DR. ROGER VILA (Orcid ID : 0000-0002-2447-4388)
DR. LEONARDO DAPPORTO (Orcid ID : 0000-0001-7129-4526)

Article type : Original Article

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

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

[1] ZEN Lab, Dipartimento di Biologia dell'Università di Firenze, via Madonna del Piano 6 50019 Sesto Fiorentino, Italy.

[2] Institut de Biologia Evolutiva (CSIC - Universitat Pompeu Fabra), Passeig Marítim de la Barceloneta 37, 08003 Barcelona, Catalonia, Spain.

[3] Institut Botànic de Barcelona (IBB, CSIC-Ajuntament de Barcelona), Passeig del Migdia s/n, 08038 Barcelona, Catalonia, Spain.

[4] Centre for Biodiversity & Environment Research, University College London, Gower Street, London, WC1E 6BT.

[5] Ecology and Genetics Research Unit, University of Oulu, PO Box 3000, Oulu, 90014, Finland.

[6] ZOOLAB, Dipartimento di Scienze della Vita e Biologia dei Sistemi, Università degli Studi di Torino, Via Accademia Albertina 13, 10123 Turin, Italy.

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

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

Running title: Endemicity centres on Alps and Apennines

ABSTRACT

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

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

INTRODUCTION

Endemicity is a central concept in biogeography and conservation biology, denoting the condition of a taxon to be exclusively distributed in a given area (Anderson, 1994). Consequently, endemics cannot be identified without defining a region within vaster areas. The centres of endemism are regions where several endemics co-occur (Harrison & Noss, 2017) given a combination of geographical, historical and ecological processes (Sandel et al., 2020; Zuloaga, Currie, & Kerr, 2019). They are typically limited by barriers and characterised by stable and often singular climatic conditions, so that genetic and faunistic divergence can accumulate (Ohlemüller et al., 2008; Sandel et al., 2020; Zuloaga et al., 2019). In this respect, the concept of centres of endemism largely overlaps with that of refugia (Keppel et al., 2012). While a main objective of conservation biology is to identify and protect centres of endemism —because they host peculiar and irreplaceable biodiversity elements and function as refuges from ongoing environmental changes (Brooks et al., 2015; Harrison & Noss, 2017)— a major goal of evolutionary ecology is to understand the mechanisms that produced centres of endemism (Crother & Murray, 2011; Keppel et al., 2012).

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

Another pathway generating clusters of endemics arise when, after major environmental changes, species widely distributed track their habitats to reduced "safe havens" (Crother & Murray, 2011). These taxa, evolved in different areas, end up co-occurring in habitat remnants to which we refer as ecological endemicity centres (ECOcs, Figure 1). ECOcs coincide with the *ex situ* refugia (Keppel et al., 2012) and ecological refuges (Davis et al., 2013). A paradigm for ECOcs is the co-occurrence of boreo-alpine species in areas formerly covered by ice-sheets (Mutanen et al., 2012). These mechanisms are not mutually exclusive and, in heterogeneous environmental settings, composite endemicity centres assembled after evolutionary and ecological processes can emerge (Crother & Murray, 2011). The methods to identify centres of endemism are well established;

conversely the mechanisms behind their emergence have often remained unresolved. Their understanding requires the application of eclectic approaches combining palaeoecological reconstructions, high-resolution occurrence data, phylogeographic assessments and species functional traits for large homogenous taxa (Brooks et al., 2015; Davis et al., 2013; Keppel et al., 2012; Zuloaga et al., 2019). We hypothesised that endemic assemblages formed under different processes show different features. EVOc endemics have likely evolved under similar processes, they should have a similar degree of genetic divergence from closest relatives and should spread all along the existing phylogenetic diversity. A completely different pattern is expected in ECOcs representing secondary sympatry areas. Here endemics with dissimilar evolutionary histories met and are expected to encompass a higher variability in divergence time (Moritz et al., 2009). Moreover, ECOc endemics should belong to a reduced subset of families or genera strictly adapted to the environmental settings rarefied after historical changes. In this respect they are expected to show higher phylogenetic clustering and lower variance in phenotypic traits.

Recognising endemic entities is a challenging task. Indeed, macroecological studies usually employ entities recognised by taxonomists at the species level. On the other hand, phylogeography is rooted on the pervasive evidence that most species encompass a wide variation of spatially-structured diversity both as cryptic taxa and as genetic lineages. Genetic lineages are not recognised in taxonomic catalogues, are usually excluded by macroecological studies, are not protected and there are no protocols to include them in conservation plans (e.g. IUCN Red List) (Brooks et al., 2015). This exclusion results in a significant loss of the signal of the Quaternary processes they convey and discards their fundamental contribution to biogeography and conservation (Brooks et al., 2015; Vodă, Dapporto, Dincă, & Vila, 2015).

Here, we identified entities in the entire butterfly fauna (269 species) occurring along the Alps, the Italian Peninsula and surrounding islands (Figure 2a) based on an “and/or” approach, where an entity is represented by a group of individuals recognised as a species by taxonomists and/or based on a phylogenetic-based species delimitation approach (GMYC) (Figure 2b). GMYC is increasingly used in macroecology (Fujisawa, Vogler, & Barraclough, 2015) and sometimes it proved to be more effective than taxonomic assessments in documenting eco-evolutionary processes (Liu et al., 2018). Using this dataset, we investigate whether the two different kinds of centres of endemism can be identified. This region represents an ideal system to test this hypothesis because it is located in the centre of the Mediterranean, a major hotspot where particularly high biodiversity has emerged from the interplay between Africa and Eurasia and the possibility for many species to persist during the Pleistocene (Bonelli et al., 2018). The continuous S-shaped mountain-hill system comprising the Alps and the Apennines (37º to 48º of latitude), encompasses the glacial refugium of the Italian Peninsula (Dapporto et al., 2019; Hewitt, 1999, Taberlet et al., 1998) and mountain areas covered by ice caps during glacial maxima (Figure 2a). Typically, the region is considered as a single biogeographic unit: the Italian refugium (Drovetski et al., 2018; Hewitt, 1999; Petit et al., 2003; Taberlet et al., 1998), but based on paleogeographic and paleoclimatic evidence (Figure 2a), we hypothesise that distinct EVOcs and ECOcs occur in this region. To test this, we 1) combine a massive dataset of COI sequences and occurrence data to evaluate if different centres are recognisable by regionalisation analysis, 2) verified whether the endemics from the potential ECOc have stricter requirements in key ecological traits and encompass a lower phylogenetic diversity and 3) verified whether endemics from potential EVOc show similar and shorter divergence times, being mostly represented by intraspecific genetic lineages. Answering these questions can provide fundamental insights for understanding the ecological and evolutionary processes generating endemicity in biodiversity hotspots and informs towards more effective conservation strategies.

METHODS

Sampling and datasets

The study area includes the Alps (www.alpconv.org), the Italian Peninsula, Sicily and the small Italian islands closer to this land than to any other (Figure 2a). We obtained 307,228 records for butterfly species as recognised in Wiemers et al. (2018) within the study area for cells of 0.5x0.5 degrees of latitude and longitude, corresponding to 1277 km2 in the centre of the study area (Rome) (sources described in Appendix S1). We generated occurrence maps for each species and compared them with the distribution of European butterflies (Kudrna, 2019) with the goal to remove possible misplaced records. After filtering unique occurrences for each cell, we counted 27,123 records (available in Dryad https://doi.org/10.5061/dryad.tb2rbnzzf). We gathered 26,557 COI (standard barcode, 658 bp) sequences from 519 species occurring in the Western Palearctic (Dryad). Among these, 23,563 COI sequences belong to the 269 species occurring in the study area (DS-ALPAPENN BOLD datasets).

Phylogeny and GMYC

We collapsed the COI dataset to unique haplotypes using the "haplotype" function of the R package "pegas" (https://cran.r-project.org/web/packages/pegas/index.html). We used BEAST 1.8 (Drummond, Rambau, & Suchard, 2013) to reconstruct five ultrametric phylogenetic trees, one for each butterfly family (the single European Riodinidae was merged with Lycaenidae) (available in Dryad). The number of haplotypes was 6459 (3232 Nymphalidae, 644 Pieridae, 561 Hesperiidae, 247 Papilionidae and 1775 Lycaenidae-Riodinidae). Each dataset included one outgroup for each of the other families. Two independent chains of 100 million generations were run in BEAST for each dataset. The substitution model was set to GTR+I+G with six gamma rate categories. A coalescent tree prior was set. Divergence times were estimated by applying a strict clock and a normal prior distribution centred on the mean between two widely used substitution rates of 1.5% uncorrected pairwise distance per million years (Quek, Davies, Itino, & Pierce, 2004), and 2.3 % (Brower, 1994). Values were sampled every 10% of the run length and convergence was inspected in Tracer v.1.6 (http://tree.bio.ed.ac.uk/software/tracer/). We applied the general mixed Yule-coalescent model (GMYC, Fujisawa & Barraclough, 2013) for each family tree to identify evolutionary significant units (ESUs) using the R package "splits" (https://cran.r-project.org/web/packages/SplitSoftening/index.html) with default settings.

We identified entities as taxa recognised by the taxonomic list of Wiemers et al. (2018) and/or as haplotypes belonging to different GMYC ESUs (Figure 2b). According to the GMYC results each species identified by Wiemers et al. (2018) could be 1) "single entity species (SE)": all haplotypes of a species belong to a single GMYC ESU, 2) "multiple entity species (ME)": haplotypes belong to two or more ESUs, 3) "lumped entities (LE)": two or more species are recovered as a single ESU, and 4) "lumped + multiple entities (LME): species are split in multiple ESUs and lumped with other species (Figure 2b).

For SE and LE all occurrences were attributed to the original species while for ME and LME, we attributed species occurrence to their most probable ESU by using "biodecrypt" ("recluster" R package, https://rdrr.io/github/leondap/recluster/). The function creates concave hulls based on the distribution of the sequences attributed to a given ESU and uses the relative hull geometries to attribute unknown occurrence data to a given species (Platania et al., 2020)(see Appendix S1, Figures S1 and S2 for details). The "biodecrypt" function also provides a measure for hull overlap as an evaluation of sympatry among cryptic entities.

We identified as endemics those entities for which all COI sequences occurred exclusively within the study area.

1) Which are the centres of endemism?

To locate the centres of endemism we ran regionalisation analyses for the occurrence data of endemics in 0.5x0.5 cells. We used the "recluster.region" function in the R package "recluster" (https://cran.r-project.org/web/packages/recluster/index.html) specifically designed to retrieve biogeographic regions at the intracontinental scale. We obtained clustering solutions from 2 to 8 centres based on two indices of beta-diversity suited to identify regions based on vicariant patterns of distribution: (1) the Simpson turnover index, accounting for species replacement in terms of faunistic elements and (2) the species replacement component of the phylogenetic beta diversity index PhyloSor (Leprieur et al. 2012), which also accounts for the phylogenetic dissimilarity among communities. As a phylogenetic reference, we used the time-calibrated phylogenetic tree for all 496 species of European butterflies, based on 14 mitochondrial and nuclear genes (Wiemers, Chazot, Wheat, Schweiger, & Wahlberg, 2020). The PhyloSor index has been calculated using the "betapart" R package (https://cran.r-project.org/web/packages/betapart/index.html). The "recluster.region" function also calculates the silhouette width and the explained dissimilarity, evaluating how cells resemble those of their own centre (cohesion) compared to other centres (separation). Once the centres were obtained, we identified their exclusive endemics using the "indval" function in the "labdsv" R package (https://cran.r-project.org/web/packages/labdsv/labdsv.pdf).

2) Are endemics characterised by different ecological traits and phylogenetic diversity in different centres?

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

To assess differences in traits we applied a phylogenetic ANOVA, using the "aov.phylo" function of the R package "geiger" (https://cran.r-project.org/web/packages/geiger/index.html). As a reference phylogeny we used the time-calibrated phylogenetic tree of European butterflies (Wiemers et al., 2020). We carried out pairwise comparisons through sequential Bonferroni corrections. We log transformed the number of host plants to improve its normality. To investigate if the traits show different variances among groups, we carried out tests of variance homogeneity (followed by pairwise comparisons with sequential Bonferroni correction) through the non-parametric Fligner-Killeen test, using the "check_homogeneity" function of the R package "performance" (https://cran.r-project.org/web/packages/performance/index.html).

To understand if the communities of each centre showed a reduced phylogenetic diversity compared to the entire European butterfly fauna, we compared phylogenetic distances between all entities recorded in each centre and the phylogenetic tree of European butterflies. As measures we used the mean pairwise distances (MPD, i.e. the mean distances between all species in each community) and the mean nearest taxon distance (MNTD, i.e. the mean distance separating each species in the community from its closest relative) (Webb, Ackerly, McPeek, & Donoghue, 2002). Then, we compared phylogenetic distances of the endemics of each centre against the tree of the whole community they belong to, by pruning the European butterfly tree to include only the species of each centre. Finally, we tested if the phylogenetic distances among endemics of each centre differ from those showed in the whole European butterfly fauna. We compared the distances among the tested community to those obtained for 10,000 communities of the same richness randomly selected in the phylogenetic tree using the "ses.mpd" and the "ses.mntd" functions of the "picante" R package (https://cran.r-project.org/web/packages/picante).

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

For each entity, we obtained genetic divergence from its closest entity in the five family trees using the function "distTips" of the R package "adephylo" (https://cran.r-project.org/web/packages/adephylo/index.html). We compared divergence between endemics from different centres and between endemics and non-endemics of the same region using ANOVAs and comparison of variance as described above. We also compared divergence among types of entity using the same method. We did not apply a correction for phylogenetic autocorrelation because genetic distances are exactly the variable compared here. We compared the incidence of the SE, ME, LE and LME endemics among centres through a Chi Square test.

RESULTS

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

1) Which are the centres of endemism?

The 69 endemics were not homogeneously distributed along the study area. The number of endemics showed a single peak of richness with about 20 endemics per cell over the Alps (Figure 3a,b, Figure S4-72 Appendix S1 for individual distributions). When the effect of local richness was removed by calculating percentages of endemics, Sicily emerged as a main endemicity hotspot (Figure 3c), showing endemicity percentages around 15%, similar to Alps (Figure 3d). Regionalisation among cells containing at least two endemics revealed the same solution in taxonomic and phylogenetic beta-diversity for k=2 and k=3 clusters. For k=2, the solution showed a silhouette value of 0.556 and 0.651 and explained dissimilarity of 68.9% and 73.0% for the Simpson and the nestedness component of PhyloSor indices, respectively. Both solutions separated the Alps from the Italian Peninsula, Sicily and surrounding islands (hereafter Peninsula-Sicily centre) (Figure 4a). The "indval" function showed that 62 of 69 endemics exclusively occur in the Alps or in Peninsula-Sicily, suggesting a strong turnover (Simpson index = 0.79) (diamonds in Figure 4a). A solution of k=3, identical for the Simpson and PhyloSor indices, split the Peninsula-Sicily centre into Italian Peninsula and Sicily (red and yellow regions in Figure 4b) with a higher silhouette (0.623 and 0.710) and a substantial increase of explained dissimilarity to 79.4% and 84.1% for the Simpson and PhyloSor indices, respectively. This higher silhouette can be explained by a moderate distinction in endemics between the Italian Peninsula and Sicily, including 17 species exclusive of a single region and 16 shared (Simpson index = 0.33, diamonds in Figure 4b). A partition of k=4 showed different patterns when using the Simpson and PhyloSor indices. By using the Simpson index, recluster.region separated the Western and Eastern Alps (Figure 4b), with a lower silhouette of 0.558 and an explained dissimilarity of 84.6%, but the two Alpine regions showed a low turnover (16 exclusive vs 26 shared species with a low Simpson turnover index = 0.04). For the PhyloSor index a partition for k=4 separated Northern from Central-Southern Apennines, again with a lower silhouette (0.567) and a moderate increase of explained dissimilarity (89.4%) (Figure 4c). Also in this case the turnover between the two Apennine regions was low (4 exclusive vs 6 shared species) (diamonds in Figure 4c). Partitions for higher k gradually lowered silhouette values (always <0.5) and lost geographic coherence. For this reason, the solution with k=2 (Alps, Peninsula-Sicily) is preferred for the highest turnover, followed by k=3 (Alps, Peninsula, Sicily). The Alpine centre of endemism is about 210,000 km² while the Peninsula-Sicily centre is about 270,000 km².

2) Are endemics characterised by different traits in different centres?

The PCA identified one function from the phenological traits showing eigenvalues higher than one (Figure S73-74, Appendix S1), mostly correlated with flight period and voltinism. The PCA for distribution traits extracted a component positively correlating with distribution and altitudinal ranges, and a component positively correlated with minimum and maximum altitudes.

When comparing traits of endemics exclusive of centres obtained for k=2, Alps endemics showed a significantly shorter flight period compared to both Peninsula-Sicily endemics and Alpine non-endemics, occurred over smaller areas, at higher altitudes and with narrower altitude ranges. Notably, there were no significant differences between the traits of Peninsula-Sicily endemics and non-endemics (Table 2, Figure 5). In many cases the variances also differed (Table 2, Figure 5b,f) and alpine endemics significantly showed lower variance for wingspan, host plant specialisation and phenology. No comparisons have been made using three centres (Alps, Peninsula, Sicily) because the entities would have been too few to obtain reliable results. MPD and MNTD tests revealed that the whole faunas of the Alps and Peninsula-Sicily have similar phylogenetic distances compared to the entire European butterfly fauna (Table 3). Similarly, endemics from Peninsula-Sicily did not show significant differences in both measures compared to random subsets of European taxa and of the community they belong to (Table 3). Conversely, the Alpine endemics revealed a significantly lower phylogenetic diversity compared to random subsets of European and Alpine butterflies for both MPD and MNTD (Table 3), underlying phylogenetic clustering.

In fact, the 42 Alpine endemics represented only four families and 12 genera. No endemic species of Pieridae occurred and a single genus was represented among endemic Hesperiidae and Papilionidae (*Pyrgus* and *Parnassius*). Moreover, 16 entities (about 38% of Alpine endemics) belonged to the single genus *Erebia*. Finally, some endemics were lumped (see below), therefore lacking differentiation on the phylogenetic tree (*Pyrgus carlinae, Erebia styx, E. stiria, E. tyndarus* and *Melitaea aurelia*). On the other hand, the 34 Peninsula-Sicily endemics spread over five families and 22 genera and the most represented genus (*Melitaea*) only includes 6 entities (about 17% of peninsular endemics) (Supplementary Figures S4-S72).

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

Phylogenetic ANOVA showed that different types of endemics showed different divergence (Df=3, Sum. Sq.=151.74, F=15.406, $P<0.001$) and post-hoc comparisons revealed that SE have the highest divergence, followed by ME and then by LE and LME (with similar divergence) (Figure S75 for p-values). Endemics showed a lower divergence from their closest relatives compared to non-endemics (Table 2, Figure 5f). Moreover, Peninsula-Sicily endemics showed lower variance in divergence, compared to both Alpine endemics and non-endemics from their centre with a particularly high frequency around 2 MY (Figure 5f). SE, LE, LME showed higher frequencies in the Alps (13, 8 and 4, respectively) compared to Peninsula-Sicily (3, 5, 0, respectively), while ME were less frequent in the Alps compared to Peninsula-Sicily (17 vs 26). A Chi Square test of independence of the frequencies showed a highly significant effect (χ^2= 12.118, df = 3, p=0.007).

DISCUSSION

Understanding the mechanisms generating centres of endemicity is crucial for the comprehension of the origin of spatial patterns of biodiversity on Earth and instrumental for their protection. Based on the distribution of endemic species and GMYC ESUs and their ecological traits we provide support for the existence of two main types of endemicity centres along the European region formed by the continuous mountain chain of the Alps and the Apennines. The endemics of the two centres showed differences in their ecological traits, in their phylogenetic distances and in the variance of genetic divergences that align with predictions for the existence of two different endemicity centres: an ecological centre, originated as an *ex situ* refugium after the recent occupation of the formerly glaciated Alps and an evolutionary centre, originated as an *in situ* refugium in the Italian Peninsula and Sicily, which remained suitable for butterflies during Pleistocene glacial cycles.

Two centres of endemism in the Alps-Apennines region

Our result challenges the common perception of the Alpine-Apennine area as a single unit, known as the “Italian refugium” (Hewitt, 1999; Petit et al., 2003; Taberlet et al., 1998). Indeed, we objectively identified two centres of endemism: Alps and Peninsula-Sicily. The two centres shared only 7 endemics, versus 35 endemics exclusive from the Alps and 27 from Peninsula-Sicily. This occurred despite the continuity of the mountain-hill chain, the presence of high-altitude areas in the Apennines (a maximum of 2912m in mainland and of 3324m in Sicily) and the inclusion in the Alps centre of many low-altitude cells. Moreover, endemics from the Alps and Peninsula-Sicily are characterised by different traits, phylogenetic representation and variance of genetic divergences. In the Peninsula-Sicily region, the phylogenetic and ecological spectra of endemics are variegated since they belong to five families and 22 genera comprising both strictly Mediterranean species (e.g. *Hipparchia leighebi*, *Hipparchia blachieri*, *Zerynthia cassandra*, *Melanargia arge, Pyronia cecilia*) and also typical mountain taxa (e.g. *Erebia pluto*, *Erebia montana*, *Melitaea varia*). Conversely, their genetic divergence is less variable: most taxa seem to have differentiated during the Pleistocene and are recognised as deeply diverging intraspecific lineages (ME). This pattern agrees with the current view of European Quaternary phylogeography, deeply impacted by long cold periods, when most central-northern Europe, Alps and Pyrenees were covered by ice sheets surrounded by permafrost and tundra belts (Ehlers, Ehlers, Gibbard, & Hughes, 2011). During cold pulses, many temperate species persisted in separated glacial refugia (notably the peninsulas of Iberia, Italy and Balkans, the Mediterranean islands, and the Maghreb); during interglacials, they dispersed towards higher latitudes and altitudes (Hewitt, 1999; Petit et al., 2003; Schmitt, 2007). Virtually all European taxa showed differentiation among these areas and signal of (repeated) post-glacial poleward expansion (forest plants, Petit et al., 2003; butterflies, Schmitt, 2007; Dapporto et al., 2019; mammals, Seddon, Santucci, Reeve, & Hewitt, 2001; springtails, Fiera, Habel, Kunz, & Ulrich, 2017). The high incidence of ME endemics dated to the onset of the Pleistocene and limited to Peninsula-Sicily fits with the definition of neoendemics, described as recently diverged species that failed to disperse out of their ancestral area (Flantua et al., 2020).

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

The Alpine centre is richer than Peninsula-Sicily in number of entities, included the endemics. Species richness in a given biome depends on the area the biome occupied along historical time (Jetz & Fine, 2012). Due to the large extension of tundra, steppic and subarctic biotas in Europe during most Pleistocene, the continent hosts many cold-adapted species that likely suffered range contractions during interglacials, including the current one. Among them, *Erebia* is the largest butterfly genus in Europe (58 species) and has a main centre of diversification on this continent (Peña, Witthauer, Klečková, Fric, & Wahlberg, 2015). It is likely that most *Erebia* had wider distributions during glacial periods and contracted their ranges to mountain and northern European areas during the last interglacial. This genus alone contributes 38% to Alpine endemics determining, together with other mountain specialist endemics, a significant phylogenetic clustering of Alpine endemics compared to the whole Alpine and European faunas. Phylogenetic clustering of high altitude communities is known in Lepidoptera (Brehm, Strutzenberger & Fiedler, 2013) and it aligns with the hypothesis of an Alpine endemic fauna formed after habitat tracking of a reduced set of genera specialised to the peculiar tundra-like environment, now limited to mountain areas in southern parts of Europe.

Generally, boreo-alpine species show low intraspecific differentiation between regions because of the recent geographic split which explains the low incidence of ME we found in the Alps (Mutanen et al., 2012). The high incidence of SE endemics showing high variance in divergence time in the Alps is also in line with the hypothesis of an ecological refugium amassing species with different evolutionary histories. The large number of SE and LE endemics in the Alps showing higher (SE) and lower (LE) divergence than ME endemics, contributed to the higher variance in genetic divergence of Alpine compared to Peninsula-Sicily endemics. The high incidence of LE and LME also fits with the mechanisms at the basis of EVOc formation since they generally represent diverged species that exchanged mitochondrial DNA by introgression following secondary sympatry. Their incidence in the Alps could have contributed to the absence of a significant higher genetic differentiation of Alpine compared to Peninsula-Sicily endemics. In general, the endemics of the Alps fit with the definition of palaeoendemics, described as relict species whose ranges became spatially restricted (Flantua et al., 2020).

The mechanism hypothesised here is likely responsible for the distribution of most alpha diversity in Europe since formerly glaciated areas in the Alps, Pyrenees and Balkans are currently the richest areas for butterflies in the continent (Hawkins, 2010). Without doubt, the Apennines also functioned as an *ex situ* refugium, as indicated by the presence of several cold-adapted species (12 *Erebia* spp., 2 *Parnassius* spp., many Lycaenidae spp.) and because mountain areas host the richest butterfly communities of the Peninsula-Sicily centre. However, this phenomenon has involved different entities from the Alps since only three high-altitude endemics are shared between the two centres (*Polyommatus damon*, *Melitaea varia*, *Erebia montana*). Other shared ME endemics belong to altitude generalist taxa (*Lycaena alciphron*, *Melitaea aurelia* and *Melitaea cinxia*) and to a Mediterranean taxon (*Lycaena thersamon*). The extinction of several high-altitude species in the Apennines during the last interglacial could also account for the high turnover with the Alps. Indeed, Apennine high-altitude refugia could be too small, warm and isolated to allow the persistence of cold-adapted species (Marta, Lacasella, Cesaroni, & Sbordoni, 2019). Currently, several cold-adapted taxa show small, isolated populations in the Apennines that are considered on the brink of extinction (*Erebia pandrose* and *Erebia montana* in Northern-Central Apennines*, Erebia gorge* in Southern Apennines), locally declining (*P. mnemosyne* and *P. apollo*) or have gone recently extinct (*Erebia aethiops* and probably *E. gorge* in Northern Apennines) (Balletto, Bonelli, & Cassulo, 2007; Cini et al., 2020; Piazzini & Favilli, 2020). This is also the reason why it has been suggested that distinctions should be applied in IUCN assessments for butterflies populations from Alps and Apennines (Bonelli et al., 2018).

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

Signal for dispersal along and across Alps and Apennines: implications for conservation

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

We can thus reject the hypothesis that the Alps and the Apennines represent a corridor for most butterflies (Dapporto, Fattorini, Vodă, Dincă, & Vila, 2014) and presumably for other insect species, with two main breaks located close to the geographical boundary between the Alps and the Apennines (west Liguria) and on the Strait of Messina. The identification of these breaks has important consequences for the conservation of the populations living in the study area. Indeed, the strong turnover existing between the Alps, Peninsula and Sicily demonstrates that each of these areas represents an independent management unit and needs specific protection. The endemics identified in this study occurred in two centres of similar size: about 200,000 km². Due to the large size of these centres, future studies should focus on identifying smaller portions representing key biodiversity areas for potential conservation actions (Brooks et al., 2015). An analysis of butterfly richness weighted by their risk of extinction (IUCN assessments) ranked the Alps, the southern tip of the Italian Peninsula and eastern Sicily as the most important areas for butterfly conservation in Italy (Girardello, Griggio, Whittingham, & Rushton, 2009). Several

National Parks protect important areas of the peninsula, while only four regional parks with more limited funding are located in Sicily. In particular, while the Calabrian side of the Strait of Messina is protected by the Aspromonte National Park, on the Sicilian side, the Peloritani mounts are completely unprotected.

The importance of integrating taxonomical and genetic approaches: methodological implications

The method we used to identify endemic taxa may have strong implications for future studies. The introduction of phylogenetic diversity and endemicity (Faith, 1992; Rosauer, Laffan, Crisp, Donnellan, & Cook, 2009) added the evolutionary dimension to the study of communities and to conservation biology (Laity et al., 2015). Such methods assume that ancient divergence or wider phylogenetic representation of communities have higher value in identifying areas of endemism and key areas for conservation (Laity et al., 2015). Alternatively, we used a qualitative approach generating a list of endemics for regionalisation and comparison of species traits. Our evaluation, based on an “and/or” approach, allowed to include several species that were lumped in a COI-based GMYC analysis, likely due to events of post-speciation mitochondrial introgression (Dincă et al., 2015). In particular, 18 endemic entities, widely recognised as good species by butterfly specialists also based on nuclear markers (Wiemers et al., 2018), were recovered as LE and LME. On the other hand, without the ME highlighted by GMYC, the Peninsula-Sicily region could not have been identified due to the low incidence of SE and LE endemics. Accordingly, a study at species level on the Italian hotspots for Lepidoptera, Carabidae, amphibians and reptiles identified most irreplaceable areas in the Alps, while none was recovered in Apennines and in Sicily (Balletto et al., 2010).

Recent reviews indicate key areas of endemism for conservation should be recognised also based on the intraspecific genetic divergence they encompass (Brooks et al., 2015). If we consider the GMYC entities as units of genetic divergence, most ME endemics of the study area only show two or three ESUs across the whole West Palaearctic; these fractions indicate that species defining the Alps and Peninsula-Sicily as centres of endemism encompass a considerable fraction (33-50%) of the whole genetic differentiation of the species they belong to (Figure S3).

Currently, only data based on single mitochondrial markers (COI) are available to investigate highly diversified taxa at the continental and sub-continental scale. However, with increasing sequencing capacity we expect that, in the near future, massive genomic comparative data will provide higher resolution to phylogenetic assessments of endemic taxa.

Conclusions

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

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

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

Acknowledgements

Support for this research was provided by 'La Caixa' Foundation (ID 100010434) to MM (grant LCF/BQ/DR20/11790020), the Ramón y Cajal program of the Spanish Ministry of Science and Innovation to GT (RYC2018-025335-I), the Academy of Finland to VD (Academy Research Fellow, decision no. 328895), by project PID2019-107078GB-I00 / AEI / 10.13039/501100011033 to RV and by the projects "Ricerca e conservazione sui lepidotteri diurni di sei Parchi Nazionali dell'Appennino centro-settentrionale", "Ricerca e conservazione sugli Impollinatori dell'Arcipelago Toscano e divulgazione sui Lepidotteri del parco" and "Servizio di attuazione delle azioni per la protezione degli impollinatori e diffusione dell'entomofauna del Parco Nazionale dell'Alta Murgia" to LD. We thank Francesca Barbero, Luca Pietro Casacci, Sylvain Cuvelier, Stefano Scalercio, Raluca Vodă, and other people providing help in data collection with a special reference to iNaturalist and Observado users.

REFERENCES

Anderson, S. (1994). Area and endemism. *The Quarterly Review of Biology*, *69*(4), 451–471.

Balletto, E, Bonelli, S., & Cassulo, L. (2007). insecta Lepidoptera Papilionoidea (Rhopalocera). *Checklist and Distribution of the Italian Fauna. Ministero Dell'ambiente e Della Tutela Del Territorio, Direzione per La Protezione Della Natura. Cd-RoM*.

Balletto, Emilio, Bonelli, S., Borghesio, L., Casale, A., Brandmayr, P., & Vigna Taglianti, A. (2010). Hotspots of biodiversity and conservation priorities: A methodological approach. *Italian Journal of Zoology*, *77*(1), 2–13.

Beck, H. E., Zimmermann, N. E., McVicar, T. R., Vergopolan, N., Berg, A., & Wood, E. F. (2018). Present and future Köppen-Geiger climate classification maps at 1-km resolution. *Scientific Data*, *5*, 180214.

Bonelli, S., Casacci, L. P., Barbero, F., Cerrato, C., Dapporto, L., Sbordoni, V., … Balletto, E. (2018). The first red list of Italian butterflies. *Insect Conservation and Diversity*, *11*(5). doi: 10.1111/icad.12293

Brehm, G., Strutzenberger, P., & Fiedler, K. (2013). Phylogenetic diversity of geometrid moths decreases with elevation in the tropical Andes. *Ecography*, 36(11), 1247-1253.

Brooks, T. M., Cuttelod, A., Faith, D. P., Garcia-Moreno, J., Langhammer, P., & Pérez-Espona, S. (2015). Why and how might genetic and phylogenetic diversity be reflected in the identification of key biodiversity areas? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *370*(1662), 20140019.

Brower, A. V. (1994). Rapid morphological radiation and convergence among races of the butterfly Heliconius erato inferred from patterns of mitochondrial DNA evolution. *Proceedings of the National Academy of Sciences*, *91*(14), 6491–6495.

Cini, A., Barbero, F., Bonelli, S., Bruschini, C., Casacci, L. Pietro, Piazzini, S., … Dapporto, L. (2020). The decline of the charismatic *Parnassius mnemosyne* (L.)(Lepidoptera: Papilionidae) in a Central Italy national park: a call for urgent actions. *Journal of Insect Biodiversity*, *16*(2), 47–54.

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

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

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

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

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

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

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

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

Faith, D. P. (1992). Conservation evaluation and phylogenetic diversity. *Biological Conservation*, *61*(1), 1–10.

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

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

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

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

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

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

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

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

Hülber, K., Kuttner, M., Moser, D., Rabitsch, W., Schindler, S., Wessely, J., … Dullinger, S. (2020). Habitat availability disproportionally amplifies climate change risks for lowland compared to alpine species. *Global Ecology and Conservation*, e01113.

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

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

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

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

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

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

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

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

Middleton-Welling, J., Dapporto, L., García-Barros, E., Wiemers, M., Nowicki, P., Plazio, E., … Shreeve, T. (2020). A new comprehensive trait database of European and Maghreb butterflies, Papilionoidea. *Scientific Data*, *7*(1), 1–10.

Moritz, C., Hoskin, C. J., MacKenzie, J. B., Phillips, B. L., Tonione, M., Silva, N., … Graham, C. H. (2009). Identification and dynamics of a cryptic suture zone in tropical rainforest. *Proceedings of the Royal Society B: Biological Sciences*, *276*(1660), 1235–1244.

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

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

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

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

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

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

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

Quek, S., Davies, S. J., Itino, T., & Pierce, N. E. (2004). Codiversification in an ant-plant mutualism: stem texture and the evolution of host use in *Crematogaster* (Formicidae: Myrmicinae) inhabitants of Macaranga (Euphorbiaceae). *Evolution*, *58*(3), 554–570.

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

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

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

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

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

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

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

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

Wiemers, M., Chazot, N., Wheat, C. W., Schweiger, O., & Wahlberg, N. (2020). A complete time-calibrated multi-gene phylogeny of the European butterflies. *ZooKeys*, *938*, 97.

Zuloaga, J., Currie, D. J., & Kerr, J. T. (2019). The origins and maintenance of global species endemism. *Global Ecology and Biogeography*, *28*(2), 170–183.

Data Accessibility Statement:

R scripts and data to replicate the analysis are available in Dryad DOI:

https://doi.org/10.5061/dryad.tb2rbnzzf. Previously published and newly generated COI data are also available in Dryad and in the DS-ALPAPENN BOLD project at https://www.boldsystems.org/ where the GenBank accession codes are also reported.

Author Contributions:

LD, MM, GT, AC, VS and RVi designed the research. LD, MM, AC, LP, VS validated records on iNaturalist. EB and SB collected literature records, all the authors collected field data on butterflies. VD, LD, MM, RVi and LP curated COI sequencing. MM, LD and AC collected trait data. LD and GT carried out the analyses. All the authors contributed to the writing of the paper.

Table 1. Ecological traits describing alpha and beta niches.

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., 2020) 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 km2 occupied in Europe (range size)
		The maximum altitude reported
		The minimum altitude reported
		Altitudinal range

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

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	**0.009**
	Host plants	11.692	11.692	15.542	**<0.001**	0.046	**0.010**
	Phenology PC1	0.085	0.085	34.730	**<0.001**	**0.001**	**<0.001**
	Distribution PC1	0.130	0.130	53.014	**<0.001**	**0.001**	0.076
	Distribution PC2	0.205	0.205	96.017	**<0.001**	**0.001**	0.402
	Closest relative	233.000	232.957	5.254	**0.023**	-	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	**0.012**	-	**0.001**
E	Wingspan	0.002	0.002	1.777	0.188	0.354	**0.020**

nd	Alps	Host plants	2.377	2.377	4.837	0.032	0.137	0.301
		Phenology PC1	0.035	0.035	34.407	**<0.001**	**<0.001**	**0.001**
		Distribution PC1	0.044	0.044	25.668	**<0.001**	**<0.001**	0.406
		Distribution PC2	0.137	0.137	64.671	**<0.001**	**<0.001**	0.847
		Closest relative	6.800	6.802	1.209	0.276		**0.006**

Table 3. Comparison of medium phylogenetic distances (MPD) and mean nearest taxon distance (MNTD) between 1-2) the tree of European butterflies and the subsets represented by species occurring in the Alps and Peninsula-Sicily main centres (Alps-Eur and PS-Eur); 3-4) the tree of Alps and Peninsula butterflies and the subsets represented by endemics from Alps and Peninsula-Sicily main centres (Alps End-Alps and PS End-PS); 5-6) the tree of European butterflies and the subsets represented by endemics from Alps and Peninsula-Sicily (Alps End-Eur tree and PS End-Eur tree). Taxa, number of entities; MPD Obs and MNTD Obs, mean observed phylogenetic distance and mean nearest taxon distance; MPD Rand and MNTD Rand, mean phylogenetic distance and mean nearest taxon distance obtained in 10000 null models; MPD Z and MNTD Z, Z-values; MPD P and MNTD P, p-values (significant results are in bold).

Comparison	Taxa	MPD Obs	MPD Rand	MPD Z	MPD P	MNTD Obs	MNTD Rand	MNTD Z	MNTD P
1 Alps-Eur	319	172.942	169.718	1.993	0.983	11.737	11.969	-0.296	0.384
2 PS-Eur	242	174.995	169.705	2.550	0.997	13.621	14.153	-0.496	0.313
3 Alps End-Alps	42	140.203	172.972	-5.913	**<0.001**	17.865	38.092	-3.741	**<0.001**
4 PS End-PS	34	169.639	175.008	-0.893	0.181	38.049	45.746	-1.143	0.130
5 Alps End-Eur	42	140.203	169.706	-4.606	**<0.001**	17.865	38.548	-3.918	**<0.001**
6 PS End-Eur	34	169.639	169.710	-0.010	0.466	38.049	43.208	-0.790	0.215

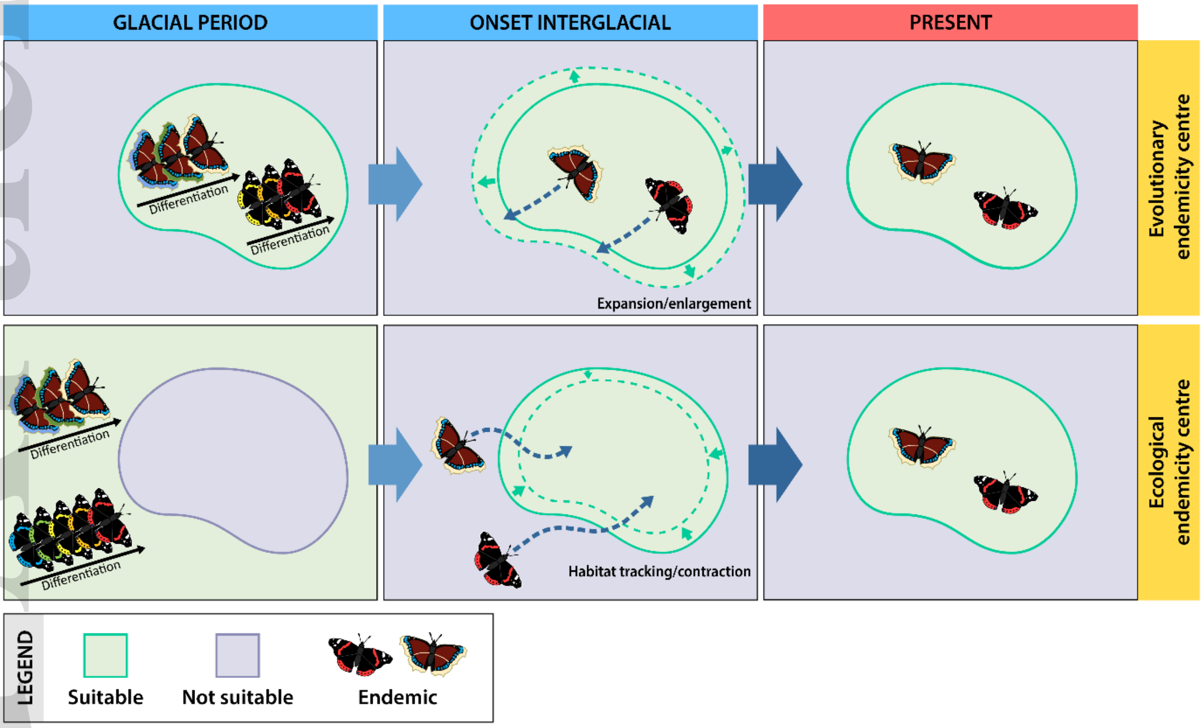


Figure 1. A model representing the mechanisms generating evolutionary and ecological endemicity centres for two endemic taxa. In evolutionary endemicity centres, species differentiate during similar time lags in suitable areas (typically glacial refugia), which may expand (dotted line) following climatic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

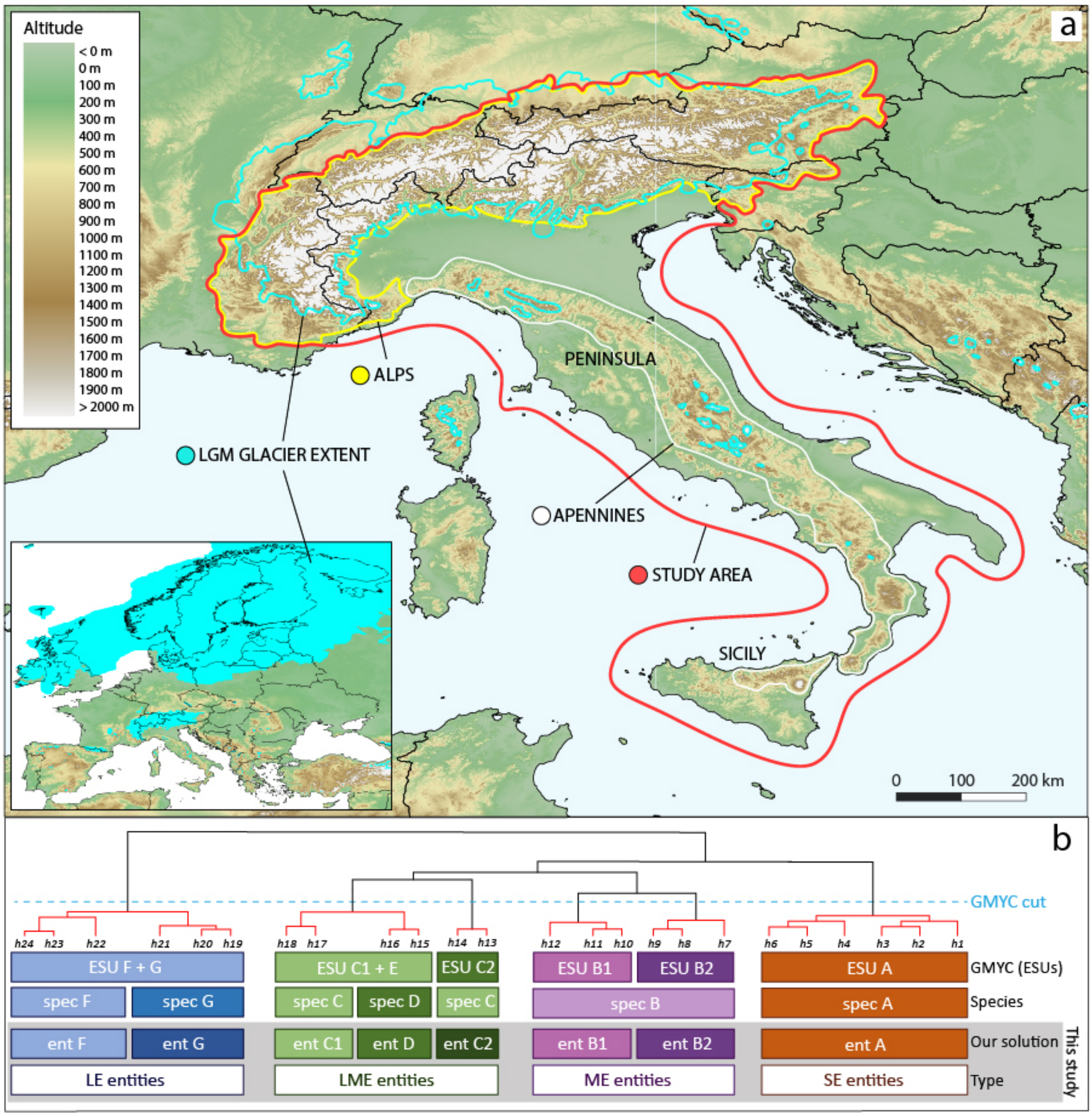


Figure 2. (a) The study area (red polygon), including the boundaries of the Alps (yellow polygon), the Apennines (white polygon), and those areas covered by the ice cap during the last glacial maximum (LGM) (cyan polygons). The inset shows the distribution of ice sheets during the LGM in Europe. (b) A scheme summarising how entities are identified as an "and/or" combination of taxonomy (species) and GMYC delimitations (ESUs). Abbreviations of the entity types: SE, single entity species; ME, multiple entity species; LME, lumped + multiple entities; LE, lumped entities.

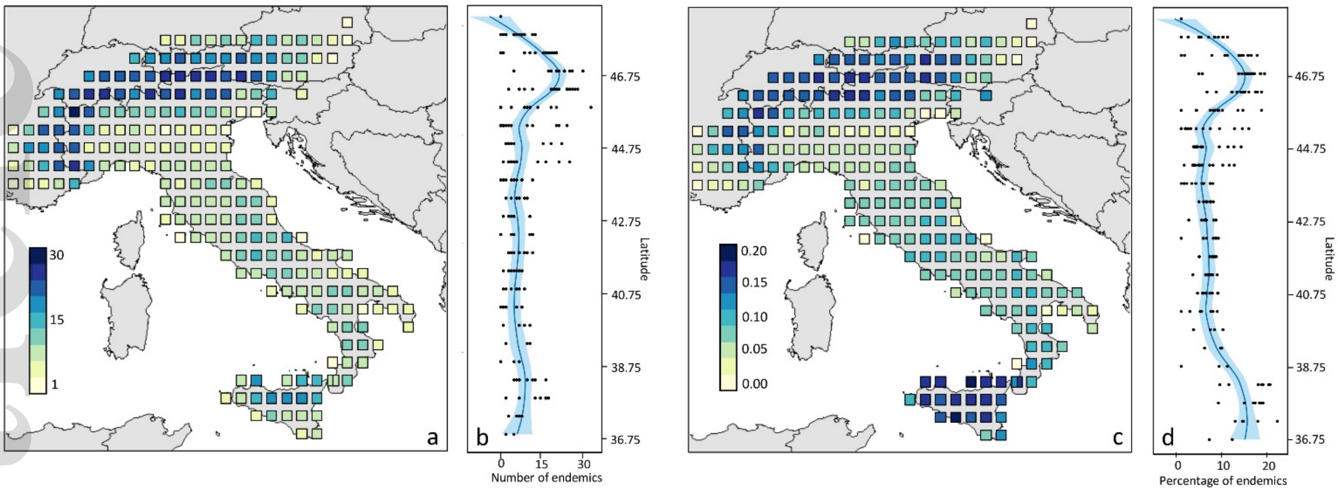


Figure 3. (a) Number of endemic entities recorded in each 0.5x0.5 cell and (b) latitudinal trend obtained by loess regression. (c) Percentage of endemic over observed richness (d) showing two peaks in Sicily and Alps.

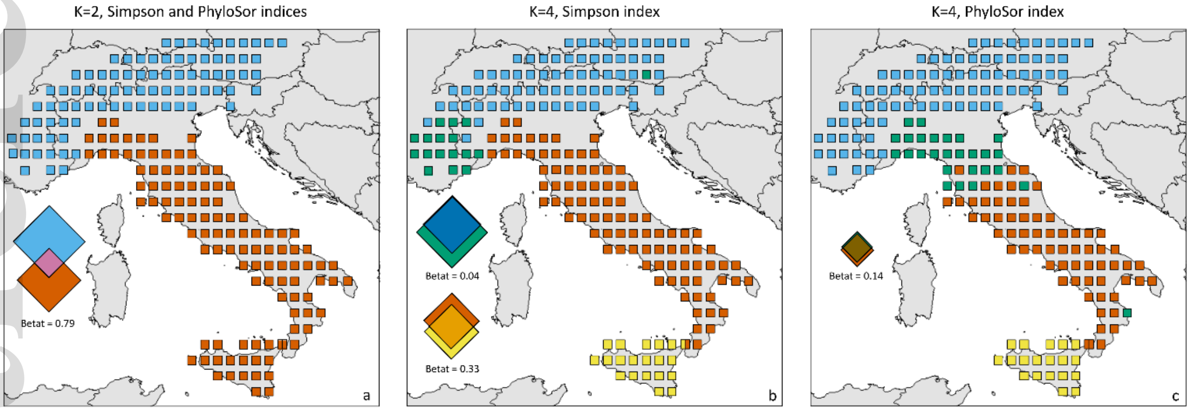


Figure 4. The solutions for k=2 (a) and k=4 (b, c)obtained by applying the recluster.region algorithm to the occurrence of endemic taxa by using the Simpson and the PhyloSordiversity indices. The size of diamonds in figures (a) and (b) is proportional to the number of endemics occurring in each centre,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

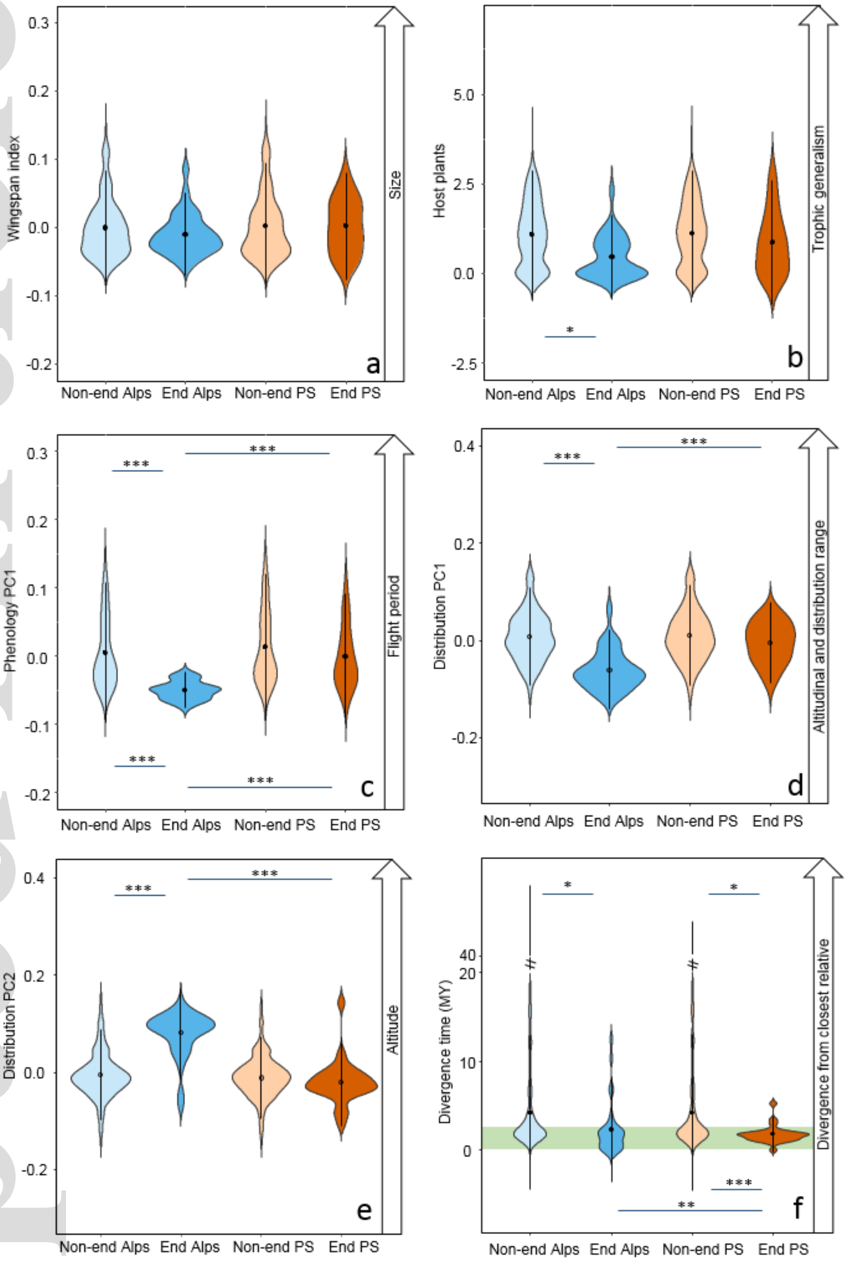
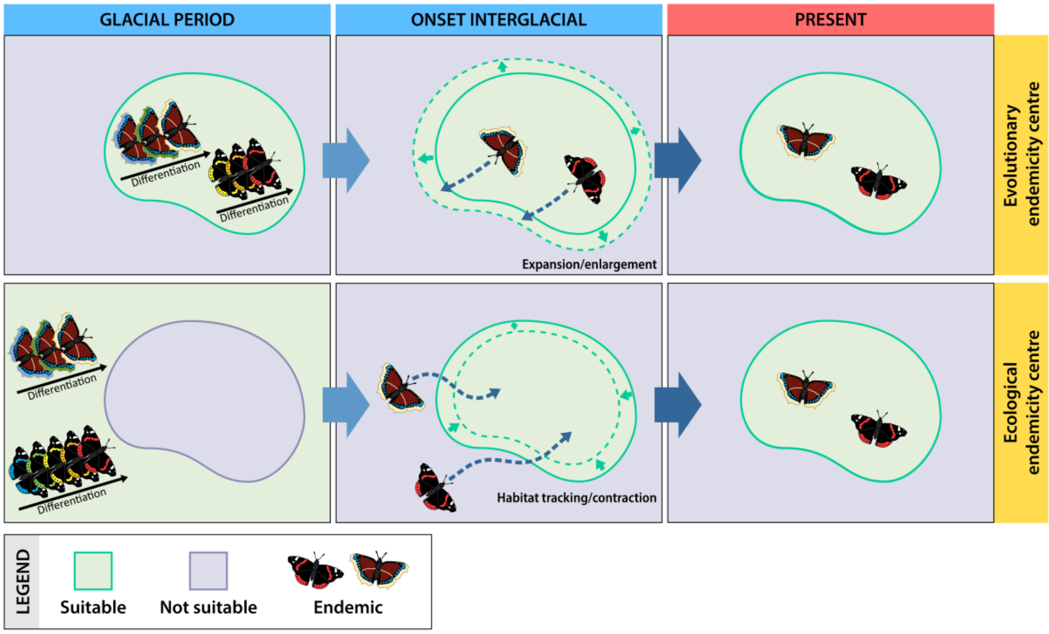
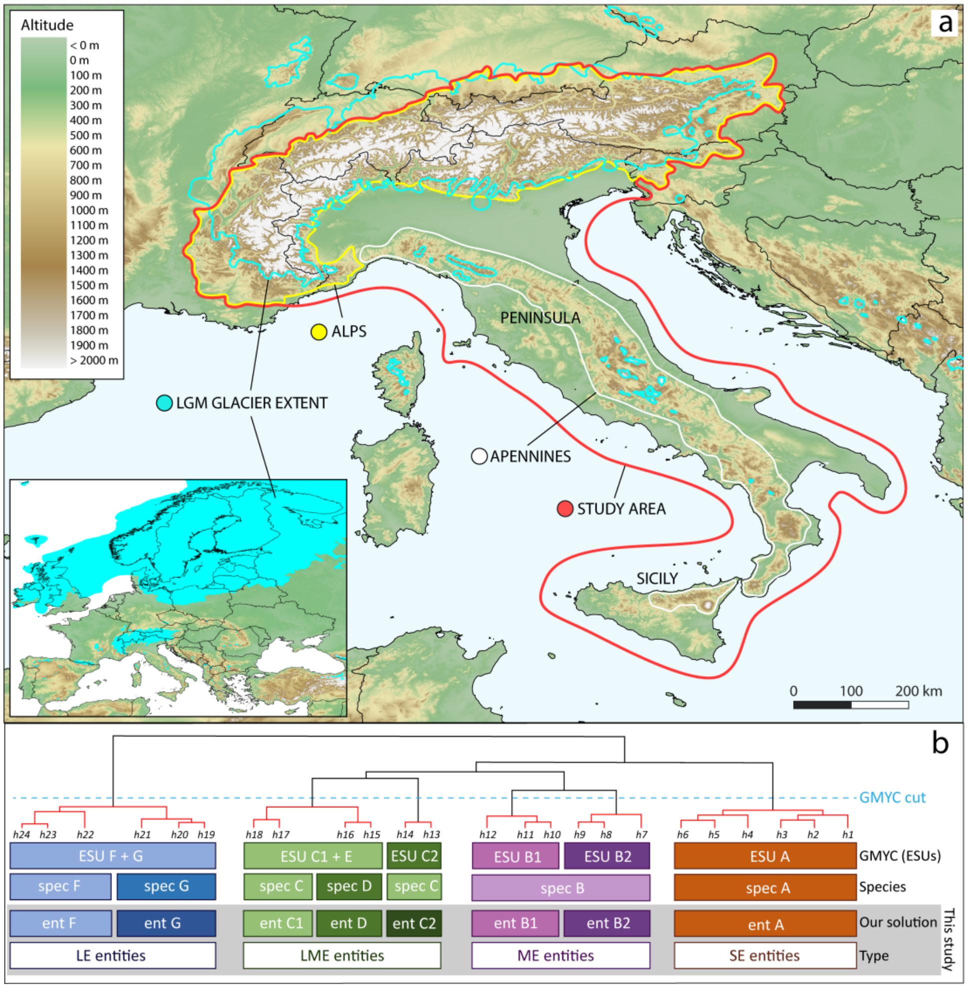


Figure 5. Violin plots comparing the distribution of the features of taxonomic species involved in non-endemic taxa occurring on Alps (Non-end Alps) and on Peninsula-Sicily (Non-end PS), with endemics from the Alpine (End Alps) and Peninsular-Sicily (Non-end PS) centres. Above the plots we reported significant pairwise comparison of means as obtained by phylogenetic ANOVAs; below the plots we reported significant pairwise comparisons between variances (Fligner-Killeen

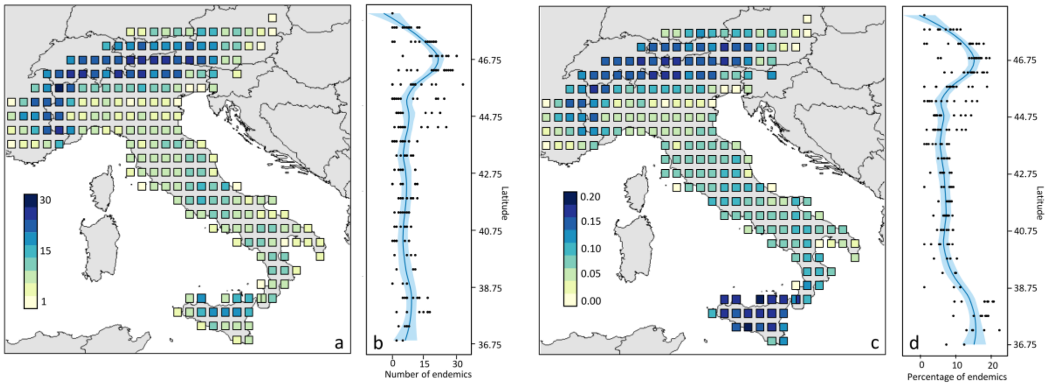
test). All significance assessed after sequential Bonferroni correction. The green area in (f) represents the Pleistocene and Holocene (2.6MY-present).



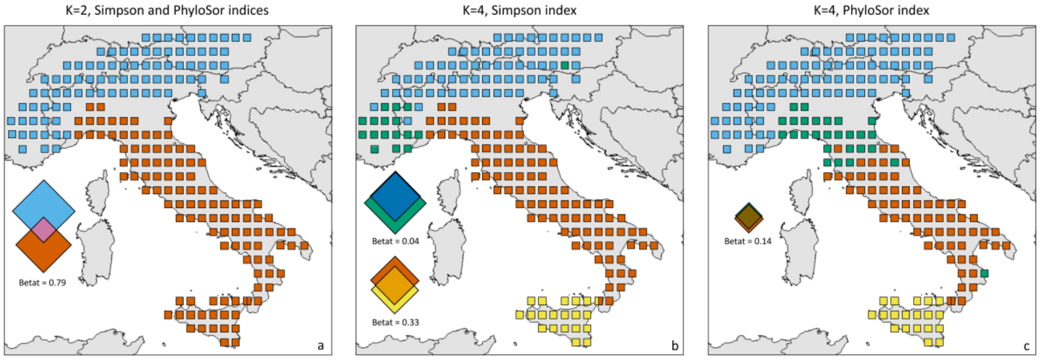
mec_15795_f1.tif



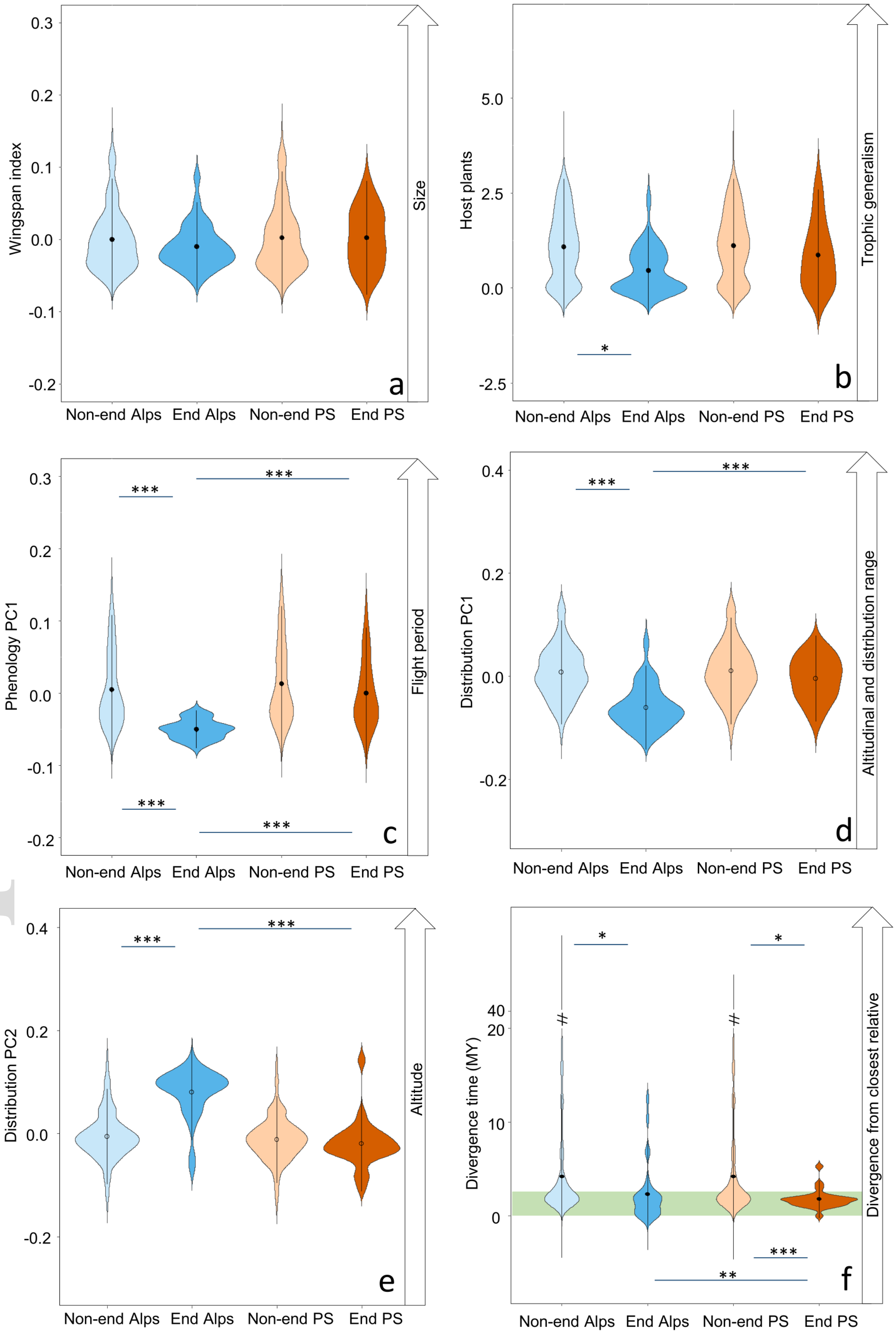
mec_15795_f2.tif



mec_15795_f3.tif



mec_15795_f4.tif



mec_15795_f5.tif