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 factors.
ecological processes (Sandel et al., 2020; Zuloaga et al., 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 et al., 2013). A paradigm for EVOcs is the diversification occurred in warm refugia during Pleistocene cold periods (Brooks et al., 2015; Hewitt, 1999; Taberlet et al., 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 (Mutansen 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 paleoecological 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 probably evolved under similar processes, they should have a similar degree of genetic divergence from closest relatives and should spread all along the existing

**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) [Colour figure can be viewed at wileyonlinelibrary.com]
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, macro-ecological 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 Lists; 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; Vodá et al., 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 et al., 2015) and sometimes it proved to be more effective than taxonomic assessments in documenting coevolutionary processes (Liu et al., 2018). Using this data set, 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 (37° to 48° 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 EVOc and ECOc occur in this region. To test this, we (a) combined a massive data set of cytochrome c oxidase subunit 1 (COI) sequences and occurrence data to evaluate if different centres were recognisable by regionalisation analysis; (b) verified whether the endemics from the potential ECOc have stricter requirements in key ecological traits and encompass a lower phylogenetic diversity; and (c) 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 endemism in biodiversity hotspots and informs towards more effective conservation strategies.

2 | MATERIALS AND METHODS

2.1 | Sampling and data sets

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 0.5 × 0.5 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.tb2brnzzf). 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 data set, dx.doi.org/10.5883/DS-ALPAPENN).

2.2 | Phylogeny and GMYC

We collapsed the COI data set 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 (Suchard et al., 2018) 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 data set included one outgroup for each of the other families. Two independent chains of 100 million generations were run in BEAST for each data set. 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 et al., 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 (a) “single entity (SE)”: all haplotypes of a species belong to a single GMYC ESU; (b) “multiple entity (ME)”: haplotypes belong to two or more ESUs; (c) “lumped entities (LE)”: two or more species are recovered as a single ESU; and (d) “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 entity and uses the relative hull geometries to attribute unknown occurrence data to a given entity (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.

### 2.2.1 | Which are the centres of endemism

To locate the centres of endemism we ran regionalisation analyses for the occurrence data of endemics in 0.5 × 0.5 cells. We used the "recluster.region" function in the R package "recluster" ([https://cran.r-project.org/web/packages/recluster/index.html](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: (a) the Simpson turnover index, accounting for species replacement in terms of faunistic elements; and (b) the species replacement component of the phylogenetic beta diversity index PhyloSor ([Leprieur et al., 2012](https://doi.org/10.1186/1471-2148-12-420)), 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 et al., 2020](https://doi.org/10.1007/s10293-019-00825-5)). The PhyloSor index has been calculated using the "betapart" R package ([https://cran.r-project.org/web/packages/betapart/index.html](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](https://cran.r-project.org/web/packages/labdsv/labdsv.pdf)).

### 2.2.2 | 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 nonendemics from the same centre. We used a series of 10 ecological traits for European butterflies ([Middleton-Welling et al., 2020](https://doi.org/10.1371/journal.pbio.2003203); [Platania et al., 2020](https://doi.org/10.3390/genes11030305)). 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](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](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](https://doi.org/10.1007/s10293-019-00825-5)). 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 nonparametric Fligner–Killeen test, using the "check_homogeneity" function of the R package "performance" ([https://cran.r-project.org/web/packages/performance/index.html](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 et al., 2002](https://doi.org/10.1093/molbev/msg030)). 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 entities of each centre. Finally, we tested if the phylogenetic distances among endemics of each centre differ

<table>
<thead>
<tr>
<th>Type</th>
<th>Trait</th>
<th>Measure description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alpha traits</td>
<td>Trophic generalism (feeding trait)</td>
<td>The number of host plant genera</td>
</tr>
<tr>
<td></td>
<td>Mobility (morphology trait)</td>
<td>The wingspan index (<a href="https://doi.org/10.1371/journal.pbio.2003203">Middleton-Welling et al., 2020</a>) based on multiple bibliographic measurements of wingspan</td>
</tr>
<tr>
<td></td>
<td>Phenology (life history trait)</td>
<td>The number of months during which adults fly in Europe</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Beta traits</th>
<th>Distribution and environmental preferences</th>
<th>The number of 30 × 30 km² cells occupied in Europe (range size)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>The maximum altitude reported</td>
</tr>
<tr>
<td></td>
<td></td>
<td>The minimum altitude reported</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Altitudinal range</td>
</tr>
</tbody>
</table>
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).

2.2.3 Do the endemics from EVOc 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 nonendemics 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 by a chi-squared test.

3 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 Palearctic 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 ± 9.98 SD% geographic overlap (Figure S3b).

3.1 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, Appendix S1: Figures S4–S72 for individual distributions). When the effect of local richness was removed by calculating percentages of endemics, Sicily emerged as a main endemcity 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 $k = 1$ 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 index and the turnover component of the PhyloSor index, 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 (four exclusive vs six 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 km$^2$ while the Peninsula-Sicily centre is about 270,000 km$^2$.

3.2 | Are endemics characterised by different traits in different centres?

The PCA identified one function from the phenological traits showing eigenvalues higher than one (Appendix S1: Figures S73 and S74), 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 endemics occurring on Alps represented only four families and 12 genera. No endemic species of Pieridae occurred and a single genus was represented among endemic Hesperidae 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, Erebia stria, Erebia 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 six entities (about 17% of peninsular endemics; Figures S4–S72).

3.3 | Do the endemics from EVOc show lower variance in genetic divergence?

Phylogenetic ANOVA showed that different types of endemics showed different divergence (df = 3, Sum. Sq. = 151.74, $F = 15.406$, $p < .001$) 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 nonendemics (Table 2, Figure 5f). Moreover, Peninsula-Sicily endemics showed lower variance in divergence, compared to both Alpine endemics and nonendemics from their centre with a particularly high frequency around 2 million years ago (Ma) (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-squared test of independence of the frequencies showed a highly significant effect ($\chi^2 = 12.118$, $df = 3$, $p = .007$).

### 4 | DISCUSSION

Understanding the mechanisms generating centres of endemity 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 endemity 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 phylogeographic distances and in the variance of genetic divergences that align with predictions for the existence of two different endemity 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.

#### 4.1 | 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 seven 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 2912 m in mainland and of 3324 m 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 liegehi, Hipparchia blachieri, Zerynthia cassandra, **TABLE 2** Phylogenetic ANOVA and homogeneity of variance tests comparing species traits and divergence from the closest relative among endemics (End) and nonendemics (Non-end) and between entities from the two centres

<table>
<thead>
<tr>
<th>Groups</th>
<th>Feature</th>
<th>Sum-Sq</th>
<th>Mean-Sq</th>
<th>$F$</th>
<th>$p$ (typ)</th>
<th>$p$ (phy)</th>
<th>$p$ (var)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-end Alps versus End Alps</td>
<td>Wingspan</td>
<td>0.003</td>
<td>0.003</td>
<td>1.955</td>
<td>.163</td>
<td>.541</td>
<td>.009</td>
</tr>
<tr>
<td></td>
<td>Host plants</td>
<td>11.692</td>
<td>11.692</td>
<td>15.542</td>
<td>&lt;.001</td>
<td>.046</td>
<td>.010</td>
</tr>
<tr>
<td></td>
<td>Phenology PC1</td>
<td>0.085</td>
<td>0.085</td>
<td>34.730</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>Distribution PC1</td>
<td>0.130</td>
<td>0.130</td>
<td>53.014</td>
<td>&lt;.001</td>
<td>.001</td>
<td>.076</td>
</tr>
<tr>
<td></td>
<td>Distribution PC2</td>
<td>0.205</td>
<td>0.205</td>
<td>96.017</td>
<td>&lt;.001</td>
<td>.001</td>
<td>.402</td>
</tr>
<tr>
<td></td>
<td>Closest relative</td>
<td>233.000</td>
<td>232.957</td>
<td>525.4</td>
<td>.023</td>
<td>—</td>
<td>.373</td>
</tr>
<tr>
<td>Non-end PS versus End PS</td>
<td>Wingspan</td>
<td>0.000</td>
<td>0.000</td>
<td>0.009</td>
<td>.925</td>
<td>.928</td>
<td>.931</td>
</tr>
<tr>
<td></td>
<td>Host plants</td>
<td>1.252</td>
<td>1.252</td>
<td>1.606</td>
<td>.206</td>
<td>.181</td>
<td>.758</td>
</tr>
<tr>
<td></td>
<td>Phenology PC1</td>
<td>0.004</td>
<td>0.004</td>
<td>1.446</td>
<td>.230</td>
<td>.205</td>
<td>.376</td>
</tr>
<tr>
<td></td>
<td>Distribution PC1</td>
<td>0.005</td>
<td>0.005</td>
<td>1.810</td>
<td>.180</td>
<td>.144</td>
<td>.375</td>
</tr>
<tr>
<td></td>
<td>Distribution PC2</td>
<td>0.002</td>
<td>0.002</td>
<td>0.899</td>
<td>.344</td>
<td>.308</td>
<td>.555</td>
</tr>
<tr>
<td></td>
<td>Closest relative</td>
<td>274.200</td>
<td>274.215</td>
<td>634.4</td>
<td>.012</td>
<td>—</td>
<td>.001</td>
</tr>
<tr>
<td>End Alps versus End PS</td>
<td>Wingspan</td>
<td>0.002</td>
<td>0.002</td>
<td>1.777</td>
<td>.188</td>
<td>.354</td>
<td>.020</td>
</tr>
<tr>
<td></td>
<td>Host plants</td>
<td>2.377</td>
<td>2.377</td>
<td>4.837</td>
<td>.032</td>
<td>.137</td>
<td>.301</td>
</tr>
<tr>
<td></td>
<td>Phenology PC1</td>
<td>0.035</td>
<td>0.035</td>
<td>34.407</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>Distribution PC1</td>
<td>0.044</td>
<td>0.044</td>
<td>25.668</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>.406</td>
</tr>
<tr>
<td></td>
<td>Distribution PC2</td>
<td>0.137</td>
<td>0.137</td>
<td>64.671</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>.847</td>
</tr>
<tr>
<td></td>
<td>Closest relative</td>
<td>6.800</td>
<td>6.802</td>
<td>1.209</td>
<td>.276</td>
<td>—</td>
<td>.006</td>
</tr>
</tbody>
</table>

$p$-values in bold indicate significant results. Sample size is 319 for Non-end Alps versus End Alps; 242 for Non-end Peninsula-Sicily (PS) versus End PS; 62 for End Alps versus End PS.

Abbreviations: Sum-Sq, sum of squares; Mean-Sq, mean squares; $F$, $F$-value; $p$ (typ), $p$-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.
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 et al., 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, Dapporto et al., 2019; Schmitt, 2007; mammals, Seddon et al., 2001; springtails, Fiera et al., 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 endemity 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 (a) species showing genetic differentiation have lower dispersal capability and stronger ecological impediments to dispersal; and that (b) 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, steppe and subarctic biotas in Europe during most Pleistocene, the continent hosts many cold-adapted species that probably 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 et al., 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 et al., 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 paleoendemics, described as relict species whose ranges became spatially restricted (Flantua et al., 2020).

The mechanism hypothesised here is probably 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 altitudinal generalist taxa (*Lycaena alcyphon*, *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 et al., 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.memosyne* and *P.apollo*) or have gone recently extinct (*Erebia aethiops* and probably *E.gorge* in Northern Apennines; Balletto et al., 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 butterfly populations from Alps and Apennines (Bonelli et al., 2018).

**TABLE 3 Comparison of medium phylogenetic distances (MPD) and mean nearest taxon distances (MNTD) between 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); 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); 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)**

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Taxa</th>
<th>MPD Obs</th>
<th>MPD Rand</th>
<th>MPD Z</th>
<th>MPD p</th>
<th>MNTD Obs</th>
<th>MNTD Rand</th>
<th>MNTD Z</th>
<th>MNTD p</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Alps-Eur</td>
<td>319</td>
<td>172.942</td>
<td>169.718</td>
<td>1.993</td>
<td></td>
<td>11.737</td>
<td>11.969</td>
<td></td>
<td>0.296</td>
</tr>
<tr>
<td>2. PS-Eur</td>
<td>242</td>
<td>174.995</td>
<td>169.705</td>
<td>2.550</td>
<td></td>
<td>9.97</td>
<td>13.621</td>
<td>14.153</td>
<td>0.496</td>
</tr>
<tr>
<td>3. Alps End-Alps</td>
<td>42</td>
<td>140.203</td>
<td>172.972</td>
<td>−5.913</td>
<td>&lt;.001</td>
<td>17.865</td>
<td>38.092</td>
<td>−3.741</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>4. PS End-PS</td>
<td>34</td>
<td>169.639</td>
<td>175.008</td>
<td>−0.893</td>
<td></td>
<td>1.81</td>
<td>38.049</td>
<td>45.746</td>
<td>−1.143</td>
</tr>
<tr>
<td>5. Alps End-Eur</td>
<td>42</td>
<td>140.203</td>
<td>169.706</td>
<td>−4.606</td>
<td>&lt;.001</td>
<td>17.865</td>
<td>38.548</td>
<td>−3.918</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>6. PS End-Eur</td>
<td>34</td>
<td>169.639</td>
<td>169.710</td>
<td>−0.010</td>
<td></td>
<td>0.466</td>
<td>38.049</td>
<td>43.208</td>
<td>−0.790</td>
</tr>
</tbody>
</table>

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 10,000 null models; MPD Z and MNTD Z, Z-values; MPD p and MNTD p, p-values (significant results are in bold).
Butterfly species with endemic genetic lineages and high intra-specific 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 stronger phonology of Alpine species is probably 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 probably too high and cancelled any possibility for dispersal.

4.2 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 et al., 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 et al., 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.

4.3 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 endemcity (Faith, 1992; Rosauer et al., 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, probably 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 subcontinental 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.
5 | 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 endemity centre in the Alps, and an evolutionary endemity 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 subcentres. This result challenges the demics and neoendemics, respectively. Peninsula and Sicily mostly determined by the occurrence of paleoenvironments, 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 1000010434) 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 Scalerio, Raluca Vodă, and other people providing help in data collection with a special reference to iNaturalist and Observado users.

AUTHOR CONTRIBUTIONS


DATA AVAILABILITY STATEMENT

R scripts and data to replicate the analysis are available in Dryad https://doi.org/10.5061/dryad.tb2rnnzzf. 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.

ORCID

Mattia Menchetti https://orcid.org/0000-0002-0707-7495
Gerard Talavera https://orcid.org/0000-0003-1112-1345
Alessandro Cini https://orcid.org/0000-0003-0355-2188
Vlad Dincă https://orcid.org/0000-0003-1791-2148
Leonardo Platania https://orcid.org/0000-0002-1289-5564
Simona Bonelli https://orcid.org/0000-0001-5185-8136
Roger Vila https://orcid.org/0000-0002-2447-4388
Leonardo Dapporto https://orcid.org/0000-0001-7129-4526

REFERENCES


SUPPORTING INFORMATION
Additional supporting information may be found online in the Supporting Information section.