Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Dissecting an endemism hotspot: Spatial phylogenetics of the Iberian angiosperm flora

Alberto J. Coello
albjcoello@gmail.com

National Museum of Natural History
Ignacio Ramos-Gutiérrez
Universidad Autónoma de Madrid
Rafael Molina-Venegas
Estación Biológica de Doñana
Antoni Buira
Real Jardín Botánico
Mario Mairal
Universidad Complutense de Madrid
Pablo Vargas
Real Jardín Botánico
Rosa A. Scherson
University of Chile
Brent D. Mishler
University of California, Berkeley
Juan Carlos Moreno-Saiz
Universidad Autónoma de Madrid
Mario Fernández-Mazuecos
Real Jardín Botánico

Research Article

Keywords: angiosperms, Balearic Islands, biodiversity, CANAPE, evolution, Iberian Peninsula, Mediterranean biodiversity hotspot, phylogenetic diversity, phylogenetic endemism, spatial phylogenetics

Posted Date: November 3rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11235030/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

Spatial phylogenetics has emerged as a powerful tool to identify evolutionarily important areas for biodiversity conservation through mapping of phylogenetic diversity and endemism, including centers of neoendemism and paleoendemism. Although Europe is not generally regarded as a major center of plant diversity, recent global spatial phylogenetic studies have identified the Iberian Peninsula, in southwestern Europe, as a center of plant phylogenetic endemism at the global scale. Here we conducted a fine-scale spatial phylogenetic analysis of the angiosperm flora of the Iberian Peninsula and the Balearic Islands, which comprises c. 5,500 native species, more than 1,200 of which are strictly endemic. To this end, we combined a comprehensive distribution database with the most exhaustive megaphylogeny published to date, comprising 53% of the species. Phylogenetic diversity and endemism metrics were calculated, and Categorical Analysis of Neo- And Paleo-Endemism (CANAPE) was conducted, followed by an evaluation of the use of phylogenetic imputation to incorporate unplaced taxa. Abiotic factors potentially driving spatial phylogenetic patterns were tested.

Results

The distribution of phylogenetic diversity in Iberian angiosperms shows higher values in the Eurosiberian than in the Mediterranean floristic region. CANAPE revealed centers of endemism in mountain ranges (particularly the Baetic System, the Pyrenees, and the Cantabrian Mountains), the Balearic Islands, the semi-arid southeast of the Iberian Peninsula, and some coastal areas. Centers dominated by neoendemism are concentrated in areas combining recent geographic isolation, mediterranean-type climate, and diverse edaphically stressful substrates (calcareous, dolomitic, ultramafic, gypsic), whereas centers dominated by paleoendemism are more scattered and occur under a wider range of abiotic conditions. Superendemism centers occur in the highest mountain ranges (Sierra Nevada and the Pyrenees), the Balearic Islands, and the semi-arid southeast, functioning both as centers of recent speciation and as refugia for phylogenetically isolated lineages. Phylogenetic imputation proved useful in uncovering biologically meaningful spatial phylogenetic patterns that would otherwise remain undetected.

Conclusions

The Iberian Peninsula is a globally relevant center of endemism. Geographic patterns of phylogenetic diversity and endemism reveal evolutionary and biogeographic processes shaping plant assemblages in the western Mediterranean biodiversity hotspot, while providing a robust framework for informing future conservation strategies.

1. Introduction

The increase in biodiversity loss in recent decades has been well documented in the literature and demands prompt action (e.g., [1, 2]), but the limited resources available for conservation require prioritization [3]. Strategies for selection of protected areas have been traditionally focused on measurements of species richness [4], which do not account for evolutionary processes. Species equally contribute to them regardless of their evolutionary history, even though not all species have the same conservation value from an evolutionary perspective [5, 6]. The field of spatial phylogenetics addresses this limitation by quantifying and mapping biological diversity in an evolutionary framework. The first metric to incorporate evolutionary information from phylogenetic trees was *phylogenetic diversity* (PD) [7], which quantifies the diversity represented by a set of taxa in a location by adding up the phylogenetic branch lengths that connect them. This was followed by the development of *phylogenetic endemism* (PE) [8], a modification of PD to which range-restricted branches contribute more than widespread ones. Note that there are two distinct meanings of "endemism" – PE is concerned with *relative endemism*, a continuous value inversely related to the range size of a species or lineage, as compared to the more traditional *absolute endemism*, a binary value indicating whether or not a species or lineage is completely restricted to some defined geographic region [9]. We mostly use the former meaning in this paper, and specify absolute (or strict) endemism when we use that meaning.

The early stages of spatial phylogenetics ran into technical drawbacks that limited its progress [10]. A breakthrough in the field was the addition of a statistical framework and the development of Categorical Analysis of Neo- And Paleo-Endemism (CANAPE) by Mishler et al. [11], which enables the identification of key areas for biodiversity conservation based on their distinct evolutionary histories. These include areas of neoendemism, characterized by the accumulation of range-restricted short branches and typically indicating recent in-situ diversification; areas of paleoendemism, characterized by the accumulation of range-restricted long branches and potentially representing refugia for lineages that have experienced high extinction; and areas of mixed or superendemism, which accumulate range-restricted branches without a predominance of short or long branches. The concepts of neoendemism and paleoendemism are rooted in classical plant biogeography works [9, 12, 13], but spatial phylogenetics places them within a robust statistical and phylogenetic framework. In the last decade, numerous studies around the world have shown how CANAPE and other spatial phylogenetics methods can be applied to detect critical areas for conservation of evolutionary history (e.g., [14–16]). This progress has been fostered by the growing availability of biodiversity data from global repositories [10], including species occurrence records (primarily from the Global Biodiversity Information Facility; GBIF) and DNA sequences (primarily from GenBank), the latter of which have enabled the development of deeply sampled megaphylogenies (e.g., the megaphylogeny of seed plants [17]).

The European continent is not usually thought of as a key area for vascular plant diversity at the global scale, given its relatively low species richness compared with tropical and subtropical regions [18] (but see [19]). However, the southern European peninsulas are part of the Mediterranean Basin, one of the world's major biodiversity hotspots, where a remarkable concentration of strictly endemic species is undergoing exceptional loss of habitat [20]. In particular, multiple global spatial phylogenetic analyses of vascular plants coincide in highlighting the Iberian Peninsula, in southwestern Europe, as an

endemism center of global relevance [21–23]. Specifically, the global species-level analysis of seed plants by Cai et al. [21] robustly identified mainland Spain, which comprises most of the Iberian Peninsula, as a center of neoendemism. In contrast, the global genus-level analysis of angiosperms by Qian et al. [23] and the species-level analysis of ferns by Qian et al. [22] recovered the same area as a center of paleoendemism. Finer-scale spatial phylogenetic analyses of the Iberian vascular flora would greatly contribute to understanding the status of this region as an endemism center at the global scale.

The Iberian Peninsula is among the most important regions for the European vascular flora, harboring over 5,500 native species [24, 25], which represent about 54% of European species [26]. Remarkably, this includes 27 genera and more than 1,200 species that are endemic in the absolute sense. The Iberian Peninsula comprises several glacial refugia for plant diversity [27, 28] and represents an important portion of the Mediterranean Basin biodiversity hotspot [20]. The diversity and singularity of the Iberian flora have usually been interpreted as the result of multiple interacting factors. These include the high environmental heterogeneity of the Iberian Peninsula (climate, soil, topography), together with its location between Africa and the rest of Europe, and a complex geological and climatic history since the Miocene [29, 30]. This history comprises major events such as the Messinian salinity crisis in the late Miocene, the establishment of a mediterranean-type climate across most of the Peninsula in the Pliocene, and the Quaternary climatic cycles [31].

The vascular flora of the Iberian Peninsula is exceptionally well documented, having been studied since the 16th century [26]. A modern, complete flora has recently been finalized [24] and a comprehensive occurrence database for native species at 10-km resolution is available [25]. Additionally, more than 50% of native species have been sequenced and are included in megaphylogenies [17, 32], while the remaining ones can be incorporated based on taxonomy using phylogenetic imputation strategies [33]. Non-phylogenetic geographic patterns of diversity and endemism in the Iberian flora have been extensively studied using diverse approaches [25, 26, 29, 30, 34–39], but very few studies have incorporated a phylogenetic component. An early spatial phylogenetic study by Buira et al. [40] analyzed patterns of phylogenetic endemism, but it was based on a phylogenetic tree including only species strictly restricted to Iberia (absolute endemics) and did not incorporate approaches such as CANAPE. Therefore, the location of centers of plant neoendemism, paleoendemism and mixed/superendemism in the Iberian Peninsula remains uncertain, limiting the application of evolutionary criteria to area prioritization.

The aim of this study was to investigate spatial phylogenetic patterns for the entire Iberian angiosperm flora (c. 95% of the Iberian vascular flora) and to relate these patterns to abiotic factors. Specifically, we addressed the following objectives: (i) to infer patterns of phylogenetic diversity and endemism for native Iberian angiosperms; (ii) to assess the effects of several sources of uncertainty, including the incorporation of species lacking phylogenetic information through imputation strategies, the exclusion of taxonomically contentious apomictic genera, and the analysis of strictly endemic species rather than the entire flora; and (iii) to relate spatial phylogenetic patterns to climatic, edaphic and topographic variables. Based on previous studies in the Iberian Peninsula and elsewhere, we hypothesized that neoendemism patterns in the Iberian vascular flora are associated with a combination of climatic and edaphic conditions, whereas paleoendemism patterns are primarily associated with climatic conditions [40–42].

2. Materials and methods

2.1. Study region

The Iberian Peninsula comprises continental Portugal and Spain, as well as Andorra and Gibraltar. In addition, our study region includes the Balearic Islands and several minor offshore archipelagos, all of which are geologically and floristically related to the Iberian Peninsula [26, 43] (Fig. 1). The flora of this entire region is collectively referred to as the Iberian flora. This territory displays pronounced topographic heterogeneity, ranging from sea level to 3,479 m (Mulhacén Peak in the Sierra Nevada, Baetic System [44]; Fig. 1). It is crossed by several east-west mountain ranges that may have restricted latitudinal migration and favored the isolation of lineages during glacial-interglacial periods [30, 45]. In fact, the importance of Iberian mountains for species richness and endemism has been highlighted by several studies [25, 26, 29, 30]. Most of the Iberian Peninsula belongs to the Mediterranean floristic region [46], characterized by hot, dry summers and a strongly seasonal precipitation regime [47]. Climate varies across Mediterranean Iberia, with more continental conditions in the interior plateaus and semi-arid conditions in the southeast. In contrast, the northern and northwestern regions belong to the Eurosiberian floristic region and display a temperate oceanic climate [46] (Fig. 1). The Iberian Peninsula also exhibits considerable geological and edaphic diversity, which has influenced plant diversification [40], broadly divided into two main domains: acidic siliceous substrates in the western half, and basic calcareous and dolomitic substrates in the eastern half. More isolated ultramafic, gypsum, saline, and volcanic substrates further contribute to edaphic and floristic diversity.

2.2. Species list and distribution data

For the native angiosperm flora of the study region, the list of species was retrieved from the AFLIBER database [25], which is largely based on *Flora iberica* [24] but incorporates subsequent taxonomic updates. For nomenclatural standardization, we used the names accepted in the World Checklist of Vascular Plants [48] to harmonize names with those used in the megaphylogeny [17, 32] (see below). A list showing the correspondence between names in the distribution and phylogenetic datasets (*species_dictionary.csv*) can be downloaded from the Zenodo repository (https://doi.org/10.5281/zenodo.23045722) In addition, we merged subspecific taxa into their respective species, resulting in a final dataset of 5,390 angiosperm species belonging to 1,008 genera. Distribution data were also retrieved from the AFLIBER database, which provides the most up-to-date distribution dataset available for vascular plants of the study area (available at https://doi.org/10.5061/dryad.gmsbcc2kv; [25]). The database integrates 1.8 million records from numerous sources at a spatial resolution of 10×10 km and has been validated by taxonomic experts to ensure data quality (see extended methods in [25] for details). The 10-km resolution was selected to maximize spatial detail after preliminary spatial phylogenetic analyses at 10-, 20- and 50-km resolution revealed consistent patterns. Small grid cells occurring along the boundaries between UTM zones, with an area smaller than 50 km^2, were merged with the nearest complete adjacent grid cell. After grid-cell merging and subsetting to include only angiosperm taxa, our dataset comprised 1,714,051 occurrences (Table 1), distributed across 6,142 grid cells, with distribution information available for all species.

Table 1
Datasets for the angiosperm flora of the Iberian Peninsula and the Balearic Islands used in spatial phylogenetic analyses. For each dataset, the numbers of species and occurrences are provided.

Dataset	Species	Occurrences
Comprehensive	5,390	1,714,051
Core	2,858	1,135,570
Apomict-excluded	2,759	1,127,206
Strictly-endemic	449	34,106

2.3. Phylogenetic trees

Spatial phylogenetic analyses require a phylogenetic tree encompassing as many species as possible that occur in the study region and are included in a spatial dataset. As a comprehensive phylogenetic tree of Iberian angiosperms is not yet available, our study was based on the previously published, time-calibrated spermatophyte GBOTB megaphylogeny ("GenBank taxa with a backbone provided by Open Tree of Life") [17, 32]. Therefore, no DNA sequences were newly generated from wild or cultivated plants for this study, and no collection permits were required. The GBOTB megaphylogeny was first pruned to include only Iberian angiosperms, resulting in a tree containing 53% of known Iberian native species (2,858) and 75% of native genera (756). Our main analyses were conducted using this *core dataset*, containing only species included in the GBOTB megaphylogeny (one phylogenetic tree). The distribution data (*core_distribution.csv*) and phylogenetic tree (*core_phylogeny.tree*) for this dataset can be downloaded from Zenodo. Additionally, we assessed the impact of different sources of uncertainty in the data using three alternative datasets, including distribution data and phylogenetic trees (Table 1):

1. A *comprehensive dataset* containing all known Iberian angiosperm species generated using imputation (100 phylogenetic trees). We performed phylogenetic placement imputation to increase phylogenetic coverage using the R package *randtip* [33], with the GBOTB megaphylogeny as the backbone tree. This software inserts phylogenetically unplaced taxa (PUTs, hereafter) into an ultrametric backbone tree based on their taxonomy, and addresses several drawbacks of previous tools (see [33] for details). Specifically, each PUT was inserted into the backbone tree at a random position along a branch selected within the clade delimited by taxa belonging to the same genus or, when congeneric taxa are missing, the same family, with the probability of branch selection being proportional to branch length. For example, the European beech, *Fagus sylvatica*, which was not included in the backbone tree, was inserted at a randomly defined position within the clade defined by the other *Fagus* species present in the backbone tree. Whenever congeneric taxa of a PUT formed a polyphyletic group, we inserted it into the largest monophyletic cluster of congeneric taxa (i.e., the "largest" option in the *randtip* workflow). Additionally, if the congeneric taxa of a PUT formed a paraphyletic group, it was bound while preserving this topology. In this way, all Iberian angiosperms included in the species list but missing from the GBOTB tree were incorporated. To account for uncertainty arising from random placement, the *randtip* procedure was repeated 100 times, yielding a set of 100 trees encompassing the entire Iberian angiosperm flora and capturing uncertainty in the phylogenetic placement of PUTs. The distribution data (*comprehensive_distribution.csv*) and phylogenetic trees (*comprehensive_phylogenies.tre*) can be downloaded from Zenodo.
2. An *apomict-excluded* dataset excluding species belonging to apomictic and taxonomically complex genera with large numbers of Iberian endemics (i.e., *Alchemilla*, *Hieracium*, *Limonium*, and *Taraxacum* [29]; restricted to only species included in the GBOTB tree; one phylogenetic tree). This dataset was designed to assess potential biases associated with highly polymorphic groups in which apomixis (asexual production of seeds), hybridization and reticulate evolution are common, resulting in numerous microspecies whose taxonomic delimitation and distribution ranges are often uncertain, following Cai et al. [21]. The distribution data (*apomict-excluded_distribution.csv*) and phylogenetic tree (*apomict-excluded_phylogeny.tre*) can be downloaded from Zenodo.
3. A *strictly-endemic dataset* comprising only species that are endemic to the Iberian Peninsula in the absolute sense, and which were also included in the GBOTB tree (one phylogenetic tree). This dataset addressed the issue that some centers of endemism may occur near the boundaries of the study region, partly due to lineages that are range-restricted within the study region but have wider distribution ranges outside [10]. The distribution data (*strictly-endemic_distribution.csv*) and phylogenetic tree (*strictly-endemic_phylogeny.tre*) can be downloaded from Zenodo.

2.4. Spatial phylogenetic analyses

For each dataset (Table 1) and grid cell, two non-phylogenetic metrics were calculated: species richness (SR) and weighted endemism (WE). Spatial phylogenetic analyses were conducted using the R package *canaper* [49] on the high-performance computing cluster at *Universidad Autónoma de Madrid* (CCC-UAM). First, observed phylogenetic diversity (PD) and phylogenetic endemism (PE) were calculated and mapped. Then, we calculated significance levels of PD, PE, relative phylogenetic diversity (RPD) and relative phylogenetic endemism (RPE) based on comparison of observed values against a null model (curveball, [50]) generated by spatial randomization (999 random communities and 200,000 iterations). According to significance levels, values were categorized and mapped as significantly high ($p > 0.975$ and $p > 0.99$), significantly low ($p < 0.025$ and $p < 0.01$), or non-significant. RPD significance reflects concentrations of short or long branches, while RPE significance reflects concentrations of range-restricted short or long branches. Finally, endemism categorization was performed using the CANAPE approach [11], which distinguishes four categories of endemism centers: neoendemism, paleoendemism, mixed endemism and superendemism.

For the *comprehensive dataset*, non-phylogenetic metrics (SR and WE) were calculated and compared with those obtained from the *core dataset* based on absolute difference and percentage similarity values. Spatial phylogenetic analyses were conducted for each of the 100 trees of the *comprehensive dataset* generated using *randtip* to account for uncertainty in the placement of PUTs. Results across the 100 replicates were summarized using different approaches

depending on the metric. For PD and PE, which are continuous metrics, we calculated the mean, standard deviation, maximum and minimum values for each grid cell. For the significance levels of PD and RPD, as well as for CANAPE, the percentage of trees yielding each category was calculated for each grid cell. For CANAPE, superendemism cells were counted among the mixed endemism ones, as they are a subset of them. Significance levels and CANAPE results across the 100 trees were further summarized using a majority-rule approach, whereby a category was assigned to a grid cell when it was obtained for > 50% of trees. In the case of CANAPE, for a small number of grid cells the four endemism categories together accounted for > 50% of replicates, but no single category exceeded 50%. These grid cells were assigned to an additional category, referred to as “undetermined endemism”.

2.5. Abiotic drivers of spatial phylogenetic patterns

To visually explore the relationships between spatial phylogenetic patterns and abiotic variables (climatic, edaphic, and topographic) across the Iberian Peninsula, we generated scatter plots following Thornhill et al. [51]. We focused on two spatial phylogenetic metrics: RPD, categorized according to its significance level, and RPE, categorized according to CANAPE results [11]. Both RPD and RPE values were obtained from the *core dataset* (Table 1). Among abiotic variables, we selected six that broadly capture the environmental variability of the Iberian Peninsula. These included four climatic variables: mean annual temperature and annual precipitation as general climatic descriptors; precipitation of the warmest quarter, as summer drought distinguishes mediterranean-type from oceanic climate [52]; and continentality, which captures the influence of oceanic vs. continental conditions. Data for mean annual temperature, annual precipitation and precipitation of the warmest quarter were downloaded from CHELSA v1.2 [53], while continentality data were obtained from ENVIREM [54]. As an edaphic variable, we included data for surface soil pH (0–5 cm) from SoilGrids 2.0 [55], as this variable has previously been identified as a driver of plant diversification in the Iberian Peninsula [40]. Finally, we included roughness data from EarthEnv [56] to represent topographic variability. The dataset used for the analyses (abiotic_drivers.*csv*) can be downloaded from Zenodo.

To statistically test the relationships between endemism categories and abiotic variables, we fitted multinomial logistic regression models using the ‘multinom’ function implemented in the *nnet* R package [57]. The response variable consisted of the CANAPE categories ("neoendemism", "paleoendemism", "mixed endemism", and "superendemism"), with non-significant cells used as the reference category. Predictor variables included the six climatic, edaphic and topographic variables listed above. All predictor variables were standardized (mean = 0, standard deviation = 1) prior to analysis to facilitate comparison of regression coefficients among predictors. Statistical significance of model coefficients was assessed using Wald z-statistics calculated from the estimated coefficients and their standard errors, and corresponding two-tailed p-values were obtained assuming a normal distribution. To evaluate the potential influence of spatial structure on the observed environmental associations, a second model was fitted including longitude and latitude as additional covariates. Model fit was compared using Akaike’s Information Criterion (AIC).

3. Results

3.1. Spatial phylogenetic patterns

When considering only species included in the GBOTB megaphylogeny (*core dataset*; see Table 1), SR showed the highest values along the major mountain ranges of the study region (Cantabrian Mountains, Pyrenees, Central System, and Baetic System) and along the eastern coastline (Figure S1a), where values were influenced by a higher density of floristic records. WE reached its highest values in the two highest mountain ranges (Sierra Nevada, in the Baetic System, and the Pyrenees), whereas the remainder of the territory exhibited generally low values (Figure S2a).

Raw PD displayed the highest values along the mountain ranges and eastern coastline, similar to SR patterns (Figure S3a). Significantly high PD values were predominantly concentrated along the northern coastline of the Eurosiberian floristic region, whereas the remaining territory, including the Mediterranean floristic region, was dominated by significantly low PD values (Fig. 2a). RPD showed significantly high values, indicating a concentration of long branches, along the northern and northwestern coastline and in the southern half of the Peninsula, while significantly low values, indicating a concentration of short branches, were found in the Northern Inner Plateau, the Iberian System, the Ebro Valley, the Guadiana Valley, and the Baetic System (Fig. 2b).

Raw PE showed a very high value in a single grid cell in the Cantabrian Mountains, driven by the presence of *Nuphar pumila* exclusively in this grid cell. Major mountain ranges, including the Pyrenees, the Cantabrian Mountains and the Baetic System, also exhibited higher values than the rest of the territory (Figure S3b).

CANAPE (Fig. 2c) revealed a concentration of centers dominated by neoendemism in the Baetic System and the Balearic Islands. Centers dominated by paleoendemism were dispersed across the study region, with a slightly higher concentration in the northwestern quarter of the Peninsula, the semi-arid southeast, the Pyrenees and certain Mediterranean and Atlantic coastal wetlands. Centers of mixed endemism were primarily located along three major mountain ranges (the Cantabrian Mountains, the Pyrenees, and the Baetic System) as well as in the Balearic Islands, the semi-arid southeast and some coastal areas (e.g., Doñana and Marismas del Odiel in southwestern Spain). Superendemism was largely restricted to the highest mountains (Sierra Nevada, in the Baetic System, and the Pyrenees) and to parts of the semi-arid southeast and the Balearic archipelago.

3.2. Effects of phylogenetic imputation

When using the *comprehensive dataset*, which incorporates the entire known angiosperm flora of the study region (see Table 1), the SR pattern closely matched that observed for the *core dataset* (Figure S1). Similarly, WE showed highly consistent patterns between the *comprehensive dataset* and the *core dataset* (Figure S2).

Spatial phylogenetic patterns were similar across the 100 replicates derived from the *randtip*-generated trees of the *comprehensive dataset* (Figs. 3, S4-S8). Furthermore, mean values (PD in Figure S4a; PE in Figure S5a) and majority-rule maps (PD significance in Figure S6f; RPD in Figure S7f; CANAPE in Fig. 4a) were similar to those obtained from the *core dataset* (Figs. 2, S3). Nevertheless, some differences were observed between the two datasets. For example, PE

did not show the extremely high value in a single grid cell of the Cantabrian Mountains (Figure S5), and significantly high values of RPD increased in certain areas (Figure S7f). Overall, the majority-rule CANAPE result from the *comprehensive dataset* was largely congruent with that from the *core dataset*, and the number of grid cells recovered as undetermined endemism was very small (0.47%). Nevertheless, the significance or endemism category changed for a subset of grid cells. In particular, a substantial proportion of neo- and paleoendemism grid cells in the *core dataset* were classified as mixed endemism in the *comprehensive dataset*, while some mixed endemism grid cells shifted to superendemism (Fig. 4b, Table S1). Changes included, for example, the recovery of a neoendemism center in northwestern Spain (central Galicia), an increase in superendemism across the Balearic Islands, a reduction of the paleoendemism area in the semiarid southeast, and a wider presence of mixed endemism centers in the Southern Inner Plateau (La Mancha) (Fig. 4a).

3.3. Exclusion of apomictic genera

Spatial phylogenetic results obtained using the *apomict-excluded dataset* (Table 1; Figs. 5a and S9) were very similar to those derived from the *core dataset* (Figs. 2, S1a, S2a and S3). Relevant differences in CANAPE results included some shifts from neoendemism to mixed endemism or superendemism in parts of the Balearic Islands, as well as an expansion of neoendemism in the western Baetic System (Fig. 5a).

3.4. Strictly endemic species only

The *strictly-endemic dataset* (Table 1) contained a much smaller number of species, which were mainly concentrated in mountain ranges, whereas several low-altitude grid cells (e.g., in the Northern and Southern Inner Plateaus) showed a complete absence of endemic species (Figure S10a). Raw values of PD and PE declined, and the number of grid cells with significantly low or high values of PD and RPD was greatly reduced (Figure S10c – f) compared with results obtained from the *core dataset* (Figs. 2a – b, S3). Endemism centers were largely restricted to the Balearic Islands, the Pyrenees, and the Baetic System (Fig. 5b). Relative to the *core dataset*, endemism centers in the Pyrenees and the Baetic System were reduced and showed a predominance of mixed endemism. In contrast, patterns in the Balearic Islands were more similar to those derived from the *core dataset*.

3.5. Relationships with abiotic variables

Scatter plots suggested relationships between RPD and several abiotic variables (Fig. 6a-f). Significantly low RPD values were primarily associated with low annual precipitation (< 1000 mm; Fig. 6b), low precipitation of the warmest quarter (Fig. 6c), and high continentality (Fig. 6d), whereas significantly high RPD values occurred across a broader range of these same variables. For RPE and endemism center categorization (Fig. 6g-l), additional patterns were observed. Neoendemism centers were predominantly associated with low annual precipitation (Fig. 6h), low precipitation of the warmest quarter (Fig. 6i), and high pH (Fig. 6k), whereas paleoendemism and, particularly, mixed endemism centers spanned a wider range of values for these same variables. Superendemism centers were mainly associated with intermediate values of continentality (Fig. 6j), but showed bimodal distributions, with concentrations at both low and high values, for mean annual temperature (Fig. 6g), precipitation of the warmest quarter (Fig. 6i), and roughness (Fig. 6l).

The non-spatial multinomial logistic regression model recovered the following patterns (Table 2): neoendemism was positively associated with pH and roughness, and negatively with precipitation of the warmest quarter and continentality; paleoendemism was positively associated with mean annual temperature and roughness, and negatively with continentality; mixed endemism was positively associated with pH and roughness, and negatively with precipitation of the warmest quarter and continentality; and superendemism was positively associated with pH and roughness, and negatively with mean annual temperature, annual precipitation and continentality.

Table 2

Relationships between endemism categories for the Iberian angiosperm flora and abiotic variables estimated using multinomial logistic regression models. R are shown for both a non-spatial and a spatial model, the latter including longitude and latitude as additional covariates. Regression coefficients are presente the following significance codes for predictor coefficients: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

Model	Endemism category	(Intercept)	Mean annual temperature (°C)	Annual precipitation (mm)	Precipitation of warmest quarter (mm)	Continentality (°C)	pH	Roughness	Longitude	Latit
Non-spatial	Neoendemism	-4.200	-0.153	0.343	-1.327***	-1.173***	1.303***	0.972***	–	–
	Paleoendemism	-3.691	0.487***	0.016	0.114	-0.665***	0.207	0.244*	–	–
	Mixed endemism	-2.511	0.060	0.066	-0.392***	-1.284***	0.936***	0.733***	–	–
	Superendemism	-5.935	-0.763***	-0.841***	-0.194	-2.639***	1.915***	1.504***	–	–
Spatial	Neoendemism	-4.870	-1.168***	1.658***	-2.338***	-1.436***	0.095	0.005	2.255***	-1.34
	Paleoendemism	-3.686	0.257	0.181	0.125	-0.703***	0.015	0.049	0.296	-0.39
	Mixed endemism	-2.685	-0.645***	0.589***	-0.177	-1.427***	0.424***	0.127	0.783***	-1.26
	Superendemism	-7.960	-2.174***	1.174***	-1.056***	-3.570***	0.861*	0.467**	2.923***	-2.63

The spatial model, including longitude and latitude as additional covariates, recovered the following patterns: neoendemism was positively associated with annual precipitation and longitude, and negatively with mean annual temperature, precipitation of the warmest quarter, continentality and latitude; paleoendemism was negatively associated with continentality and latitude; mixed endemism was positively associated with annual precipitation, pH and longitude, and negatively with mean annual temperature, continentality and latitude; and superendemism was positively associated with annual precipitation,

pH, roughness and longitude, and negatively with mean annual temperature, precipitation of the warmest quarter, continentality and latitude. AIC values supported this spatial model (AIC = 5694.966) against the non-spatial model (AIC = 6504.031).

4. Discussion

Previous analyses of biodiversity in the Mediterranean Basin have primarily focused on species richness, paying little attention to the evolutionary history of species within specific areas. Here, our species-level spatial phylogenetic analysis reveals fine-scale patterns of angiosperm evolutionary diversity across the Iberian Peninsula and the Balearic Islands, in the western Mediterranean biodiversity hotspot, including a significant representation of paleo-, neo-, mixed, and superendemism.

4.1. Distribution of phylogenetic diversity and endemism

The geographic pattern for raw values of angiosperm phylogenetic diversity across the Iberian Peninsula and the Balearic Islands (Figure S3a) mirrors that of species richness (Figure S1; see also [25, 26]), as reported in previous studies from other regions of the world [10, 58]. As expected, higher values are observed along mountain ranges, as well as in those regions with higher sampling intensity in the AFLIBER distribution database [25]. This tight relationship between PD and richness is expected, given that PD is an additive measure – each species sampled adds a branch to the phylogeny [7]. It therefore serves as the null hypothesis for the randomization test used here to identify locations where PD is surprisingly high or low for a given richness.

The distribution of statistically significant PD values based on the randomization tests (which controls for the effect of taxon richness; Fig. 2a) shows a clear geographic pattern. Along the northern coast of the Iberian Peninsula, corresponding to low and mid elevations within the Eurosiberian floristic region, significantly high PD values (indicating phylogenetic overdispersion) are recovered for many grid cells (Fig. 2a), potentially an indication of older communities where competition has had time to act. Due to long-term environmental stability, the Cantabrian coast appears to have acted as a refugium in which distantly related evolutionary lineages have persisted [59]. In contrast, significantly low values (indicating phylogenetic clustering) clearly predominate across the remaining territory, including the entire Mediterranean region and the main Eurosiberian mountain ranges of Iberia (the Cantabrian Mountains and the Pyrenees; Fig. 2a). These results are consistent with previous analyses of the phylogenetic structure of forest communities [59]. The environmental filter imposed by the mediterranean-type climate means that only a subset of plant lineages has been able to persist and diversify under these conditions. Mediterranean flora and vegetation formed relatively recently in association with the establishment of the mediterranean-type climate in the Pliocene [60] (c. 3.2 Ma), and have subsequently evolved adaptations to withstand the summer drought and recurrent fires that characterize this climate [47, 61–63]. This has promoted recent diversification within specific parts of the phylogeny in Mediterranean Iberia, resulting in comparatively low PD. Previous studies have linked phylogenetic clustering in Mediterranean plant communities to environmental filtering associated with fire regimes and aridity [59, 64, 65]. Consistent with this interpretation, the northern coast of the Iberian Peninsula is also thought to have acted as a refugium for non-angiosperm plant groups with higher water requirements and greater sensitivity to drought (Molino *et al.*, in prep.). An important exception to this general pattern is the significantly low PD observed in the main Eurosiberian mountain ranges (Fig. 2a), which may be explained by the strong environmental filtering imposed by low temperatures at high elevations, relative to the milder conditions prevailing at low and mid altitudes within the Eurosiberian floristic region [59].

Relative phylogenetic diversity displays a different spatial pattern, with significantly high values more widely distributed, not only in the Eurosiberian floristic region but also across Mediterranean Iberia, particularly in the southern half of the Peninsula (Fig. 2b). In contrast, significantly low RPD values are largely restricted to dry and continental areas (Figs. 2b, 6a-f). This pattern suggests that both oceanic and milder mediterranean-type climates have acted as refugia for the persistence of ancient evolutionary lineages, whereas harsher, more continental mediterranean-type climates are dominated by assemblages resulting from recent diversification, regardless of their range sizes. These results are broadly consistent with patterns reported in other regions of the world, such as California [51] and Australia [66], where more humid climates are associated with high RPD. Nevertheless, this result should be taken with caution given that many cells with significant RPD values may be floristically undersampled (see Figure S1).

Regarding phylogenetic endemism, the main mountain ranges (Fig. 1) display the highest raw values in mainland Iberia (Figure S3b), associated with high weighted endemism (Figure S2; see also [40]). The CANAPE results identified the Pyrenees, the Baetic System, parts of the Cantabrian Mountains, the semi-arid southeast and some coastal wetlands as the main centers of endemism in mainland Iberia (Figs. 2c, 3, 4a, 5a; see details below). In fact, the Pyrenees and the Baetic System were still identified as endemism centers, although less strongly, when analyses were restricted to Iberian endemics only (Fig. 5b). These results highlight the importance of topographically complex mountain ranges for Iberian angiosperm endemism [26, 29, 35, 36, 67, 68]. Given the predominantly east-west orientation of Iberian mountain ranges (Fig. 1), they have traditionally been interpreted as barriers to the movement of organisms during Quaternary climatic oscillations, leading to the formation of multiple isolated refugia and promoting speciation [69]. This has resulted in a high diversity of range-restricted taxa in these mountain ranges, which explains their strong concentration of PE (Fig. 2c). Nevertheless, the scarcity of significant endemism in mountain ranges of the inner Iberian Peninsula (Central System, Iberian System) is noteworthy (see also [39]). This pattern may reflect their more extreme continental conditions, which may have reduced their capacity to act as refugia during Quaternary climatic oscillations [29]. In addition, the low concentration of endemism in the Central System may be partially explained by its relatively uniform siliceous lithology, which may have provided fewer opportunities for diversification driven by local edaphic specialization. Meanwhile, the north-south orientation and central location of the Iberian System may have provided migratory routes between northern and southern mountain ranges and limited geographic isolation and diversification. Finally, although both the Central System and the Iberian System reach considerable elevations, they are generally more environmentally homogeneous and may have offered fewer stable microrefugia than the Pyrenees or the Baetic ranges, potentially limiting the persistence and diversification of narrow endemic lineages.

Our analyses also highlight the Balearic archipelago (Fig. 1) as a hotspot of phylogenetic endemism. Despite its relatively low species richness compared to the mainland (Figure S1), the prevalence of endemic lineages in the Balearic flora [29, 70, 71] led to the identification of endemism centers across almost the entire Balearic archipelago (Figs. 2c, 3, 4a, 5), consistent with the geographic isolation of lineages [72]. The separation of the Balearic archipelago from the

Baetic System (of which it is geologically considered a continuation) occurred after the Messinian salinity crisis [73, 74], thereby isolating the Balearic flora prior to the establishment of a mediterranean-type climate [60] (c. 3.2 Ma). This long-term isolation appears to have promoted evolutionary divergence between some mainland and island lineages, leading to the current concentration of PE across the archipelago. Nevertheless, eustatic sea-level falls during the Quaternary glaciations reduced the distance separating the Iberian Peninsula and the Balearic Islands [75], potentially favouring colonization. In addition to strictly endemic Balearic lineages, PE in the archipelago is enhanced by range-restricted species shared with the eastern Iberian coast [76] and with the islands of Corsica and Sardinia (outside of the study region) [77].

4.2. Distribution of endemism categories

The three types of endemism centers (neo-, paleo- and mixed/superendemism) are heterogeneously distributed across the Iberian Peninsula (Figs. 2c, 3, 4a, 5a). Neoendemism is concentrated in several mountain ranges (*sierras*) of southern Iberia, particularly in the western Baetic System (Serranía de Ronda, including Sierra del Aljibe, Sierra de Grazalema, Sierra Bermeja, and Sierra de las Nieves) and the eastern Baetic System (Sierra de Cazorla, Sierra de Segura, and Sierra de Filabres). It is also prominent in the Balearic Islands, especially in southeastern Mallorca, and on the highest peak of the archipelago (Puig Major in Serra de Tramuntana, Mallorca) (Figs. 2c, 3a, 4a, 5a). This distribution of neoendemism appears to be related to geographic isolation and topographic complexity, in combination with abiotic variables such as summer precipitation and soil pH, in agreement with our starting hypothesis (Fig. 6g-l). Specifically, neoendemism centers are clustered in stressful environments characterized by mediterranean-type climates (Fig. 6i) and basic substrates (Fig. 6k) that have promoted recent speciation, as suggested in a previous analysis of the Iberian endemic flora [40]. The environmental filters imposed by the establishment of mediterranean-type summer drought since the Pliocene [47, 60] and stressful soils mean that only a subset of the Iberian angiosperm flora has been able to persist and diversify in these regions of Mediterranean Iberia. Under these conditions, high evolutionary turnover, driven by recurrent local extinction followed by rapid recolonization and speciation, provides a plausible explanation for the predominance of neoendemism [78]. Our non-spatial regression model supported this interpretation, as shown by the significant associations of neoendemism with a low precipitation during the warmest quarter (characteristic of mediterranean-type climates) and a high soil pH, among other variables (Table 2). However, the relationship with pH disappeared in the spatial model. This may be explained by the strong longitudinal pattern of soil pH across the Iberian Peninsula, with a clear predominance of acidic soils in the west and basic soils in the east, which makes it difficult to disentangle the effects of geography and soil pH.

Focusing on specific regions, in the Balearic archipelago neoendemism appears to have been driven primarily by geographic isolation [71, 79], whereas in the Baetic System it has likely been promoted by a combination of factors. These include a mediterranean-type climate, geographic isolation among *sierras* resulting from a fragmented orography of recent origin [80], and edaphic specialization on diverse substrates (calcareous, dolomitic, ultramafic) [37, 81–86]. The role of ultramafic (serpentine) substrates in the western Baetic System (Sierra Bermeja) is particularly noteworthy [87, 88], resembling the well-documented influence of serpentine soils on endemism and speciation in the mediterranean-type region of California [89, 90]. Interestingly, an ultramafic area in the Eurosiberian floristic region (central Galicia in northwestern Spain [91]) is also recovered as a neoendemism center when analyzing the *comprehensive dataset* (Figs. 3a, 4a), supporting a role of these substrates in recent speciation regardless of climatic context. Additionally, gypsum substrates, which are also stressful for plants and known to promote speciation and endemism [82, 92, 93], explain some scattered neoendemism centers in the Southern Inner Plateau and Ebro Valley (Figs. 2c, 3a) (see [94, 95]).

The distribution of paleoendemism centers is more scattered across the study region (Figs. 2c, 3b, 4a, 5a) and appears to be driven by diverse processes, as indicated by their occurrence across a wider range of abiotic conditions than neoendemism centers (Fig. 6g-l). In the non-spatial regression model, paleoendemism was associated with warmer and less continental climates (Table 2). This is consistent with the expectation that paleoendemic lineages persist in climatically buffered refugial areas [96] and agrees with our starting hypothesis. Nevertheless, the relationship with mean annual temperature disappeared in the spatial model, probably as a result of the latitudinal temperature gradient, which makes it difficult to disentangle the effects of geography and temperature.

A slight concentration of paleoendemism is found in the Eurosiberian floristic region, specifically in northwestern Iberia and in scattered grid cells in the Pyrenees. The angiosperm flora of the Eurosiberian floristic region is adapted to more humid conditions [97] and is therefore phylogenetically more closely related to the flora of central Europe [26]. The presence of species whose close relatives occur outside the study region tends to increase branch lengths in the regional phylogenetic tree [49], which, when combined with a restricted distribution within Iberia, may lead to the inference of paleoendemism. An example of this is *Soldanella villosa*, which appears to contribute to paleoendemism in the western Pyrenees. A similar pattern, although under contrasting climatic conditions, is observed in the semi-arid region of southeastern Iberia, which harbors steppic species with restricted distributions within the Iberian Peninsula and close relatives in North Africa or the eastern Mediterranean Basin [98]. Likewise, coastal wetlands are inhabited by aquatic and halophytic plants that are often phylogenetically isolated and occur in only a small number of grid cells within the Iberian context (e.g., the halophytic *Tetraena alba* in Delta de l'Ebre, northeastern Spain) [99, 100]. Certain additional centers of paleoendemism scattered across western and southern Iberia may be associated with the persistence of narrowly distributed relict lineages in climatically stable, coastal, or edaphically specialized habitats that have been proposed as refugia [27]. Finally, some scattered paleoendemism grid cells in the inner Iberian Peninsula are influenced by the presence of certain phylogenetically isolated and narrowly distributed endemic species, such as *Gyrocaryum oppositifolium* [5].

Mixed endemism (including superendemism) centers are widespread across the study region but concentrated in mountain ranges, islands, semi-arid areas, and certain coastal and inland lowlands (Figs. 2c, 3c-d, 4a, 5a). Associations with abiotic variables in the regression models (Table 2) were complex and hard to interpret, likely reflecting the coexistence of range-restricted lineages associated with long, short and intermediate branches in these areas. Among mountain ranges, the Pyrenees, Cantabrian Mountains, and Baetic System display the greatest extent of mixed- and superendemism. The Pyrenees are dominated by superendemism, whereas mixed endemism predominates in the Cantabrian Mountains, with superendemism restricted to some of the highest peaks (e.g., Picos de Europa for the *comprehensive dataset*; Figs. 3d, 4a) (see [101]). In the Baetic System, mixed endemism is widespread and intermingled with neoendemism (see above), with superendemism largely restricted to Sierra Nevada, the highest mountain range of the Iberian Peninsula. The general drivers of endemism discussed above operate in these areas, with no predominance of either neo- or paleoendemism, suggesting a dual role as both centers

of recent speciation and refugia for phylogenetically isolated lineages, including the presence of arctic-alpine taxa [26]. This dual role is particularly evident in the highest mountain ranges with wide altitudinal gradients, such as Sierra Nevada [102] and the Pyrenees, characterized by high topographic heterogeneity and diverse habitats, supporting high species richness [44, 103]. Sierra Nevada hosts the highest richness of narrow endemics in the study region [29, see also 81] and is considered a major hotspot of plant endemism in the Mediterranean context [28, 104]. The floristic distinctiveness of Sierra Nevada within the Baetic System is likely driven by its pronounced topographic isolation and exceptional elevation. Its highest vegetated zones, locally known as *borreguiles*, receive a continuous supply of snowmelt throughout the year and support numerous narrow endemics. Interestingly, siliceous substrates predominate there, in contrast to other Baetic *sierras* dominated by calcareous, dolomitic and ultramafic substrates, and identified as centers of neoendemism (see above; [81]). In the Pyrenees, mixed and superendemism centers are concentrated in the central-southern sector. This is a topographically and geomorphologically complex region proposed as an important glacial refugium harboring putative relict species (e.g., *Ramonda myconi*, *Dioscorea pyrenaica*) and high levels of endemism, including many short-branched monocots [40]. Additionally, the mixed/superendemism pattern is reinforced by an edge effect associated with central European species reaching the southern limits of their distributions in this mountain range [103, 105] (see below).

In the Balearic Islands, mixed endemism is intermingled with neoendemism, and superendemism predominates in the mountains of northwestern Mallorca (Serra de Tramuntana), except for the neoendemism center on the highest peak (see above; see also [79]). In these mixed/superendemism centers, long-branched endemics, such as the monotypic genera *Naufraga* and *Femeniasia*, coexist with short-branched ones. Finally, another important area of mixed endemism is the semi-arid southeast of the Iberian Peninsula, in this case intermingled with paleoendemism (see above). Notably, the most arid zones of this region, which are also the most arid zones of Europe (Desierto de Tabernas, Cabo de Gata), are recovered as superendemism centers, reflecting a combination of species phylogenetically isolated in the Iberian context, with their closest relatives in North Africa and the eastern Mediterranean Basin, and recently diversified species from Mediterranean lineages [98].

4.3. Limited effects of confounding factors

One of the limitations of spatial phylogenetic analyses applied to specific territories is an edge effect, in which endemism centers may be inferred as a result of lineages that are range-restricted in the study region, typically around its edges, but widespread outside [106]. As discussed above, this pattern is relevant to the Iberian flora, given that some European species reach the Pyrenees as the southern limit of their distribution [103], while some African species reach southern Iberia as the northern limit of their distribution [26]. These taxa contribute to the prevalence of endemism centers in these areas (Fig. 2c), which is not necessarily an issue for the inference of regional-level patterns and their application to conservation policies [10]. Nevertheless, we explored whether the Pyrenees and southern Iberia (particularly the Baetic System) remain areas of phylogenetic endemism after partially accounting for this edge effect by evaluating spatial phylogenetic patterns using strictly endemic species, that is, species exclusive to the study region (Figs. 5b, S10). This approach is not without limitations, as it does not fully remove the edge effect given that deeper lineages also affect PE (i.e., a species may be strictly endemic but have widespread close relatives outside the region). The approach also greatly reduces the number of species available for analysis, which may obscure biogeographic patterns in the study region. In fact, the number of statistically significant grid cells was greatly reduced for all metrics (Figs. 5b, S10) compared with analyses including non-endemic species (Figs. 2, S3). Importantly, CANAPE still detected endemism centers in both the Pyrenees and the Baetic System (Fig. 5b), although these were less extensive and concentrated primarily at their highest peaks in the Central Pyrenees and Sierra Nevada. This result supports the role of the two highest mountain ranges for the evolutionary diversity of the angiosperm flora, although the predominance of mixed endemism when analyzing strictly endemic species is consistent with a limited ability of CANAPE to discriminate between neo- and paleoendemism based on this dataset. In contrast, the pattern in the Balearic Islands (Fig. 5b) is similar to that found when non-endemic species are included (Fig. 2c). This is associated with the high number of Balearic endemics [29, 70], which is expected in an insular system [72].

Another potential confounding factor affecting our results is the presence of several genera with large numbers of Iberian endemics that reproduce by apomixis, specifically *Alchemilla*, *Hieracium*, *Limonium*, and *Taraxacum*. These are among the thirty plant genera with the highest numbers of Iberian endemics [29]. Because of the rapid diversification of apomictic genera and the strongly overlapping distributions typically exhibited by the resulting taxa [107], they were expected to contribute to centers of neoendemism, but their contentious taxonomy introduces uncertainties. After excluding these genera from spatial phylogenetic analyses, the general pattern (Fig. 5a) did not change significantly (Fig. 2c), highlighting the robustness of the results. The most relevant change was the reduction of neoendemism centers in the Balearic Islands, particularly in southeastern Mallorca (Fig. 5a), which is associated with the high richness in *Limonium* endemics in this area [73].

4.4. The use of phylogenetic imputation in spatial phylogenetics

To the best of our knowledge, this is the first study to explicitly test the effect of phylogenetic imputation on the inference of spatial phylogenetic patterns. Due to sampling limitations, spatial phylogenetic studies are often restricted to relatively small taxonomic groups that allow deep species-level sampling (e.g., [11, 14, 108]), or are conducted at higher taxonomic scales, such as the genus level (e.g., [23, 109]). Species-level studies focused on large floras are typically limited to analyzing a subset of the known diversity [21, 110] or require substantial resources to sample the highest possible proportion of species (e.g., [16]). In the present study, we were able to analyze the known angiosperm flora of the Iberian Peninsula (*comprehensive dataset*, 5,390 species) by combining the previously-published GBOTB megaphylogeny [17, 32] with a phylogenetic imputation strategy that incorporates unsampled species based on taxonomic information [33]. This approach allowed us to use distribution data for the entire flora and to explore spatial phylogenetic patterns of Iberian angiosperms at high (species level) resolution (Fig. 4a), while explicitly accounting for uncertainty in the placement of unsampled species (Fig. 3).

Our phylogenetic imputation approach produced generally congruent results across replicates (Figs. 3, 4b, S4-S8). For the endemism center categorization (CANAPE), only 29 grid cells (0.47% of the territory) were categorized as undetermined endemism (i.e., all endemism categories combined accounted for more than half of the replicates, but no single category reached a majority; Fig. 4a). Moreover, the spatial patterns observed in the majority-rule CANAPE map derived from the imputation replicates (Fig. 4a) are very similar to those obtained when analyzing only the species included in the GBOTB megaphylogeny (*core dataset*, Fig. 2c), with most grid cells retaining the same endemism category (Fig. 4b). These results demonstrate the feasibility of comprehensively

analyzing regional floras (or faunas) when available phylogenetic data broadly represent extant biodiversity. In our case, the GBOTB megaphylogeny includes a substantial proportion of the Iberian angiosperm flora (53% of species and 75% of genera).

Of course, observing overall congruence between spatial phylogenetic patterns obtained with and without phylogenetic imputation is not necessarily a reason to use imputation. Some differences were detected, which should be interpreted with caution. There could be artifacts introduced if the sampling of clades in the backbone phylogeny is biased, or if taxonomic problems exist at the genus or family level. However, some differences may be due to the imputed species adding a signal that is missing in the taxa with existing molecular data. For example, a neoendemism center is recovered in northwestern Spain (central Galicia) when using the *comprehensive dataset* (Figs. 3a, 4a), but not when using the *core dataset* (Fig. 2c). This neoendemism center is consistent with the presence of ultramafic lithologies in this area, mirroring the association between this type of substrate and neoendemism observed in southern Iberia (Sierra Bermeja). In fact, this area of central Galicia harbors several narrowly distributed serpentinophytes that are absent from the GBOTB megaphylogeny but are incorporated into the *comprehensive dataset* through phylogenetic imputation (*Armeria merinoi*, *Leucanthemum gallaecicum*; see [39]). Similarly, the analysis of the *comprehensive dataset* contributes to support some isolated endemism centers in mountain ranges of the inner Iberian Peninsula, including the Central System (mixed endemism in Sierra de Gredos, Figs. 3c, 4a, see [111]; neoendemism in Sierra de Guadarrama, Figs. 3a, 4a) and the Iberian System (neoendemism in Sierra de Javalambre, Figs. 3a, 4a). These examples illustrate the usefulness of phylogenetic imputation in uncovering biologically meaningful spatial phylogenetic patterns that would otherwise remain undetected [112].

The Iberian vascular flora constitutes an ideal system for exploring the utility of phylogenetic imputation in spatial phylogenetics, as it is well documented in terms of taxonomy [24] and distribution ranges [25], and it is well represented in global megaphylogenies [17, 32]. Therefore, while species-level phylogenetic sampling remains incomplete, phylogenetic imputation approaches (e.g., *randtip* [33]) provide a viable solution for conducting detailed spatial phylogenetic analyses without restricting them to a subset of the known diversity, to small taxonomic groups (e.g., [14]), or to higher taxonomic levels such as genera (e.g., [109]). Nevertheless, the sequencing of additional taxa is ongoing, which will improve accuracy and reduce uncertainty in future analyses, and will help test results obtained through phylogenetic imputation.

5. Conclusions

This study sheds light on the spatial phylogenetic patterns of the angiosperm flora of the Iberian Peninsula and the Balearic Islands, a key region within the Mediterranean biodiversity hotspot and a plant endemism center at the global scale. The Balearic archipelago, major mountain ranges (the Pyrenees, the Baetic System, and the Cantabrian Mountains), coastal wetlands, and the semi-arid southeast of the Iberian Peninsula emerge as prominent areas of phylogenetic endemism, with contrasting prevalences of neoendemism, paleoendemism and mixed/superendemism shaped by climatic and edaphic conditions. In addition, this study represents the first exploration of phylogenetic imputation as a means to achieve virtually complete species-level sampling in a spatial phylogenetic analysis of an entire flora, while explicitly accounting for uncertainty in the placement of unsampled species. Overall, results were highly congruent with those obtained without phylogenetic imputation, while also revealing additional patterns that would otherwise remain undetected. Together, the geographic patterns of phylogenetic diversity and endemism uncovered here provide novel insights into the evolutionary and biogeographic processes shaping plant diversity in the western Mediterranean, and offer a robust framework to inform future biodiversity management and conservation strategies (see also [113]).

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declared no conflict of interest.

Supporting information

Additional information can be found online in the Supporting information section. All datasets used in this study can be found in the Zenodo repository: https://doi.org/10.5281/zenodo.23045722.

Funding

This study is part of the *HISPAPHYLO* project (ref. PID2021-124234NA-I00), funded by MCIN/AEI/10.13039/501100011033/ERDF, EU. Additional support was provided by the *DAHPNE* project (ref. PID2024-158450NB-I00), funded by MICIU/AEI/10.13039/501100011033 and by ESF+. AJC was also supported by the

National Museum of Natural History (Smithsonian Institution) through a Peter Buck Postdoctoral Fellowship 2023.

Author Contribution

Conceptualization: all authors; Data curation: AJC, IRG; Formal analysis: AJC, IRG, RMV; Funding acquisition: MFM; Investigation: AJC, IRG, MFM; Supervision: MFM; Visualization: AJC; Writing – original draft: AJC, MFM; Writing – review & editing: all authors.

Acknowledgement

The authors thank Virginia Valcárcel, Nagore G. Medina and Andrew Thornhill for their support and useful comments; and the Center for Scientific Computation at Universidad Autónoma de Madrid (CCC-UAM) for providing computing resources and support.

Data Availability

All datasets used in this study can be found in the Zenodo repository: https://doi.org/10.5281/zenodo.23045722.

References

1. Ceballos G, Ehrlich PR, Barnosky AD, García A, Pringle RM, Palmer TM. Accelerated modern human–induced species losses: Entering the sixth mass extinction. Sci Adv. 2015;1:e1400253. https://doi.org/10.1126/sciadv.1400253.
2. Cowie RH, Bouchet P, Fontaine B. The Sixth Mass Extinction: fact, fiction or speculation? Biol Rev. 2022;97:640–63. https://doi.org/10.1111/brv.12816.
3. McNeely JA, Miller KR, Reid WV, Mittermeier RA, Werner TB. Conserving the world's biological diversity. Gland and Washington, DC: IUCN, WRI, CI, WWF-US, and the World Bank; 1990.
4. Fleishman E, Noss RF, Noon BR. Utility and limitations of species richness metrics for conservation planning. Ecol Indic. 2006;6:543–53. https://doi.org/10.1016/j.ecolind.2005.07.005.
5. Vargas P, Jiménez-Mejías P, Fernández-Mazuecos M. Endangered living fossils' (ELFs): Long-term survivors through periods of dramatic climate change. Environ Exp Bot. 2020;170:103892. https://doi.org/10.1016/j.envexpbot.2019.103892.
6. Vane-Wright RI, Humphries CJ, Williams PH. What to protect?—Systematics and the agony of choice. Biol Conserv. 1991;55:235–54. https://doi.org/10.1016/0006-3207(91)90030-D.
7. Faith DP. Conservation evaluation and phylogenetic diversity. Biol Conserv. 1992;61:1–10. https://doi.org/10.1016/0006-3207(92)91201-3.
8. Rosauer D, Laffan SW, Crisp MD, Donnellan SC, Cook LG. Phylogenetic endemism: a new approach for identifying geographical concentrations of evolutionary history. Mol Ecol. 2009;18:4061–72. https://doi.org/10.1111/j.1365-294X.2009.04311.x.
9. Favarger C, Contandriopoulos J. Essai sur l'endémisme. Ber Schweiz Bot Ges. 1961;71:384–408.
10. Mishler BD. Spatial phylogenetics. J Biogeogr. 2023;50:1454–63. https://doi.org/10.1111/jbi.14618.
11. Mishler BD, Knerr N, González-Orozco CE, Thornhill AH, Laffan SW, Miller JT. Phylogenetic measures of biodiversity and neo- and paleo-endemism in Australian *Acacia*. Nat Commun. 2014;5:4473. https://doi.org/10.1038/ncomms5473.
12. Wulff EV, Waltham. Mass: Chronica Botanica Co; 1943. https://doi.org/10.5962/bhl.title.6284.
13. Stebbins GL, Major J. Endemism and speciation in the California flora. Ecol Monogr. 1965;35:1–35. https://doi.org/10.2307/1942216.
14. Nitta JH, Mishler BD, Iwasaki W, Ebihara A. Spatial phylogenetics of Japanese ferns: Patterns, processes, and implications for conservation. Am J Bot. 2022;109:727–45. https://doi.org/10.1002/ajb2.1848.
15. Omollo WO, Rabarijaona RN, Ranaivoson RM, Rakotoarinivo M, Barrett RL, Zhang Q, et al. Spatial heterogeneity of neo- and paleo-endemism for plants in Madagascar. Curr Biol. 2024;34:1271–e12834. https://doi.org/10.1016/j.cub.2024.02.023.
16. Zhang X-X, Ye J-F, Laffan SW, Mishler BD, Thornhill AH, Lu L-M, et al. Spatial phylogenetics of the Chinese angiosperm flora provides insights into endemism and conservation. J Integr Plant Biol. 2022;64:105–17. https://doi.org/10.1111/jipb.13189.
17. Smith SA, Brown JW. Constructing a broadly inclusive seed plant phylogeny. Am J Bot. 2018;105:302–14. https://doi.org/10.1002/ajb2.1019.
18. Sabatini FM, Jiménez-Alfaro B, Jandt U, Chytrý M, Field R, Kessler M, et al. Global patterns of vascular plant alpha diversity. Nat Commun. 2022;13:4683. https://doi.org/10.1038/s41467-022-32063-z.
19. Valente LM, Savolainen V, Vargas P. Unparalleled rates of species diversification in Europe. Proc R Soc B: Biol Sci. 2010;277:1489–96. https://doi.org/10.1098/rspb.2009.2163.
20. Myers N, Mittermeier RA, Mittermeier CG, da Fonseca GAB, Kent J. Biodiversity hotspots for conservation priorities. Nature. 2000;403:853–8. https://doi.org/10.1038/35002501.
21. Cai L, Kreft H, Taylor A, Schrader J, Dawson W, Essl F, et al. Climatic stability and geological history shape global centers of neo- and paleoendemism in seed plants. Proc Natl Acad Sci USA. 2023;120:e2300981120. https://doi.org/10.1073/pnas.2300981120.
22. Qian H, Kessler M, Qian S, Zhang J. Patterns and drivers of taxonomic and phylogenetic endemism in regional fern floras across the world. Biol Conserv. 2024;291:110506. https://doi.org/10.1016/j.biocon.2024.110506.

23. Qian H, Mishler BD, Zhang J, Qian S. Global patterns and ecological drivers of taxonomic and phylogenetic endemism in angiosperm genera. Plant Divers. 2024;46:149–57. https://doi.org/10.1016/j.pld.2023.11.004.
24. Castroviejo S, editor. Flora iberica. Madrid: Real Jardín Botánico, CSIC; 1986–2021.
25. Ramos-Gutiérrez I, Lima H, Pajarón S, Romero-Zarco C, Sáez L, Pataro L, et al. Atlas of the vascular flora of the Iberian Peninsula biodiversity hotspot (AFLIBER). Glob Ecol Biogeogr. 2021;30:1951–7. https://doi.org/10.1111/geb.13363.
26. Aedo C, Buira A, Medina L, Fernández-Albert M. The Iberian vascular flora: Richness, endemicity and distribution patterns. In: Loidi J, editor. The vegetation of the Iberian Peninsula. Volume 1. Cham: Springer; 2017. pp. 101–30. https://doi.org/10.1007/978-3-319-54784-8_4.
27. Médail F, Diadema K. Glacial refugia influence plant diversity patterns in the Mediterranean Basin. J Biogeogr. 2009;36:1333–45. https://doi.org/10.1111/j.1365-2699.2008.02051.x.
28. Médail F, Quezel P. Hot-spots analysis for conservation of plant biodiversity in the Mediterranean Basin. Ann Missouri Bot Gard. 1997;84:112–27. https://doi.org/10.2307/2399957.
29. Buira A, Aedo C, Medina L. Spatial patterns of the Iberian and Balearic endemic vascular flora. Biodivers Conserv. 2017;26:479–508. https://doi.org/10.1007/s10531-016-1254-z.
30. Lobo JM, Castro I, Moreno JC. Spatial and environmental determinants of vascular plant species richness distribution in the Iberian Peninsula and Balearic Islands. Biol J Linn Soc. 2001;73:233–53. https://doi.org/10.1111/j.1095-8312.2001.tb01360.x.
31. Postigo Mijarra JM, Barrón E, Gómez Manzaneque F, Morla C. Floristic changes in the Iberian Peninsula and Balearic Islands (south-west Europe) during the Cenozoic. J Biogeogr. 2009;36:2025–43. https://doi.org/10.1111/j.1365-2699.2009.02142.x.
32. Jin Y, Qian H. V.PhyloMaker2: An updated and enlarged R package that can generate very large phylogenies for vascular plants. Plant Divers. 2022;44:335–9. https://doi.org/10.1016/j.pld.2022.05.005.
33. Ramos-Gutiérrez I, Lima H, Vilela B, Molina-Venegas R. A generalized framework to expand incomplete phylogenies using non-molecular phylogenetic information. Glob Ecol Biogeogr. 2023;32:1707–16. https://doi.org/10.1111/geb.13733.
34. Gómez-Campo C, Bermúdez-de-Castro L, Cagiga MJ, Sánchez-Yélamo MD. Endemism in the Iberian Peninsula and Balearic Islands. Webbia. 1984;38:709–14. https://doi.org/10.1080/00837792.1984.10670341.
35. Parga IC, Moreno Saiz JC, Humphries CJ, Williams PH. Strengthening the Natural and National Park system of Iberia to conserve vascular plants. Bot J Linn Soc. 1996;121:189–206. https://doi.org/10.1111/j.1095-8339.1996.tb00753.x.
36. Domínguez Lozano F, Galicia Herbada D, Moreno Rivero L, Moreno Saiz JC, Sainz Ollero H. Areas of high floristic endemism in Iberia and the Balearic islands: an approach to biodiversity conservation using narrow endemics. Belg J Entomol. 2000;2:171–85.
37. Buira A, Cabezas F, Aedo C. Disentangling ecological traits related to plant endemism, rarity and conservation status in the Iberian Peninsula. Biodivers Conserv. 2020;29:1937–58. https://doi.org/10.1007/s10531-020-01957-z.
38. Sainz Ollero H, Moreno Saiz JC. Flora vascular endémica española. In: Pineda FD, de Miguel JM, Casado MA, Montalvo J, editors. La diversidad biológica de España. Madrid: Pearson Educación; 2002. pp. 175–95.
39. Moreno Saiz JC, Donato M, Katinas L, Crisci JV, Posadas P. New insights into the biogeography of south-western Europe: spatial patterns from vascular plants using cluster analysis and parsimony. J Biogeogr. 2013;40:90–104. https://doi.org/10.1111/j.1365-2699.2012.02774.x.
40. Buira A, Fernández-Mazuecos M, Aedo C, Molina-Venegas R. The contribution of the edaphic factor as a driver of recent plant diversification in a Mediterranean biodiversity hotspot. J Ecol. 2021;109:987–99. https://doi.org/10.1111/1365-2745.13527.
41. Earl C, Belitz MW, Laffan SW, Barve V, Barve N, Soltis DE, et al. Spatial phylogenetics of butterflies in relation to environmental drivers and angiosperm diversity across North America. iScience. 2021;24:102239. https://doi.org/10.1016/j.isci.2021.102239.
42. Harrison S, Noss R. Endemism hotspots are linked to stable climatic refugia. Ann Bot. 2017;119:207–14. https://doi.org/10.1093/aob/mcw248.
43. Sàbat F, Gelabert B, Rodríguez-Perea A, Giménez J. Geological structure and evolution of Majorca: Implications for the origin of the Western Mediterranean. Tectonophysics. 2011;510:217–38. https://doi.org/10.1016/j.tecto.2011.07.005.
44. Molero J, Marfil JM. Betic and Southwest Andalusia. In: Loidi J, editor. The vegetation of the Iberian Peninsula. Volume 2. Cham: Springer; 2017. pp. 143–247. https://doi.org/10.1007/978-3-319-54867-8_4.
45. Hewitt GM. Some genetic consequences of ice ages, and their role in divergence and speciation. Bot J Linn Soc. 1996;58:247–76. https://doi.org/10.1111/j.1095-8312.1996.tb01434.x.
46. Rivas-Martínez S, Penas Á, Díaz González TE, Cantó P, del Río S, Costa JC, et al. Biogeographic units of the Iberian Peninsula and Baelaric Islands to district level. A concise synopsis. In: Loidi J, editor. The vegetation of the Iberian Peninsula. Volume 1. Cham: Springer; 2017. pp. 131–88. https://doi.org/10.1007/978-3-319-54784-8_5.
47. Lionello P, Malanotte-Rizzoli P, Boscolo R, Alpert P, Artale V, Li L, et al. The Mediterranean climate: An overview of the main characteristics and issues. In: Lionello P, Malanotte-Rizzoli P, Boscolo R, editors. Developments in Earth and environmental sciences. Volume 4. Chantilly: Elsevier; 2006. pp. 1–26. https://doi.org/10.1016/S1571-9197(06)80003-0.
48. Govaerts R, Nic Lughadha E, Black N, Turner R, Paton A. The World Checklist of Vascular Plants, a continuously updated resource for exploring global plant diversity. Sci Data. 2021;8:215. https://doi.org/10.1038/s41597-021-00997-6.
49. Nitta JH, Laffan SW, Mishler BD, Iwasaki W. canaper: Categorical analysis of neo- and paleo-endemism in R. Ecography. 2023;2023:e06638. https://doi.org/10.1111/ecog.06638.
50. Strona G, Nappo D, Boccacci F, Fattorini S, San-Miguel-Ayanz J. A fast and unbiased procedure to randomize ecological binary matrices with fixed row and column totals. Nat Commun. 2014;5:4114. https://doi.org/10.1038/ncomms5114.

51. Thornhill AH, Baldwin BG, Freyman WA, Nosratinia S, Kling MM, Morueta-Holme N, et al. Spatial phylogenetics of the native California flora. BMC Biol. 2017;15:96. https://doi.org/10.1186/s12915-017-0435-x.
52. Kottek M, Grieser J, Beck C, Rudolf B, Rubel F. World Map of the Köppen-Geiger climate classification updated. Meteorol Z. 2006;15:259–63. https://doi.org/10.1127/0941-2948/2006/0130.
53. Karger DN, Conrad O, Böhner J, Kawohl T, Kreft H, Soria-Auza RW, et al. Climatologies at high resolution for the earth's land surface areas. Sci Data. 2017;4:170122. https://doi.org/10.1038/sdata.2017.122.
54. Title PO, Bemmels JB. ENVIREM: An expanded set of bioclimatic and topographic variables increases flexibility and improves performance of ecological niche modeling. Ecography. 2018;41:291–307. https://doi.org/10.1111/ecog.02880.
55. Hengl T, de Jesus JM, Heuvelink GBM, Gonzalez MR, Kilibarda M, Blagotić A, et al. SoilGrids250m: Global gridded soil information based on machine learning. PLoS ONE. 2017;12:e0169748. https://doi.org/10.1371/journal.pone.0169748.
56. Amatulli G, Domisch S, Tuanmu M-N, Parmentier B, Ranipeta A, Malczyk J, et al. A suite of global, cross-scale topographic variables for environmental and biodiversity modeling. Sci Data. 2018;5:180040. https://doi.org/10.1038/sdata.2018.40.
57. Venables WN, Ripley BD. Modern applied statistics with S. New York: Springer; 2002.
58. Forest F, Grenyer R, Rouget M, Davies TJ, Cowling RM, Faith DP, et al. Preserving the evolutionary potential of floras in biodiversity hotspots. Nature. 2007;445:757–60. https://doi.org/10.1038/nature05587.
59. Padullés Cubino J, Lososová Z, Bonari G, Agrillo E, Attorre F, Bergmeier E, et al. Phylogenetic structure of European forest vegetation. J Biogeogr. 2021;48:903–16. https://doi.org/10.1111/jbi.14046.
60. Suc J-P. Origin and evolution of the Mediterranean vegetation and climate in Europe. Nature. 1984;307:429–32. https://doi.org/10.1038/307429a0.
61. Rundel PW, Arroyo MTK, Cowling RM, Keeley JE, Lamont BB, Vargas P. Mediterranean biomes: Evolution of their vegetation, floras, and climate. Annu Rev Ecol Evol Syst. 2018;47:383–407. https://doi.org/10.1146/annurev-ecolsys-121415-032330.
62. Rundel PW, Arroyo MTK, Cowling RM, Keeley JE, Lamont BB, Pausas JG, et al. Fire and plant diversification in Mediterranean-climate regions. Front Plant Sci. 2018;9:851. https://doi.org/10.3389/fpls.2018.00851.
63. Vargas P. The Mediterranean floristic region: High diversity of plants and vegetation types. In: Goldstein MI, DellaSala DA, editors. Encyclopedia of the world's biomes. Volume 3. Oxford: Elsevier; 2020. pp. 602–16. https://doi.org/10.1016/B978-0-12-409548-9.12097-4.
64. Verdú M, Pausas JG. Fire drives phylogenetic clustering in Mediterranean Basin woody plant communities. J Ecol. 2007;95:1316–23. https://doi.org/10.1111/j.1365-2745.2007.01300.x.
65. Santana VM, Alday JG, Adamo I, Alloza JA, Baeza MJ. Climate, and not fire, drives the phylogenetic clustering of species with hard-coated seeds in Mediterranean Basin communities. Perspect Plant Ecol Evol Syst. 2020;45:125545. https://doi.org/10.1016/j.ppees.2020.125545.
66. Thornhill AH, Mishler BD, Knerr NJ, González-Orozco CE, Costion CM, Crayn DM, et al. Continental-scale spatial phylogenetics of Australian angiosperms provides insights into ecology, evolution and conservation. J Biogeogr. 2016;43:2085–98. https://doi.org/10.1111/jbi.12797.
67. Loidi J, Campos JA, Herrera M, Biurrun I, García-Mijangos I, García-Baquero G. Eco-geographical factors affecting richness and phylogenetic diversity patterns of high-mountain flora in the Iberian Peninsula. Alp Bot. 2015;125:137–46. https://doi.org/10.1007/s00035-015-0149-z.
68. Loidi J, Biurrun I, García-Mijangos I, García-Baquero G, Herrera M, Campos JA. The high mountain flora and vegetation. In: Loidi J, editor. The vegetation of the Iberian Peninsula. Volume 2. Cham: Springer; 2017. pp. 433–58. https://doi.org/10.1007/978-3-319-54867-8_9.
69. Gómez A, Lunt DH. Refugia within refugia: Patterns of phylogeographic concordance in the Iberian Peninsula. In: Weiss S, Ferrand N, editors. Phylogeography of southern European refugia. Dordrecht: Springer; 2007. pp. 155–88.
70. Aedo C, Medina L, Fernández-Albert M. Species richness and endemicity in the Spanish vascular flora. Nord J Bot. 2013;31:478–88. https://doi.org/10.1111/j.1756-1051.2012.00009.x.
71. Vilà M, Muñoz I. Patterns and correlates of exotic and endemic plant taxa in the Balearic islands. Ecol Mediterr. 1999;25:153–61. https://doi.org/10.3406/ecmed.1999.1880.
72. Whittaker RJ, Fernández-Palacios JM. Island biogeography: Ecology, evolution, and conservation. Oxford: Oxford University Press; 2007.
73. Llorens L, Gil L. The Balearic Islands. In: Loidi J, editor. The vegetation of the Iberian Peninsula. Volume 2. Cham: Springer; 2017. pp. 3–33. https://doi.org/10.1007/978-3-319-54867-8_1.
74. Médail F. Plant biogeography and vegetation patterns of the Mediterranean islands. Bot Rev. 2022;88:63–129. https://doi.org/10.1007/s12229-021-09245-3.
75. Fraile-Jurado P, Mejías-García JC, Roldán-Muñoz ME, Borja-Barrera C. Reconstructing the emerged areas of the Mediterranean during the Last Glacial Period: Incorporating bathymetric data and sea level fluctuations. Geogr Fis Din Quat. 2023;46:241–56. https://doi.org/10.4454/hrffie9.
76. Bezares L, Fernández-Mazuecos M, Cano E, Borràs J, Cerca J, Cursach J, et al. Ibero-Balearic plant disjunctions: genomic support for rapid and recurrent long-distance colonizations of the endangered *Diplotaxis ibicensis* (Brassicaceae) despite no dispersal syndromes. Plant Biol. 2026;28:1560–72. https://doi.org/10.1111/plb.70227.
77. Carnicero P, Schönswetter P, Fraga Arguimbau P, Garcia-Jacas N, Sáez L, Galbany-Casals M. Phylogeography of western Mediterranean *Cymbalaria* (Plantaginaceae) reveals two independent long-distance dispersals and entails new taxonomic circumscriptions. Sci Rep. 2018;8:18079. https://doi.org/10.1038/s41598-018-36412-1.
78. Morlon H, Potts MD, Plotkin JB. Inferring the dynamics of diversification: A coalescent approach. PLoS Biol. 2010;8:e1000493. https://doi.org/10.1371/journal.pbio.1000493.

79. Guardiola M, Sáez L. Are Mediterranean island mountains hotspots of taxonomic and phylogenetic biodiversity? The case of the endemic flora of the Balearic Islands. Plants. 2023;12:2640. https://doi.org/10.3390/plants12142640.
80. Braga JC, Martín JM, Quesada C. Patterns and average rates of late Neogene–Recent uplift of the Betic Cordillera, SE Spain. Geomorphology. 2003;50:3–26. https://doi.org/10.1016/S0169-555X(02)00205-2.
81. Mota JF, Pérez-García FJ, Jiménez ML, Amate JJ, Peñas J. Phytogeographical relationships among high mountain areas in the Baetic Ranges (South Spain). Global Ecol Biogeogr. 2002;11:497–504. https://doi.org/10.1046/j.1466-822X.2002.00312.x.
82. Mota JF, Garrido-Becerra JA, Merlo ME, Medina-Cazorla JM, Sánchez-Gómez P. The edaphism: Gypsum, dolomite and serpentine flora and vegetation. In: Loidi J, editor. The vegetation of the Iberian Peninsula. Volume 2. Cham: Springer; 2017. pp. 277–354. https://doi.org/10.1007/978-3-319-54867-8_6.
83. Rajakaruna N. The edaphic factor in the origin of plant species. Int Geol Rev. 2004;46:471–8. https://doi.org/10.2747/0020-6814.46.5.471.
84. Rajakaruna N. Lessons on evolution from the study of edaphic specialization. Bot Rev. 2018;84:39–78. https://doi.org/10.1007/s12229-017-9193-2.
85. Peñas J, Pérez-García FJ, Mota JF. Patterns of endemic plants and biogeography of the Baetic high mountains (south Spain). Acta Bot Gallica. 2005;152:347–60. https://doi.org/10.1080/12538078.2005.10515494.
86. Medina-Cazorla JM, Garrido-Becerra JA, Mendoza Fernández A, Pérez-García FJ, Salmerón E, Gil C, et al. Biogeography of the Baetic ranges (SE Spain): A historical approach using cluster and parsimony analyses of endemic dolomitophytes. Plant Biosyst. 2010;144:111–20. https://doi.org/10.1080/11263500903367983.
87. Pérez-Latorre AV, Hidalgo-Triana N, Cabezudo B. Composition, ecology and conservation of the south-Iberian serpentine flora in the context of the Mediterranean basin. Jard Bot Madr. 2013;70:62–71. https://doi.org/10.3989/ajbm.2334.
88. Pérez-Latorre AV, Keen N, Casimiro-Soriguer F, Goncalves E, Hidalgo-Triana N. An updated checklist of serpentinophytes for research and conservation in ultramafic ecosystems on the southern Iberian Peninsula (Spain). Ecol Res. 2024;39:543–62. https://doi.org/10.1111/1440-1703.12478.
89. Safford HD, Viers JH, Harrison SP. Serpentine endemism in the California flora: A database of serpentine affinity. Madroño. 2005;52:222–57. https://doi.org/10.3120/0024-9637(2005)52%5B222:SEITCF%5D2.0.CO;2.
90. Anacker BL, Whittall JB, Goldberg EE, Harrison SP. Origins and consequences of serpentine endemism in the California flora. Evolution. 2011;65:365–76. https://doi.org/10.1111/j.1558-5646.2010.01114.x.
91. Silva-Pando FJ. Las plantas endémicas y subendémicas de Galicia. Bol BIGA. 2008;3:9–150.
92. Escudero A, Palacio S, Maestre FT, Luzuriaga AL. Plant life on gypsum: A review of its multiple facets. Biol Rev. 2015;90:1–18. https://doi.org/10.1111/brv.12092.
93. Palacio S, Cera A, Escudero A, Luzuriaga AL, Sánchez AM, Mota JF, et al. Recent and ancient evolutionary events shaped plant elemental composition of edaphic endemics: a phylogeny-wide analysis of Iberian gypsum plants. New Phytol. 2022;235:2406–23. https://doi.org/10.1111/nph.18309.
94. Martínez-Hernández F, Medina-Cazorla JM, Mendoza-Fernández A, Pérez-García FJ, Sánchez-Gómez P, Garrido-Becerra JA, et al. Preliminary essay on the chorology of the Iberian gypsicolous flora: rarity and richness of the gypsum outcrops. Acta Bot Gallica. 2009;156:9–18. https://doi.org/10.1080/12538078.2009.10516138.
95. Martínez-Hernández F, Pérez-García FJ, Garrido-Becerra JA, Mendoza-Fernández AJ, Medina-Cazorla JM, Martínez-Nieto MI, et al. The distribution of Iberian gypsophilous flora as a criterion for conservation policy. Biodivers Conserv. 2011;20:1353–64. https://doi.org/10.1007/s10531-011-0031-2.
96. Abellán P, Svenning J-C. Refugia within refugia – patterns in endemism and genetic divergence are linked to Late Quaternary climate stability in the Iberian Peninsula. Bot J Linn Soc. 2014;113:13–28. https://doi.org/10.1111/bij.12309.
97. Moreno JM, Pineda FD, Rivas-Martinez S. Climate and vegetation at the Eurosiberian-Mediterranean boundary in the Iberian Peninsula. J Veg Sci. 1990;1:233–44. https://doi.org/10.2307/3235660.
98. Mendoza-Fernández A, Pérez-García FJ, Martínez-Hernández F, Medina-Cazorla JM, Garrido-Becerra JA, Calvente MEM, et al. Threatened plants of arid ecosystems in the Mediterranean Basin: A case study of the south-eastern Iberian Peninsula. Oryx. 2014;48:548–54. https://doi.org/10.1017/S0030605313000495.
99. Molina JA. Aquatic and wetland vegetation of the Iberian Peninsula. In: Loidi J, editor. The vegetation of the Iberian Peninsula. Volume 2. Cham: Springer; 2017. pp. 355–96. https://doi.org/10.1007/978-3-319-54867-8_7.
100. Salazar-Mendías C, Lendínez ML. Mediterranean halophytic flora and vegetation in the Iberian Peninsula (Spain and Portugal). In: Grigore M-N, editor. Handbook of halophytes. Cham: Springer; 2020. pp. 1–44. https://doi.org/10.1007/978-3-030-17854-3_12-1.
101. Jiménez-Alfaro B, Carlón L, Fernández-Pascual E, Acedo C, Alfaro-Saiz E, Alonso Redondo R, et al. Checklist of the vascular plants of the Cantabrian Mountains. Mediterr Bot. 2021;42:e74570. https://doi.org/10.5209/mbot.74570.
102. Molina-Venegas R, Aparicio A, Lavergne S, Arroyo J. Climatic and topographical correlates of plant palaeo- and neoendemism in a Mediterranean biodiversity hotspot. Ann Bot. 2017;119:229–38. https://doi.org/10.1093/aob/mcw093.
103. Ninot JM, Carrillo E, Ferré A. The Pyrenees. In: Loidi J, editor. The vegetation of the Iberian Peninsula. Volume 1. Cham: Springer; 2017. pp. 323–66. https://doi.org/10.1007/978-3-319-54784-8_8.
104. Heywood VH. The jewel in the crown. Plant Talk. 2003;32:42–3.
105. Pironon S, Gómez D, Font X, García MB. Living at the limit in the Pyrenees: Peripheral and endemic plants are rare but underrepresented in protection lists. Divers Distrib. 2022;28:930–42. https://doi.org/10.1111/ddi.13487.
106. Mishler BD, Guralnick R, Soltis PS, Smith SA, Soltis DE, Barve N, et al. Spatial phylogenetics of the North American flora. J Syst Evol. 2020;58:393–405. https://doi.org/10.1111/jse.12590.

107. Hojsgaard D, Hörandl E. Apomixis as a facilitator of range expansion and diversification in plants. In: Pontarotti P, editor. Evolutionary biology: Biodiversification from genotype to phenotype. Cham: Springer; 2015. pp. 305–27. https://doi.org/10.1007/978-3-319-19932-0_16.
108. Miller JT, Prentice E, Bui EN, Knerr N, Mishler BD, Schmidt-Lebuhn AN, et al. *Banksia* (Proteaceae) contains less phylogenetic diversity than expected in Southwestern Australia. J Syst Evol. 2023;61:957–66. https://doi.org/10.1111/jse.13019.
109. Scherson RA, Thornhill AH, Urbina-Casanova R, Freyman WA, Pliscoff PA, Mishler BD. Spatial phylogenetics of the vascular flora of Chile. Mol Phylogenet Evol. 2017;112:88–95. https://doi.org/10.1016/j.ympev.2017.04.021.
110. Allen JM, Germain-Aubrey CC, Barve N, Neubig KM, Majure LC, Laffan SW, et al. Spatial phylogenetics of Florida vascular plants: The effects of calibration and uncertainty on diversity estimates. iScience. 2019;11:57–70. https://doi.org/10.1016/j.isci.2018.12.002.
111. Vargas P, García B. Plant endemics to Sierra de Gredos (central Spain): taxonomic, distributional, and evolutionary aspects. Jard Bot Madr. 2008;65:353–66. https://doi.org/10.3989/ajbm.2008.v65.i2.298.
112. Rangel TF, Colwell RK, Graves GR, Fučíková K, Rahbek C, Diniz-Filho JAF. Phylogenetic uncertainty revisited: Implications for ecological analyses. Evolution. 2015;69:1301–12. https://doi.org/10.1111/evo.12644.
113. Ramos-Gutiérrez I, Pipins S, Molina-Venegas R, Fernández-Mazuecos M, Jiménez-Mejías P, Moreno-Saiz JC, et al. Priority areas for preserving angiosperm evolutionary history in the Iberian Peninsula. Divers Distrib. 2026;32:e70224. https://doi.org/10.1111/ddi.70224.
114. Fick SE, Hijmans RJ. WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. Int J Climatol. 2017;37:4302–15. https://doi.org/10.1002/joc.5086.

Figures

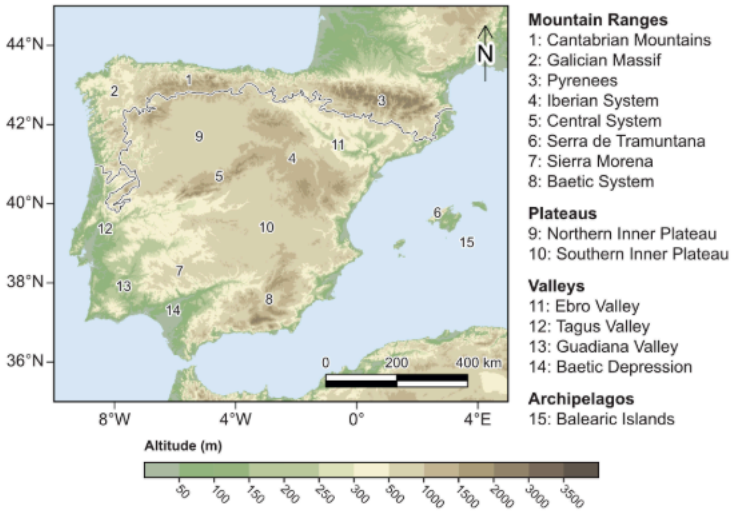


Figure 1

Topographic map of the Iberian Peninsula (color gradient from lower to higher altitudes: green, yellow and brown) based on topographic data from WorldClim 2.1 [114]. Numbers indicate the main geomorphological units. The black line indicates the approximate limit between the Eurosiberian floristic region (north) and the Mediterranean floristic region (south) based on Rivas-Martínez et al. [46].

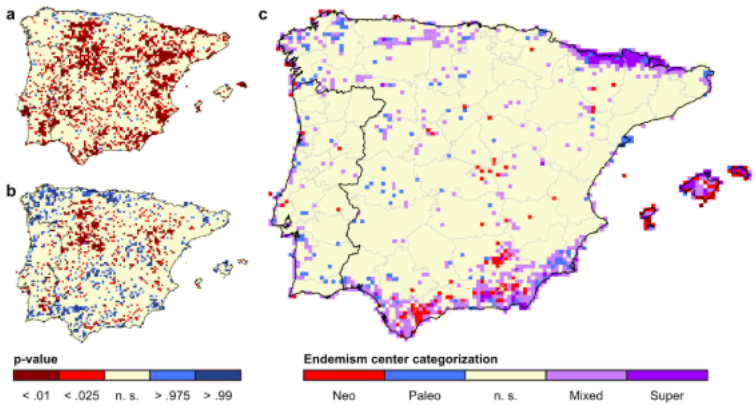


Figure 2

Spatial phylogenetic patterns for the Iberian angiosperm flora based on the *core dataset* (see Table 1), including (a) phylogenetic diversity (PD) significance, (b) relative phylogenetic diversity (RPD) significance, and (c) endemism center categorization according to Categorical Analysis of Neo- And Paleo-

Endemism (CANAPE). Significant grid cells (a, b) are colored according to *p*-values: $p < 0.01$ in dark red, $p < 0.025$ in red, $p > 0.975$ in blue and $p > 0.99$ in dark blue. Centers of endemism in c are colored according to the endemism categorization: neoendemism in red, paleoendemism in blue, mixed endemism in light purple and superendemism in dark purple. Non-significant grid cells are pale yellow. Continuous black lines indicate country borders and light grey lines indicate administrative divisions in Spain and Portugal.

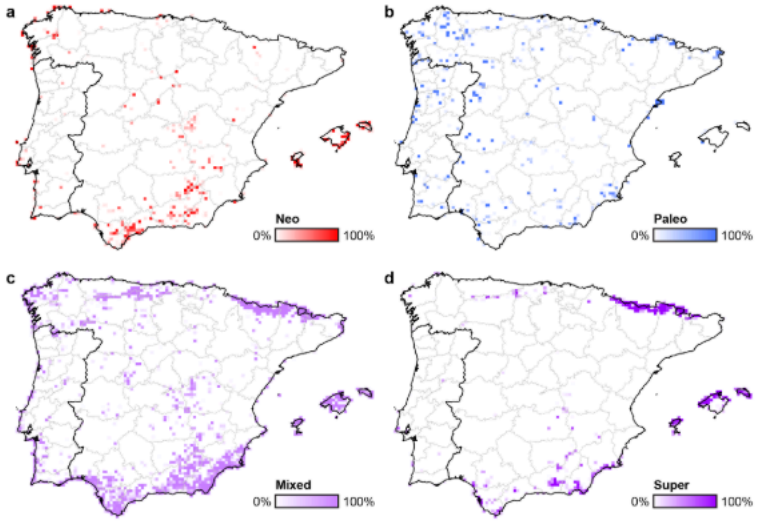


Figure 3

Endemism center categorization (CANAPE) for the Iberian angiosperm flora based on the *comprehensive dataset* (see Table 1) and accounting for uncertainty in the placement of unsampled species (100 *randtip* replicates). For every grid cell, maps represent the percentage of replicates in which each endemism category was recovered: (a) neoendemism, (b) paleoendemism, (c) mixed endemism, and (d) superendemism. Continuous black lines indicate country borders and light grey lines indicate administrative divisions in Spain and Portugal.

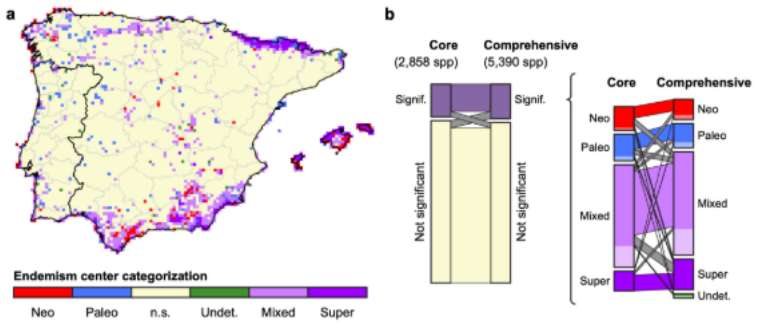


Figure 4

(a) Majority-rule consensus map resulting from the endemism center categorization (CANAPE) for the Iberian angiosperm flora based on the *comprehensive dataset* (see Table 1) and accounting for uncertainty in the placement of unsampled species (100 *randtip*replicates). Each grid cell is colored according to the endemism category (neoendemism in red, paleoendemism in blue, mixed endemism in light purple, superendemism in dark purple, undetermined endemism in green and not significant in pale yellow) recovered for > 50% of replicates (see Materials and methods for details). Continuous black lines indicate country borders and light grey lines indicate administrative divisions in Spain and Portugal. (b) Comparison of endemism categorization between the *core dataset* and the *comprehensive dataset* (see Table 1) for the significant and non-significant grid cells (left) and for only significant grid cells (right; those grid cells only significant for one of those datasets lack connections and display a lighter color). The comparison matrix between the *core dataset* and the *comprehensive dataset* is shown in Table S1 in Supporting information.

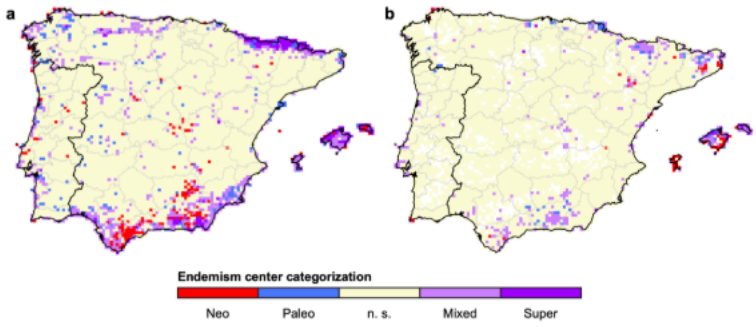


Figure 5

Endemism categorization (CANAPE) for the Iberian angiosperm flora according to (a) the *apomict-excluded dataset*and (b) the *strictly-endemic dataset* (see Materials and methods and Table 1 for dataset details). Centers of endemism are colored according to the endemism categorization: neoendemism in red, paleoendemism in blue, mixed endemism in light purple and superendemism in dark purple. Non-significant grid cells are pale yellow. Continuous black lines indicate country borders and light grey lines indicate administrative divisions in Spain and Portugal.

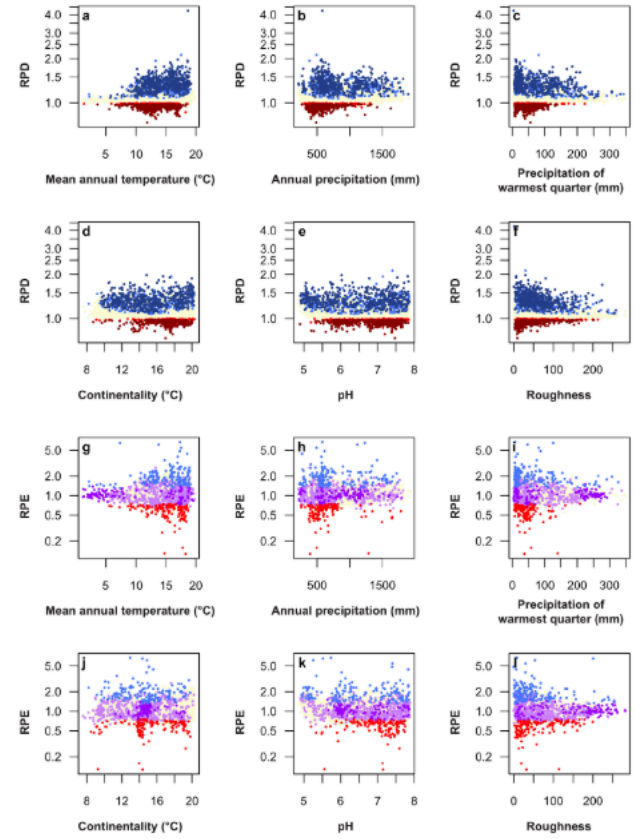


Figure 6

Scatter plots showing the relationships between spatial phylogenetic metrics for the Iberian angiosperm flora and abiotic variables. Values of relative phylogenetic diversity (RPD; a – f) are categorized according to their significance ($p < 0.01$ in dark red, $p < 0.025$ in red, $p > 0.975$ in blue, $p > 0.99$ in dark blue, not significant in pale yellow). Values of relative phylogenetic endemism (RPE; g – l) are categorized according to the endemism center categorization (neoendemism in red, paleoendemism in blue, mixed endemism in light purple, superendemism in dark purple).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supportinginformation.docx