sq, sum of squares; Mean-Sq, mean squares; F, F value; P, p-value  
668 without considering phylogeny (typical ANOVA); P(phy), p-value adjusted for phylogeny  
669 (Phylogenetic ANOVA); P(var), p-value associated to the Fligner-Killeen test for  
670 homogeneity of variances. P-values in bold indicate significant results. Sample size is 319 for  
671 Non-end Alps vs End Alps; 242 for Non-end Peninsula-Sicily (PS) vs End PS; 62 for End  
672 Alps vs End PS.  
673  
674

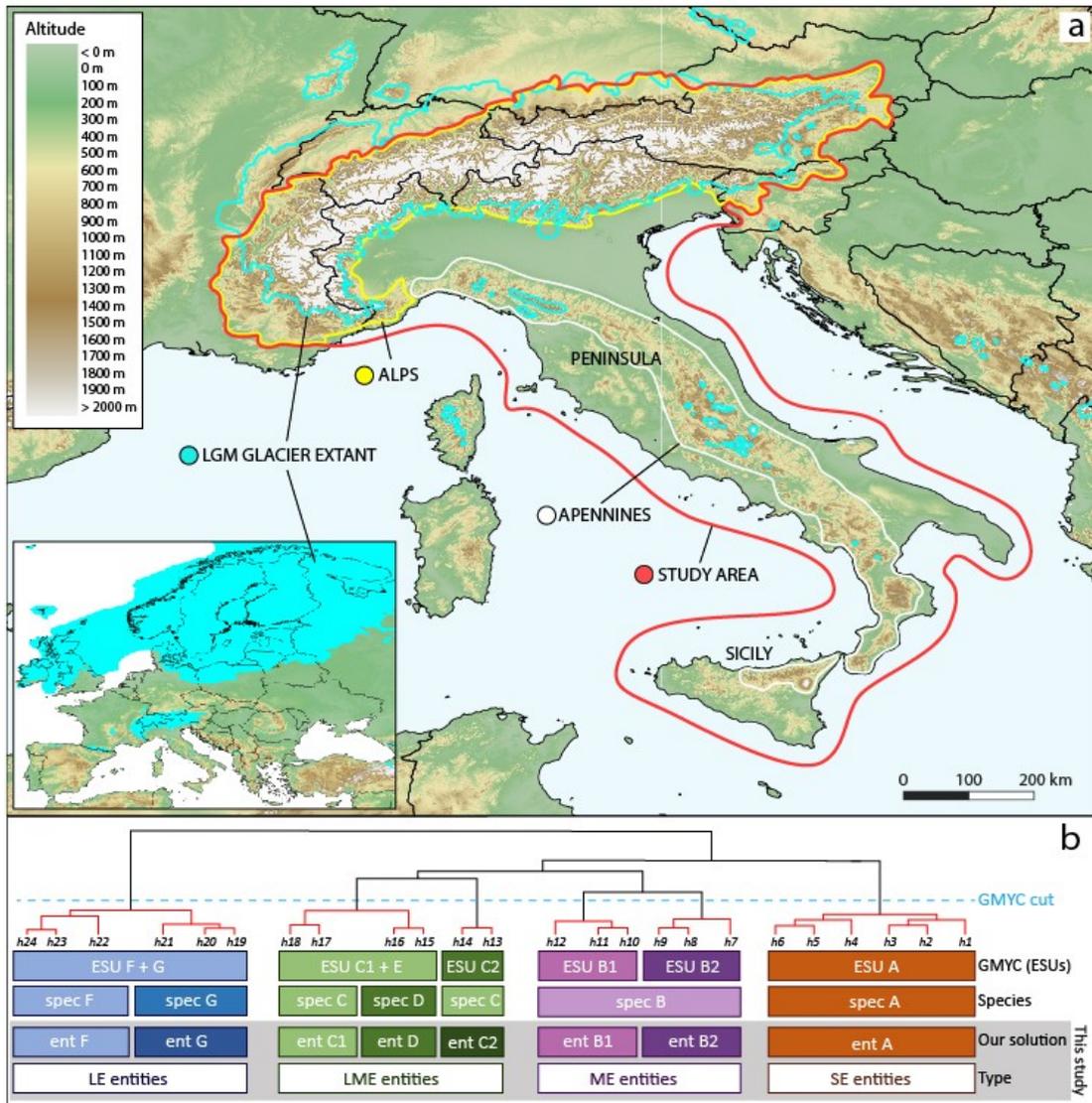
Groups	Feature	Sum-Sq	Mean-Sq	F	P	P(phy)	P(var)
Non-end Alps vs End Alps	Wingspan	0.003	0.003	1.955	0.163	0.541	<b>0.009</b>
	Host plants	11.692	11.692	15.542	<b>&lt;0.001</b>	0.046	<b>0.010</b>
	Phenology PC1	0.085	0.085	34.730	<b>&lt;0.001</b>	<b>0.001</b>	<b>&lt;0.001</b>
	Distribution PC1	0.130	0.130	53.014	<b>&lt;0.001</b>	<b>0.001</b>	0.076
	Distribution PC2	0.205	0.205	96.017	<b>&lt;0.001</b>	<b>0.001</b>	0.402
	Closest relative	233.000	232.957	5.254	<b>0.023</b>	-	0.373
Non-end PS vs End PS	Wingspan	0.000	0.000	0.009	0.925	0.928	0.931
	Host plants	1.252	1.252	1.606	0.206	0.181	0.758
	Phenology PC1	0.004	0.004	1.446	0.230	0.205	0.376
	Distribution PC1	0.005	0.005	1.810	0.180	0.144	0.375
	Distribution PC2	0.002	0.002	0.899	0.344	0.308	0.555
	Closest relative	274.200	274.215	6.344	<b>0.012</b>	-	<b>0.001</b>
End Alps vs End PS	Wingspan	0.002	0.002	1.777	0.188	0.354	<b>0.020</b>
	Host plants	2.377	2.377	4.837	0.032	0.137	0.301
	Phenology PC1	0.035	0.035	34.407	<b>&lt;0.001</b>	<b>&lt;0.001</b>	<b>0.001</b>
	Distribution PC1	0.044	0.044	25.668	<b>&lt;0.001</b>	<b>&lt;0.001</b>	0.406
	Distribution PC2	0.137	0.137	64.671	<b>&lt;0.001</b>	<b>&lt;0.001</b>	0.847
	Closest relative	6.800	6.802	1.209	0.276		<b>0.006</b>

675

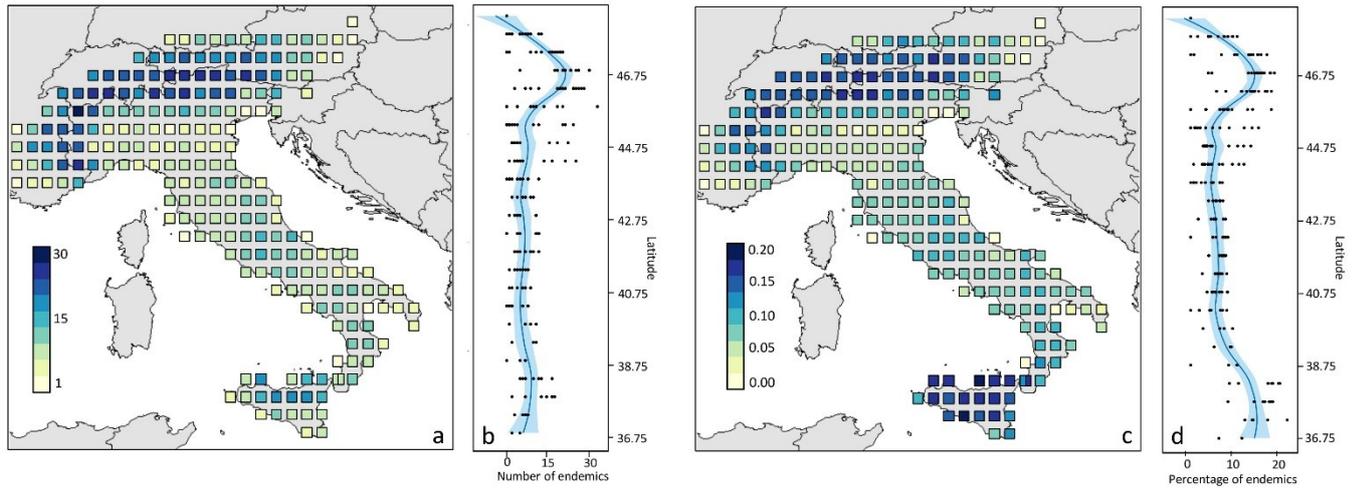


676  
677

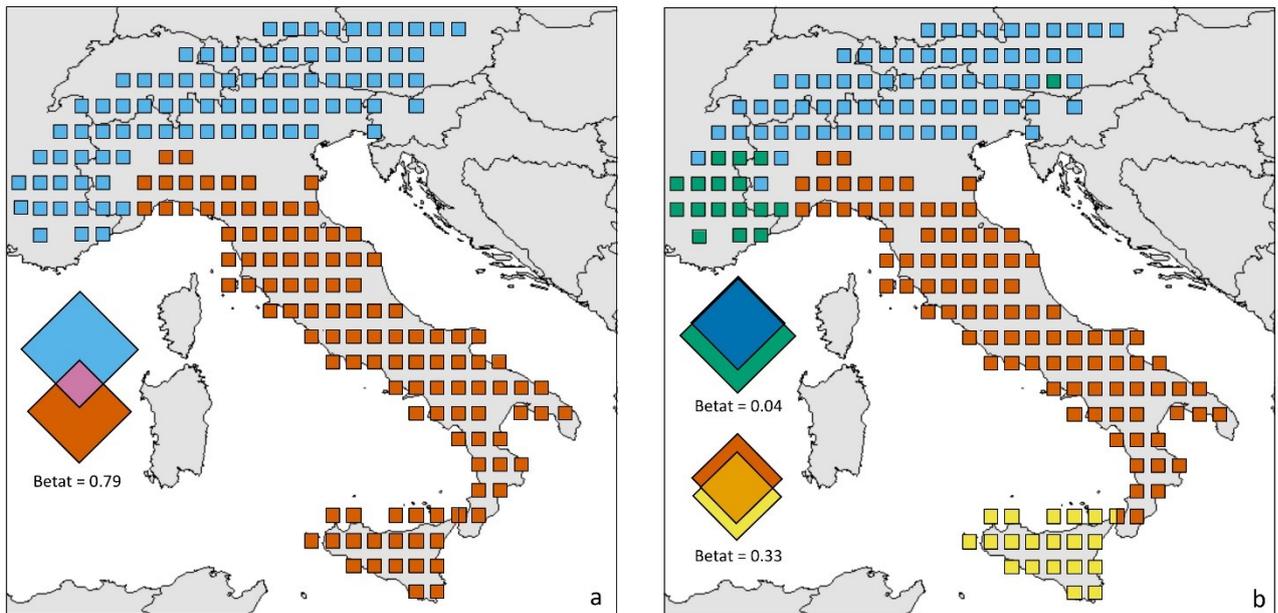
678 Figure 1. A model representing the mechanisms generating evolutionary and ecological  
679 endemism centres for two endemic taxa. In evolutionary endemism centres, species  
680 differentiate during similar time lags in suitable areas (typically glacial refugia), which may  
681 expand (dotted line) following climatic changes (typically during interglacial periods). In the  
682 ecological endemism centres, species that evolved over large areas with different degrees of  
683 divergence (typically during glacial periods) converge in particular areas after major  
684 environmental changes (e.g. during interglacials) and reduce their ranges as long as the  
685 environmental process operates (dotted line).  
686



687 Figure 2. (a) The study area (red polygon), including the boundaries of the Alps (yellow  
 688 polygon), the Apennines (white polygon), and those areas covered by the ice cap during the  
 689 last glacial maximum (LGM) (cyan polygons). The inset shows the distribution of ice sheets  
 690 during the LGM in Europe. (b) A scheme summarising how entities are identified as an “or-  
 691 or” combination of taxonomy (species) and GMYC delimitations (ESUs).  
 692  
 693

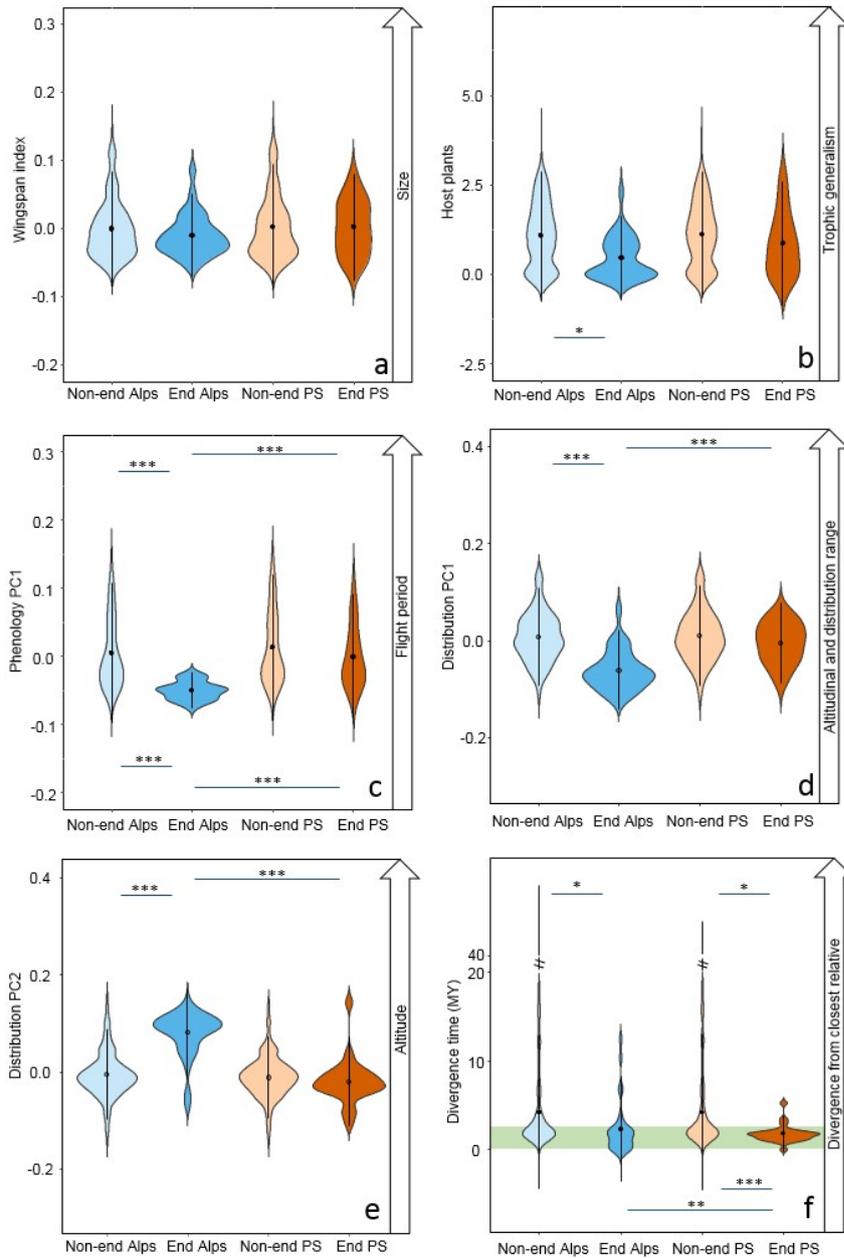


694  
 695 Figure 3. (a) Number of endemic entities recorded in each 0.5x0.5 cell and (b) latitudinal  
 696 trend obtained by loess regression. (c) Percentage of endemic over observed richness (d)  
 697 showing two peaks in Sicily and Alps.  
 698



700  
 701  
 702  
 703  
 704  
 705  
 706

Figure 4. The solutions for  $k=2$  (a) and  $k=4$  (b) obtained by applying the `recluster.region` algorithm to the occurrence of endemic taxa. The size of diamonds in figures a and b is proportional to the number of endemics occurring in each centre, and their overlap represents the number of shared species. Values of turnover index (Betat) calculated on these numbers are also provided.



707  
708

709 Figure 5. Violin plots comparing the distribution of the features of taxonomic species  
 710 involved in non-endemic taxa occurring on Alps (Non-end Alps) and on Peninsula-Sicily  
 711 (Non-end PS), with endemics from the Alpine (End Alps) and Peninsular-Sicily (Non-end PS)  
 712 centres. Above the plots we reported significant pairwise comparison of means as obtained by  
 713 phylogenetic ANOVAs; below the plots we reported significant pairwise comparisons  
 714 between variances (Fligner-Killeen test). All significance assessed after sequential Bonferroni  
 715 correction. The green area in (f) represents the Pleistocene and Holocene (2.6MY-present).