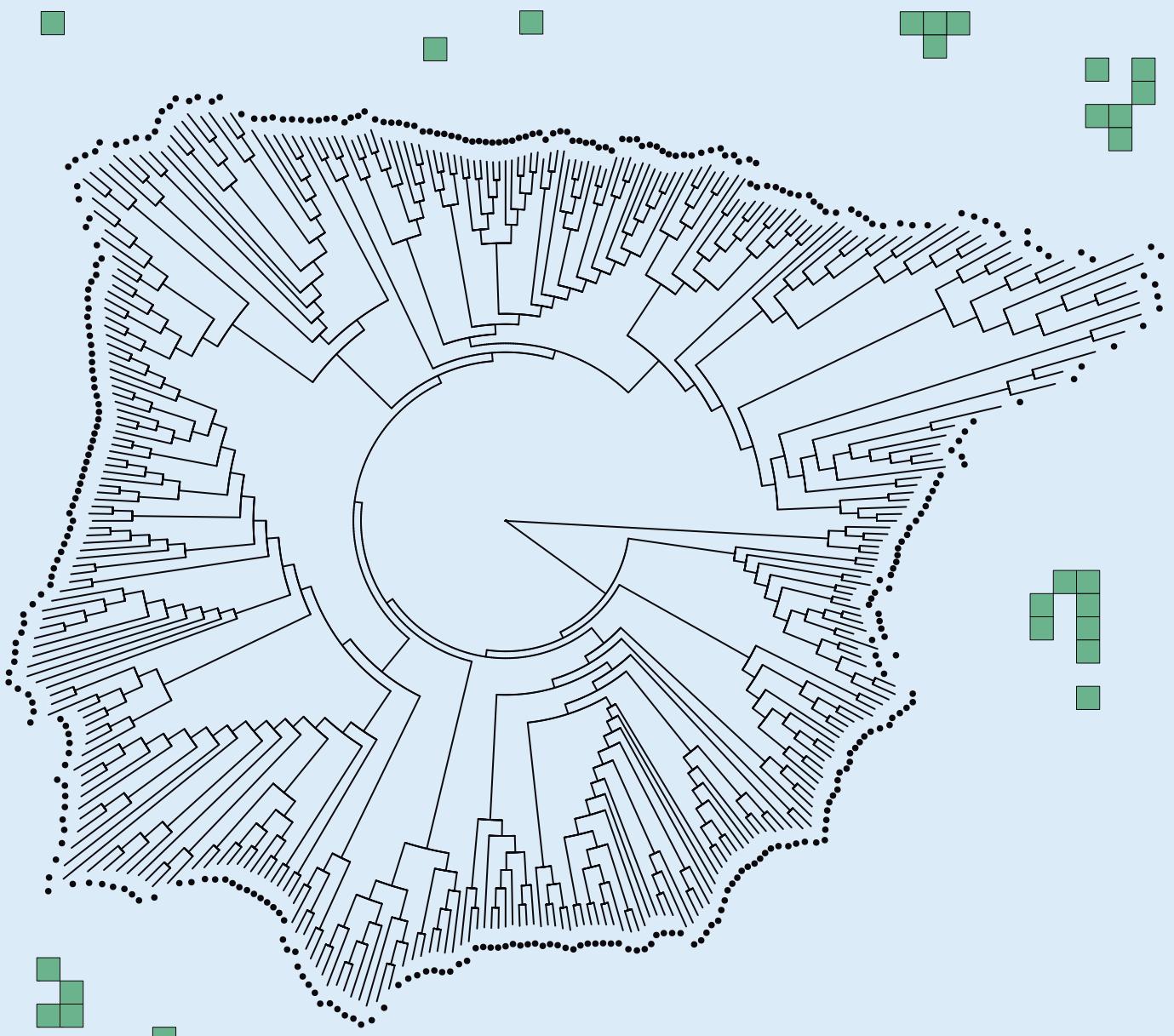


New insights on Spatial and Evolutionary patterns in the Iberian vascular flora: novel methodologies and applications in biogeography and conservation



Ignacio Ramos-Gutiérrez

Madrid, 2024



Universidad Autónoma
de Madrid

Escuela
de Doctorado

New insights on Spatial and Evolutionary patterns in the Iberian vascular flora: novel methodologies and applications in biogeography and conservation

Programa de Doctorado en Ecología

DOCTORAL THESIS

Submitted for the degree of Doctor by:

Ignacio Ramos-Gutiérrez

Under the supervision of:

Dr. Juan Carlos Moreno Saiz

Dr. Rafael Molina Venegas

Madrid, 2024

UNIVERSIDAD AUTÓNOMA DE MADRID
Programa de Doctorado en Ecología



**New insights on Spatial and Evolutionary patterns in
the Iberian vascular flora: novel methodologies and
applications in biogeography and conservation**

Ignacio Ramos-Gutiérrez

Title:

New insights on Spatial and Evolutionary patterns in the Iberian vascular flora: novel methodologies and applications in biogeography and conservation

Author:

Ignacio Ramos-Gutiérrez

Doctoral Programme:

Programa de Doctorado en Ecología, Universidad Autónoma de Madrid.

Thesis Supervision:

Dr. Juan Carlos Moreno Saiz, Catedrático Universitario, Departamento de Biología - Botánica, Universidad Autónoma de Madrid

Dr. Rafael Molina Venegas, Profesor Ayudante Doctor, Departamento de Ecología, Universidad Autónoma de Madrid

External Reviewers:

Dr. Matilda Brown. Kew Royal Botanic Gardens.

Dr. Alberto J. Coello. Smithsonian Institution, National Museum of Natural History.

Thesis Defense Committee:

Dr. María Begoña García González. Instituto Pirenaico de Ecología - CSIC.

Dr. Virginia Valcárcel Núñez. Universidad Autónoma de Madrid.

Dr. Angelino Carta. Università di Pisa

Dr. Isabel Marques. Universidade de Lisboa.

Dr. Ignacio Morales Castilla. Universidad de Alcalá.

*Some things in life are bad
They can really make you mad
Other things just make you swear and curse
When you're chewing on life's gristle
Don't grumble, give a whistle
And this'll help things turn out for the best...*

A Richi y Lola

Acknowledgements

Empiezo esta primera parte de mi tesis la última de todas, cómo no. Supone un reto personal dejar por escrito a las personas que, de un modo u otro, forman parte de esta tesis, y sin cuya ayuda y apoyo todo hubiera sido diferente. Terminando de escribir veo que todo esto se lo debo a tanta gente que no puedo más que sentirme un enorme privilegiado de haber podido contar con vosotros. Muchas gracias.

En primer lugar, tengo que agradecer enormemente a Juan Carlos y Rafa por haberos embarcado en la dirección de esta tesis. Y quiero agradeceros no solamente haberme dirigido, sino la manera en la que lo habéis hecho. Me he sentido muy apoyado y cuidado a nivel personal durante estos años. Y ha sido mucho también lo que he aprendido de vosotros (y lo que me falta...). Rafa, muchas gracias por todas las lecciones que he aprendido de ti. Gracias por haber estado pendiente de solicitar un contrato cuando se me terminó el mío. Gracias por tu ayuda con los manuscritos y por tenerme siempre como una prioridad. Pero, sobre todo, muchas gracias por haberme introducido en el maravilloso mundo de las filogenias, del que ya no sé si podré salir. Juan Carlos, muchas gracias por haberme acogido en tu laboratorio ya antes de la tesis, con las “Barlias” y las “Nepetas”. Por ofrecerme salir al campo siempre que era posible, y por enseñarme no sólo ciencia, sino una manera de trabajar. A ambos quiero agradecerlos la confianza, el buen trato y ambiente del que he disfrutado durante estos años. Habéis sido para mí los directores que cualquier doctorando hubiera querido pedir.

Muchas gracias también a toda la unidad de Botánica por acogerme de manera tan amigable, y especialmente a Virginia, Julia y José María con quien he tenido la suerte de haber compartido mis primeras experiencias como docente. Muchas gracias a Manuel, por ayudarme siempre de manera tan resolutiva con mis dudas burocráticas; a Juancho por tenerme en cuenta siempre como uno más, especialmente en nuestro pequeño proyecto de los paneles; a Rubén, por todas las facilidades en clase, pero sobre todo por lo que he disfrutado con un compañero *biciclista* (¿con quién iba a hablar de La Vuelta yo si no?), y por las rutas por Navamorcende y Mijares (¡y las que quedan!). Muchas gracias, Belén, por tu apoyo con la colección viva y tu entusiasmo en enseñarnos cosas a los no tan jóvenes (aunque a veces haya que escuchar locuras, como que las algas “molan”). No te perdonó sin embargo que sigas

teniendo una mimosa viva y te rías de la(s) que se me murió(-eron). Muchas gracias a Pedro (aunque desde la lejanía ahora) por las incontables resoluciones de dudas taxonómicas y por tu ayuda con esos hierbajos que son las *Carex*, pero sobre todo por los todos los buenos ratos. Cuántas cosas tendré que agradecerte, Nagore, pero especialmente tu cercanía, facilidad y confianza para cualquier cosa, y todo lo que he disfrutado en las prácticas de *Flora*. Y por supuesto, muchas gracias a Mario por todo aguantar mis dudas desde el inicio de los tiempos con las “Nepetas”, pero sobre todo por tu sentido del humor y los buenos ratos que hemos pasado.

Thank you, Félix, Seb and Hong, for including me since the very beginning as part of your fantastic lab, and for the wonderful Monday lunch conversations. I felt so fortunate for being part of such a welcoming group of people; you made me enjoy incredibly my time there at London. Thank you so much. A toda la gente que conocí en Kew (Itxaso, Mauricio, Carlos, Sandra, Federico, Elvira, Jerónimo...), y que hicieron que me sintiera como en casa durante ese tiempo. A Arnau por aguantarme fin de semana sí y fin de semana también visitando los invernaderos, y siempre tener cosas nuevas que enseñarme. Y por supuesto a Nacho y a Sara, que me acogieron allí desde el primer día, y donde surgió el precioso proyecto que es SEBOTA. A Sara, además, muchas gracias por el amor y entusiasmo por las plantitas y siempre tener un hueco para compartirlos (¡vivan los Sebotellines!).

Muchas gracias a la *chusma* de botánica. Realmente no sabéis cuánto aprecio el buen ambiente de trabajo que ha habido durante todos estos años y que me hacía (quizás excepto un rato los lunes muy por la mañana) querer ir a la universidad. Muchas gracias a todos y a todas, Marina, Celina, David, Pablo “*Musgos*”, Raúl, Ana, Sonia, Álex, Sofía... A Laura, la mejor compañera de finalización de tesis, por todas las dudas e incertidumbres que me has solucionado. A Pablo “*Carex*”, por siempre tener tiempo para hacernos a los demás pasar un buen rato. Muchas gracias, Alberto por el sentido del humor, y lo bien que lo he pasado contigo (ya desde los inicios en el Jardín), hasta en los momentos más difíciles, como puede ser cierto viaje Madrid-Plasencia. Mi cardiólogo no obstante no te perdona los sustos que me has dado en los últimos tiempos. A Angélica, muchas gracias también por apoyarme en el bonito (aunque fallido) proyecto de *Beer Friday*. Quizás con el *Wine Thursday* haya más

suerte. Muchas gracias a Jimena por ser una compañera inseparable del *Writing Club*, sin olvidarme por supuesto de haber compartido alguna aventura haciendo “cierta pancarta” de “cierto personaje” en los pasillos de la facultad. Muchas gracias a María (Leo) por tu apoyo durante toda la tesis (y ya desde antes), pero en especial en el final, por sacar siempre un rato para mirar mis cosas y por tu ayuda para mejorar esta tesis. A María (Sanz) por compartir las aventuras con (y sin) los caricólogos. A Manu, por tu entusiasmo y por siempre tener una buena cara para los demás. Gracias por todo lo que me has enseñado (no sólo de plantitas). Y por supuesto a Julia, muchas gracias por apoyarme en las frikadas, por compartir esta aventura que ha sido labeleR, por el “bici-mix” y por el género *Trorgssionis*. May the `for()`s be with you.

Tengo que agradecer a todo el mundo con los que compartí mis años en el G-III. A todos mis jefes, que con el tiempo se convirtieron en amigos, y cuyas lecciones forman parte de lo que soy. A todos mis “niños”, de los que también he aprendido mucho, y a los que he intentado transmitir todo lo que el grupo me dio. Y por supuesto, a todos con los que he compartido generaciones y jefatura (Lampe, Chaco, Markitos, Sara, Cris, Manu, Bea, Rubén, Patri, Rafa, Rosu, Jose, Alien...), sin duda la gente con la que he pasado los mejores momentos de mi vida. Muchas gracias también a TitoValen, por ser un inseparable compañero de aventuras y por ser un ejemplo para mí. A mis compañeros del máster, con los que empezó todo, y en especial a Caye, Bárbara, Jorge, Hodei, Julia, Ana, Marina, Belén y Julen, por todo lo que hemos compartido también después. Viva la Unión por el Pinar y la Dehesa. Muchas gracias a Bea por los nachos en las lonjas, y por haberme enseñado tanto de lo que es este proceso. Efectivamente, aunque sea de cuando en cuando, creo que nunca te librará de mí. Sin todos vosotros esta tesis nunca hubiera existido.

A los *idiotas*; Pablo, Santi, Viti y Guille; que sois los que más me habéis aguantado en esta etapa. Aunque os riais de mi frikismo, de R, de las plantas, del despachito tecnológico, de los límites de velocidad, y de las alguitas del museo nacional. Muchas gracias por ser el apoyo que necesito, aunque no merezco. Gracias por aguantarme cuando empecé con la pesadez de la bici, y por sumaros después. Muchas gracias por hacer sobrellevable la pandemia, por el M+1, por la sotaquía. Sois los auténticos colegas del padrón. Muchas gracias también a

Alfon, Mery y Lucas, por descubrirme y compartir conmigo el maravilloso maridaje que es el rugby con margaritas. Muchas gracias a todos vosotros por las tardes de Yola, de V.O. y de *poppy*.

Por último, pero más especialmente, a mi familia. Nada de esto habría sido posible si no hubiera sido por vosotros. Gracias a Talillo, por aguantar mis chorradas, las turras, y por no mandarme a la mierda con “la planta del día”. Gracias a Santi por todo lo que me has aguantado, y siempre ser capaz de hacer una broma más graciosa que la anterior. Y muchas gracias, Papá y Mamá, por todo lo que me habéis enseñado. Por darme herramientas para aprender y respuestas a mis preguntas. Por haberme animado siempre a hacer lo que me gustase, y por escucharme y aguantarme después. En resumidas cuentas, muchas gracias por ser un ejemplo de qué es lo que quiero ser.

Abstract

The Iberian Peninsula is one of the most diverse regions in the Mediterranean Basin hotspot, sheltering around 6,500 vascular plant species and subspecies, of which nearly one third are endemic to the region. Positioned at the convergence of the Mediterranean Sea and the Atlantic Ocean, and characterized by a varied array of topographic landforms and other climatic and edaphic factors, the region exhibits a remarkable diversity of environments. These features have prompted, on the one hand, a notable diversification of plant lineages, and, on the other hand, have favored the colonization and settlement of a great number of species from different origins, thus contributing to the outstanding plant diversity that characterizes the Iberian Peninsula.

The general objective of this thesis was characterizing the distribution patterns of the Iberian flora at the regional scale, evaluating its taxonomic and phylogenetic structure, and identifying priority areas for conservation. To do so, the following specific objectives were pursued: a) compiling a chorological database including distribution information for the Iberian flora at the finest possible scale (10x10 km); b) designing a generalized workflow to expand incomplete phylogenies; c) evaluating the impact of contrasting floristic elements (endemic taxa *versus* the entire flora), phylogenetic information, and spatial resolution on bioregionalization analyses, and d) identifying areas of the Iberian Peninsula encompassing significantly high values of threatened phylogenetic diversity to inform a list of priority areas for conservation.

In an early stage of the dissertation, information from research papers, regional atlases and global biodiversity repositories was gathered and curated to create the AFLIBER database. This dataset comprises taxonomic and chorological information for the entire vascular plant flora of the Iberian Peninsula, including 6,456 species and subspecies of plants, of which 2,142 are endemic to the region. For each taxon, distributional information using 10-km-side grid cells was recorded. This lookup was performed using both automatic and manual quality filters, and the resultant distribution maps were visually checked by experts. The distribution dataset relies on nearly 2.8 million records, of which 1.8 million correspond to unique occurrences. As a second step in the dissertation, the *randtip* framework to expand incomplete phylogenies using

non-molecular information was created and implemented in an R package. The software uses taxonomic information to designate the phylogenetic segments that most certainly contain phylogenetically uncertain taxa (PUTs) and to bind them below the corresponding clades at a randomly selected point. Beyond providing a comprehensive toolkit of functions to optimally expand incomplete phylogenies, potential pitfalls and opportunities to optimize parameter determination for PUT binding were disclosed.

Equipped with AFLIBER and *randtip*, the focus was on assessing the importance of i) using distinct floristic elements (endemic taxa *versus* entire flora), ii) including phylogenetic information, and iii) the spatial resolution, for bioregionalization analyses, identifying which environmental variables more significantly drove each scheme. Endemic species yielded a more topography-driven regionalization, while the whole flora produced bioregions more aligned with other environmental variables, and both taxonomically and phylogenetically informed analyses led to similar bioregionalization schemes. Lastly, we explored the spatial distribution of angiosperm phylogenetic diversity and provided a priority list of areas for conservation using a complementarity analysis. To do so, the evolutionary distinctiveness of the species and their probability of becoming extinct based on global threat assessments were combined into an integrative analysis that revealed 21 complementary areas for conservation, some of which extend beyond currently protected areas.

This doctoral dissertation establishes a foundational framework for exploring spatial and evolutionary dimensions of the flora of the Iberian Peninsula. The incorporation of innovative methodologies developed here offers opportunities for reassessing various macroecological and biogeographic enquiries, including bioregion delimitation and the identification of critical areas for biodiversity conservation. In summary, the materials and methodologies presented here, alongside the obtained results, represent a fresh perspective in the field of plant biogeography, providing a highly valuable framework for analyzing biogeographic and evolutionary patterns, as well as conservation imperatives, in diverse ecosystems and floras worldwide.

Resumen

La Península Ibérica es una de las regiones más diversas del punto caliente de biodiversidad de la cuenca mediterránea, albergando cerca de 6.500 especies y subespecies de plantas vasculares, cerca de un tercio de las cuales son endémicas de la región. Ubicada en la confluencia entre el Mar Mediterráneo y el Océano Atlántico, y caracterizada por una variada matriz de accidentes geográficos y otros factores climáticos y edáficos, esta región exhibe una extraordinaria diversidad de ambientes. Estas características han producido, en primer lugar, una eminentemente diversificación de linajes de plantas, y, en segundo lugar, han favorecido la colonización y establecimiento de un gran número de especies desde distintos orígenes, contribuyendo así a la excepcional diversidad vegetal que caracteriza a la Península Ibérica.

El objetivo general de esta tesis es caracterizar los patrones de distribución de la flora ibérica a una escala regional, estudiando cómo está estructurada desde un punto de vista taxonómico y evolutivo, e identificar áreas prioritarias para la conservación. Para ello, se abordaron los siguientes objetivos específicos: a) recopilar una base de datos corológica incluyendo información de distribución para la flora ibérica a la mayor resolución posible (10x10 km); b) diseñar un flujo de trabajo general para expandir filogenias incompletas; c) evaluar el impacto de distintos elementos florísticos (endémico frente a flora completa), información filogenética y resolución espacial en análisis de biorregionalización; y d) identificar áreas de la Península Ibérica que alberguen valores significativamente altos de diversidad filogenética amenazada para generar una lista de áreas prioritarias para la conservación.

En una fase inicial de la tesis, se recopiló y revisó información de artículos científicos, atlas regionales y repositorios globales de biodiversidad para crear la base de datos AFLIBER. Este recurso recoge información taxonómica y corológica para toda la flora de la Península Ibérica, incluyendo 6.456 especies y subespecies de plantas, de las cuales 2.142 son endémicas del área de estudio. Para cada taxón, se registró información de distribución utilizando cuadrículas de 10 km de lado. Esta búsqueda se realizó utilizando filtros de calidad, tanto automáticos como manuales, y los mapas finales de distribución fueron revisados individualmente por expertos. La base de datos comprende casi 2,8 millones de registros, de los cuales 1,8 millones son únicos. Como segundo paso de esta tesis, se creó *randtip*, un marco de trabajo para

expandir filogenias incompletas utilizando información no molecular, y se implementó en un paquete de R. El programa usa información taxonómica para demarcar el ámbito de la filogenia a la que pertenecen los taxones filogenéticamente inciertos (PUTs) e insertarlos en los clados correspondientes en un punto seleccionado de manera aleatoria. Más allá de ofrecer una amplia gama de herramientas para expandir filogenias de manera óptima, se discutieron posibles dificultades y oportunidades de utilizar distintos parámetros para añadir PUTs.

Mediante el uso de AFLIBER y *randtip*, se buscó evaluar la importancia de i) estudiar distintos elementos florísticos (taxones endémicos en contraposición a la flora completa), ii) incluir información filogenética, y iii) la resolución espacial en análisis de biorregionalización, identificando qué variables ambientales resultaban más significativas para guiar cada propuesta. El uso de especies endémicas lleva a una regionalización más basada en factores topográficos, mientras la flora al completo delimita regiones más relacionadas con otras variables ambientales, y tanto los análisis basados en información taxonómica como filogenética resultaban en propuestas de biorregiones similares. Por último, estudiamos la distribución espacial de la diversidad de angiospermas y generamos una lista de áreas prioritarias para la conservación utilizando un análisis de complementariedad. Para ello, se combinó la singularidad evolutiva de las especies y su probabilidad de extinguirse basadas en evaluaciones globales de amenaza en un análisis integrado que mostró 21 áreas complementarias de conservación, algunas de las cuales se encuentran fuera de espacios actualmente protegidos.

Esta tesis doctoral establece un marco básico para estudiar las dimensiones espacial y evolutiva de la flora de la Península Ibérica. La incorporación de metodologías innovadoras desarrolladas aquí brinda la oportunidad de reevaluar distintas preguntas en macroecología y biogeografía, incluyendo la delimitación de regiones biológicas y la detección de áreas fundamentales para la conservación de biodiversidad. En resumen, los materiales y métodos presentados, junto con los resultados obtenidos, suponen una perspectiva original en el campo de la biogeografía de plantas, aportando un valioso marco de trabajo para analizar patrones biogeográficos y evolutivos, así como necesidades de conservación, en diferentes ecosistemas y floras de distintas partes del planeta.

Table of Contents

Acknowledgements	xi
Abstract	xv
Resumen	xvii
1 General introduction	1
1.1 Biogeography	1
1.2 Shortfalls on biodiversity knowledge in the era of big data	2
1.3 Geographic characterization of the Iberian Peninsula	3
1.4 The Flora of the Iberian Peninsula	7
1.5 Floristic regions in the Iberian Peninsula	8
1.6 Conservation of the Iberian flora	10
1.7 Outline of the dissertation	14
Outline of the dissertation	14
1.7.1 Objectives	14
1.7.2 Structure of the dissertation	15
General Methods	16
1.7.3 Area of study	16
1.7.4 Expansion of backbone phylogenies	16
2 Atlas of the vascular flora of the Iberian Peninsula biodiversity hotspot (AFLIBER)	19
2.1 Introduction	20
2.2 Methods	22
2.2.1 Data Compilation	23
2.2.2 Analysis of completeness	24
2.3 Results	25
2.3.1 Species list and distribution data	25
2.3.2 Database completeness	27
2.4 Discussion	27
3 A generalized framework to expand incomplete phylogenies using non-molecular phylogenetic information	31

3.1	Introduction	32
3.2	General workflow	35
3.2.1	Input files	36
3.2.2	Selecting MDCCs and binding PUTs	39
3.3	The “Taxon list” and “backbone” modes of randtip	41
3.4	Newly designed features for PUT binding	43
3.4.1	Polyphyletic, paraphyletic and singleton groups of PPCR species	45
3.4.2	Manual definition of MDCCs	46
3.4.3	Respecting monophyletic and paraphyletic groups	47
3.4.4	Clumping PUTs	47
3.4.5	Non-ultrametric phylogenies	49
3.4.6	Customizing a subset of branches to randomize PUTs	50
3.5	Concluding remarks	50
4	The role of endemism and evolutionary history in defining biological regions	53
4.1	Introduction	54
4.2	Methods	57
4.2.1	Distributional data	57
4.2.2	Phylogenetic data	57
4.2.3	Regionalization analyses	59
4.2.4	Environmental characterization of bioregions	59
4.3	Results	60
4.3.1	Determination of number of bioregions	60
4.3.2	Bioregion delimitation	61
4.3.3	Environmental characterization of bioregions	63
4.4	Discussion	65
4.5	Conclusions	69
5	Regional application of a global conservation approach: spatial patterns of EDGE angiosperms in the Iberian Peninsula	71
5.1	Introduction	72
5.2	Methods	74
5.2.1	Taxonomic and chorological data	74
5.2.2	Phylogenetic data	75
5.2.3	Global extinction risk data	75
5.2.4	EDGE score calculation	76
5.2.5	Spatial analyses	78
5.3	Results	80
5.4	Discussion	82
6	Discussion	91
6.1	An updated and comprehensive distributional dataset	91
6.2	Phylogenetic expansion	93
6.3	Method selection in phylogenetic expansion	94

6.4	Disparities and resemblance with previous Iberian plant bioregion proposals	97
6.5	Understanding the variability in Iberian bioregion proposals	98
6.6	Phylogenetic bioregionalization in the Iberian Peninsula	99
6.7	Integration of phylogenetic information in conservation planning	100
6.8	Concluding remarks	102
7	Conclusions	103
7	Conclusiones	105
	References	107
	Appendices	123
Chapter 2.	AFLIBER	124
2a.	Publication	124
2b.	Extended Methods	131
2c.	Supplementary Figures	137
2d.	Supplementary References	139
2e.	Compilation data sources	145
2f.	GBIF references	161
Chapter 3.	randtip	179
3a.	Publication	179
3b.	Tutorial of use	189
3c.	Functions implemented in randtip	213
Chapter 4.	Bioregions	217
4a.	Supplementary Information	217
Chapter 5.	EDGE	227
5a.	Supplementary Information	227

Chapter 1

General introduction

1.1 Biogeography

Biogeography, in the words of Lomolino et al. (2010), is the science dedicated to documenting and understanding spatial patterns of biological diversity. While its roots trace back to the 18th and 19th centuries during the era of European exploration, it has evolved into the modern discipline we recognize today over recent decades. Pioneering explorers such as Buffon (1761), Forster (1778) or Alexander von Humboldt (1807) chronicled their discoveries from voyages around the globe, noting the apparent natural rules governing the distribution of plants and animals. These included Buffon's Law on species assemblages and observations regarding latitudinal and altitudinal trends among others, laying the foundations of traditional biogeography. Following the Era of Exploration, significant theoretical advancements emerged, including the evolutionary perspectives of Darwin (1859) and Wallace (1870) or the inclusion of the continental drift theory (Wegener, 1912) to recognize relationships among biotas distributed in different continents. These approximations allowed biogeographers to leverage several disciplines such as botany, zoology, geography, paleontology, genetics, or ecology, shaping modern biogeography in the interphase of all of them. Since the

latter decades of the 20th century, this field has witnessed an exponential increase in the volume of publications, propelled by the incorporation of comparative methods, phylogenetic analyses, and advancements in computational capacity and spatial data visualization techniques to explore geographic patterns (Lomolino et al., 2010). This uprising trend continues to our days (Posadas et al., 2013), boosted by the increase of available spatial data and the combination of analyzing techniques.

1.2 Shortfalls on biodiversity knowledge in the era of big data

In the current era of big data (Mayer-Schönberger and Cukier, 2013), the world has witnessed an unprecedent surge in data creation, processing, and storage, with the scientific community being no exception to this trend. The accessibility of publishing and accessing such information proves unvaluable for scientific endeavors, driving widespread interest in initiatives aimed at compiling biodiversity data for various exploratory purposes, including taxonomy (<https://www.itis.gov/>), genetics (<https://www.ncbi.nlm.nih.gov/>), species distribution mapping (Jetz et al., 2012; <https://www.gbif.org/>), or functional traits (<https://www.try-db.org/>), among others, all of which play key roles in advancing ecological research (Riddle et al., 2011). However, despite the growing availability of valuable scientific data, our understanding on biodiversity remains incomplete (Hortal et al., 2015) and biased (Gaston and Rodrigues, 2003). These gaps in knowledge constitute the so-called shortfalls, encompassing a diverse array of topics for which information is lacking. Hortal et al. (2015) identified three major shortfalls that impede our capacity to perform macroecological analyses. Foremost among them is the ‘Linnean shortfall’ (Lomolino, 2004), which refers to the discrepancy between known species and the total existing species. Adjacent to this,

there exist two additional extrinsic shortfalls (*sensu* Hortal et al., 2015), which form integral components of this doctoral dissertation. On the one hand, the ‘Wallacean shortfall’ refers to gaps in information concerning species distributions (Lomolino, 2004), reflecting the disparity between documented and actual species occurrences. On the other hand, the ‘Darwinian shortfall’ highlights our limited understanding of the evolutionary relationships among species (Diniz-Filho et al., 2013). Altogether, these knowledge gaps regarding the organization of biodiversity hinders our ability to effectively safeguard it (Riddle et al., 2011).

1.3 Geographic characterization of the Iberian Peninsula

The Iberian Peninsula is a unique geomorphological land mass placed in the westernmost tip of the Mediterranean Basin (see Figure 1.1a). Its relative isolation from the rest of the continent, coupled with its environmental heterogeneity, has facilitated the emergence and persistence of a significant biodiversity, rendering it a focal point for biogeographic and evolutionary investigations. This region encompasses not only the main peninsular land but also a collection of small archipelagoes and the Balearic Islands, extending as an offshore prolongation of the Baetic range (Sàbat et al., 2011). Situated between the Cantabrian Sea to the north, the Atlantic Ocean to the west, and the Mediterranean Sea to the east, it is geographically isolated from the remainder of the Eurasian continent by the Pyrenees mountain range (see Figure 1.1b). Although a body of water, namely the Gibraltar strait, lies between the Iberian Peninsula and the African continent, the former remains relatively close, with a separation of less than 15 km. The Iberian Peninsula is crisscrossed by numerous mountain ranges, predominantly aligned in an east-to-west orientation. This configuration gives rise to



Figure 1.1: Geographic context (a) and topography of the Iberian Peninsula (b). The Iberian Peninsula (contoured) comprises mainland Spain and Portugal, Andorra and the Balearic Islands, as well as several minor offshore islands (not shown). Political boundaries between countries are indicated by a dashed line in b), where major mountain ranges and river basins are also depicted.

flat river basins in lower areas, manifesting as plateaus (Tagus and Douro rivers) or depressions (Ebro, Guadiana, and Guadalquivir rivers) (see Figure 1.1b).

The Iberian Peninsula is renowned for its climatic diversity, presenting substantial differences in several bioclimatic factors that profoundly influence the spatial distribution of plants (Sainz Ollero and Sánchez de Dios, 2011). Firstly, significant differences in mean annual temperature values can be observed, typically higher across plateaus and depressions, while remaining lower along the northern Cantabrian coast and in mountainous areas (Figure 1.2a). Additionally, a gradient in annual precipitation further complements this climatic profile. The northwestern coast of the Iberian Peninsula presents higher precipitation values due to the prevailing winds originating from the Atlantic Ocean. Conversely, the continental zones in the southern slopes of the Cantabrian and Central mountain ranges experience notably drier conditions due to the Föhn effect (Figure 1.2b). The most arid regions along this gradient are found in the southeastern corner of Iberian Peninsula, encompassing Murcia and Almería provinces, where near-desert conditions persist (Kidron and Lázaro, 2020), with annual precipitation barely surpassing 240 mm. Furthermore, these two bioclimatic variables exhibit significant seasonality, contributing to the prevalence of pronounced Mediterranean conditions (i.e. dry and warm summers, cf. Rivas Martínez, 2005) across much of the Iberian Peninsula, whereas the northern regions experience temperate oceanic conditions (see Figure 1.2c). The western half of the Iberian Peninsula, along with the Pyrenees, predominantly show acidic soils, while the eastern half and major alluvial plains present alkaline conditions (although stained by acidic islands matching mountain massifs; see Figure 1.2d). This heterogeneity in geomorphology, climate, and soil composition, along with various historical factors, renders the Iberian Peninsula

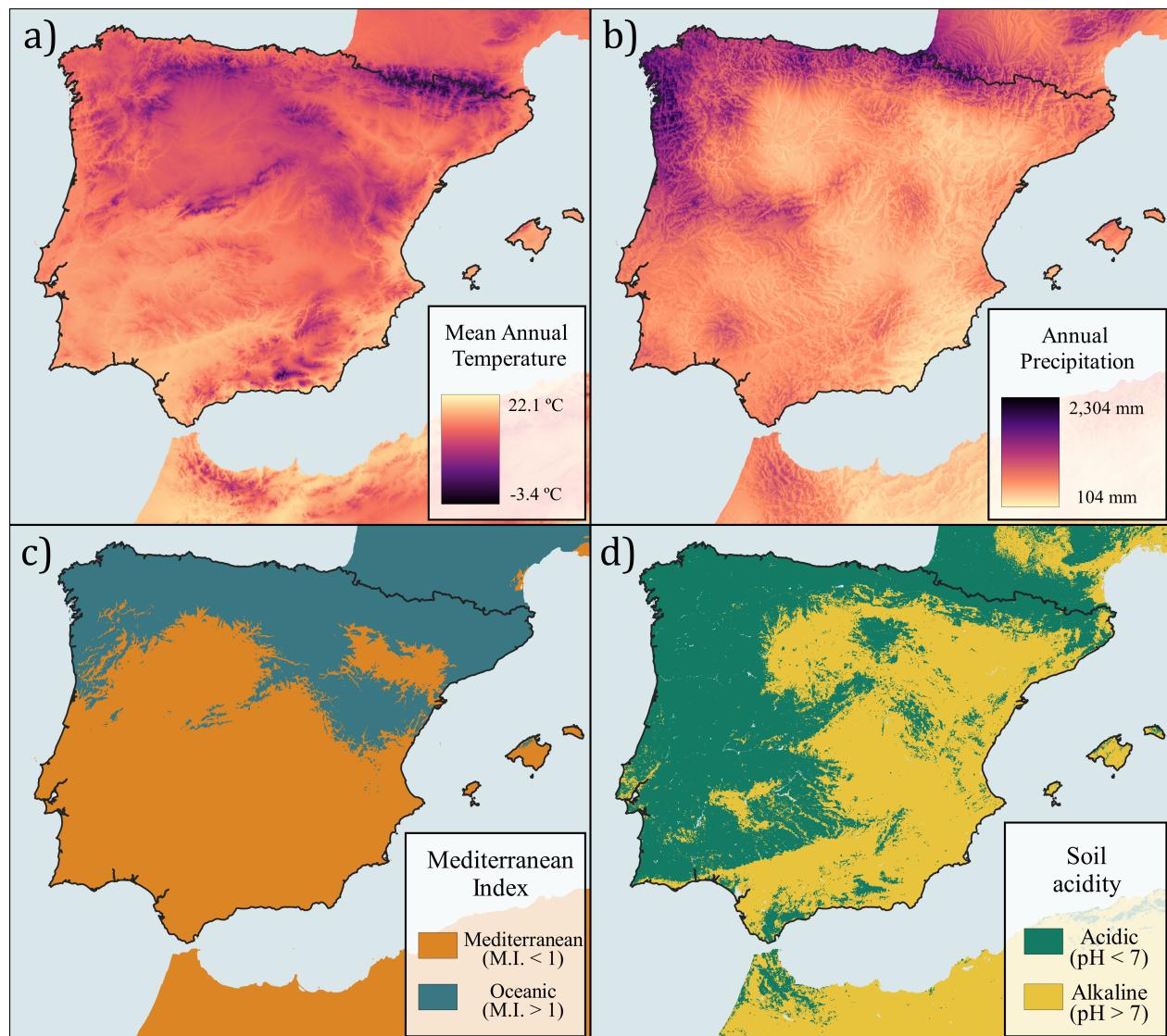


Figure 1.2: Spatial distribution of bioclimatic and edaphic conditions throughout the Iberian Peninsula. Panel a) shows continuous variation in mean annual temperature, while panel b) depicts annual precipitation, where warm colours stand for more arid conditions. Panel c) divides the peninsula in regions attending to their Mediterranean Index (*sensu* Rivas Martínez, 2005). Panel d) divides the Iberian Peninsula attending to the soil acidity, namely acidic regions (presenting a soil pH under 7) or alkaline regions (having pH values over 7). Spatial raster data were obtained from CHELSA 2.1 database (Karger et al., 2017) and SoilGrids (Hengl et al., 2017).

into a diverse puzzle of environmental conditions, which has resulted in a rich floristic region within the Mediterranean hotspot (Médail and Quézel, 1997).

1.4 The Flora of the Iberian Peninsula

The Iberian Peninsula harbors almost 6,500 species and subspecies of vascular plants, from which more than 33% are endemic to the region (Ramos-Gutiérrez et al., 2021), arising as one of the richest areas within the Mediterranean hotspot (Médail and Quézel, 1997). This richness is particularly concentrated around the Baetic and Pyrenean mountain ranges. The remarkable floristic diversity found within the territory has been extensively studied and attributed primarily to climatic, edaphic and topographical factors, as well as historical influences.

Firstly, the coexistence of two distinct climatic regions, namely Mediterranean and temperate-oceanic zones (see Figure 1.2c) endows the Iberian Peninsula with taxa from both climatic realms, giving rise to two major floristic elements: the Mediterranean and Euro-Siberian, with transition zones in between (Sánchez de Dios et al., 2009). This, altogether with the prevalence of alkaline soils in the east and acidic soils in the west (Figure 1.2d) establishes an eastern-western frontier that demarcates the Iberian flora. There is also a strong aridity gradient throughout the Iberian Peninsula, where areas near to the Atlantic Ocean present a humid climate while the southeastern region is semi-desertic (Figure 1.2a-b). This phenomenon leads to the appearance of temperate-affiliated species in the northern areas, and the prevalence of arid-adapted taxa associated with northern African lineages (Molina-Venegas et al., 2015) in the southeastern Iberian region. Topography also plays an important role in shaping the flora of the Iberian Peninsula. Firstly, the geographic isolation of the Balearic

Islands has fostered the emergence of numerous endemic taxa therein. Additionally, the altitudinal heterogeneity of mountain ranges entails a diverse array of microclimates within narrow areas, providing a mosaic of distinct habitats (Lobo et al., 2001; Rahbek et al., 2019). Moreover, the predominant east-to-west orientation of Iberian mountain ranges not only generates distinctions between the drier and warmer southern slopes compared to the cooler and more humid northern ones, but has also acted as a barrier to latitudinal migration during climatic oscillation events (Moreno Saiz and Lobo, 2008). This natural barrier promotes the isolation of small plant populations, thereby facilitating the emergence of endemic taxa (Moreno Saiz, 2011; Sainz Ollero and Sánchez de Dios, 2011). Lastly, the geological history of the Iberian Peninsula insight into the diversity observed today, with the presence of paleotropical, saharo-sindian, irano-turanian, and boreoalpine elements enriching the region with taxa associated with tropical, xerophytic, steppic and alpine environments, respectively (Buira et al., 2017; Moreno Saiz, 2011; Thompson, 2020). These elements result from plant colonization and establishment during past geologic events such as eustatic changes (in the case of saharo-sindian and irano-turanian elements) or climatic oscillations (as seen in the paleotropical and boreoalpine elements; Willis and Whittaker, 2000).

1.5 Floristic regions in the Iberian Peninsula

Just as taxonomy and systematics utilize the concept of species as a fundamental unit for studying biodiversity, biogeography organizes the global biota into distinct territories known as bioregions, facilitating detailed exploration (Hengeveld, 1992; Kreft and Jetz, 2010; Mackey et al., 2008). Since the pioneering work of early biogeographers such as Sclater (1858), Wallace (1876), and de Candolle (1855), numerous bioregionalizations

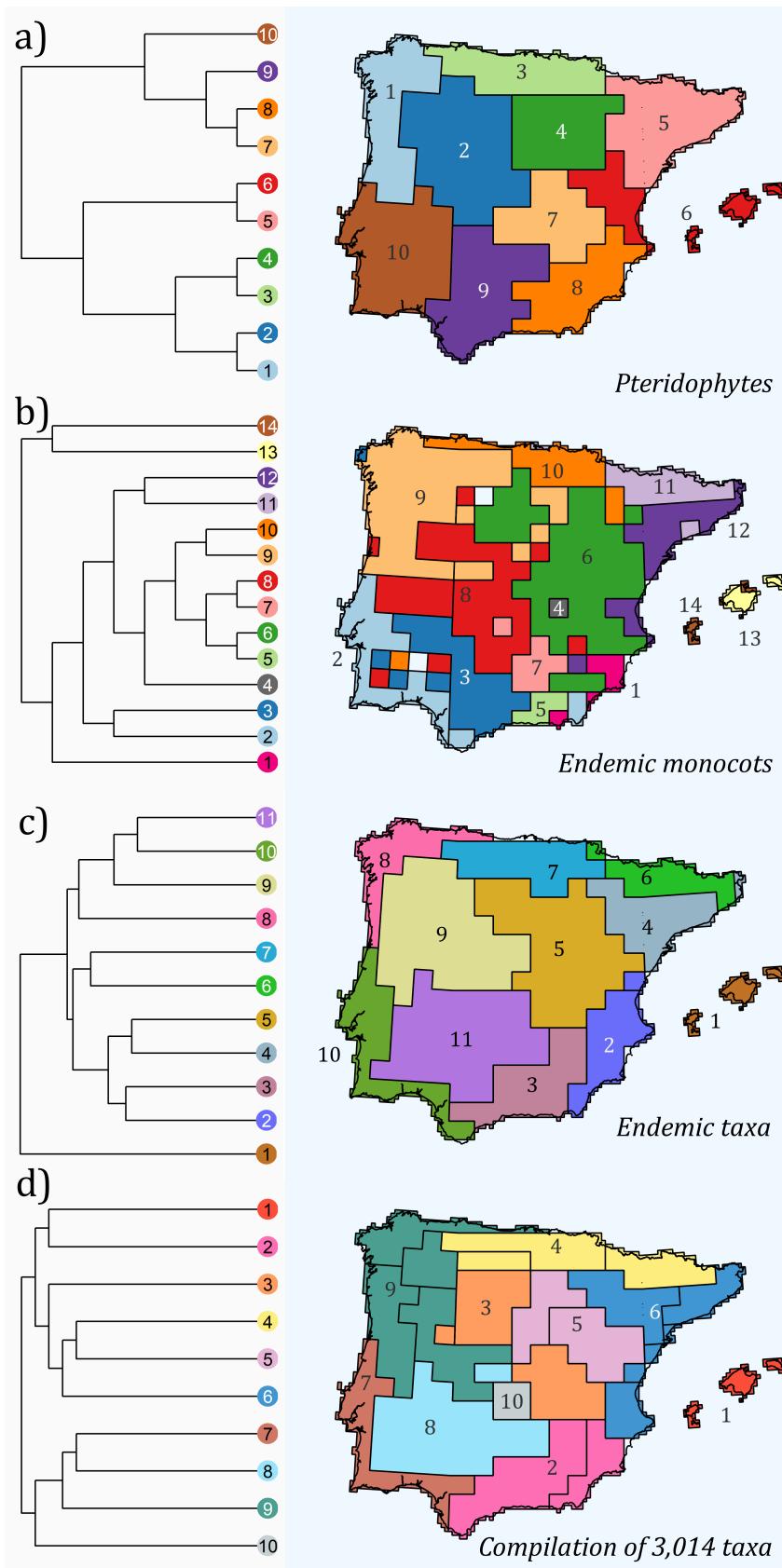


Figure 1.3: Previous bioregion proposals based on Iberian vascular flora using a 50 km UTM square area; modified from Moreno Saiz and Lobo, 2008 (a); Moreno Saiz et al., 1998 (b); Buira et al., 2017 (c) and Moreno Saiz et al., 2013 (d). On the right side of the figures, the bioregion maps are depicted, while the dendograms in the left show the dissimilarities among them. Colours in each figure match between the dendrogram and the bioregion maps.

have been undertaken, employing various taxa and classification methodologies (see Kreft and Jetz, 2010 for a comprehensive overview). These bioregionalization efforts are conducted across different spatial scales, encompassing continental and country-level approaches (e.g. Droissart et al., 2018; González-Orozco et al., 2013; Pataro et al., 2021), and the latter can significantly influence conservation initiatives (Habitat Directive, 1992; IBRA, 2000).

The Iberian Peninsula has long captured the attention botanists due to its remarkable floristic richness and heterogeneity, making a focal point for numerous bioregionalization analyses. Despite employing various clustering algorithms across different spatial scales, these studies have typically utilized only a fraction of the entire flora. For example, floristic bioregions have been proposed based on endemic dicotyledon species (Sainz Ollero and Hernández Bermejo, 1985), endemic monocotyledon species (Moreno Saiz et al., 1998), all endemic taxa (Buira et al., 2017), and pteridophytes (Márquez et al., 2001; Moreno Saiz and Lobo, 2008). Lastly, Moreno Saiz et al. (2013) conducted the most taxonomically extensive plant regionalization of the Iberian Peninsula to date, using 3,041 species. While the pteridophyte bioregionalization revealed a strong latitudinal component in region delimitation, most analyses underscored the importance of the acidic-alkaline frontier in defining floristic differences among regions, as well as highlighting the unique floristic composition of the Balearic Islands (Figure 1.3).

1.6 Conservation of the Iberian flora

As stated before, the Iberian Peninsula boasts an exceptional diversity of vascular plants, harboring nearly half of all European flora (Aedo et al., 2017), and providing habitat for over two thousand endemic taxa, solidifying its status as a hub of

plant diversity. Consequently, preserving this unique natural heritage is imperative from both regional and global perspectives. Presently, Spain records 1,196 species as threatened (Moreno Saiz et al., 2019), encompassing around 16% of its flora, while Portugal lists 381 (12% of its flora; Carapeto et al., 2020; Sequeira et al., 2011), and Andorra identifies 307 (19%; Carrillo et al., 2008) (see Figure 1.4). However, these figures are somewhat inaccurate for the Iberian Peninsula as a whole due to political considerations. Each constituting country maintains its own regional list, and Spain's Red Lists incorporating species from the Canary Islands, whose flora shares more affinities with the Afro-tropical region. Furthermore, only 21% and 19% of the Spanish and Portuguese plant species, respectively, have undergone formal Red List assessments. Globally, a mere 142 Iberian species are classified as threatened out of the 993 studied using a global approach (IUCN, 2024).

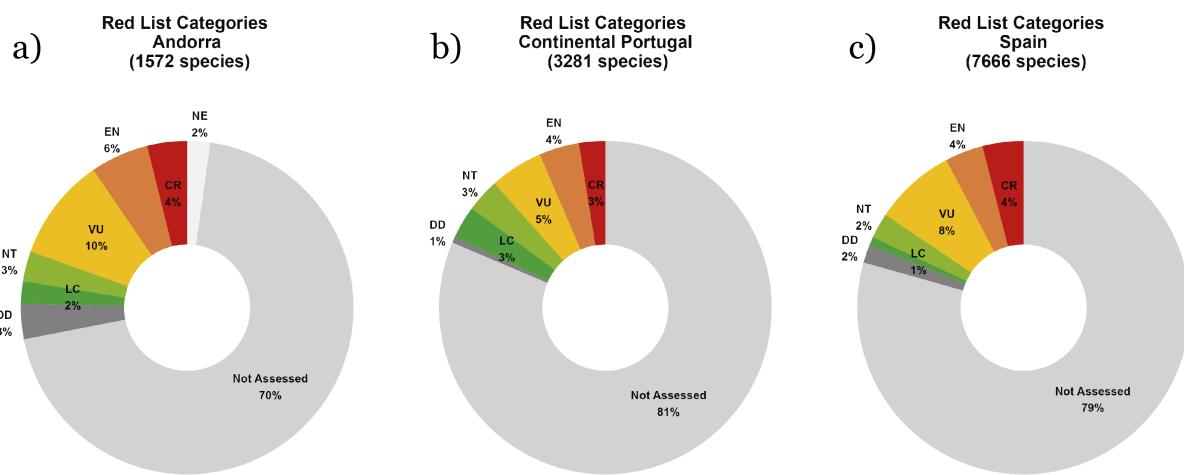


Figure 1.4: Distribution of vascular plant Red List Categories for Iberian territories, namely Andorra (a); Continental Portugal (b); and Spain (including Canary Islands) (c).

In their efforts to conserve their natural heritage, both Portugal and Spain have designated significant portions of their territory under different protection schemes. However, the location of zones of natural interest straddling the two countries often

creates an intricate web of interests, with different protection designations assigned to each national area, leading to an uneven distribution of conservation schemes (see Figure 1.5). Furthermore, international legislation, particularly within the Natura 2000 Network, adds another layer of complexity. While these areas encompass a substantial amount of land under protection, safeguarding plant diversity does not always count among the objectives of the Network's sites (Birds Directive, 1979; Habitats Directive, 1992). Evaluating the effectiveness of natural protected areas in biodiversity preservation is crucial for policymaking (Araújo et al., 2007), a topic extensively explored by researchers in Spain and Portugal. These studies identify priority areas for conservation and gaps in protection networks. Castro Parga et al. (1996) conducted a comprehensive analysis of existing national and natural parks, revealing discrepancies between irreplaceable biodiversity areas and protected zones. They suggested that nearly a hundred 10x10 km grid cells should be incorporated into protected areas. Similarly, Domínguez Lozano et al. (2000) highlighted the lack of representation of narrow endemics within protected areas. Despite these earlier findings, Muñoz-Rodríguez et al. (2016) reported that less than half of the threatened species in the Iberian Peninsula were covered by the national and natural park network, falling short of the targets set by the Global Strategy for Plant Conservation (COP, 2003). Consequently, it is evident that there is a deficiency in attention to plant conservation in land management.

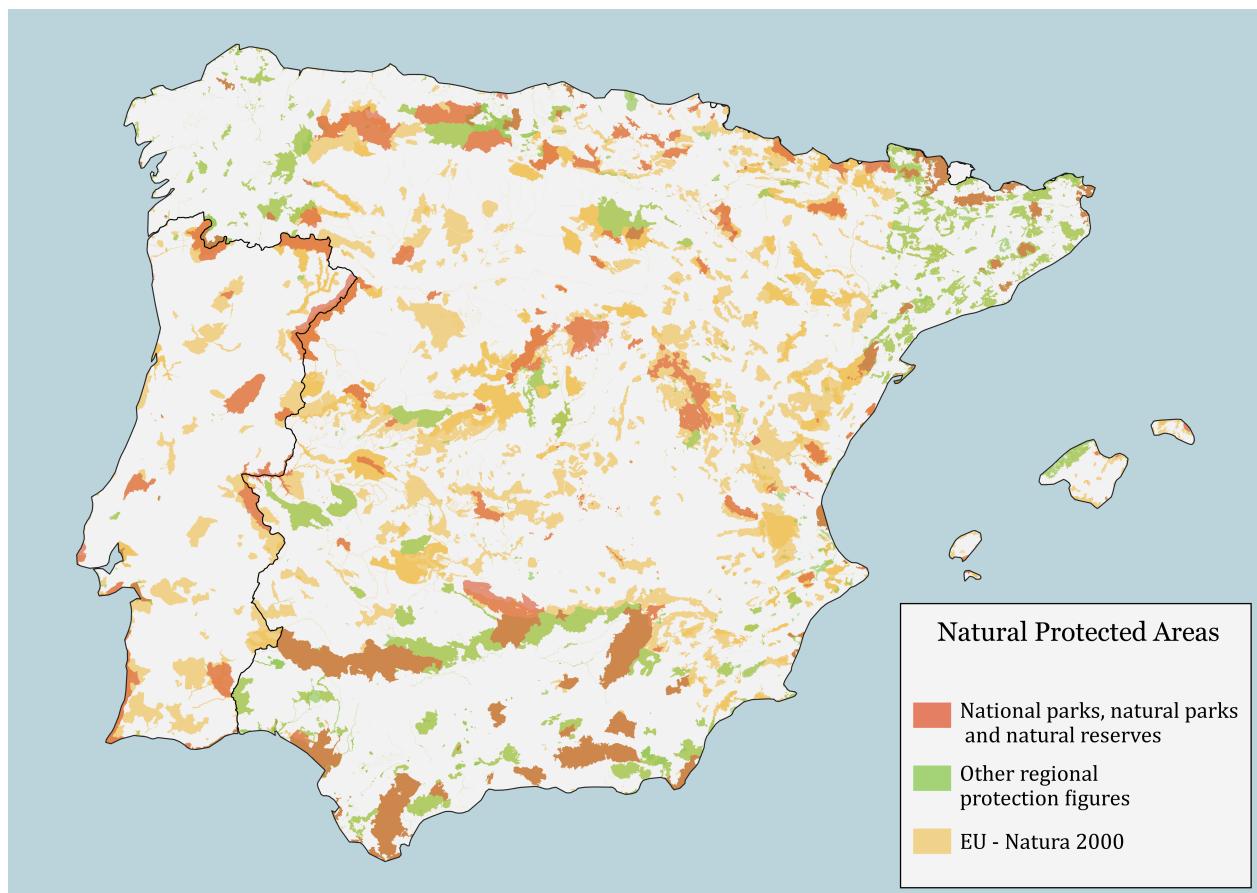


Figure 1.5: Natural Protected Areas of the Iberian Peninsula. National parks, natural parks and natural reserves are plotted in red, while other regional protection figures are depicted in green, and Natura 2000 sites are shown in yellow.

1.7 Outline of the dissertation

1.7.1 Objectives

The main objectives of this thesis are twofold: a) to address the two primary shortfalls that hinder the advancement of biogeographic inquiries in the Iberian Peninsula (namely ‘Wallacean’ and ‘Darwinian’ shortfalls); and b) to incorporate previously proposed solutions to conduct cutting-edge biogeographic and conservation analyses. Thus, we have developed a framework that enables integrating evolutionary information into biogeographic analyses, while using an extensive dataset of chorological data for the vascular flora of the Iberian Peninsula. This has facilitated the implementation of state-of-the-art bioregion delimitation techniques and the identification of priority areas for conservation. To achieve these objectives, we outline the following specific items:

1. Develop a high resolution (10x10 km) database consolidating a comprehensive collection of published and unpublished chorological data for all terrestrial vascular plants in the Iberian Peninsula.
2. Devise a methodological pipeline for systematically expanding incomplete phylogenies for biogeographic and evolutionary analyses, providing an R-based software implementation for practical application.
3. Investigate the significance of endemicity, evolutionary history, and spatial resolution in the delineation of floristic bioregions within the Iberian Peninsula.
4. Examine the distribution of plant phylogenetic diversity and establish a prioritization scheme for areas aimed at effectively conserving evolutionary history across the Iberian Peninsula.

1.7.2 Structure of the dissertation

This doctoral thesis comprises four manuscripts organized in a paper-like format, two of which have already been published in a prestigious international, peer-reviewed scientific journal, while the remaining two are undergoing preparation for submission. The thesis begins with a comprehensive introduction (Chapter 1), providing a theoretical foundation for the subsequent discussions, and the manuscripts presented afterwards address distinct objectives. Chapter 2 outlines the process employed to compile distributional information regarding the entire vascular flora of the Iberian Peninsula, emphasizing the potential of this database for further spatial analyses (Ramos-Gutiérrez et al., 2021). Chapter 3 presents a methodological approach to incorporate missing phylogenetic information into evolutionary analyses using a bespoke R package (Ramos-Gutiérrez et al., 2023). Chapter 4 builds upon the groundwork laid in the preceding chapters to explore the significance of endemism, phylogeny and scale in the delineation of bioregions (Ramos-Gutiérrez et al. *in preparation*). Lastly, Chapter 5 aims to generate a priority list of areas for the optimal conservation of plant phylogenetic diversity in the Iberian Peninsula by integrating spatial, phylogenetic and conservation data (Ramos-Gutiérrez et al. *in preparation*). The general methodology section provides a concise description of the study area and offers insights into phylogeny expansion, while specific methods employed in each work are detailed within the respective chapters to enhance the clarity of this dissertation. A wrap-up discussion of the results obtained throughout the thesis is presented in Chapter 6, followed by a set of conclusions (Chapter 7).

General Methods

1.7.3 Area of study

The Iberian Peninsula, located in the southwestern corner of the European continent, encompasses the mainland territories of Spain and Portugal, Andorra, the Balearic Islands, and several smaller archipelagoes (e.g. Columbretes Islands in the Mediterranean Sea, and Cíes or Berlengas Islands in the Atlantic Ocean). Altogether, it covers an area of 587,303 km². Within this region, a total of 6,316 land 10x10 km MGRS terrestrial grid cells will serve as the basis for spatial analyses conducted in this thesis.

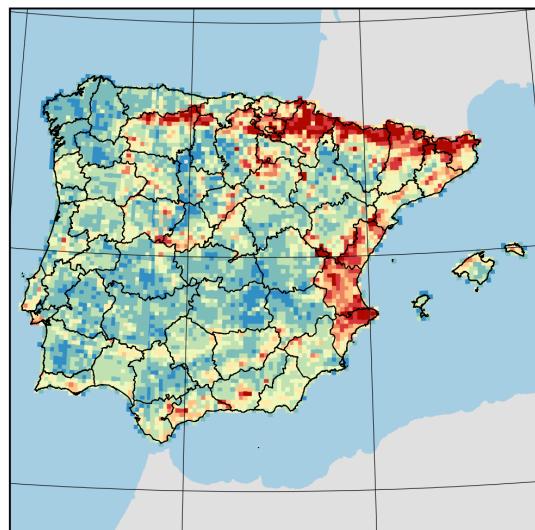
1.7.4 Expansion of backbone phylogenies

Two different approaches to phylogenetic expansion were undertaken for chapters 4 and 5, respectively. Both approaches utilized the GenBank-Open Tree of Life (GBOTB) phylogeny provided by Smith and Brown (2018) as backbone tree. However, chapter 4 incorporated an extension to include ferns and lycophytes (GBOTB.extended; see Jin and Qian, 2019). In both cases, 100 phylogenies were generated to assess the impact of phylogenetic uncertainty in the results. In Chapter 4, phylogenetic expansion was conducted using the ‘randtip’ package (Ramos-Gutiérrez et al., 2023). This process involved binding 3,638 phylogenetically uncertain taxa (PUTs, *sensu* Rangel et al., 2015) to the GBOTB.extended tree at specific or subspecific levels. The resulting phylogenies included the 6,456 Iberian taxa. Conversely, for Chapter 5 ‘V.PhyloMaker’ (Jin and Qian, 2019) was utilized to obtain fully comprehensive angiosperm phylogenies. In this case, 260,053 PUTs at species level were bound using scenario 2, resulting in phylogenies with 329,798 tips. Polytomies present across the internal nodes were converted to binary bifurcations using a bespoke function.

The scripts developed for the analyses conducted in each chapter are publicly available online in a GitHub repository, accessible at https://github.com/iramosgutierrez/PhD_code.

Chapter 2

Atlas of the vascular flora of the Iberian Peninsula biodiversity hotspot (AFLIBER)



Chapter Published in:

Ramos-Gutiérrez, I., Lima, H., Pajarón, S., Romero-Zarco, C., Sáez, L., Pataro, L., Molina-Venegas, R., Rodríguez, M. Á., & Moreno-Saiz, J. C. (2021). Atlas of the vascular flora of the Iberian Peninsula biodiversity hotspot (AFLIBER). *Global Ecology and Biogeography*, 30(10), 1951–1957. <https://doi.org/10.1111/geb.13363>

2.1 Introduction

With nearly 6,500 native vascular plant species and subspecies, of which 28% are endemics (Buira et al., 2017), the Iberian Peninsula (western Mediterranean) and its eastern Baetic offshore prolongation (the Balearic archipelago, Sàbat et al., 2011) comprise about a fourth of the total vascular taxa of the Mediterranean basin ($\sim 25,000$ species), the world's third-richest plant biodiversity hotspot (Fady and Concord, 2010; Myers et al., 2000). This outstanding plant diversity prompted the *Flora iberica* (FI) project in the mid-1980s, a long-lasting initiative aimed at synthesizing the systematic botany knowledge of the Iberian-Balearic region. Since then, 22 volumes have been published describing $\sim 5,600$ species and subspecies (Castroviejo, 1986), providing valuable information on whether a taxon is endemic to the region and in which administrative provinces it occurs. Yet, FI is still a work in progress for two main reasons. First, the last two volumes devoted to the grass family Poaceae (~ 500 taxa in the region) are still underway. Second, there is a continuous publication of new plant occurrences and revisions of taxonomical treatments based on both classical and molecular systematics. Consequently, FI also tackles new botanical publications to collate newly described taxa and plant occurrences on its web site (<http://www.floraiberica.es>).

Besides, the Spanish plant information system Anthos (<http://www.anthos.es/>) was created soon after the FI project to host floristic occurrences, and it currently comprises ~ 2 million plant records georeferenced according to the standard European UTM-grid at 10-km resolution. A similar structure has been used for the equivalent database of Portugal Flora-On (<https://flora-on.pt>) and many subnational biodiversity

databases, atlases and/or red books on plant distributions across this region (e.g. Bañares et al., 2004; Villar et al., 1997). Importantly, all plant records included in Anthos, plus ~13 million extra records from other databases, have been uploaded to the Global Biodiversity Information Facility (GBIF, <https://www.gbif.org/>), the largest source of Iberian plant distributional data hitherto.

Although we acknowledge the extraordinary value of open access biodiversity repositories, it should be noted that some records are often duplicated, imprecise or of dubious nature (Gaiji et al., 2013). Also, many records remain hardly accessible in non-digitized bibliographic sources (e.g. data from red books or floristic atlases) and restricted regional electronic repositories. Moreover, floristic information is often dated and riddled with errors (Anderson et al., 2016; Maldonado et al., 2015), which may have hampered fine-grained plant macroecological studies in the western Mediterranean. For instance, inflated omission error rates often force spatial analyses to be conducted at coarse grains (e.g. Buira et al., 2017; Lobo et al., 2001; Moreno Saiz et al., 2013) that might not be appropriate to inform conservation planning and policy making.

In this datapaper we provide AFLIBER, a comprehensive database connecting the up-to-date species list of the Iberian-Balearic vascular native flora with its matching occurrence records rescaled to a standard UTM 10 km grain resolution. The outstanding characteristics of this database are three-fold: (1) it has been generated according to the latest knowledge in floristics and taxonomy published for the region; (2) it has reduced geographical biases by incorporating both subregional databases broadly absent from previous wide-ranging repositories and distributional information published in monographs, grey literature and floristic articles; and (3) it provides occurrence records that were filtered conservatively to minimize identification and georeferencing errors

(see Methods) by means of a combined automatic and manual protocol undertaken by professional botanists specialized in the Iberian flora. Thus, this dataset represents an unprecedented effort to support botanical, biogeographic and macroecological studies in the western Mediterranean by providing comprehensive and carefully curated plant distributional information across the Iberian Peninsula.

2.2 Methods

The study area includes the Iberian Peninsula (mainland Portugal, Spain and Andorra) and the adjacent archipelagos of Berlengas (0.99 km^2), Columbretes (0.19 km^2) and Balearic Islands ($4,992 \text{ km}^2$), altogether comprising 6,316 terrestrial 10-km side UTM grid-cells (see Figure S2.1 in Appendix 2c).

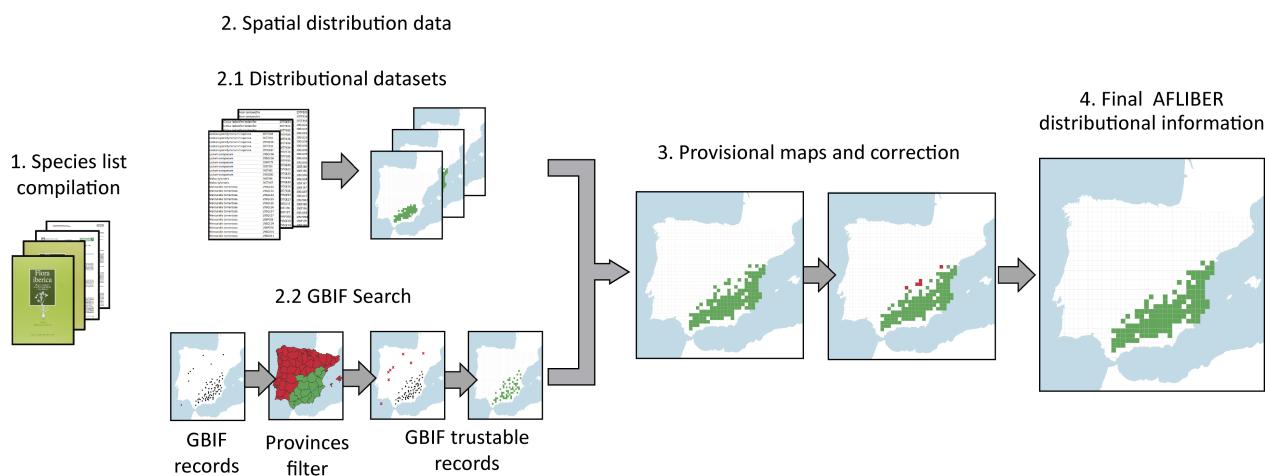


Figure 2.1: Workflow for creation of AFLIBER. (1) We compiled all species described in Flora iberica (FI) and later monographs and publications. (2) For each species in the list, we retrieved spatial distribution data in 10 km UTM cells, following a twofold procedure: (2.1) we retrieved information from reliable datasets (see Appendix 2d); and (2.2) the GBIF occurrence records were imported, disregarding occurrences with low coordinate resolution or that mapped away from the known distribution of the species (provinces filter). (3) These data were plotted as provisional maps that were visually explored and corrected by experts to achieve (4) the final result

2.2.1 Data Compilation

First, we compiled all native vascular plant species and subspecies included in the published volumes of FI (Castroviejo, 1986). Species in the Poaceae family (unpublished volumes) were obtained from different sources, including a comprehensive monograph by Romero-Zarco (2015), generic treatments published in recent years (see supplementary bibliography in Appendix 2d) and personal communications from experts involved in the preparation of the FI Poaceae volumes (i.e. Acedo, *personal communication*; Devesa, *personal communication*). Second, we conducted an exhaustive literature review to retrieve all newly published taxa (until July 2020) that were missing in the previous sources. Specifically, we revised the “new taxa” tab of the FI website (www.floraiberica.es/miselania/nuevos_taxones.php), updated generic monographies, and articles with newly recorded or described species in the study area (see Appendix 2d). Taxa were considered endemic whenever their distribution was limited to the study area including the northern slope of the Pyrenees (border with France), thus following previous biogeographical accounts of the Iberian flora (e.g. Buira et al., 2021; Moreno-Saiz and Sainz-Ollero, 1992). All distributional information at 10 km resolution was retrieved from high-quality, trustable databases available from electronic repositories and published research papers (Appendix 2e), including an unpublished database that was carried out over 30 years of data compilation from non-digitized sources (IberBal-Flora; see Appendix 2e). In order to complement the survey, GBIF plant occurrence records with a minimum resolution of 10 km were imported using the ‘rgbif’ R package (Chamberlain et al., 2019). It is important to note that GBIF includes large amounts of unverified records that do not reflect native distribution of species (e.g. corresponding to ex situ cultivation, botanical gardens). Thus, in

order to tease apart doubtful and/or potentially erroneous information conservatively, we filtered GBIF records to include only those georeferenced within the limits of the administrative provinces of Spain, Portugal and Andorra where the species occur according to FI and other later reliable publications. Clearly outlying observations were removed using the *over* function of the ‘sp’ R package (Pebesma and Bivand, 2005). References on occurrence records retrieved from GBIF are shown on Appendix 2f. Furthermore, the resultant distributional maps were explored visually one by one by the professional botanists of the team (i.e. C.R.Z., I.R.G., J.C.M.S., L.S. and S.P.) in search for erroneous or doubtful occurrences and to amend potential information gaps manually on the basis of expert knowledge of the study region (Figure 2.1). A list including 29,600 disregarded records is provided in Supplement 3 (available online at <https://doi.org/10.5061/dryad.gmsbcc2kv>).

2.2.2 Analysis of completeness

We used the non-parametric Chao2 estimator (Colwell and Coddington, 1994) to detect unequally sampled areas across the territory. To do so, the study area was divided into administrative provinces following the FI scheme (see Figure S2.2 in Appendix 2b). Each province containing less than fifty 10 km grid cells was merged with its nearest neighbouring one to avoid extremely distinct sizes among them, making a total of 56 provinces (ranging from 4,500 to 21,700 km²). The analysis was performed within provinces and for the complete territory as a single unit, using the *specpool* function as implemented in the ‘vegan’ R package (Oksanen et al., 2019).

Table 2.1: Data files variable names and description (data can be found on Dryad at: <https://doi.org/10.5061/dryad.gmsbcc2kv>)

Variable	Definition
(a)AFLIBER_Species_list.csv	
Taxon	Genus, species and subspecies (if applicable) epithets
Scientific_Name	Accepted scientific name
Endemic	Endemism of the Balearic Islands and/or the Iberian Peninsula (including French northern Pyrenean slope)
Genus	Adopted taxonomical category
Species	Adopted taxonomical category
Subspecies	Adopted taxonomical category
Class	Taxonomical category according to the National Center for Biotechnology Information (www.ncbi.nlm.nih.gov/guide/taxonomy/)
Order	Taxonomical category (according to NCBI)
Family	Taxonomical category (according to NCBI)
GBIF_id	Taxonomic numerical identifier in the Global Biodiversity Information Facility (https://www.gbif.org/)
POW_Name	Accepted taxonomic name in Plants of the World online (www.plantsoftheworldonline.org)
(b)AFLIBER_Distributions.csv	
Taxon	Genus, species and subspecies (if applicable) epithets
UTM.cell	UTM 10-km-side grid cell where the taxon is recorded
References	Sources from which the occurrence data were obtained. Numerical references correspond to those shown in the Supporting Information (Supplement 4b)

2.3 Results

2.3.1 Species list and distribution data

A total of 6,456 native species and subspecies were compiled (5,681 species after collapsing subspecies), of which 2,142 are endemic to the study area (Table 2.1a). We gathered a total of 2,716,018 distributional records (Table 2.1b; Figure 2.2a) comprising 1,824,549 unique (i.e. non-duplicated) occurrences at the 10-km-side grid-cell resolution (Figure 2.1b). Plant occurrences were distributed across 6,303 grid-cells (99.8% of total cells), of which 6,177 included species that are endemic to the study area (97.8% of the total) (Figure 2.2c). Empty cells (0.2%) generally corresponded to coastal cells only partially covering inland area, or to areas located between UTM transition zones.

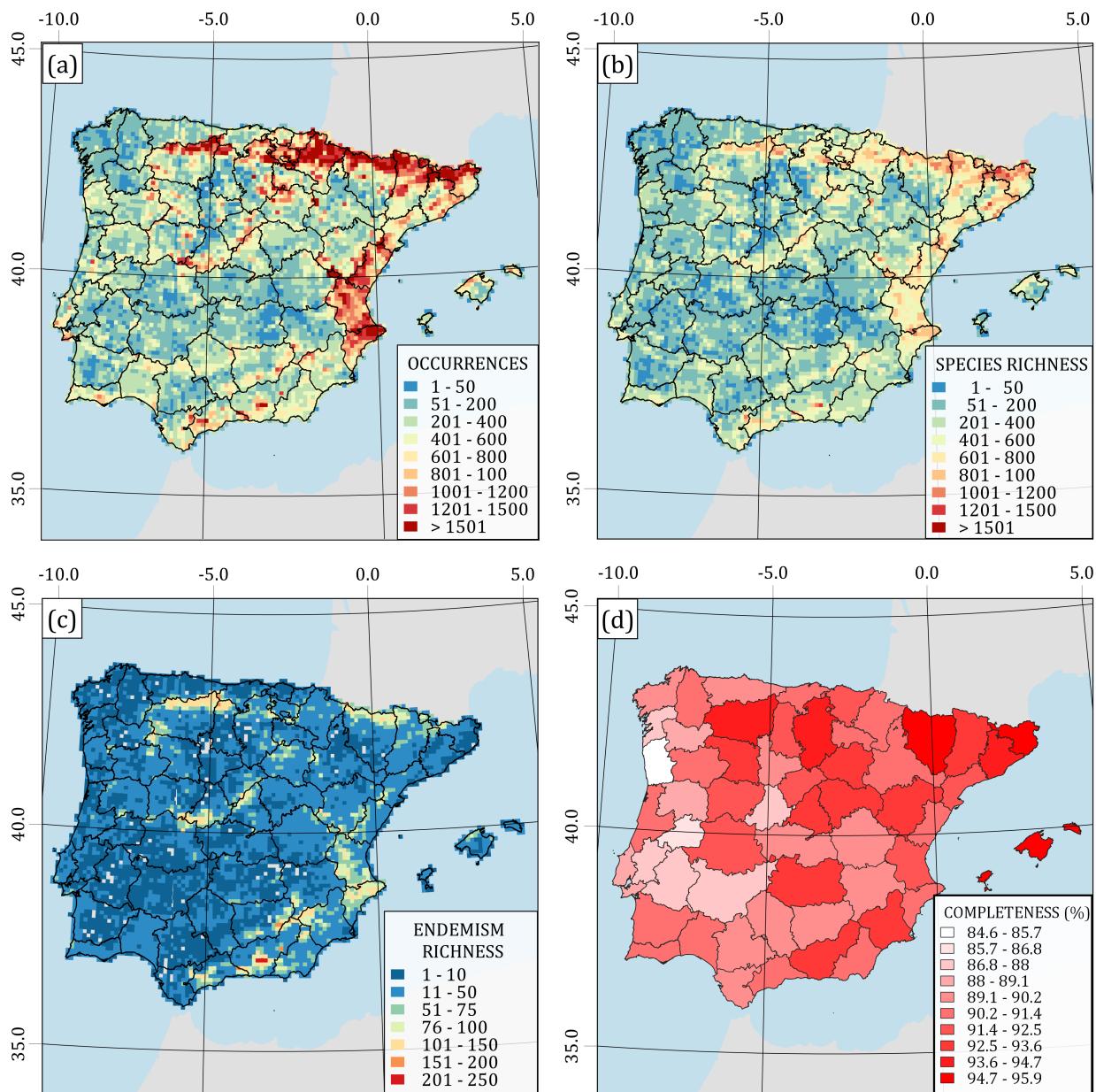


Figure 2.2: AFLIBER database number of (a) occurrences; (b) species richness (i.e., non-duplicated occurrences); (c) endemic species richness per 10 km UTM cell; and (d) percentage of regional floristic completeness.

2.3.2 Database completeness

Floristic completeness for the entire study area was strikingly high (98.57%), suggesting that the flora in the Iberian Peninsula is quite already known and well represented in our database. Indeed, when investigated separately throughout administrative provinces, completeness estimates remained high overall, with values ranging between 85 and 96% (cf. Figure 2.2d).

2.4 Discussion

The outstanding plant diversity and geographic extension of the Iberian Peninsula within the Mediterranean basin hotspot makes it an ideal setting to advance our understanding of the Mediterranean flora (Nieto Feliner, 2014), which has been largely hampered by the lack of a comprehensive source of regional plant distribution. Given its size and exhaustive data curation, the AFLIBER database represents a major step towards filling in this central gap of floristic and distributional knowledge in the western Mediterranean, having the potential to become the new Iberian reference for future plant macroecological, biogeographic and conservation studies. Indeed, with its 6,456 species and subspecies, this database not only approaches the 6,500 taxa estimated by Buira et al. (2017), but also elevates the number of endemic Iberian taxa to 2,142 species and subspecies. These numbers posit a higher figure than earlier accounts of Iberian plant diversity (Buira et al., 2017, 2020; Sainz-Ollero and Moreno-Saiz, 2002), representing an increase in endemics above 30% and thus equaling Iberian plant endemicity level to that of the Anatolian Peninsula in the Eastern Mediterranean (Davis et al., 1988). Interestingly, this pattern is much in line with earlier studies that recognized both ends of the Mediterranean basin as major centers of plant endemism

(Médail and Quézel, 1997).

It is important to note that, despite all the efforts to detect mistakes and misidentifications, the AFLIBER database might include some erroneous records and taxonomic treatments that are still subject to debate. Yet, bearing in mind that our knowledge of the Iberian flora remains incomplete, AFLIBER is certainly the most exhaustive and refined Iberian plant database published hitherto, representing a major improvement over other popular repositories of distributional information widely used in the past. For instance, the latest update of Anthos, the main source of reference for Iberian plant distribution so far, dates back to 2016 (L. Medina, *personal communication*).

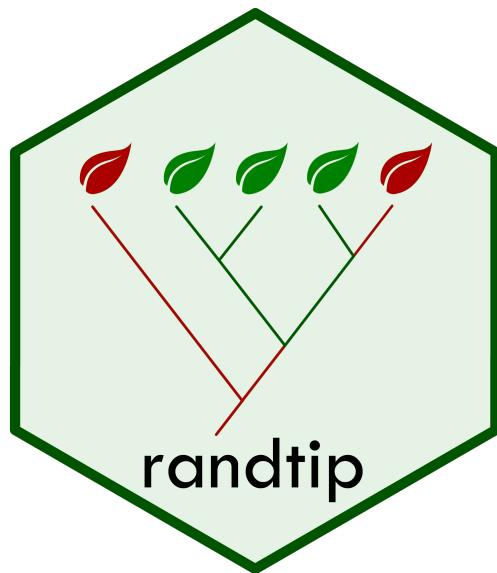
Although floristic completeness across the study area was acceptable (see Figure 2.2d), we note that some territories have been less prospected (Figure 2.2a) or lack distributional information in standard UTM grid cells (i.e. atlases), particularly towards western and southern areas. Such gaps of floristic knowledge may lead to unbalanced accounts of distributional information, which could be addressed by both by future prospections conducted in poorly sampled areas and by georeferencing extant information treasured in herbaria. Nevertheless, distinct levels of plant diversity across Iberian territories may also respond to natural processes. For example, while plant species richness peaked in mountainous regions (see 2.2b and Castro Parga et al., 1996; Lobo et al., 2001), human modification of natural ecosystems has led to depauperated levels of diversity over the centuries in certain areas, such as inner depressions and fertile river basins. A similar pattern can be observed for endemic species, which are particularly abundant in mountain regions such as the Pyrenees, the Cantabrian range and, especially, in the Sierra Nevada (Baetic range), the hottest spot of floristic diversity in the Mediterranean (Blanca et al., 1998; Buira et al., 2021; Heywood, 2003).

The essence of the AFLIBER database lies in taxonomic botany, a discipline that is slowly disappearing from academic curricula (Crisci et al., 2020). This ongoing botany crisis is attributable, in part, to the bibliometrics that are used to judge the impact of scientific research, which often undermine plant taxonomy, despite the latter is at the basis of many downstream disciplines such as ecology, biogeography and conservation biology (Crisci et al., 2020; Woodland, 2007). Nonetheless, professional and amateur Iberian botanists continue generating priceless floristic and taxonomic data, and the recently constituted Spanish Botanical Society will take up the challenge of maintaining taxonomic botany alive as a foundational task, including data curation of the AFLIBER database and the incorporation of the two lasting volumes of *Flora iberica* (Poaceae) and other relevant essays in due course.

Chapter 3

A generalized framework to expand incomplete phylogenies using non-molecular phylogenetic information

3



Chapter Published in:

Ramos-Gutiérrez, I., Lima, H., Vilela, B., & Molina-Venegas, R. (2023). A generalized framework to expand incomplete phylogenies using non-molecular phylogenetic information. *Global Ecology and Biogeography*, 32(10), 1707–1716. <https://doi.org/10.1111/geb.13733>

3.1 Introduction

The past two decades have seen an explosive interest in incorporating evolutionary history into ecological analyses (Cavender-Bares et al., 2009; Mouquet et al., 2012; Webb et al., 2002), boosting several disciplines such as community ecology (Davies, 2021), macroecology (Lamsdell and Congreve, 2021) and conservation biology (Molina-Venegas et al., 2020). This eco-phylogenetic revolution was driven by the increased availability of molecular information (Sayers et al., 2020) and sophisticated tools for inferring phylogenetic trees (Smith and Walker, 2019), which have lifted our understanding of species evolutionary relationships to unprecedent levels. However, and despite the phylogeny of certain groups, such as mammals, is nearly completed (Upham et al., 2019), phylogenetic relationships remain vastly uncertain—particularly shallow ones (i.e., infra-family)—for many groups. For example, one of the largest global phylogenies of angiosperm plants published to date includes only ~12.5% of the species in the group (Janssens et al., 2020), and recent accounts of terrestrial arthropod biodiversity showed that up to 80% of insect species are yet to be discovered (Stork, 2018). These bleak figures suggest that evolutionary biologists will have to tackle phylogenetic uncertainty for a long time to come.

Conscious of the limited extent of molecular phylogenetic information, Rangel et al. (2015) developed a theoretical foundation to systematically account for phylogenetic uncertainty in quantitative analyses. Roughly, the procedure starts with the identification of phylogenetically uncertain taxa (PUTs), that is, taxonomic units (e.g., species, subspecies) that are well delineated in the continuum of biodiversity but remain missing from available phylogenies. Then, all acceptable taxonomic, morphological, or

behavioural information on the PUTs is used to conservatively define their most derived consensus clades (MDCCs), that is, the less inclusive phylogenetic nodes that most certainly contain them. Finally, each PUT is assigned to a random point along one randomly selected branch of its corresponding MDCC, and the procedure is replicated a high number of times to obtain a distribution of possible trees that can be used in downstream analyses iteratively. While the “true” phylogenetic hypothesis will most certainly remain unsampled, the workflow allows exploring the parameter space, thereby quantifying the extent to which phylogenetic uncertainty has a significant impact in the analyses (e.g., Calatayud et al., 2019; Molina-Venegas et al., 2021). Rangel et al. (2015) accompanied their framework with the software SUNPLIN, a set of algorithms for randomly expanding phylogenies using the aforementioned procedure (Martins et al., 2013).

Although Rangel et al. (2015) suggested that the identification of MDCCs should be based on expert taxonomic evaluation, such knowledge is in practice beyond the reach of most researchers, particularly when dealing with very large phylogenies that often encompass a wide spectrum of taxonomic groups and thousands of species. In a valuable attempt to automatize the identification of MDCCs, Jin and Qian (2019) developed ‘V.PhyloMaker’, an R package that can generate large phylogenies of vascular plants (recently updated as U.PhyloMaker to include vertebrate animals; Jin and Qian, 2023). PhyloMaker is based on the seminal idea of the classical software Phylomatic (Webb and Donoghue, 2005), which uses a taxonomically informed backbone mega-tree to automatically define MDCCs (in the case of PhyloMaker, genus or family nodes in case the former are not available) and bind the PUTs to the selected clades. Beyond covering features that were already implemented in Phylomatic, PhyloMaker provides

an option to insert PUTs in randomly chosen nodes below the crown node of the corresponding MDCCs, so that a distribution of possible phylogenies can be generated with relatively little effort (Jin and Qian, 2019).

However, we note that current available tools for the insertion of PUTs, while valuable, have some important drawbacks. For example, PhyloMaker uses a pure node-based approach to insert PUTs, and thus the simulations often lead to the formation of polytomies even if a fully bifurcated backbone tree is used. In contrast, SUNPLIN allows the insertion of PUTs along randomly selected branches, but the user must manually set all the MDCCs for the simulations (Martins et al., 2013). PhyloMaker circumvents this limitation at the cost of requiring an “annotated” backbone mega-tree (a linkage between all the species represented in the backbone tree and their taxonomic genus and family) that is provided by the developers of the software, and thus the user is forced to use the backbone trees for which the software was implemented. Also, the definition of MDCCs on the basis of a few taxonomic ranks (e.g., PhyloMaker only considers genus or family nodes otherwise) might be excessively conservative and hence suboptimal under certain circumstances. For example, large taxonomic families often include taxonomic ranks between the family and genus level that may represent putative MDCCs (e.g., subfamilies, tribes and subtribes in the Asteraceae, Poaceae and Fabaceae plant families). Finally, there are shortcomings that are transversal to all available software for PUT binding, including the disregard of paraphyletic groups (Hörandl and Stuessy, 2010) and the impossibility to fully customize the space of phylogenetic edges for the insertion of PUTs among other issues.

Here, we present a completely generalized and flexible framework to expand incomplete phylogenies. The framework is implemented in the R package ‘randtip’, a

toolkit of functions that was designed to randomly bind PUTs in backbone phylogenies through a fully customizable procedure that uses automatically retrieved and arranged taxonomic data as a major source of phylogenetic information. Although randtip can generate fully operative phylogenies for any group of organisms using just a list of species and a backbone tree, we discuss a variety of circumstances that may require customizing simulation parameters beyond default settings to optimally expand the trees, including a detailed step-by-step tutorial that was designed to provide guidelines to non-specialist users (see Supporting Information, Appendix 3b).

3.2 General workflow

In this section, we describe the general workflow of randtip to expand phylogenies. Roughly, given a list of taxa (typically Linnean binomials) for which a phylogeny is to be obtained and a backbone tree (provided by the user), the software identifies putative MDCCs for the PUTs in the list. MDCCs are defined based on taxonomic ranks, including genus, subtribe, tribe, subfamily, family, superfamily, order and class, and by default the software will select the less inclusive among the available. Once each PUT is assigned to a MDCC, randtip will automatically bind them to the backbone tree according to the parameters that are set for the simulations, and a phylogeny including all the taxa in the user's list is returned (Figure 3.1). The workflow can be customized using a variety of parameters that are either passed through the whole simulation or adjusted independently for each PUT (see Supporting Information for detailed step-by-step examples).

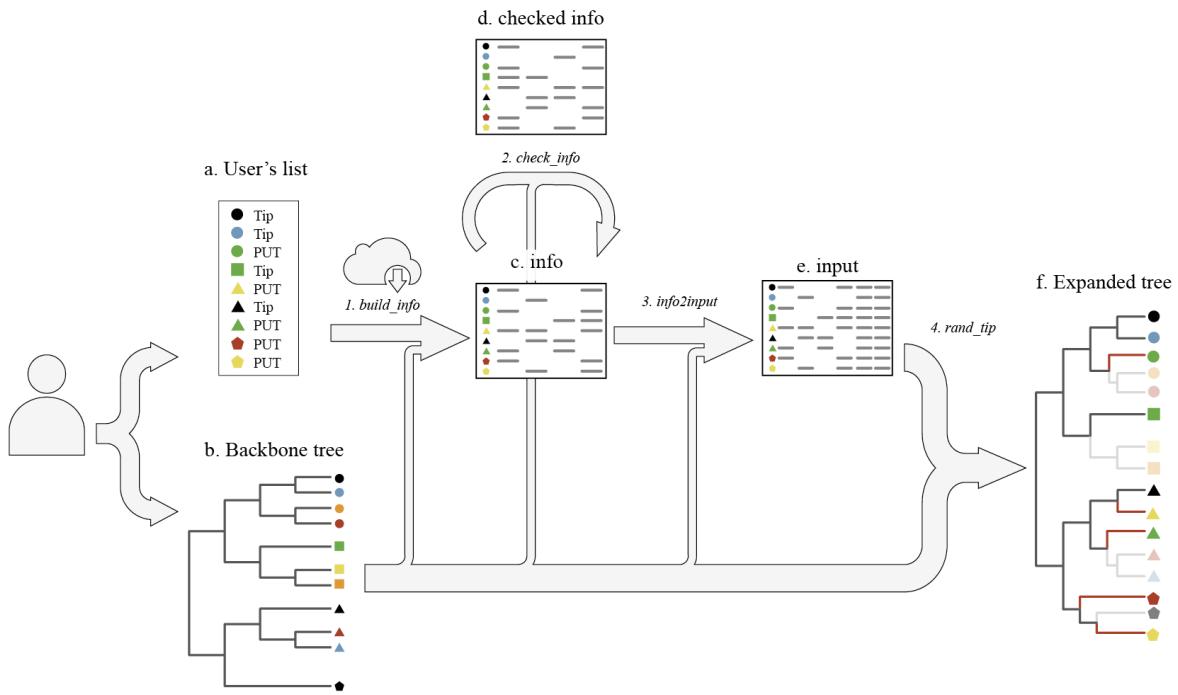


Figure 3.1: Schematic workflow of `randtip`. The user provides a backbone phylogeny and a list of taxa that are to be bound to the former (some of these are already placed in the tree while others represent phylogenetically uncertain taxa or PUTs). The function `build_info` creates the template `info` and retrieves taxonomic information for the listed taxa (and for those represented in the phylogeny if the “backbone” mode of `randtip` is set to TRUE) from web repositories. The resultant dataframe (`info`) can be evaluated with the function `check_info`. Once the user has edited `info` according to the particularities of each PUT, the dataframe is passed through `info2input` to create the `input` object for the `rand_tip` function, which in turn will expand the backbone phylogeny (in red, terminal branches subtending newly bound PUTs).

3.2.1 Input files

The workflow of `randtip` is guided by a dataframe R object (hereafter “`info`”) and the instructions that are passed through the main function of the package (`rand_tip`). The dataframe `info` is a template with 21 columns—20 variables of type character or logical plus one integer variable for internal use—that must contain, as a minimum, all the taxa in the user’s list (column 1) and their genus rank (column 2). Optionally, the user may provide supra-generic taxonomic ranks and set parameter values specifically for individual PUTs. For simplicity, we will consider the most common scenario in

the ecological literature where the operative taxa represent Linnean binomials (genus and species with or without subspecific epithets), although genus-level phylogenies are also supported. The info template can be created automatically using the auxiliary function *build_info*, which is fed with species names in a character vector or single-column dataframe. Besides, *build_info* can interact with a suite of taxonomic repositories—currently implemented for “ncbi” (default), “itis”, “gbif” and “bold” via the classification function of ‘taxize’ R package (Chamberlain et al., 2020)—to automatically retrieve and arrange taxonomic information that will be used to identify putative suprageneric MDCCs for the PUTs (note that information to define genus-level MDCCs is intrinsically contained in the scientific names of the species). This can be done by setting the argument “find.ranks” of *build_info* to TRUE (default). We recommend providing at least one supra-generic rank (e.g., taxonomic family) for all the species in info, which will be used to define MDCCs whenever the genus of the PUTs is missing in the phylogeny (otherwise the PUTs will not be bound). Often the user will need to further edit info once the template is created (for example, to customize binding parameters for certain PUTs or to amend taxonomic mistakes in web repositories). This can be done directly in R using the auxiliary function *edit_info* or exporting the dataframe as a spreadsheet (e.g., csv or xlsx) and importing it back into R once all the edits are completed.

The user must provide a backbone phylogeny as a phylo R object. Although randtip can identify MDDCs on the sole basis of taxonomic ranks of the species that are included both in the user’s list and the backbone tree (hereafter “taxon list” mode), MDDCs can also be identified based on taxonomic ranks of all the species that are represented in the tree regardless of their presence in the user’s list (hereafter

“backbone” mode). Both approaches have pros and cons (see Section 3.3), but they will perform identically whenever the genus of the PUTs is represented by at least one species in the backbone tree. To use the “backbone” mode of `randtip`, the argument “mode” of `build_info` must be set to “backbone” (default) for the software to include all the species in the phylogeny as rows in the info dataframe (otherwise, only the species that appear both in the phylogeny and the user’s list will be included), so that their taxonomic information can be automatically retrieved and arranged (if the argument “`find.ranks`” of `build_info` is set to TRUE).

Once the dataframe `info` is assembled, we strongly recommend checking the incidence of PUTs in the user’s list and their putative MDCCs. This can be done with the auxiliary function `check_info`, which will inform on the PUT status of the species, the presence of possible spelling errors, putative MDCCs, and the phyletic nature of the set of species that are included in each MDCC and share taxonomic ranks (e.g., congenerics, contribals, confamiliars) with the corresponding PUT—hereafter phylogenetically placed and co-ranked (PPCR) species. Also, the tip labels of the backbone tree are checked out for duplicates (e.g., *Ziziphora taurica taurica* and *Ziziphora taurica*), and the software evaluates if the tree is ultrametric or not. By default, `check_info` will make use of parallel processing to speed up the search for possible spelling errors and the identification of the phyletic nature of PPCR species, which is convenient for very large datasets. The auxiliary functions `get_clade` and `plot_clade` can in turn be used to extract and plot any subtree representing putative MDCCs, so that the user can visually explore them using the R graphic window (PPCR and non-PPCR species of the PUT are shown in contrasting colours, see Supporting Information for examples). Exploring MDCCs is particularly recommended to optimize

PUT binding, and particularly when PPCR species form polyphyletic groups (see Section 3.2.2). Alternatively, subtrees can be exported in Newick format to visualize them using auxiliary software such as Dendroscope (Huson and Scornavacca, 2012), which may be convenient for very large clades. Once the MDCCs are defined and the user has optionally customized parameter values for individual PUTs, the wrapping function *info2input* is fed with the dataframe info and the backbone phylogeny to create a final dataset that will be passed through the *rand_tip* function to expand the tree. This final dataset ensures consistent structure for use in *rand_tip* and allows generating as many trees as desired without the need to search for putative MDCCs in info repeatedly. This is done by *info2input* just once, a computationally intense task that is, by default, expedited using parallel processing.

3.2.2 Selecting MDCCs and binding PUTs

The binding of PUTs is conducted with the function *rand_tip*, which includes a variety of parameters that are passed through the whole simulation (Table S3.1, Appendix 3c). However, all the parameter arguments of *rand_tip* can be adjusted independently for each PUT by editing in the corresponding slots of info, which makes the framework completely flexible and customizable.

Randtip will always try to find the less inclusive MDCC of each PUT according to the taxonomic ranks that are provided in info, starting from genus level and up to class level until a MDCC is found. Regardless of the mode of randtip that is set by the user (“backbone” or “taxon list”), the software will always first attempt to define genus-level MDCCs as the most recent common ancestor (MRCA) of all the species in the backbone tree that are congeneric to the PUTs. However, MDCCs above

the genus level may differ between the two modes of `randtip`. On “taxon list” mode, supra-generic MDCCs are defined as the MRCA of all the species in the user’s list that are PPCR with the target PUT (e.g., contribals, consubfamiliars, confamiliars). In contrast, the “backbone” mode (default) defines supra-generic MDCCs as the MRCA of all the species in the backbone phylogeny (regardless of their presence in the user’s list) that are PPCR with the target PUT (see Figure 3.2 and Section 3.3 for an extended discussion).

By default, `rand_tip` will bind each PUT to a randomly selected branch below the crown node of the corresponding MDCC, the probability of being added along any branch being directly proportional to the length of the branch—if the argument “prob” is set to TRUE (default). Alternatively, branches can be selected on the basis of equal probability, and in either case the user can decide to add the stem branch of the clade to the pool of candidate branches—if the argument “use.stem” is set to TRUE (default is FALSE). The exact point to insert the PUT in the selected branch is sampled from a uniform distribution. Importantly, the extent to which the default behaviour of `rand_tip` to insert PUTs represents an optimal scenario may depend on the phyletic nature of their PPCR species. These can represent monophyletic (whenever the MDCC is exclusively shaped by species that are PPCR with the target PUT), singleton (terminal branch), paraphyletic (whenever the species that map within the MDCC but are not PPCR with the PUT form either a monophyletic or singleton group) or polyphyletic (set of PPCR species that does not fit any of the previous categories) groups (see Section 3.4.1 and Figure 3.3). The PPCR species of a given PUT could form a polyphyletic group simply because one of them maps clearly away from the main (monophyletic) cluster of PPCR species—for example, because the

outlying PPCR species is labelled in error (Pentinsaari et al., 2020)—in which case the default behaviour of *rand_tip* to bind the PUT (i.e., any branch below the crown node of the largest monophyletic cluster) would be reasonable. However, the polyphyletic nature of the PPCR species could also be due to “intruder” species that map within an otherwise monophyletic cluster, in which case the default behaviour of *rand_tip* could be suboptimal because the evidence that the largest monophyletic cluster of the group includes the PUT is less conclusive (Figure 3.3). As we discuss in Section 3.4, *randtip* allows the user to optimize the binding of PUTs according to the specifics of each case.

It is important to note that the user can always decide to what extent they want to rely on the retrieved taxonomic ranks for the automatic identification of MDCCs. For example, if the taxonomic affiliation of a PUT to a given genus is controversial, the user may edit the dataframe *info* to change the genus-rank of the PUT into “NA”, in which case *randtip* will use the taxonomic rank immediately above to find a new MDCC.

3.3 The “Taxon list” and “backbone” modes of *randtip*

The first decision the user will have to tackle is choosing between the “taxon list” and “backbone” modes of *randtip*. As we stated earlier, both approaches will perform identically as long as the genus of the PUTs is represented by at least one species in the backbone phylogeny, yet supra-generic MDCCs may differ between the two modes. For example, it might happen that some of the PPCR species of a given PUT (let us say *confamiliars*) are missing in the user’s list but are represented in the backbone phylogeny. Thus, in case these PPCR species were phylogenetically external to the *confamiliars* of

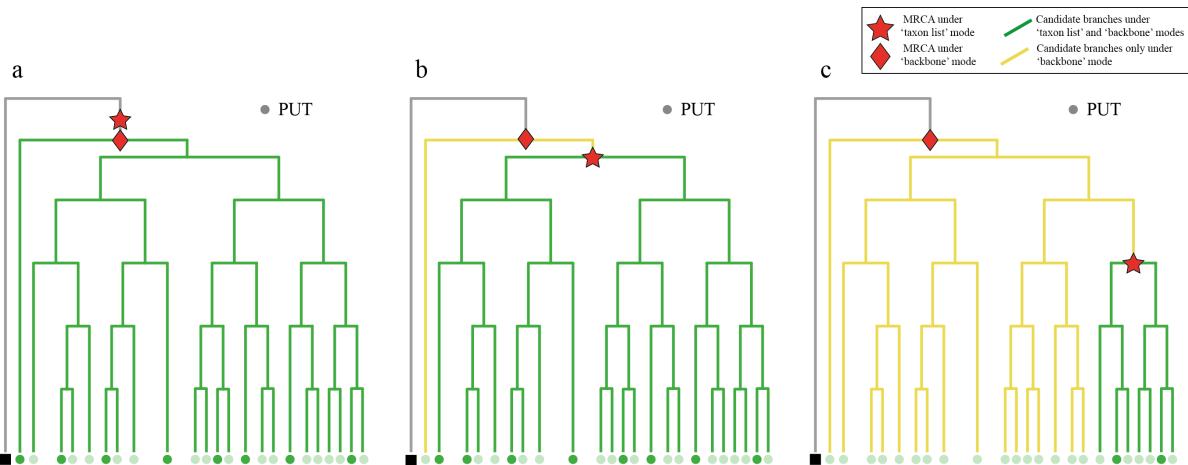


Figure 3.2: Scenarios of increasing divergence in the performance between the “taxon list” and “backbone” modes of randtip. The circle symbols on the phylogenetic tips represent *phylogenetically placed and co-ranked* (PPCR) species (e.g., confamilials) of the PUT, and the highlighted ones are those included in the user’s list in each scenario. The diamond red symbol (hereafter “diamond node”) indicates the crown node of the *most derived consensus clade* (MDCC) that is identified for the PUT when taxonomic information is available for all the species in the backbone phylogeny (i.e., under “backbone” mode), and the star red symbol (hereafter “star node”) indicates the crown node of the MDCC that is identified when taxonomic information is available only for the species in the backbone phylogeny that are also included in the user’s list (i.e., under “taxon list” mode). In the first scenario (a), the diamond and star nodes are coincident, and thus both modes of randtip will use the same space of branch lengths (in green) to bind the PUT. In the second scenario (b), the *most recent common ancestor* (MRCA) of the subset of PPCR species that are represented in the user’s list includes all PPCR species but one, and therefore the branch subtending the latter (in yellow) will never be selected under “taxon list” mode. In the third scenario (c), a higher number of PPCR species are missing from the user’s list, resulting in a smaller space of branch lengths to bind the PUT under “taxon list” mode. Note that under “backbone” mode, both the green and yellow branches would be candidates to bind the PUT.

the PUT that are included in the user’s list, the “backbone” mode of randtip would define an older MDCC than “taxon list” (Figure 3.2). It follows that the extent of the divergence in the functioning between both modes (whenever a supra-generic MDCC is to be defined) depends on the phylogenetic placement of the PPCR species that are included in the user’s list. In sum, the “backbone” mode works based on the “true” supra-generic MDCCs (but note that these may neither represent the actual MDCCs as the backbone phylogenies are often not fully comprehensive) with the trade-off that it is a more time-consuming approach than “taxon list”. In contrast, the latter might define younger supra-generic MDCCs (meaning more restricted parameter space to bind PUTs) under some circumstances (Figure 3.2). We recommend considering the

“backbone” mode as a first option (default) and use “taxon list” only when there is a low incidence of PUTs requiring supra-generic binding and/or low mismatch in the nodes defining supra-generic MDCCs between both approaches (see Figure 3.2 and Annex 3b for an extended discussion).

3.4 Newly designed features for PUT binding

As discussed above, *rand_tip* will by default bind PUTs to randomly selected branches below the crown node of the corresponding MDCCs. However, this default behaviour can be modified using a variety of arguments that are implemented in *rand_tip*. For example, if the user is not interested in generating a distribution of possible phylogenies but one single tree without randomizing the PUTs, the argument “rand.type” of *rand_tip* can be set to “polytomy” (default is “random”) for the function to insert the PUTs as polytomies at the crown nodes of their corresponding MDCCs instead. This is the only binding option that was implemented in the seminal software Phyloomatic (Webb & Donoghue, 2005), and it might still be convenient for extremely resource-consuming phylogenetic analyses where using a distribution of possible trees could be computationally prohibitive. Alternatively, the user may want to bind the PUTs following the default behaviour of *rand_tip* but still inserting some of them as polytomies in their corresponding MDCCs. To do so, the user can set the corresponding slots of the column “rand.type” of *info* to “polytomy” while keeping the argument “rand.type” of *rand_tip* to “random”.

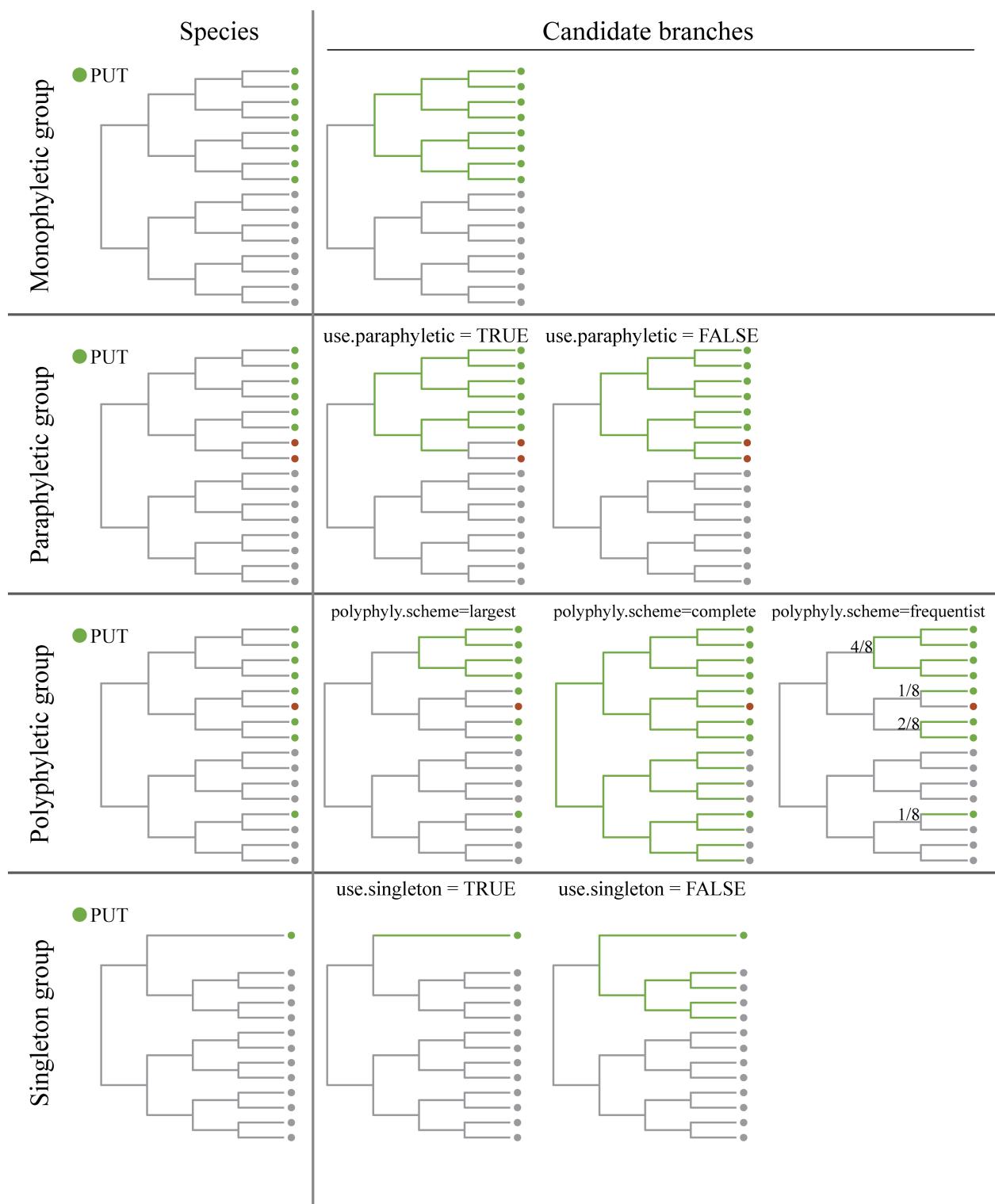


Figure 3.3: Types of phyletic groups formed by *phylogenetically placed and co-ranked* (PPCR) species (green circle symbols) and possible scenarios for PUT binding within each type. Non-PPCR species are in grey, and non-PPCR “intruder” species are in red. The candidate branches to bind the PUT in each scenario are in green (the vertical segments of the trees are purely aesthetic and were coloured to ease clade visualization). The fractions close to the phylogenetic nodes indicate the probability for the candidate clades to be selected under the scenario “frequentist”.

3.4.1 Polyphyletic, paraphyletic and singleton groups of PPCR species

While PUT randomizations within monophyletic groups of PPCR species will always follow the same scheme (i.e., by default, randomly selected branches below the crown node of the corresponding MDCCs), the user must choose between different scenarios for polyphyletic, paraphyletic and singleton groups. In case the MDCC of a PUT is shaped by a polyphyletic group of PPCR species, the software allows the user to choose between three different binding scenarios using the “polyphyly.scheme” argument. If the default option “largest” is set, *rand_tip* will pick the largest monophyletic cluster of PPCR species among the available to insert the PUT (less conservative scenario; Figure 3.3). If the option “frequentist” is set, *rand_tip* will first pick one of the constituent clusters of PPCR species that conform the polyphyletic group, the probability of being selected being proportional to the size of the cluster, and then the PUT will be inserted in the selected cluster. If the option “complete” is set, *rand_tip* will bind the PUT to a randomly selected branch below the crown node of the MDCC (most conservative scenario).

In case the MDCC of a PUT is defined by a paraphyletic group of PPCR species, two different scenarios are eligible. If the argument “use.paraphyletic” is set to TRUE (default), the candidate branches are those that keep the paraphyletic nature of the group unchanged after the binding (Figure 3.3). Otherwise, the randomization will be conducted as if the MDCC were defined by a monophyletic group of species. Importantly, certain taxonomic groups such as the Olacaceae s.l. plant family are paraphyletic (Chase et al., 2016), and thus randomizing PUTs at any point below the crown node of this family (i.e., setting “use.paraphyletic” to FALSE) may result in an excessively conservative parameter space that would encompass almost the entire

Santalales order (Malécot and Nickrent, 2008).

In case the MDCC of a PUT is defined by one single PPCR species (Figure 3.3), *rand_tip* will by default bind the PUT to the terminal branch subtending the only PPCR species, and whenever the MDCC is no longer singleton (because at least one PUT was already bound), *rand_tip* will consider the entire newly formed clade (same height as the original singleton clade) to sample candidate branches. We will refer to this procedure as “bind-to-singleton” hereafter. However, if the argument “use.singleton” is set to FALSE (default is TRUE), the parent node of the singleton PPCR species will be defined as the MDCC of the PUT instead (Figure 3.3). Although the latter scheme is more conservative than the former, it may lead to suboptimal solutions under some circumstances. For example, the parameter space to randomize a PUT whose MDCC is shaped by one single species that is the only representative of a subfamily in the phylogeny can be drastically increased in case the subfamily is the sister group to the rest of the family. Note that all these parameters can be specifically set for individual PUTs by filling in the corresponding slots of info.

3.4.2 Manual definition of MDCCs

Although *randtip* was conceived to automatize the definition of MDCCs based on taxonomic ranks, the user can manually define MDCCs for the PUTs. This can be done by filling in the corresponding slots of the columns “taxa1” and “taxa2” of *info*. As long as these slots are not set to “NA” (default), the MDCCs of the PUTs will be defined on the basis of this information instead. For example, if the slots “taxa1” and “taxa2” of a PUT are filled in with different species names, the PUT will be bound to a randomly selected branch below the MRCA of the two given species. If both slots

are filled in with the same species name, *rand_tip* will follow the bind-to-singleton procedure to insert the PUT as sister to the so defined species, and in case the same genus is provided the PUT will be inserted as sister to the clade defined by the MRCA of all the species in that genus.

3.4.3 Respecting monophyletic and paraphyletic groups

By default, *rand_tip* will never bind a PUT to a branch that results in breaking the monophyletic or paraphyletic nature of a group (of any taxonomic rank) unless the arguments “respect.mono” and “respect.para” are set to FALSE (default is TRUE). Thus, while previous software followed either approach (e.g., Phylomaker always respects monophyletic genera but SUNPLIN does not), randtip offers the user the possibility to choose between both options, either by setting the arguments of the *rand_tip* function or on a customized basis for individual PUTs by filling in the corresponding slots of info.

3.4.4 Clumping PUTs

Some genera may not be represented in the phylogeny, and thus their representative species will likely form a polyphyletic group if they are to be bound randomly below the crown node of the corresponding supra-generic MDCC. However, the user could be certain in that a group of congeneric PUTs whose genus is missing in the phylogeny is monophyletic. Thus, if the argument “clump.puts” is set to TRUE (default), *rand_tip* will first bind one of the congeneric PUTs, and then the rest will be bound following the bind-to-singleton procedure. Similarly, it may happen that supra-generic taxonomic groups are not represented in the phylogeny, in which case *rand_tip* will clump the PUTs as described above and following the taxonomic hierarchy so that the missing

taxonomic groups will form monophyletic clusters once all the PUTs are bound. As any other randomization parameter of `randtip`, the user may decide the PUTs that will be clumped in this way by setting the “`clump.puts`” option individually in the corresponding slots of `info`.

Trinomials representing infra-specific taxa (e.g., subspecies) are also supported. If “`clump.puts`” is set to TRUE, `rand_tip` will clump PUTs with infra-specific information according to their specific epithets (i.e., second name in the trinomial). To do so, `rand_tip` will first check if any of the trinomial PUTs that share specific epithet are represented in the phylogeny. This search also takes into account the type subspecies of the species, which will be detected in either trinomial (e.g., *Ablepharus chernovi chernovi*) or binomial (e.g., *Ablepharus chernovi*) nomenclature. In case one or more PPCR subspecies are found in the backbone tree, `rand_tip` will define a MDCC for the infra-specific PUTs following the standard procedures described in Section 3.4.1. Finally, if none of the trinomials in the group are found, `rand_tip` will first bind any of them to the tree, and then all the others will be bound following the bind-to-singleton procedure.

We note that some available phylogenies use, likely in error, both the binomial and trinomial form of a species to label different tips. For example, the GBOTB.tre mega-tree (Smith and Brown, 2018) includes *Ziziphora taurica taurica* and *Ziziphora taurica* as two different tips, and the GBOTB.extended.tre mega-tree (Jin and Qian, 2019) includes both *Saxifraga serpyllifolia* and *Saxifraga serpyllifolia serpyllifolia*. In these cases, `rand_tip` will randomly select either tip as the actual type subspecies and ignore the other. Although the `check_info` function will warn the user about the existence of possible duplicate taxa in the backbone tree (see Appendix 3b for

an example), we strongly recommend the user to visually revise tip labelling before expanding any backbone tree.

3.4.5 Non-ultrametric phylogenies

Previous software for PUT binding were conceived to be used with either ultrametric phylogenies (trees with branch lengths where all tips are equidistant from the root) or phylogenies without branch lengths. However, non-ultrametric trees where branch length is not proportional to time but character distance are also subject of ecological analyses (e.g., Mishler et al., 2014). The *check_info* function will warn the user in case the backbone phylogeny is non-ultrametric, and *rand_tip* will force non-ultrametric trees to be ultrametric —following the *extend*¹ method as implemented in ‘phytools’ R package (Revell, 2012) —if the argument “forceultrametric” is set to TRUE (default is FALSE). It is important to note that forcing phylogenies to be ultrametric in this way should not be taken as a formal statistical approach for inferring an ultrametric tree but a method to be deployed whenever a genuinely ultrametric phylogeny read from file fails due to issues related to numerical precision (Revell, 2012). Thus, we strongly recommend the user to visually explore phylogenetic trees that fail the ultrametricity test of *check_info* before assuming the failure is due to numerical precision of computer machinery.

If the backbone tree is non-ultrametric and the “forceultrametric” argument is set to FALSE, *rand_tip* will simulate the new branch lengths by sampling from a negative exponential distribution EX ($1/\lambda$), where λ is the inverse of the mean terminal branch length in the backbone tree. In case a backbone phylogeny without branch lengths

¹This was modified from “nlls” (in the original paper) to “extend”, as the first approach often returned polytomic trees.

is provided, *rand_tip* will output a phylogeny without branch lengths as well (i.e., topological information only). Hence, the only condition for *rand_tip* to accept a phylogeny is that it is rooted.

3.4.6 Customizing a subset of branches to randomize PUTs

The node-based workflow of randtip should suffice to cover most situations in PUT binding exercises. However, the distribution of possible branches for the simulation might not be drawn via MDCCs under some circumstances. For example, taxa of hybrid origin often appear as the sister species of either parent depending on the set of molecular markers that are used for the inference (Wang et al., 2014), in which case phylogenetic uncertainty may pertain to only two singleton putative MDCCs (assuming that the identity of the parents is known and both are represented in the backbone tree). Using the auxiliary function *custom_branch*, the user can customize specific subsets of branches to bind PUTs across any segment of the phylogeny.

3.5 Concluding remarks

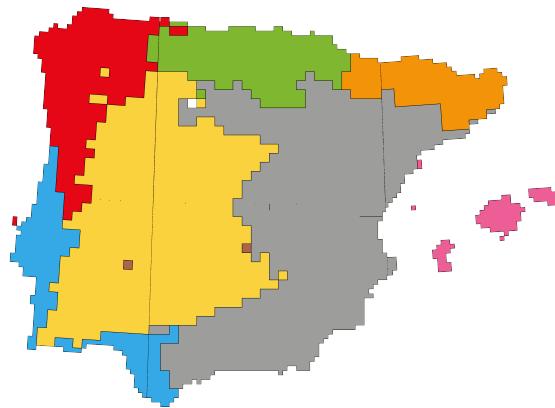
Randtip is, to our knowledge, the only framework for PUT binding that is completely flexible and generalized, thus addressing several shortcomings of previous designs and offering new opportunities to optimize parameter space in tree expansion exercises. Although randtip can generate fully operative phylogenies using default settings, we stress that accounting for phylogenetic uncertainty should not be conceived as a “black box” procedure for the immediate generation of phylogenies. Indeed, previous studies have documented inaccuracies in the generation of such “quick-and-dirty” phylogenies due to the “blind” use of software packages (Gastauer and Meira-Neto,

2013). Phylogenetic uncertainty should always be tackled with caution and restraint, for there is a variety of circumstances that may require customizing simulation parameters for specific PUTs if we are to avoid suboptimal solutions. Beyond providing newly designed tools to expand phylogenetic trees, the framework presented here will help evolutionary biologists to get the most out of the evolutionary information that can be used to guide tree expansion exercises.

Chapter 4

The role of endemism and evolutionary history in defining biological regions

4



Chapter in preparation for publication:
Ramos-Gutiérrez, I., Molina-Venegas, R., Moreno-Saiz, J.C. (2024). The role of endemism and evolutionary history in defining biological regions.

4.1 Introduction

Identifying biogeographic units, that is, areas with homogeneous biotas, dates back to the foundation of biogeography (Escalante, 2009; Hengeveld, 1992; Sclater, 1858; Wallace, 1876). However, far from being outdated, bioregions are still used for their potential to test biogeographic hypotheses (Kreft and Jetz, 2010) and for conservation purposes at regional scales (Lomolino, 2004). However, the methodology to define bioregions has substantially changed since the beginning of the discipline in the early 19th century. In early stages, the operative criteria were largely based on subjective (and incomplete) knowledge, which often led to disagreement among scholars (a well-known case is the separation between the Oriental and Australian bioregions, with different boundaries drawn by Wallace (1863), Huxley (1868) or Lydekker (1896), among others; see Ali and Heaney, 2021). Subsequently, more objective criteria were applied, as Croizat's superposition of species distribution maps (1958) or Takhtajan's hierarchical endemism (1986) among others. Lastly, highly reproducible quantitative methods have yielded dozens of bioregionalization proposals in recent years, varying in their methodology, geographical scope, and taxa under consideration (see Kreft and Jetz, 2010). However, bioregionalization exercises conducted at the regional scale are frequently based on endemic taxa solely, that is, species that are found only within the focal region, which leaves aside much of the biodiversity that characterises the territories. Moreover, evolutionary relations among taxa have been widely ignored until very recently (Daru et al., 2017a).

Some authors argue that endemic species are biological indicators which more likely reflect past biogeographical events and present local conditions than widespread ones,

hence their convenience for delimiting bioregions in regional scale analysis (Bradshaw et al., 2015; Stebbins and Major, 1965). Despite this, the notion of endemic taxa lacks uniformity, with different factors playing a role in whether a taxon achieves endemic status (Hengeveld, 1992). Firstly, being endemic can result from several evolutionary processes, including recent speciation events (neoendemics) or widespread extinction of populations with some extant ones thriving in refugia (paleoendemics) (Stebbins and Major, 1965). This collection of different evolutionary stories may imply that grouping endemic taxa as a unique entity may result in a patchwork of disparate cases. Additionally, the concept of endemism is also dependent on the geographic context of the study. As such, taxa with distributions centred in a focal region but extending beyond are often excluded from classical bioregionalization analyses, despite their informative value. This is quite clear in cases as the Iberian Peninsula, where there is an important number of species shared with the European mountains in the Pyrenees, or species shared with northern Africa in the arid southeastern corner of Iberia, for example. Furthermore, the definition of endemic taxa often relies on administrative rather than natural boundaries, which may introduce a certain level of arbitrariness. Summarising, endemic taxa may be a biased subset of biodiversity, overrepresenting lineages showing poor effectiveness in their dispersal ability, which is a key trait to determine bioregions (Antonelli, 2017). Thus, using them as bioregion indicators might cause the detection of areas where *in situ* diversification has a great effect. Therefore, we hypothesize that endemism-based analyses may explain just a fraction of the total complexity of the system.

Although evolutionary relationships are gaining momentum in macroecological analyses (Daru et al., 2016; Lamsdell and Congreve, 2021; Molina-Venegas et al.,

2020), earlier bioregionalization schemes considered species as independent entities, hence omitting the fact that they share a common evolutionary history. For example, a catastrophic event, such as the appearance of a physical barrier which isolated populations (e.g. an increase in sea level or an orogenic event) would eventually result in floristically distinct but yet phylogenetically similar regions (Molina-Venegas et al., 2015). Moreover, the use of phylogenetic information allows comparing sites with no shared taxa by accounting for phylogenetic dissimilarity among them, which may be particularly convenient in wide scale spatial analyses (Daru et al., 2020; Graham and Fine, 2008), where different regions may not share taxa just because there has not been enough time for colonization since the species were originated.

A last issue to consider when performing biological regionalization studies is the effect of spatial resolution. Although a complete and high-resolution dataset would be optimal for any biogeographic study, such task is generally a difficult target to achieve. The spatial resolution to be used must stand as a trade-off between the informativeness of the data and its reliability and completeness (Robertson et al., 2010). Lower resolution datasets are easier to compile with a low sampling omission rate. However, in such datasets, species co-occurring within a single pixel may exhibit high heterogeneity not attributable to their biology or history, but rather to the geographic extent of the grid tile. Conversely, high resolution datasets are more informative, but are harsh to compile, and prone to omission errors (Whittaker et al., 2005). However, the exponential increase of available chorological information online both in regional (e.g. www.anthos.es; Sillero et al., 2014) and global repositories (e.g. www.gbif.org) has allowed to gather high resolution datasets to address biogeographical questions (e.g. Gillings et al., 2019; Ramos-Gutiérrez et al., 2021; Ronquillo et al., 2020).

In this chapter, we use the complete vascular flora of the Iberian Peninsula to (1) evaluate the role of endemics in biological regionalization by examining the congruence between schemes based only on endemic taxa and those including all species at different spatial resolutions, and (2) assess the importance of information provided by phylogenies in delimiting such bioregions by comparing proposals based on taxonomic versus phylogenetic dissimilarities. Additionally (3), we evaluate the main topographical, biogeochemical and climatic drivers that draw the main differences among bioregions in the studied area.

4.2 Methods

4.2.1 Distributional data

We used a comprehensive distribution dataset of the Iberian vascular flora (AFLIBER; Ramos-Gutiérrez et al., 2021) which comprises georeferenced information for 6,456 species at a 10x10 km resolution. Grid cells adjacent to a change of longitudinal zone (due to the UTM projection), and thus incomplete (i.e. with a surface smaller than 100 km²), were merged to their nearest complete cell. We upscaled the data to grid cells of 20 km and 50 km sides to avoid bias due to differential sampling effort across the territory and to identify potential scale effects in the analysis. The number of grid cells and occurrences for each resolution is shown in Table 4.1.

4.2.2 Phylogenetic data

We followed the guidelines in Ramos-Gutiérrez et al. (2023) to obtain a complete phylogeny of Iberian vascular plants by expanding the most comprehensive molecular backbone tree available to date (Jin and Qian, 2019; combining information from Zanne

Table 4.1: Number of grid cells and occurrences retrieved after small cell merging and upscaling.

Resolution	Selected taxa	Number of grid cells	Number of occurrences
10x10	All taxa	6,142	1,803,855
	Endemic taxa	6,040	152,616
20x20	All taxa	1,588	926,109
	Endemic taxa	1,587	82,897
50x50	All taxa	287	327,141
	Endemic taxa	287	33,492

et al. (2014) for ferns and lycophytes and Smith and Brown (2018) for spermatophytes) using the ‘randtip’ R software (Ramos-Gutiérrez et al., 2023). We identified 3,638 phylogenetically unplaced taxa (PUTs), which were bound to a randomly selected branch below the node defining the most recent common ancestor (MRCA) of the corresponding congeners. In those cases where congeners were not available, the PUTs were bound below the MRCA of all the species in the next available rank above in the taxonomic hierarchy (i.e. consubtribals, contribals, consubfamilials, or confamilials). Whenever the available co-ranked species of the PUTs formed paraphyletic groups, they were bound below the MRCA of all congeners excluding the diverging clade, and therefore not forming polyphyletic groups during the binding process. PUTs pertaining to polyphyletic groups were bound just within the largest cluster of co-ranked species. PUTs bound to singleton groups (i.e. they were originally represented by just one tip in the backbone tree) were bound forming a monophyletic group in the branch over the original tip. To account for phylogenetic placement uncertainty, we repeated the procedure to generate 100 trees, and results were averaged across all trees. Full details on the parameters used for binding each PUT are shown in Appendix 4c-Supplementary Data.

4.2.3 Regionalization analyses

We used the algorithm implemented in the ‘phyloregion’ R package (Daru et al., 2020) to delineate bioregions, for it was designed to optimally identify both taxonomic and phylogenetic bioregions. Specifically, we drew the bioregions using the complete vascular flora and just the endemics, using both taxonomic and phylogenetic distances in each case, hence we obtained four different bioregion schemes. For each individual analysis, we searched for an optimum number of clusters (k) which maximised interregional and minimised intraregional dissimilarity. This search was performed using the *optimal_phyloregion* function of ‘phyloregion’ R package. Taxonomic and phylogenetic distances among grid-cells were computed with the *beta_diss* function using the β sim and p β sim dissimilarity indices respectively, and phylogenetic distances were averaged across the 100 trees. We set UPGMA as the clustering algorithm, and classification dendograms and inter-cluster NMDS analyses were carried out to study the dissimilarity among the resultant clusters in each case. Regions containing five or less grid cells were flagged as possible sampling artifacts for the subsequent analyses, which were performed at the 20x20 km resolution, as it was found to be the most informative.

4.2.4 Environmental characterization of bioregions

To explore the possible drivers of the resultant bioregionalization schemes, we carried out a set of linear discriminant analyses (LDA hereafter) to identify environmental variables exhibiting the most significant differences among groups, maximizing differences between bioregion clusters, and thus most likely shaping distinctiveness among bioregions. Initially, we considered average altitude, altitude heterogeneity, pH,

continentality, annual mean temperature, temperature seasonality, annual precipitation, and precipitation seasonality. Additionally, distance to the Iberian Peninsula was considered as a *proxy* variable measuring isolation from islands to mainland (see Table S4.1 for more details). Posteriorly, we performed a correlation test among these variables and discarded those showing a correlation higher than 85%, which were mean altitude (correlated with mean temperature) and continentality (correlated with temperature seasonality) (see Supplementary Figure S4.1).

4

The LDAs were sequentially performed for each separation in the classification dendograms (except for those dichotomies that split only small bioregions (i.e. enclosing five or less grid cells). Each analysis was performed 100 times, and variable values were normalized. We used function *lda* from the ‘MASS’ package (Venables and Ripley, 2002), randomly allocating a 60% of the data for training and 40% for testing. Fitness of each model was calculated as the AUC obtained with the *roc* function from ‘pROC’ package (Robin et al., 2011). AUC values and variable importance for each individual model were averaged through the 100 iterations.

4.3 Results

4.3.1 Determination of number of bioregions

The optimum number of bioregions varied for each analysis between 4 and 40, depending both on the floristic element considered (i.e. endemic *vs* whole flora), the dimension of biodiversity under evaluation (i.e. taxonomic *vs* phylogenetic dissimilarity) and the spatial resolution (see Supplementary Material, Table S4.2). No optimal number of clusters could be found for the regionalization analyses at a 10 km resolution using the whole flora. Moreover, the regions identified at such scale were generally

discontinuous and comprised a high number of uninformative clusters. These factors hampered the use of the finest resolution for the subsequent analyses. On the other hand, results for the broadest resolution (50 km side grid cells) returned a larger number of bioregions than those obtained at 20 km, while not showing clearer units. Thus, they were discarded too as most informative regions. The 20 km resolution, however, was sufficient to capture significantly continuous bioregions for the four studied approaches while showing a high spatial completeness (only one grid cell lacked any spatial information for the endemic taxa subset).

4.3.2 Bioregion delimitation

The spatial delimitation of regionalization analyses at a 20x20km resolution, as well as classification dendrograms and ordination NMDS plots is shown in Figure 4.1. Equivalent results for the finest and coarsest resolutions are shown in supplementary figures S4.2 and S4.3, respectively.

Regarding the scheme based on endemic taxa and taxonomic distances, we found eight clusters, from which one included just two grid cells. The major difference among regions was between the Balearic Islands (cluster 8) and the rest of the Peninsula, where inter-regional differences were weaker. The Peninsula was found to be primarily divided between northern-northwestern regions (clusters 2, 5 and 7) and southern ones (1, 3, 4 and 6). In the case of the northern bioregion clade, the oriental region (7) emerged as the most dissimilar, mirroring a similar pattern observed in the southern counterparts. Regions 1 and 3 clustered together, distinct from the eastern region (6), which, in turn, grouped with the double-celled region 4.

When considering all taxa, only four clearly separate clusters were found. The most

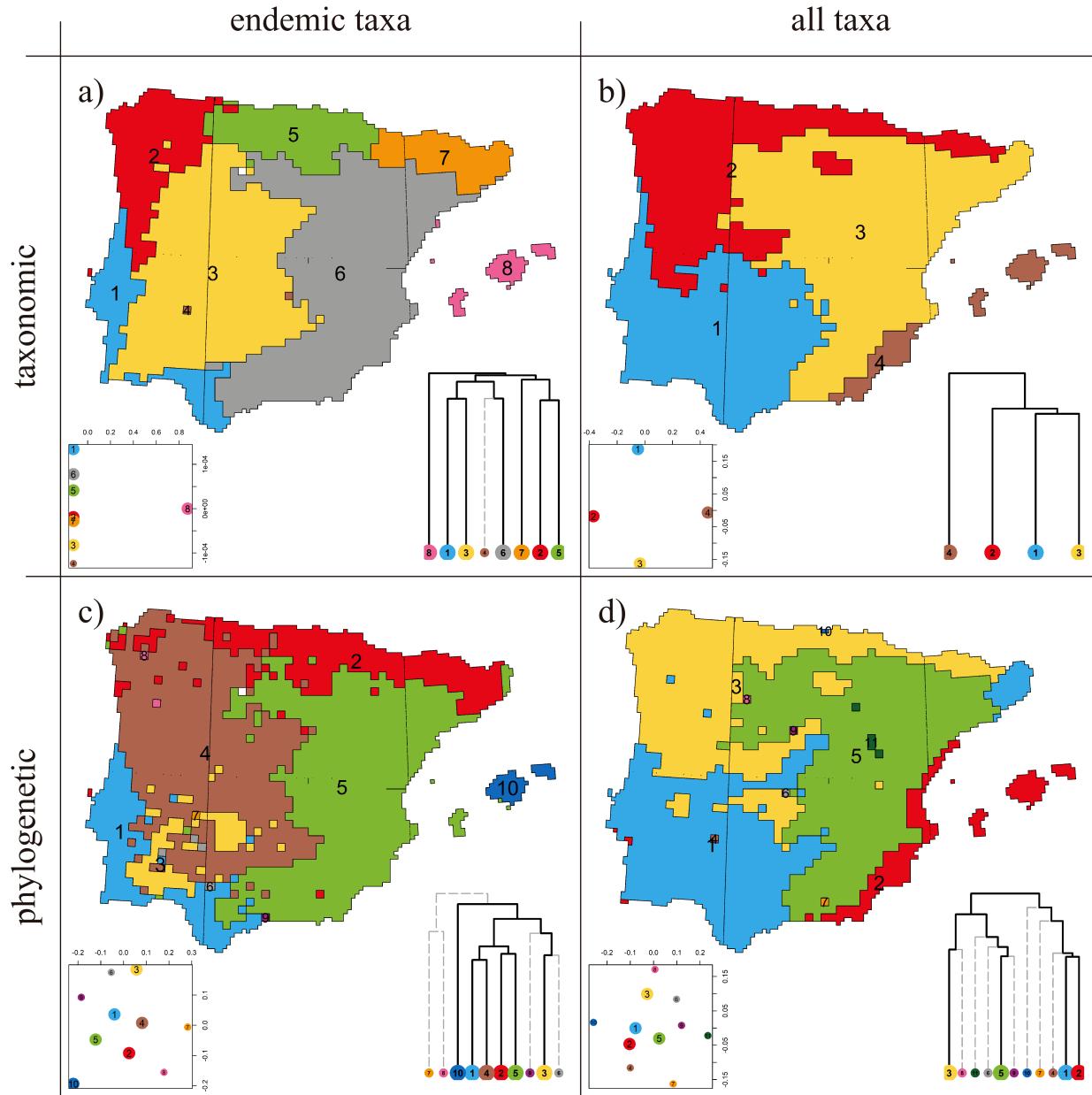


Figure 4.1: Delimitation of floristic bioregions in the Iberian Peninsula attending to their endemic species (subfigures a & c, on the left) and using the whole flora (b & d, on the right), performed using taxonomic (a & b, on the top) and phylogenetic (c & d, on the bottom) approaches. The optimal number of clusters found were 8 (a), 4 (b), 10 (c) and 11(d). Dissimilarity dendograms are plotted in the bottom-right corner of each subfigure and two-dimension NMDS analyses are plotted on the bottom-left one to visualize floristic dissimilarity among bioregions. Regions containing five or less grid cells are plotted with a small circle in the dendograms and NMDS plots. Note that nomenclature for bioregions does not stand for any biological reason; numeric names are automatically assigned in a west-to-east order of appearance.

distinct was the region encompassing the Balearic Islands and the southeastern coast of the Iberian Peninsula (region 4). However, and reproducing the pattern found for endemic taxa, there was later a division between the north-western region (2) and the southern one, which was afterwards divided in a western (1) and eastern (3) regions.

Several disparities were found when evaluating phylogenetic rather than taxonomic dissimilarity. Firstly, for the endemic taxa, the most dissimilar region was found between the easternmost Balearic Islands (Mallorca and Menorca), while the westernmost islands (Ibiza and Formentera) were classified within a bioregion encompassing eastern Iberia. Regardless of the appearance of four minor bioregions and a discontinuous southwestern region (3, most likely an outlier for its remote position in the NMDS plot), the main split among the peninsular bioregions exists between the eastern (1 and 4) and the western ones (2 and 5). Regarding the scheme based on all taxa and phylogenetic distances, two main clusters emerged. Firstly, a zone including the southwestern quarter of the Iberian Peninsula (1) and the Mediterranean coast (2), and secondly another including the northern-northwestern areas (3) and a western one (5).

4.3.3 Environmental characterization of bioregions

Regarding the scheme based on endemic taxa and taxonomic distances, the difference between the Balearic region (8) and the rest could be mainly explained by isolation from mainland (see Table 4.2). The difference between cluster of regions 2, 5 and 7, and the cluster comprising bioregions 1, 3, 4 and 6 differed mainly in their annual precipitation and seasonality, but with a lower discrimination power. Regions 1 and 4 differed from 3 and 6 mainly in the soil pH, and regions 2 and 5 differed from 7 in

Table 4.2: Summarized results of the LDA analysis for each regionalization cluster split, including only the variables summing up to 70% of the importance. Values shown in columns “AUC” and “Importance” are the averaged values of 100 individual analyses. The names of the variables (as developed in Supplementary Table S4.1) stand for the following terms. alt.heter: altitude heterogeneity; ph.mn: mean pH value; bio1.mn: mean annual temperature; bio4.mn: mean temperature seasonality; bio12.mn: annual precipitation; bio15.mn: precipitation seasonality; isolation: distance of an island from mainland. The results shown herein are a summary of those presented on Table S4.3.

Regionalization	clusters	Mean AUC	Variable	Importance (mean±sd)
Taxonomic endemics	8 vs rest	0.987	isolation	-3.46±0.54
			bio12.mn	1.131±0.048
	2,5,7 vs 1,3,4,6	0.909	bio15.mn	-0.936±0.034
			bio4.mn	-0.545±0.021
			ph.mn	2.276±0.06
	1,3 vs 4,6	0.912	alt.heter	0.642±0.036
			bio12.mn	0.541±0.068
			bio4.mn	-1.123±0.107
	7 vs 2,5	0.938	bio15.mn	0.649±0.065
			bio1.mn	-0.551±0.094
Taxonomic whole flora			isolation	-1.187±0.06
	4 vs 1,2,3	0.732	bio1.mn	-0.689±0.047
			bio12.mn	0.623±0.042
			bio4.mn	0.56±0.031
	2 vs 1,3	0.894	ph.mn	0.996±0.043
			bio12.mn	-0.725±0.053
			bio15.mn	-1.351±0.046
	1 vs 3	0.961	ph.mn	1.212±0.05
			bio1.mn	-0.514±0.049
	10 vs rest	1.000	isolation	-3.155±0.453
Phylogenetic endemics			ph.mn	-1.185±0.073
	1,4,2,5 vs 9,3,6	0.5	bio12.mn	-0.721±0.068
			bio1.mn	0.538±0.047
			bio4.mn	0.522±0.064
			ph.mn	1.811±0.044
	1,4 vs 2,5	0.893	bio15.mn	-1.099±0.036
			isolation	0.668±0.062
			alt.heter	0.62±0.029
			bio4.mn	1.513±0.045
	1 vs 4	0.927	bio15.mn	-1.366±0.07
Phylogenetic whole flora			bio12.mn	0.998±0.061
			bio12.mn	-0.718±0.087
	2 vs 5	0.903	bio15.mn	0.766±0.045
			ph.mn	0.703±0.118
			bio1.mn	0.943±0.048
	3,8,11,6,5,9 vs 10,7,4,1,2	0.908	bio12.mn	-0.721±0.043
			bio15.mn	0.57±0.042
			ph.mn	-0.502±0.052
	3,8 vs 11,6,5,9	0.954	ph.mn	2.424±0.075
			bio1.mn	1.837±0.12
4	1 vs 2	0.895	bio12.mn	-1.102±0.143
			bio15.mn	-1.308±0.104
			bio4.mn	-0.951±0.038

temperature seasonality.

However, the inclusion of the whole flora diminished the importance of isolation to define the first split, for the Balearic Islands merged to the southernmost tip of the Iberian Peninsula, while annual mean temperature and precipitation became relevant. Soil pH and annual precipitation were the variables that mostly differed between bioregion 2 and cluster comprising regions 1 and 3, which, in turn, were discriminated by precipitation seasonality and pH.

Resembling the pattern obtained for the taxonomic approach, the main split for endemic taxa using phylogenetic dissimilarities (region 10 *vs* the rest) was explained by high values of isolation. However, none of the bioclimatic variables explained the split between the cluster of regions 3, 6 and 9 and cluster 1, 2, 4 and 5 (AUC = 0.5). Within the latter cluster, regions 1 and 4 differed from 2 and 5 mainly in the soil pH and precipitation seasonality, and between them, region 1 differed from 4 mainly in precipitation regimes, while regions 2 and 5 in seasonality variables. The inclusion of the whole flora draws two clusters of bioregions differing mainly in annual mean temperature and precipitation. Within each cluster, region 3 and region 5 mainly differed in the soil pH, while differences between regions 1 and 2 could be assigned to differences between annual mean temperature and annual precipitation.

4.4 Discussion

The delimitation of biogeographic regions is influenced not only by the quality and quantity of data used (Moreno Saiz et al., 2013) and methodological approaches applied (Kreft and Jetz, 2010), but also by the targeted floristic element (endemics *versus* all species) and the dimension of biodiversity under consideration. The results obtained

herein reinforce the original hypothesis that classical analyses using just endemic taxa may be a biased subset of the broader context.

The number of optimal regions detected for each considered scheme points to differences between them. Firstly, using endemic taxa yielded a higher number of regions compared to the whole flora. This is likely an effect of using a lower number of species in the first scenario, characterized by narrow distributions, thereby enhancing dissimilarities between areas. In contrast, widely distributed species blurred this pattern when analyzing the whole flora. The inclusion of evolutionary history results in a greater number of regions too. A plausible explanation is the effect of phylogenetic uncertainty, as a high proportion of them are outliers (showing peripheral positions in the NMDS plot) and their contour is very discontinuous. Lastly, spatial resolution also affects the number of biological regions to be obtained. For the plants of the Iberian Peninsula, both the coarsest and finest resolutions resulted in a higher number of floristic regions than the intermediate one. We hypothesize the larger number of regions for the coarse resolution (50x50 km) is due to the merging of taxa into same grid cells. This would bold differences among regions as intermediate grid cells, otherwise homogenizing the region, would not have such buffering effect. Contrarily, the finest resolution (10x10 km) detects differences based on real presence and absence of taxa, and thus also shows a higher number of biological regions. These can be interpreted as nested units (e.g. sectors) within those identified at medium-size resolution.

Our bioregionalization schemes based on endemic species likely mirror differential in situ diversification within regions, for they predominantly feature recently diversified and narrowly distributed species (i.e. neo-endemics). Conversely, other ecological factors such as climatic or edaphic preferences, which typically play a significant role

in shaping the distribution of any species, are obscured. Consequently, topographical factors (primarily distance to the mainland in the case of islands, but also mountain chains or river basins) emerge as the most discriminant variables for delineating bioregions when focusing solely on endemic taxa, particularly within the taxonomic scheme. The explanatory power of topographical factors is even observable at different scales (see Figure S4.2a and Figure S4.3a). Our results align with previous findings where the authors found a strong correlation between bioregion delineation and topographical barriers for endemic taxa (Buira et al., 2017; Moreno Saiz et al., 1998). For example, these authors also described the Balearic Islands as the most distinct plant bioregion. The strong dissimilarity among floras drives distance to mainland emerge as the most important factor to divide Iberian bioregions (bioregion 8 *vs* the rest, Figure 4.1a). The Pyrenees and Cantabrian ranges appear as unique floristic regions too. These areas (along with the Balearic archipelago) contain large numbers of narrowly ranged taxa, unique to their respective territories (Guardiola and Sáez, 2023; Ninot et al., 2017). Additionally, river basins and depressions have also served as natural boundaries between endemic taxa bioregions (regions 3 and 6 in Figure 4.1a; Buira et al., 2017). Ultimately, the southwestern coastline of the Iberian Peninsula (bioregion 1) appears robustly throughout literature as a discrete bioregion. This unit may be linked to the presence of a unique littoral flora, which, in turn, is not present on the Mediterranean seashore. Most probably, the species driving the appearance of this region are coastal-specific taxa, but their distribution in the Atlantic shore impede their migration to other territories, an otherwise common scenario in the Mediterranean coastline (Médail, 2022; Molina-Venegas et al., 2015). When examining all the species, environmental variables (primarily annual temperature and pH) took precedence over

topographical factors in explaining differences among bioregions (Moreno Saiz and Lobo, 2008; Moreno Saiz et al., 2013).

The inclusion of evolutionary history, conversely, did not have a comparable impact on the delimitation of bioregions obtained. Probably the lack of such difference is greatly due to the spatial extent of the studied region, where a vast majority of lineages have been able to colonize any point of the area of study at a profound level (unlike observed for global plant phyloregionalizations; see Carta et al., 2022). Nonetheless, phylogenetic schemes revealed several interesting patterns. The use of a pure taxonomic approach defines a major division between the mediterranean and the temperate floristic elements, aligning with previous proposals (Galicia et al., 2010; Rivas-Martínez et al., 2002). The lower importance of such difference for phylogenetic regionalization, points to the existence of an abundant number of species adapted to either mediterranean or temperate conditions (i.e. the Mediterranean and Euro-Siberian floristic elements), but evenly distributed throughout the phylogeny. Moreover, we can observe the existence of a close relationship between the most adjacent Balearic Islands and the alkaline Iberian Peninsula (bioregion 5; Figure 4.1c). This pattern can be understood as a consequence of a continuous history of overseas colonisations, and a lower rate of *in situ* diversification, which can be otherwise observed in Mallorca and Menorca islands (Rosselló and Castro, 2008). This genetic flow might have been facilitated also by temporary events of shoreline approach between Iberia and the nearest Balearic Islands due to eustatic changes in glacial periods (Moreno Saiz et al., 2013). Additionally, phylogenetic regionalization allows to detect bioregions where certain lineages are evolving rapidly. This is especially interesting in a Mediterranean context, where recent and numerous diversification events were triggered by the onset of the mediterranean

climate (Thompson, 2020). An illustrative example is the alpine region 2 in Figure 4.1c, which includes the Pyrenees and Cantabrian mountains, enriched by lineages such as *Alchemilla* or *Saxifraga*, highly diversified therein. Another case of a region driven by high speciation is region 2 in Figure 4.1d, a bioregion rich in lineages of coastal mediterranean environments (such as species from genera *Teucrium*, *Allium*, *Euphorbia* or *Limonium*). This region is similar to region 1 in the same subfigure, including taxa of more humid oceanic coastal conditions, which is split in two geographically distant areas.

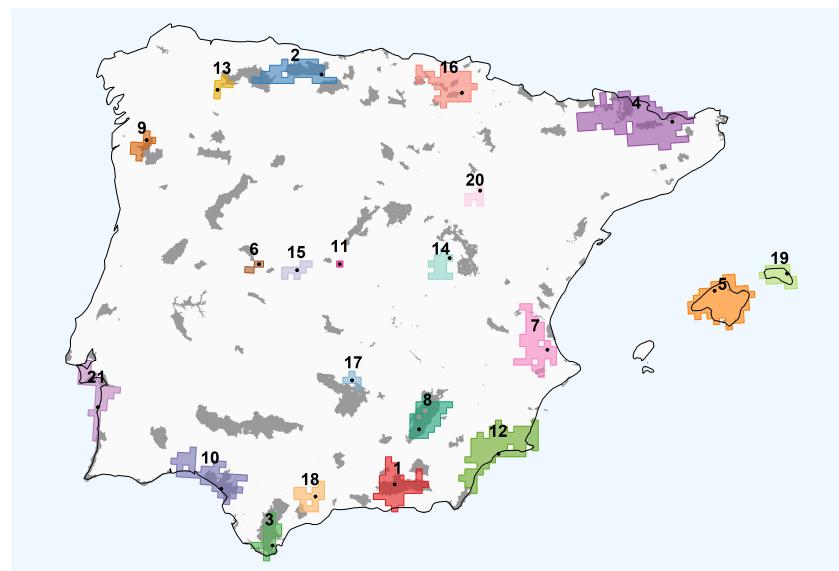
4.5 Conclusions

Our results suggest that the inclusion of non-endemic taxa should be considered, or even preferred, in bioregionalization analyses, because their omission may bias the resultant regions towards in situ diversification processes. The use of just endemic taxa is, on the one hand, more akin to delimit bioregions based on rare species, aggregating topographical entities by means of dispersal limitation and local adaptation. On the other hand, the inclusion of all species inhabiting the focal study area results in a more climatic-based bioregionalization. Also, the inclusion of evolutionary history can be incorporated to detect areas with high rates of phylogenetic clustering. However, phylogenetic bioregions may not substantially differ from a taxonomic scheme depending on the historical and spatial context of the area of study.

Chapter 5

Regional application of a global conservation approach: spatial patterns of EDGE angiosperms in the Iberian Peninsula

5



Chapter in preparation for publication:

Ramos-Gutiérrez, I., Pipins, S., Molina-Venegas, R., Fernández-Mazuecos, M., Jiménez-Mejías, P., Moreno-Saiz, J. C., & Forest, F. (2024). Regional application of a global conservation approach: Spatial patterns of EDGE angiosperms in the Iberian Peninsula.

5.1 Introduction

Earth is currently undergoing a mass extinction event that, for the first time in the planet's history, is being driven by a single species, humans (IPBES, 2019). More than 25% of all extant species are known to be on the brink of extinction, a figure that increases to nearly 45% when considering plants alone (Bachman et al., 2023; IUCN, 2023; Nic Lughadha et al., 2020). This critical situation urges conservation actions aimed at minimizing biodiversity loss caused by human activities, hence the need for analytical tools that allow optimizing the trade-off between well-informed decisions and time investment (Forest et al., 2015; Isaac and Pearse, 2018). A key component of conservation biology is prioritizing species and areas for conservation efforts, the so-called “agony of choice” (Vane-Wright et al., 1991) or “Noah’s ark problem” (Weitzman, 1998). Early attempts to prioritise species were based on either subjective or difficult to quantify criteria (e.g. charismatic or flagship species, keystone species), while priority areas were delimited based on species richness, rarity (e.g. narrowly distributed or endemic taxa), and threat status (Brooks et al., 2006; Sánchez de Dios et al., 2017). However, the availability of biodiversity data has vastly increased in the last decades, opening new opportunities to assess priority species and areas in more precise ways. For example, the rapid development of phylogenetic and genomic techniques has enabled the determination of priority areas lead by the amount of phylogenetic diversity (PD) they harbour. Phylogenetic diversity is a metric compiled by summing up the length of branches that connect a set of terminals on a phylogenetic tree (Faith, 1992). Similar approaches have been used such as prioritising species based on the fraction of unique evolutionary history they encompass (evolutionary distinctiveness; Redding and Mooers, 2006). The latter strategy may in turn facilitate

the delimitation of priority areas that harbour a high number of such evolutionarily distinct species.

Isaac et al. (2007) developed a phylogenetic metric for conservation called Evolutionarily Distinct and Globally Endangered, or EDGE. Rooted in the value-risk trade-off concept of economics (Isaac and Pearse, 2018; Weitzman, 1992), this metric combines the evolutionary distinctiveness (ED) of species with their extinction risk (GE; for Globally Endangered), so that the highest priority is given to the most evolutionarily distinct and endangered species. Therefore, although considerably related to PD, this new index elucidates the most unique portions of the Tree of Life to safeguard. The EDGE approach has been used to prioritize species across several major clades, including gymnosperms (Forest et al., 2018), chondrichthyans (Stein et al., 2018), and tetrapods (Gumbs et al., 2018), and a priority list for angiosperms is currently in progress (F. Forest et al., *personal communication*). Moreover, the framework has been successfully applied to a tangible conservation initiative that focuses on species deemed to be both threatened and evolutionarily distinct, the EDGE of Existence Programme (<https://www.edgeofexistence.org/>). The EDGE approach involves, by definition, a global assessment across all known species in a clade of interest, which has lent itself towards global prioritisation studies (Pipins et al., 2023; Safi et al., 2013), though the framework is theoretically applicable to any spatial scale. Highlighting regional priorities of EDGE species can help to link national conservation efforts to global conservation goals, such as safeguarding the Tree of Life (Carta et al., 2019; Gumbs et al., 2023a). However, to the best of our knowledge, no regional EDGE analyses incorporating an entire flora have been conducted thus far. Considering that policymaking is typically implemented at country level, regional assessments of the

geographical patterns of EDGE species may be instrumental to prioritize effective and well-informed conservation actions that take into account the evolutionary dimension of biodiversity.

Here, we conduct the first regional EDGE assessment using the angiosperm flora of the Iberian Peninsula, in the western Mediterranean basin, as a case study. This region is home to almost 5,400 angiosperm species (over 20% of the Mediterranean angiosperm flora; Ramos-Gutiérrez et al., 2021; Rundel et al., 2016), one of the most angiosperm-rich regions across the entire Mediterranean basin (Araújo et al., 2007), and an exceptional centre of endemism (Buira et al., 2021; Cai et al., 2023; Médail and Quézel, 1999), thus representing an ideal setting for our regional EDGE assessment. Specifically, we aim to identify areas of high conservation value due to the presence of large numbers of EDGE species and provide a list of priority zones identified using a complementarity approach that maximizes EDGE values across the region. Additionally, we explore how angiosperm phylogenetic diversity is distributed across the region using comprehensive phylogenies and up-to-date chorological information.

5.2 Methods

5.2.1 Taxonomic and chorological data

The global list of angiosperms was retrieved following the World Checklist of Vascular Plants (WCVP version 6, September 2021; Govaerts et al., 2021), which comprises a total number of 329,798 species. The checklist of angiosperm species native to the Iberian Peninsula and their distribution ranges were retrieved from the AFLIBER database (Ramos-Gutiérrez et al., 2021), which records the presence of vascular plant species in the study area at 10×10 km resolution (totalling 6,150 UTM grid cells). To

homogenize the taxonomic treatment considered therein, subspecific taxa were merged into their respective species, and names were synonymized to WCVP, resulting a total of 5,390 Iberian angiosperm species.

5.2.2 Phylogenetic data

Due to the lack of a phylogenetic tree for the whole angiosperm species, we used a generated set of 100 phylogenies by expanding the seed plant phylogeny of Smith & Brown (2018), after excluding non-angiosperm lineages, using the R package ‘V.Phylomaker’ (Jin and Qian, 2019). Missing species from the backbone tree (phylogenetically unplaced taxa, PUTs) were bound to a randomly selected node descending from the crown node of its genus, or family in case of missing genera (scenario 2). If the congeners of a PUT formed a polyphyletic group in the backbone tree, the PUT was imputed to a randomly selected node within the largest cluster of the genus in the backbone tree. The imputation of missing species was repeated 100 times to account for their uncertain phylogenetic placement onto the backbone tree. The phylogenies used for the analyses can be accessed online (Forest, 2023).

5.2.3 Global extinction risk data

Information on global extinction risk of all angiosperm species was retrieved from the IUCN Red List database (IUCN, 2021). For species without formal Red List assessment (which added up to the 84.9% of them), we used prediction of extinction risk using a machine-learning algorithm that allocated each species as either threatened or non-threatened species (Bachman et al., 2023). The distribution of extinction risk categories for the Iberian angiosperm species can be seen in Table 5.1, while the spatial distribution of species included as threatened (i.e. Critically Endangered, Endangered,

Table 5.1: Distribution of global extinction risk categories across the 5,390 Iberian angiosperm species. The species assessed by the IUCN Red List are assigned to the categories EX (Extinct), CR (Critically Endangered), EN (Endangered), VU (Vulnerable), NT (Nearly Threatened), and LC (Least Concern), whilst the remaining species are treated as “threatened” or “non-threatened” species based on the machine-learning predictions. Six species remained unclassified due to taxonomic mismatches between the checklists used to build phylogenies and to retrieve UICN information.

Extinction risk category	Number of species
EX	2
CR	36
EN	54
VU	50
NT	60
LC	707
Other threatened	726
Other non-threatened	3,749
Not Assessed	6

Vulnerable or assigned as threatened by the algorithm) in the Iberian Peninsula is shown in Figure 5.1.

5.2.4 EDGE score calculation

We calculated the EDGE score of every Iberian angiosperm species following the EDGE2 framework (Gumbs et al., 2023b). EDGE2 is an extension of the original EDGE framework that employs a continuous distribution of extinction probabilities instead of discrete categories. Furthermore, the EDGE2 approach accounts for PD complementarity by integrating the extinction risk of close relatives when calculating ED for a given species (Gumbs et al., 2023b). Briefly, each internal branch of the tree is scored with a probability of extinction based on the combined extinction risks of all descendant branches so that two equally evolutionarily distinct and threatened species may still show different EDGE scores if their close relatives show different threat statuses. The EDGE score calculation was performed for each of the 100 expanded phylogenies, and values for each species were averaged afterwards for posterior analyses.

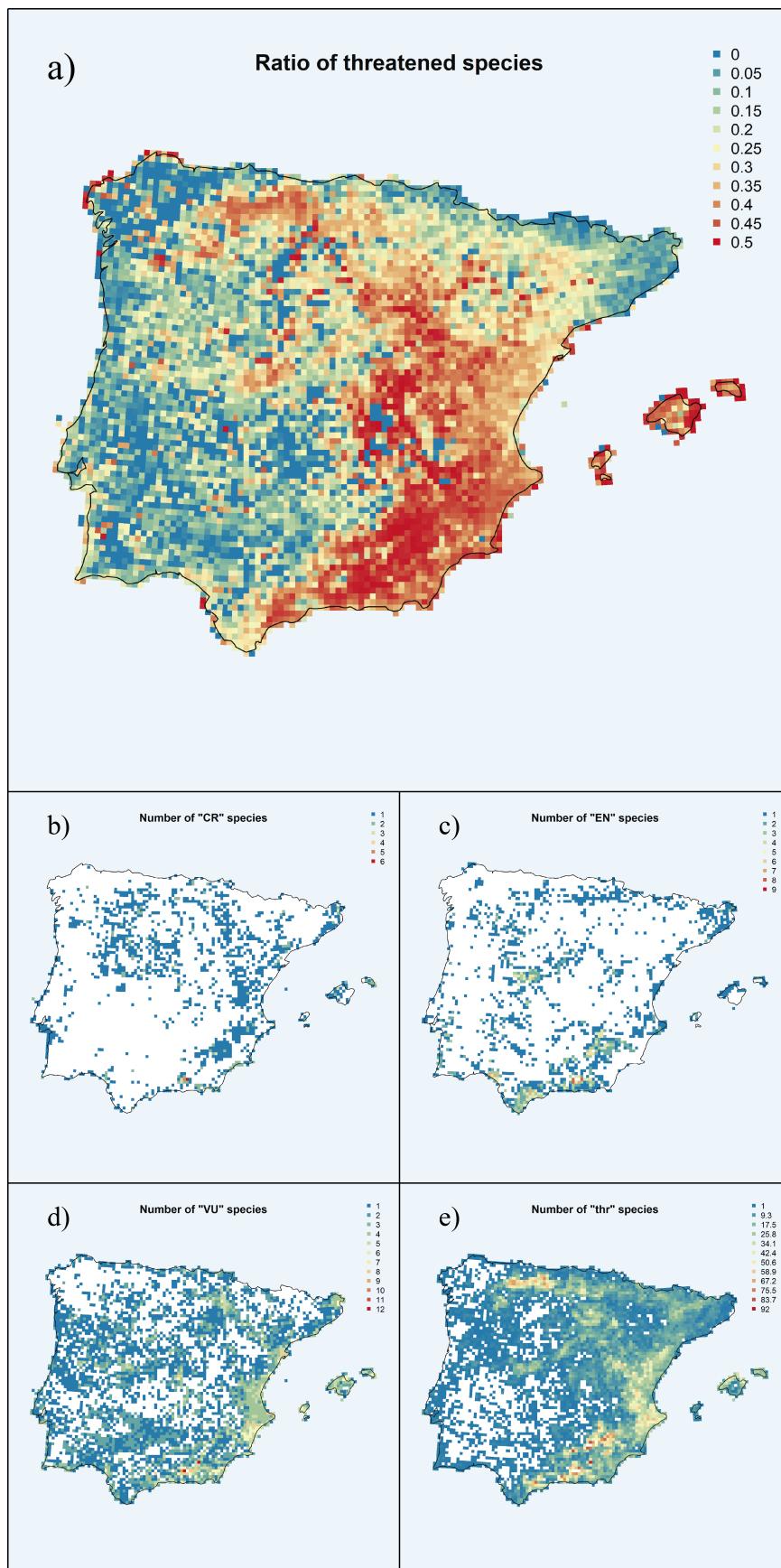


Figure 5.1: Spatial distribution of threatened species per grid cell in the Iberian Peninsula. Panel a) depicts the ratio of threatened taxa per site for all categories, calculated as the total number of threatened species divided by total number of species. Latter subfigures show the number of threatened taxa per grid cell and risk category, namely Critically Endangered (b), Endangered (c), Vulnerable (d) and detected as potentially threatened by a machine learning algorithm (e).

Species threatened with extinction (i.e. assessed as CR, EN or VU categories, or classified as threatened by machine learning), and whose EDGE value was above the median in at least 95% of the trees, were considered “EDGE species”, i.e. priority species for conservation based on their evolutionary distinctiveness and global endangerment.

5.2.5 Spatial analyses

We computed the PD of each 10x10 km grid-cell (PD_{obs}) as the minimum spanning path connecting all the species occurring within it. Because PD strongly correlates with species richness, we computed a standardized effect size (SES) score for each cell using the function *ses.pd* of the R package ‘picante’ (Kembel et al., 2010), using 999 runs of the ‘taxa shuffle’ null model approximation. Non-parametric significance values were employed with an alpha value of 0.05. Cells whose PD_{obs} appeared under the 2.5th percentile were therefore considered as significantly PD poor cells, whereas cells with a PD_{obs} over the 97.5th percentile were considered significantly PD rich cells. The EDGE value ($EDGE_{obs}$) of each grid cell was calculated by summing up the EDGE scores of the species present in it, and a SESEDGE score for each grid cell was computed following the parameters described above.

Finally, we identified the priority areas of the Iberian Peninsula to safeguard evolutionary distinctiveness through an EDGE complementarity analysis (*sensu* Vane-Wright et al., 1991; see also Pipins et al., 2023). First, we identified the top-EDGE grid-cell across the study area, and all adjacent (i.e. sharing at least one side) and floristically similar grid-cells were agglutinated to the first-rank cell. The rationale for this grid cell agglutination is that selecting areas based on single grid cells may leave adjacent areas of high EDGE value unprotected due to incomplete sampling. Two grid

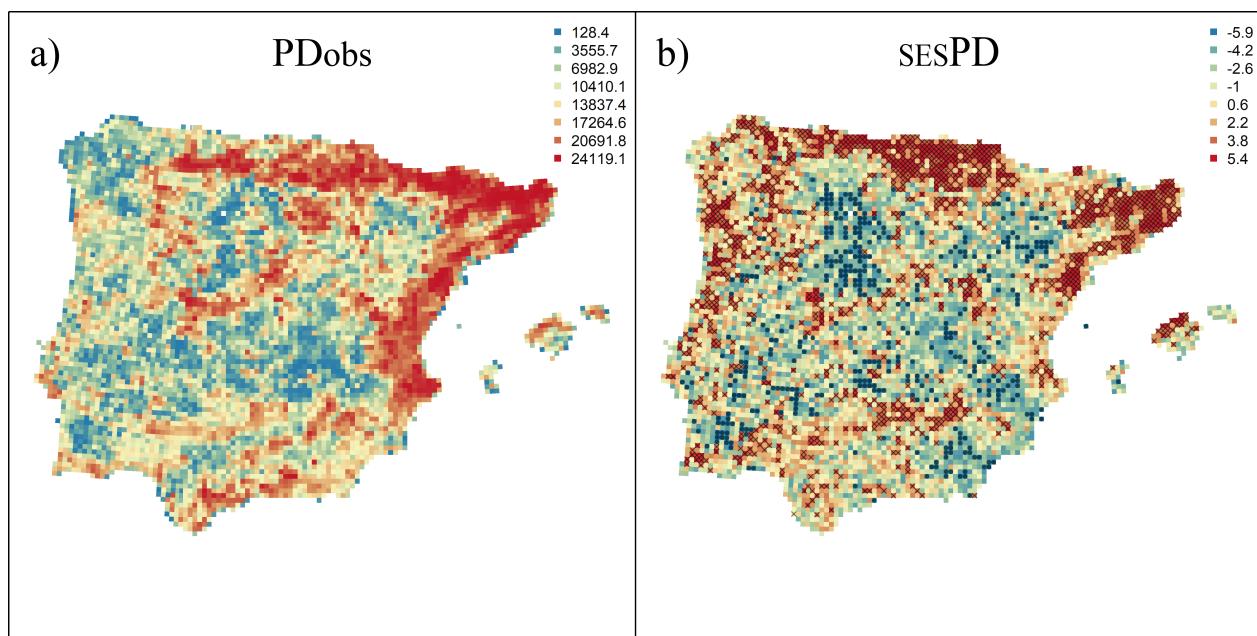


Figure 5.2: Distribution of Phylogenetic Diversity for angiosperm species in the Iberian Peninsula; a) Observed Phylogenetic Diversity; b) Standardized Phylogenetic Diversity (SES scores). Dots mark grid-cells with SES score below the 2.5th percentile (significantly low values; phylogenetic clustering), while crosses identify grid-cells above the 97.5th percentile (significantly high values; phylogenetic dispersion).

cells were deemed floristically similar if their taxonomic dissimilarity, as measured by beta turnover, laid below the 2nd percentile in the distribution of values between the focal cell and the remainder. Then, all the species included in the resulting EDGE area were removed from the dataset, and the next top-EDGE grid cell was identified, and adjacent and floristically similar cells were agglutinated as described above. The procedure continued iteratively until the complementary areas captured 90% of the total summed EDGE values of all Iberian species. Pairwise beta-turnover between grid cells was computed with the *beta.pair* function of ‘betapart’ R package (Baselga et al., 2022).

5.3 Results

Observed phylogenetic diversity (PD_{obs}) is highest over most of the mountain ranges of the Iberian Peninsula (Cantabrian Range, Pyrenees, Central and Baetic Ranges, as well as most of the Mediterranean shore; see Figure 5.2a). This pattern is relatively well correlated with the distribution of species richness (cf. Figure 5.3a). Upon correction to avoid the effect of richness (SES_{PD}), markedly elevated values of phylogenetic evenness in nearby montane and temperate areas with more stable climates are shown. In contrast, continental depressions such as plateaus and river basins, as well as the south-eastern dry region, exhibit notably lower values indicative of phylogenetic clustering (Figure 5.2b).

A total of 160 Iberian taxa identified as EDGE species (3% of the total in the region; see Table S5.1), of which 82.5% are Iberian endemics. These species belong to 41 families, with a large proportion of them found in just a few families, namely Brassicaceae (24 species), Caryophyllaceae (17), Plantaginaceae (13), Fabaceae (12) and Geraniaceae (11) (see Figure S5.1). Most Iberian EDGE species thrive in the mountainous regions of central, northern, and eastern Spain (Figure 5.3b), yet the highest SES_{EDGE} scores for the EDGE species are found along the dry southeastern coasts (Figure 5.3f). Similar patterns are found when considering the entire flora (Figure 5.3a,c,e), with high significant values clustered mainly along the south-eastern coast, except for the high SES_{EDGE} values observed in the northwestern corner of the Peninsula, indicating a the presence of species not classified as EDGE but significantly contributing to high EDGE values.

The selection of top priority regions of threatened evolutionary history – termed

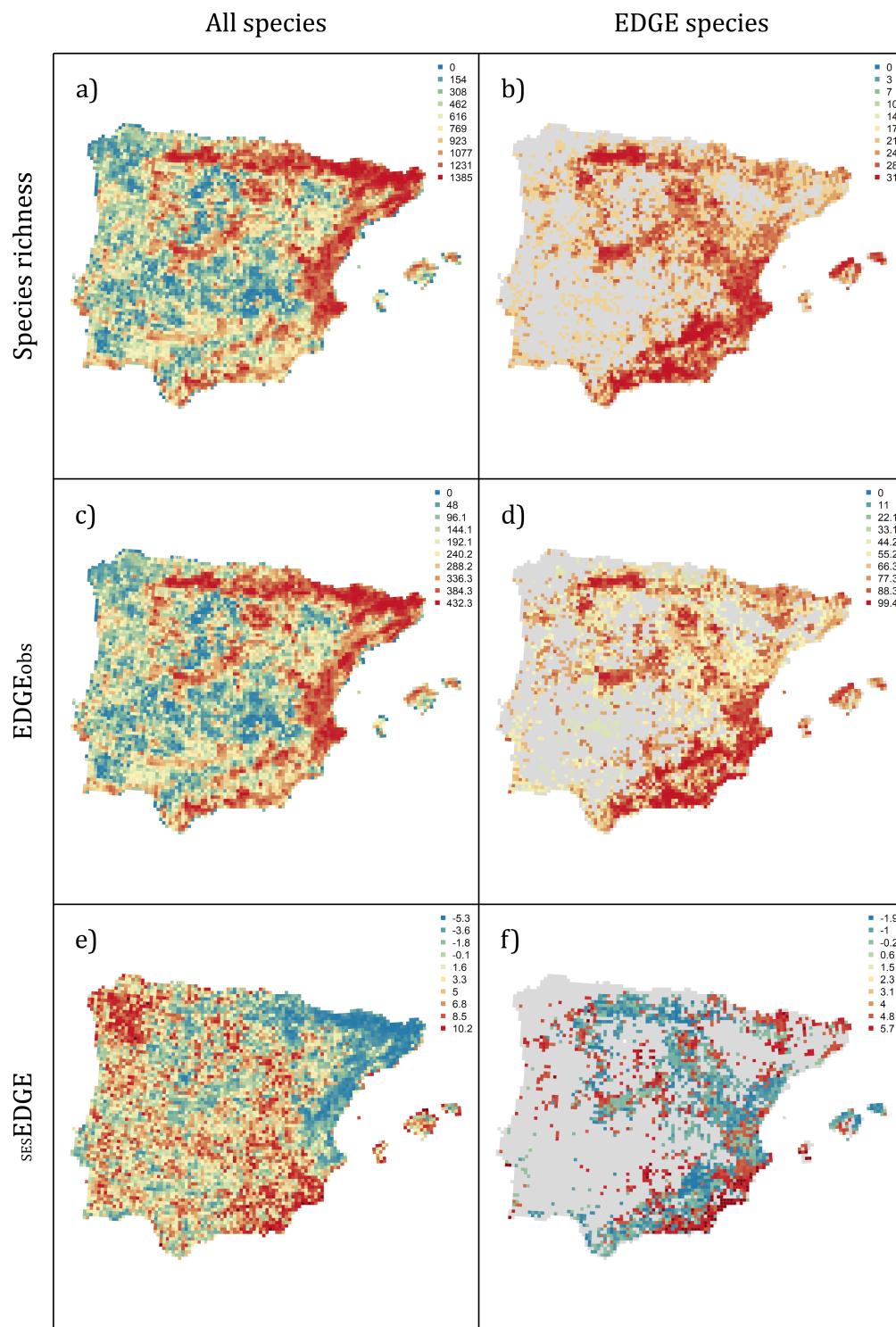


Figure 5.3: Distribution of EDGE values across the Iberian Peninsula for all species (left column) and for the 160 taxa identified as EDGE species (right column). Species richness in each grid cell is shown in panels a) and b); panels c) and d) show EDGE_{obs}, the sum of EDGE values in each grid-cell, and panels e) and f) show SESEDGE values, where dots indicate cells with statistically high SESEDGE scores.

‘EDGE zones’ (Pipins et al., 2023; Safi et al., 2013) – which secure 90% of angiosperm regional EDGE in the Iberian Peninsula resulted in 21 priority areas. Several of these areas occur along mountain ranges (e.g. EDGE zones 1, 2, 4, 6, 8, 13, 14), coastal regions (zones 7, 10, 12, or 21) or islands (zones 5 and 19 in Majorca and Minorca islands, respectively), with some exceptions (zones 3, 9, 11, 17 and 20). The appearance of each EDGE zone through the iterative process, along with the amount of regional EDGE captured and the proportion of area already protected are shown in Table 5.3. Although most of the EDGE zones partially overlap with protected areas, only 19% of the total extent of EDGE zones occurs under protected surface, and some zones (namely EDGE zones 6, 11, 14, 15, and 20) do not enjoy any preservation figure (national parks, natural parks and natural reserves) at all. The geographic occurrence of EDGE zones, along with the coverage of different figures of natural protected areas is depicted in Figure 5.4. The individual 10x10 grid cells selected in each iteration are also listed in Table S5.2 and are plotted as black dots within each EDGE zone in Figure 5.4.

5.4 Discussion

Halting biodiversity loss is a major commitment of the 2030 Agenda for Sustainable Development of the United Nations (goal 15), which demands effective conservation actions, particularly in biodiversity hotspots such as the Iberian Peninsula (<https://sdgs.un.org/goals>). Here, we present for the first time a regional EDGE assessment in one of these biodiversity hotspots, which represents a step forward in the identification of areas in need of protection within the Iberian Peninsula. Importantly, our evolutionary-based analysis can help regional Spanish and Portuguese policymakers to delimit

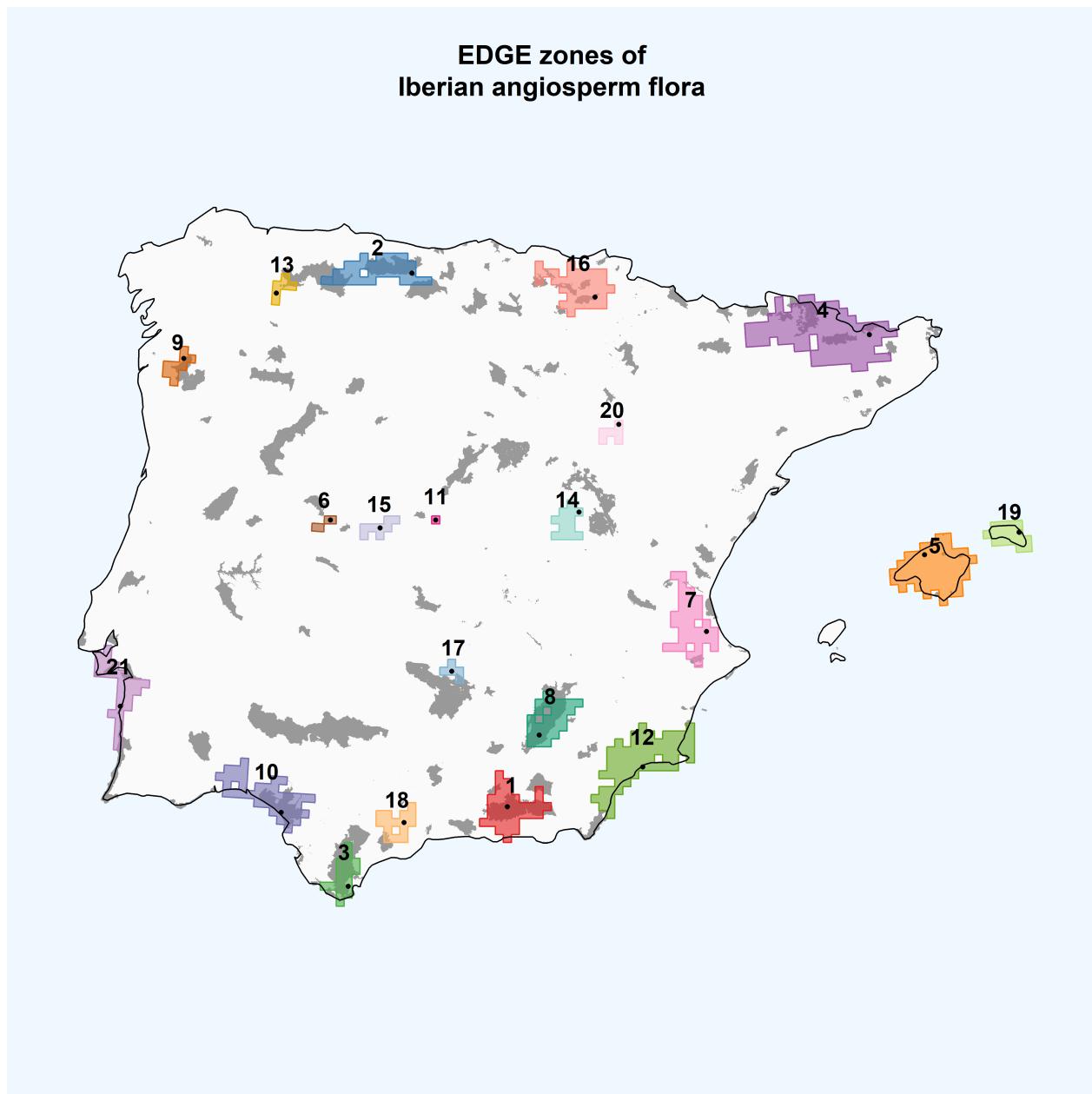


Figure 5.4: EDGE zones for the Iberian Peninsula angiosperm flora. The numbers identifying each zone match the EDGE zone numbers in Table 5.3 and Table S5.2. Black dots represent the selected grids in each EDGE zone. National parks, natural parks and natural reserves are shown in gray.

Table 5.3: List of EDGE zones and proposed names. The amount of regional EDGE captured by each zone is also shown, and the percentage of protected surface in each of them (under national park, natural park or natural reserve figures). The global percentage of EDGE zone protected under preservation figures is 19%.

EDGE zone	Proposed EDGE zone name	Individual captured EDGE (%)	Cumulative captured EDGE (%)	Protected surface (%)
1	Sierra Nevada	34.53	34.53	46.7
2	Picos de Europa	14.85	49.38	49.1
3	Alcornocales	5.41	54.79	69.2
4	Central-eastern Pyrenees	9.51	64.30	14.4
5	Mallorca island	4.85	69.15	0.9
6	Hervás	1.77	70.92	0.0
7	Valencia	3.12	74.04	3.5
8	Sierras de Cazorla y Segura	2.21	76.25	59.3
9	Peneda-Gerês	1.27	77.52	27.6
10	Doñana	2.00	79.52	27.9
11	Cadalso de Los Vidrios	1.00	80.52	0.0
12	Cartagena	1.83	82.35	0.3
13	Sierra de Ancares	1.30	83.65	24.8
14	Serranía de Cuenca	1.01	84.66	0.0
15	Sierra de Gredos	0.96	85.62	0.0
16	Basque Country	0.90	86.52	17.1
17	Puertollano	0.70	87.22	18.3
18	Sierra de las Nieves	0.97	88.19	10.3
19	Menorca island	0.92	89.11	3.2
20	Calatayud	0.45	89.56	0.0
21	Arrábida and Alentejo coastline	0.68	90.24	16.4

top-priority areas for the preservation of a huge amount of unique evolutionary history that is largely threatened. In the Iberian Peninsula, these areas of high conservation interest are found evenly distributed over the territory, often linked to coastal or montane regions, most of them poorly preserved or not covered at all under the current network of protected areas. Our results show a poor coverage of important plant areas within legally protected areas in Spain and Portugal, consistent with previous works in the subject for vascular plants (Araújo et al., 2007; Castro Parga et al., 1996) and other biological groups (Araújo, 1999; Lobo and Araújo, 2003; Sérgio et al., 2000).

Using measurable indices to prioritize areas for their long-term preservation allows

to diminish bias in political decisions. Species richness has been a major driver of the delimitation of important areas (Anderson, 2002; Plantlife, 2010; Sánchez de Dios et al., 2017), but new methodologies allow a more refined integration across the various components of biodiversity, extending the mere co-occurrence of taxa (Carta et al., 2019; Marchese, 2015). Phylogenetic diversity (PD) is a relatively straightforward index allowing such measurement, calculating the proportion of all the extant evolutionary diversity a site contains (Faith, 1992). Additionally, it can be combined with conservationist information for more efficient biodiversity safeguard policies (Isaac et al., 2007; Nee and May, 1997; but see Winter et al., 2013). In the study area, the richest areas for vascular plants are found in mountain ranges (Castro Parga et al., 1996; Lobo et al., 2001; Ramos-Gutiérrez et al., 2021), which also stand out as the most phylogenetically biodiverse regions, presenting high values of both PD_{obs} (especially in high-mountain areas) and SES_{PD} (especially in medium height montane areas). Several explanations can be put forward to explain these patterns. These areas are heterogeneous in terms of micro-climatic and geomorphological features (Antonelli et al., 2018), hence offering a wide variety of habitats in relatively small areas that may allow the coexistence of functionally disparate plant lineages at the macroecological scale, also being cradles for new species by means of radiations (Lobo et al., 2001; Rahbek et al., 2019). Ultimately, mountain regions in the area of study (both high and medium altitude ones) have been less impacted by human activities than the lowlands (Castro Parga et al., 1996), which may explain the persistence of disparate lineages in mountainous areas, resulting in high phylogenetic diversity levels.

Besides the use of PD as an index to measure the importance of investing conservation efforts in a given area to preserve the Tree of Life, the calculation of EDGE

values allows to incorporate the urge of preserving individual species on the basis of their evolutionary singularity and extinction thread. The distribution of EDGE values on the Iberian Peninsula somewhat resembles patterns to those observed for species richness, with mountain areas highlighted as EDGE species-rich regions. However, the southeastern coast of the Iberian Peninsula also appears as a priority region in terms of EDGE species. This area stands as a unique climatic region within the Peninsula, being its driest, warmest, and most arid region. This environmental harshness may explain the existence of taxa specialized to these conditions, driving to the presence of a high number of endangered taxa (see Figure 5.1 and Muñoz-Rodríguez et al., 2016) and, hence, high SESEDGE scores found there. These, along with the conversely low SESPD values found in the region, suggest the existence of a significant number of narrow-ranged neo-endemic species (Buira et al., 2021). The rarity of these narrow endemics grants them with a high global conservation interest value (i.e. high GE; Domínguez Lozano et al., 2003). However, this region harbours a few species which grant high EDGE values for their high ED. This may be the cases for species such as *Cynomorium coccineum* (Cynomoriaceae), *Gadoria falukei* (Plantaginaceae), *Patellifolia procumbens* (Amaranthaceae) or *Lafuentea rotundifolia* (Plantaginaceae), which belong to mono or bitypic genera, and hence present a great evolutionary distinctiveness. Indeed, this region is the only one containing significantly high SESEDGE values when considering the whole of Iberian angiosperms (Figure 5.3). The strikingly low SESEDGE values observed throughout the Pyrenees and northeastern Mediterranean coast might can be understood in the light of their low SESPD values (which are mainly due to a high incidence of neo-endemic monocots; see Buira et al., 2021) and the relatively low number of globally endangered taxa thriving there (see Figure 5.1). These figures,

opposite to the general pattern observed in the rest of Iberian mountains, can be attributed to the great number of species shared with other neighboring temperate or alpine regions (such as the French Massif or the Alps).

Our list of priority areas consists of 21 unique regions whose preservation would secure more than 90% of the Iberian EDGE value. These results align with previous studies regarding priority areas for plant conservation using different methodologies. Several of the EDGE zones match those considered by Castro Parga et al. (1996) and Sánchez de Dios et al. (2017). Areas described therein as important for conservation (e.g. Sierra Nevada, eastern Pyrenees, Sierra de Algeciras, Gata Cape or San Vicente in the first, or the Cantabrian and Cazorla Ranges and the arid southeastern region in the latter) were also identified as EDGE zones. However, we can observe some discrepancies among results that may arise from variations in methodological procedures. Central Pyrenees is an important plant area for its species richness, but contrarily is neither an EDGE zone nor a priority area for conservation using other indexes (Castro Parga et al., 1996). As a novelty, no previous study defined priority areas due to the inclusion of a low number of highly distinct species.

Some of the EDGE zones are to some extent included in national-level networks of protected natural areas, but only six and two of them have more than 20% and 50% of their territory covered, respectively. These zones with high proportion of protected surface mainly match with mountainous EDGE zones (except for Doñana –EDGE zone 10–, whose protection was not driven by plants but mainly animals); a bias which has already been described for different areas of the world (see Rouget et al., 2003 and references therein). Unprotected EDGE zones are mainly those based on the presence of a single high EDGE species, such as *Gyrocarpum oppositifolium* in Cadalso de los

Vidrios (EDGE zone 11) or *Pseudomisopates rivas-martinezii* in Sierra de Gredos (zone 15). These areas (see also zones 6, 17, 20), therefore, have traditionally been overlooked in conservation programmes, as they are not part of species-rich regions, while their interest rely on the presence of certain distinct taxa. Given their low extent and number of focal taxa for conservation, they emerge as ideal scenarios for microrreserve protection management (Laguna et al., 2001). Our results underscore the importance of considering multiple dimensions of biodiversity, and evolutionary distinctiveness in particular, to prioritize conservation efforts.

As previously stated, the need for prioritisation tools in conservation planning is urgent (Thompson, 2020), hence the need for this type of analyses, even if the datasets are to some extent incomplete. However, there is still room for improvement in the use of EDGE as a fundamental index for conserving biodiversity. Increasing information availability related to the conservation status of species (i.e. updating and completing the IUCN Red List, including for example regional red list categories) would be instrumental to optimise the results of EDGE analyses and making them more precise. We believe that a sensible starting point might be using machine learning approaches to predict the extinction risk of unassessed species (even common species, whose assessments surely would result in “Least Concern”), so their probability of extinction can be accounted for in a more precise way. A first logical step would be to prioritise the formal assessment of species on the EDGE species list that have not yet been assigned a Red List category (Bachman et al., 2023). Another major step forward would be the inclusion of the greatest possible number of species in the backbone phylogeny. The greater the number of phylogenetically placed taxa, the lower the phylogenetic uncertainty introduced by species missing molecular information (Ramos-

Gutiérrez et al., 2023; Rangel et al., 2015). Continuing to sequence the genomes of unsampled species is therefore key to refine these analyses and other studies using metrics derived from mega-phylogenetic trees. Nonetheless, we believe that the spatial patterns reported here are robust and unlikely to change radically, as they have been calculated using state-of-the-art methodologies and an up-to-date spatial dataset (AFLIBER) that stands out for its exceptional completeness and resolution, with few regional equivalents. The results obtained herein may help managers to select important areas for conservation of the flora they shelter. This task has already begun with the proposal submitted to the Spanish Ministry of Ecological Transition of including, among others, the top-1 EDGE species in the Iberian Peninsula (i.e. *Gyrocaryum oppositifolium*) in the list of endangered taxa, performed by the Spanish Plant Conservation Society (Moreno Saiz and Martínez García, 2023). This work represents a step forward in establishing synergies between phylogenetics, conservation biology, and policymaking at the regional level, and we hope that it will serve as a springboard for further application of the EDGE methodology in other regions of high biodiversity value.

Chapter 6

Discussion

6.1 An updated and comprehensive distributional dataset

Botanists across the Iberian Peninsula have undertaken a monumental effort in compiling and publishing plant chorological information. Project Anthos (<http://www.anthos.es/>) has long served as a primary consultation repository, including nearly 2 million occurrence records of vascular plants in the Spanish, Andorran and Portuguese territories. However, it has not been updated for a decade now, and data included therein yet include faulty records posteriorly detected and amended elsewhere. Additionally, the impossibility of downloading the whole database has impeded its generalised use to conduct biogeographic or conservationist studies. Complementary to it, there is a high number of regional repositories of distributional data, both publicly available (such as the Portuguese interactive Flora, <https://flora-on.pt/>; the Pyrenean Flora Atlas, <https://atlasflorapyrenaea.eu/> or the Valencian Community Biodiversity Data Bank, <https://bdb.gva.es/>) or access-restricted (e.g. the Threatened Flora Atlas of Andalusia, or the Atlas of Wild Flora of Castille and Leon), and originated from a range of sources, varying from ministerial information to citizen science projects. On top of that, there is a great amount of grey literature published yearly describing both

new species and occurrences that were being kept aside in atlases, as well as information available in global repositories such as GBIF (<https://www.gbif.org/>). These circumstances granted a great opportunity to design the first chapter of this dissertation and to publish the resultant dataset as an open access resource (Ramos-Gutiérrez et al., 2021).

The appearance of the AFLIBER dataset thus constituted a milestone in chorological information availability in the Iberian Peninsula, not just in terms of information compilation, but also in carefulness in the curation of the data, being all the species individually evaluated after treasuring occurrences. However, we acknowledge that without continuous curation and updates, this dataset may face the same fate as previous initiatives, losing its potential as a valuable reference tool. Efforts have been dedicated to ensuring easy accessibility for all potential

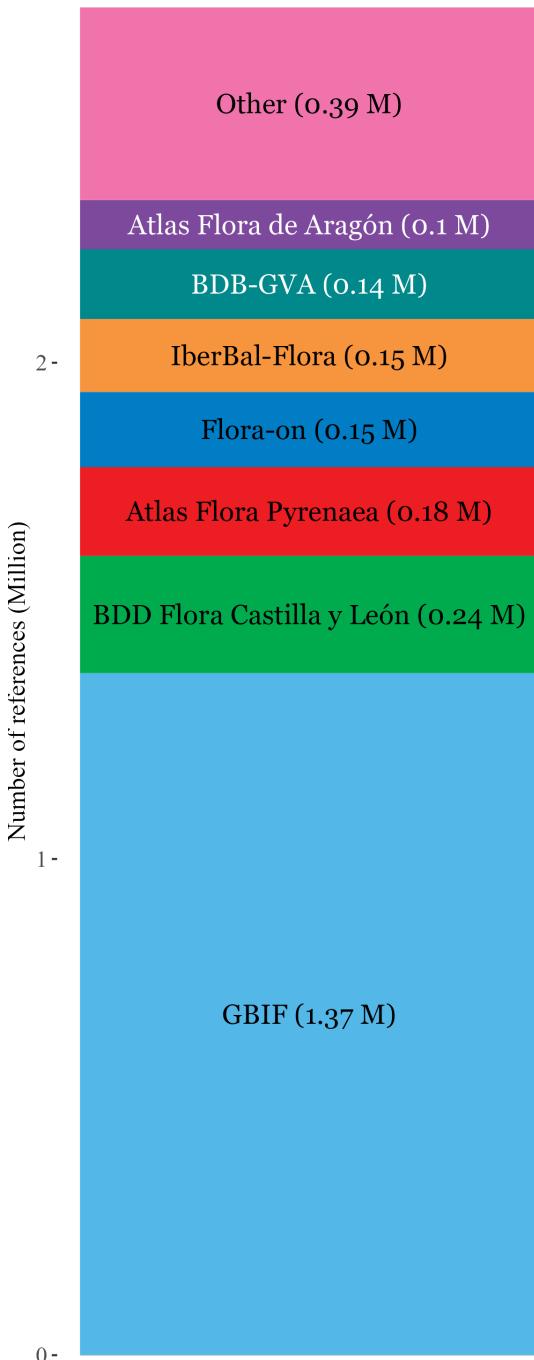


Figure 6.1: Number of occurrences regarded in AFLIBER by data source. Only the seven bulkiest datasets are shown individually, while the remaining 151 sources (ranging from 1 to 57,000 occurrences) were merged as “Other”.

users by publishing all data in open-access repositories. This includes hosting a website repository for the data (<https://iramosgutierrez.github.io/afliber/>) and an interactive *Shiny* application (<https://afliber.shinyapps.io/afliber/>) allowing users to visualize and download the data. Additionally, these efforts have led to the development of an R package for plotting plant distributions (Rodríguez-Sánchez, 2023). Furthermore, the publication of AFLIBER has spurred the establishment of a chorology working group in within the Spanish Botanical Society (<https://www.sebot.org/>), which is dedicated to the ongoing updating and correction of information contained within the atlas. Currently, a new version is in development, set to incorporate over 150,000 new occurrences.

6.2 Phylogenetic expansion

6

Integrating phylogenetic information into macroevolutionary analyses allows for the examination of working hypotheses in the light of evolution. However, a significant portion of biodiversity remains outside of phylogenetic frameworks. For instance, angiosperms, a relatively well-known group, present only ~12.5% of their known species included in the largest published phylogeny (Janssens et al., 2020). Additionally, a large proportion of biodiversity is yet to be described (May, 1988), what implies that phylogenetic uncertainty will be a constant challenge for evolutionary ecologists in the foreseeable future. Incorporating an evolutionary dimension into macroecological analyses entails two alternative strategies, each with its advantages and limitations. One approach is using non-molecular data, such as taxonomic information, to delineate a narrow phylogenetic range where species are likely to belong and then assign a random phylogenetic position (Rangel et al., 2015). Alternatively, a more straightfor-

ward method involves excluding taxa whose phylogenetic position remains unknown (FitzJohn et al., 2009; Mishler, 2023). However, adopting the latter approach often means disregarding a significant amount of data, potentially leading to a loss of representativeness and bias. Moreover, the potential tendencies introduced by randomly assigning species to a backbone phylogeny can be mitigated by generating multiple independent phylogeny replicates to test the robustness of results (Davies et al., 2012).

In the case of the Iberian Peninsula, only ~44% of all taxa (2,818 species) were phylogenetically placed. Failing to utilize the remaining taxa would mean disregarding a 35% of the total occurrence data for more than half of the Iberian species, many of which have well-documented distributions and include a significant number of endemic taxa. Consequently, the analyses conducted in this dissertation adhered to the phylogenetic expansion scheme. Additionally, for the final chapter, phylogenetic expansion was mandatory to undertake a global analysis of evolutionary distinctiveness (Gumbs et al., 2023b).

6.3 Method selection in phylogenetic expansion

Phylogenetic expansion hinges significantly on the quality and appropriateness of the input data used, which inevitably shapes the outcomes of subsequent analyses. The concept of MDCCs (Most Derived Consensus Clades), as fined by Rangel et al. (2015) denotes a clade where a Phylogenetically Uncertain Taxon (PUT) unequivocally finds its place. Therefore, each individual PUT should be evaluated separately, one at a time. Some previous software were designed to perform phylogenetic expansion automatically (as Jin and Qian, 2019, 2023), relying solely on pre-loaded taxonomic information. Critically, this approach restricts users' ability to adjust parameters

according to the specifics of each case, potentially leading to blind reliance on the output without proper scrutiny. The phyletic nature of the group where a PUT is to be bound (namely singleton or mono-, para-, or polyphyletic) can dramatically bias its final position, as several schemes can be followed (as shown in Figure 3.3). Moreover, unreal group topologies are especially common in vast backbone phylogenies, due to incorrect labelling of tips in huge datasets (Pentinsaari et al., 2020) that are not evaluated (for example, in the last plant mega-phylogeny used in Jin and Qian (2023), angiosperms appear as polyphyletic, as two ferns are incorrectly misplaced within them). To mitigate this risk, ‘randtip’ was designed to enable users to tailor parameter settings based on their expert knowledge. We advocate for a meticulous approach to defining MDCCs, emphasizing transparency and open discussion regarding the selection criteria. It is imperative for researchers to openly disclose their input data, facilitating scrutiny and ensuring the integrity of their findings.

Additionally, expanded phylogenies can substantially vary in their topology and branch-length information depending on the procedures selected for PUT binding, which in turn can have an impact on downstream analyses. For example, the increasingly popular software PhyloMaker adds species as polytomies within randomly selected nodes (scenario 2; see Jin and Qian, 2019) rather than creating new bifurcations at randomly selected branches. Although the former method is less time-consuming, it often leads to the generation of artificially elongated terminal branches. As such, subsequent analyses using these phylogenies may misestimate metrics. A clear example is Evolutionary Distinctiveness (ED), highly correlated to terminal branch lengths, and studied in this thesis’ Chapter 5. As illustrating example, we have here studied the Rosaceae family to explore the differences in ED obtained using both schemes. To

do so, we carried out 100 phylogeny replicates using each software (namely randtip; Ramos-Gutiérrez et al., 2023; and V.PhyloMaker; Jin and Qian, 2019) and calculated ED for each one. Mean ED values for each species and differences between each software are shown in Figure 6.2. A generalised overestimation is found when using a node-based approach, with just a few cases being opposite. Other widely studied phylogenetic distance indices, such as PD, MPD or MNTD, may also show under- or overestimated values in node-expanded phylogenies due to the allocation of elongated terminal branches.

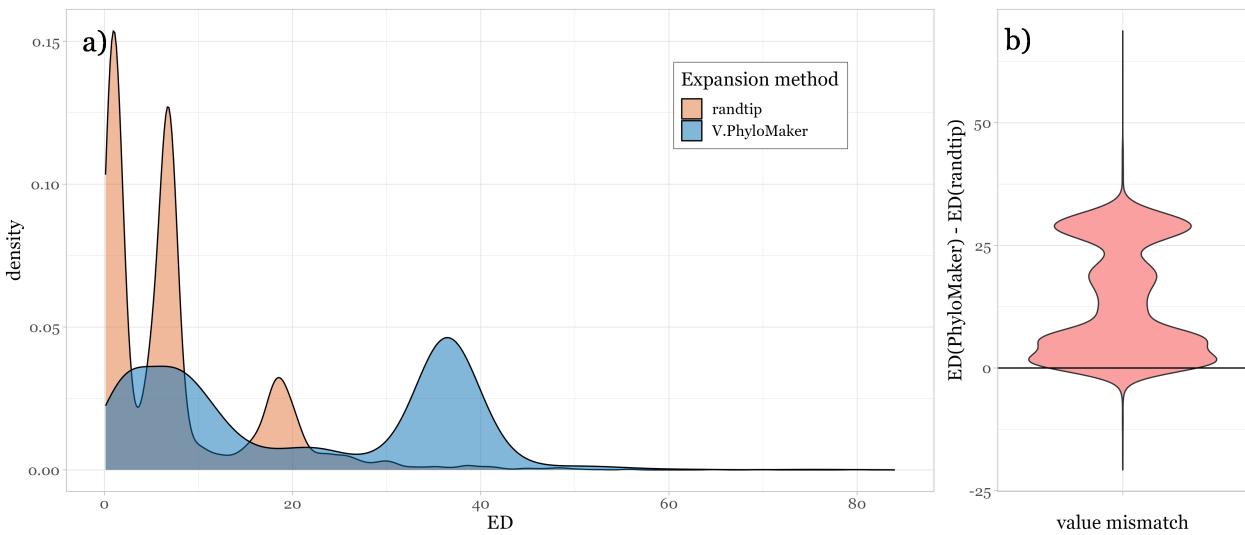


Figure 6.2: Distribution (a) and differences (b) of mean evolutionary distinctiveness (ED) values for the Rosaceae family, calculated using node-based (V.PhyloMaker; Jin and Qian, 2019; in blue) and a branch-based phylogenetic expansion (randtip; Ramos-Gutiérrez et al., 2023; in orange). Plotted values summarize the average across 100 independent phylogenetic replicates in panel a), whereas panel b) depicts the distribution of value mismatch between methodologies for each species.

6.4 Disparities and resemblance with previous Iberian plant bioregion proposals

The findings presented in Chapter 4 unveil intriguing insights into the bioregions of the Iberian Peninsula. Notably, a significant revelation pertains to the remarkable continuity observed in the Iberian alkaline Mediterranean region. Traditionally fragmented into distinct sections such as the southeastern arid region, Baetic range, and interior plateaus, this area now emerges as a cohesive and uninterrupted expanse. Conversely, the acidic Iberian region tends to be more fragmented and contingent upon the analytical procedure used. An additional result pertains to the delineation of a distinctive Mediterranean coastal bioregion, which only emerged when considering the entire flora. These findings contribute to our understanding of the boundary between the Eurosiberian (temperate) and Mediterranean bioregions, a historically contentious issue (Bolós, 1985; Galicia et al., 2010; Rivas Martínez, 2005; Sainz Ollero and Hernández Bermejo, 1985). As illustrated in Figure 4.1 of Chapter 4, the incorporation of the Central System (Hercynian Arc) into the Eurosiberian region appears to be contingent upon the taxa under study. Its exclusion from the temperate zone is likely rooted in an endemism-centric approach, while its inclusion stems from a broader perspective analysing the entire flora. These findings not only advance our understanding of Iberian biogeography but also underscore the nuanced factors influencing bioregional delineations.

Nevertheless, our results align with several previously proposed Iberian plant bioregions. One consistent element shared between our analyses and prior proposals is the floristic distinctiveness of the Balearic archipelago. Various studies, focusing on both endemic and non-endemic taxa, have identified this region as a unique floristic

territory that forms part of the Tyrrhenian Islands biodiversity hotspot (Médail and Quézel, 1999). Similarly, the Pyrenean and Cantabrian Mountains frequently emerged as distinct bioregions in the literature (e.g. Buira et al., 2017; Moreno Saiz et al., 1998), but Moreno Saiz et al. (2013) suggested to merge them into a single bioregion. Our results suggest that this discrepancy may stem from the incorporation of non-endemic taxa. The high number of narrow-ranged endemics that are restricted to single massifs contribute to their floristic uniqueness. However, these species generally co-exist with widespread alpine species. Thus, the incorporation of non-endemic taxa can obscure dissimilarities between mountain regions. The higher degree of similarity between the Iberian Peninsula and the closest islands within the Balearic archipelago (i.e. Ibiza and Formentera, the Pityusic islands) compared to the more distant islands (Mallorca and Menorca, the Gymnesic islands) had been previously noted too in Moreno Saiz et al. (1998). The recognition of these previously defined bioregions in our results underscores their robustness as biogeographic entities. Nonetheless, the variability observed between different schemes emphasized the importance of clearly defining the objectives of bioregional delimitation in each instance, enabling the selection of the most appropriate approach.

6.5 Understanding the variability in Iberian bioregion proposals

The delineation of plant bioregions within the Iberian Peninsula has sparked considerable attention, resulting in numerous classifications. However, the robustness of these proposals often come into question, as they frequently exhibit discrepancies in both number and spatial delineation across different schemes. Three primary sources contributing to these discrepancies are identified.

Firstly, the dataset itself plays a pivotal role. With advancements in chorological knowledge, researchers now have access to increasingly comprehensive and finer-resolution data, which inherently entail discrepancies in results. Secondly, the consideration of different subsets of taxa can influence bioregion delineations. While some drivers may apply across different groups of species, the unique evolutionary history of individual lineages can introduce bias. For example, regions defined based on using endemic species (e.g. Buira et al., 2017) tend to reflect the underlying evolutionary processes driving speciation patterns in the Iberian Peninsula. Furthermore, focusing on specific lineages (e.g. monocots in Moreno Saiz et al., 1998, or ferns in Moreno Saiz and Lobo, 2008) can skew results towards the evolutionary history of those particular groups. Finally, the choice of clustering algorithm can be critical in defining bioregion aggregates. While UPGMA (Unweighted Pair Group Method with Arithmetic mean) is a broadly used algorithm and generally yields consistent results (Kreft and Jetz, 2010), different metrics can greatly impact the results (Sainz Ollero and Hernández Bermejo, 1985; Ye et al., 2020). Therefore, we advocate for the integration of ordination analyses (such as PCA or NMDS) to explore and complement bioregion classifications.

6.6 Phylogenetic bioregionalization in the Iberian Peninsula

The incorporation of phylogenetic information did not yield significant differences in the delineation of bioregions within the Iberian Peninsula (see Chapter 4). This finding contrasts with the conclusions drawn by other authors in different regions, where evolutionary history played a key role in identifying geographic affinities among certain lineages (Carta et al., 2022; Daru et al., 2017b, 2020; Ye et al., 2019). We hypothesize this discrepancy between taxonomic and phylogenetic approaches can be

largely attributed to the spatial scale of our study region and its recent geological history. Previous search had explored the impact of phylogeny in extensive regions, revealing deep spatial clustering patterns among lineages (e.g. differences in regions descending from Laurasian or Gondwanan floras; Carta et al., 2022). In contrast, within our focal region, most lineages have been able to disseminate across the study area, leading to a lack of clear spatial phylogenetic clustering (as happens for European ferns; see Pataro et al., 2021). This even distribution of lineages blurs out the effect of evolutionary history at profound levels, and differences among sites are resolved based on shallow-level relationships. Therefore there is a straightforward correlation between phylogenetic and taxonomic dissimilarity among different areas. Moving forward, we propose two avenues for further investigation in this matter: i) employing a standardized methodology to cancel the correlation between taxonomic and phylogenetic patterns across different areas, which could unveil hidden evolutionary structuring within the flora in the Iberian Peninsula, and ii) incorporating a measure of each regions' phylogenetic specificity (Pataro et al., 2021).

6.7 Integration of phylogenetic information in conservation planning

The conservation of Iberian flora is paramount due to its remarkably high species richness and endemism. Not only does the region boast a significant number of unique species, but it also harbours a considerable amount of phylogenetic singularity. From a global perspective, this region is considered a centre of phylogenetic neo-endemism (Buira et al., 2017; Cai et al., 2023), characterized by recent speciation processes (Thompson, 2020). However, its southernmost position on the European continent,

coupled with its environmental heterogeneity, has facilitated the persistence of ancient lineages in climatic refugia known as paleo-endemics (Médail and Diadema, 2009). The preservation of these unique lineages thus falls under the responsibility of Spanish, Andorran and Portuguese management. While phylogenetic information has been historically overlooked, it presents a novel perspective for developing conservation programs. The utilization of numerical continuous metrics, such as EDGE (Isaac et al., 2007), provides the opportunity to prioritize taxa contributing unique evolutionary trajectories in the timeline. Efforts are currently underway to compile a priority ranking of plant species for preservation in the Iberian Peninsula, representing a crucial step for plant biodiversity preservation in the Iberian Peninsula (Fernández-Mazuecos, Jiménez Mejías et al., *personal communication*).

Axiomatically, EDGE operates on a global-scope metric (Isaac et al., 2007), as evolutionary distinctiveness is defined as the proportion of the whole diversity every species represents (Nee and May, 1997). To ensure accurate computation, it is essential that the phylogeny encompasses all taxa under scrutiny, along with their corresponding extinction risk assessments. Thus, research efforts have been predominantly concentrated on relatively small taxonomic subsets, such as mammals (including 4,182 species, Isaac et al., 2007), amphibians (encompassing 5,713 taxa, Isaac et al., 2012), or gymnosperms (containing 1,090 taxa, Forest et al., 2018). As detailed in Chapter 5, the integration of tree expansion exercises and machine learning algorithms offers a promising avenue for estimating evolutionary distinctiveness and extinction risk assessments across vast datasets, thereby facilitating the study of expansive clades like angiosperms (esteemed to include $\sim 330,000$ species). However, the use of such tools introduces two distinct sources of uncertainty. On the one hand, phylogenetic

uncertain taxa, whose impact can be mitigated by iteratively expanding backbone trees and averaging results across the entire distribution of trees. On the other hand, there is uncertainty stemming from the lack of knowledge regarding extinction probability ('Scottian' shortfall *sensu* Haelewaters et al., 2024). The reliance on global datasets poses challenges in compiling information, typically collated at regional scales. Nonetheless, global assessments can be extrapolated from regional evaluations for numerous species. This is evident in cases of endemic taxa, whose global status often aligns with regional assessments, as well as regionally least concern taxa, whose global threat status must therefore remain the same. Thus, we underscore the importance for specialists to conduct and incorporate their local assessments into global repositories.

6.8 Concluding remarks

As a key takeaway from this thesis, we provide valuable tools and insights that not only enhance our understanding of Iberian plant biogeography but also lay the groundwork for future research endeavours. By addressing both 'Wallacean' and 'Darwinian' biodiversity shortfalls, we offer distinct methodologies to tackle these pressing issues. Additionally, we shed light on the effects of various facets of diversity, including endemism and evolutionary history, on bioregion delineation. Finally, this thesis introduces a fresh perspective that can effectively guide biodiversity conservation efforts. The results presented in the last two chapters underscore the effectiveness of employing a comprehensive framework for conducting biogeographic analyses. Furthermore, they emphasise the importance of critically evaluating previously published proposals in the light of state-of-the-art data and methodologies, leading to a deeper understanding of biogeographic patterns.

Chapter 7

Conclusions

1. The AFLIBER dataset encompasses chorological information for all 6,465 registered species and subspecies of vascular plants in the Iberian Peninsula, of which 2,142 are endemic to the region. While overall completeness is high (98.6%), regional disparities exist, with certain areas being better documented than others, leading to an imbalance in regional knowledge.
2. The AFLIBER dataset will remain stored in free access repositories and should be regularly curated and updated by experts to ensure its utility for biogeographic analyses and research.
3. The implementation of *randtip* as an open-source software in R allows researchers to expand incomplete phylogenies using non-molecular data, while offering a fully customizable parameterization process. This enables expanding trees using informed parameters, rather than employing “quick-and-dirty, black-box” methodologies.
4. The delineation of biological regions of vascular plants in the Iberian Peninsula is contingent on different facets of biodiversity, such as endemictiy or evolutionary

history.

5. The use of only endemic taxa of the area of study for bioregionalization proposals results in the delimitation of floristic regions linked to topographic factors, whereas incorporating widely distributed species leads to more climatically driven bioregions. This disparity cautions against the generalization of using the endemic element as a faithful bioindicator of bioregionalization patterns.
6. In the Iberian Peninsula, phylogenetic bioregions closely resemble taxonomic ones, while emphasizing areas with phylogenetic clustering and genetic exchange between the mainland and nearby Balearic Islands. The asset of using phylogenetic data to delineate bioregions has proved to be scale-dependent, and its application should take into consideration the context of the study area.
7. The EDGE metric can be used at a regional scale to identify species warranting the highest conservation priority for preserving unique evolutionary trajectories within the focal study area. Its spatial application enables the detection of areas of high conservation interest.
8. There are 21 complementary angiosperm EDGE zones in the Iberian Peninsula, whose protection would secure 90% of the peninsular conservation priority. They are linked to areas where unique taxa thrive, such as mountains or islands, as well as others where narrowly distributed and threatened endemics occur.

Chapter 7

Conclusiones

1. La base de datos AFLIBER incluye información corológica para las 6.456 especies registradas de plantas vasculares en la Península Ibérica, 2.142 de las cuales son endémicas de la región. Mientras que la completitud general es elevada (98,6%), existen diferencias a nivel regional, con áreas mejor documentadas que otras, lo que resulta en un desequilibrio en el conocimiento regional.
2. La base de datos AFLIBER permanecerá almacenada en repositorios de acceso libre y deberá ser periódicamente revisada y actualizada por expertos para asegurar su utilidad en investigación y análisis en biogeografía.
3. La implementación de *randtip* como un programa de código abierto en R permite a los investigadores expandir filogenias incompletas usando información no molecular, ofreciendo un proceso de parametrización completamente personalizable. Esto posibilita la expansión de filogenias utilizando parámetros fundados, en vez de emplear métodos opacos.
4. La delimitación de regiones biológicas de plantas vasculares en la Península Ibérica es dependiente de distintos aspectos de la biodiversidad, como la endemidad o

la historia evolutiva.

5. El uso exclusivo de endemismos del área de estudio para propuestas de biorregionalización conlleva la demarcación de regiones florísticas relacionadas con factores topográficos, mientras que la incorporación de especies ampliamente distribuidas resulta en la delimitación de regiones más definidas por aspectos climáticos. Esta discrepancia previene de la generalización del empleo del elemento endémico como bioindicador fiel de patrones biorregionalizadores.
6. En la Península Ibérica, las biorregiones filogenéticas son similares a las taxonómicas, resaltando áreas presentando agrupamiento filogenético e intercambio genético entre el continente y las islas más cercanas del Archipiélago Balear. La ventaja del uso de información filogenética para delimitar biorregiones ha mostrado ser dependiente de la escala, y su empleo debe tener en consideración el contexto del área de estudio.
7. El índice EDGE puede ser utilizado a nivel regional para identificar qué especies poseen una prioridad mayor de conservación de cara a preservar trayectorias evolutivas singulares dentro del área de estudio. Su aplicación espacial permite detectar áreas de alto interés de conservación.
8. Existen 21 zonas EDGE complementarias para la conservación de angiospermas en la Península Ibérica, cuya preservación supondría proteger el 90% de la prioridad conservacionista a nivel peninsular. Estas áreas están ligadas a regiones donde viven especies singulares, como montañas o islas, pero también a otras donde se encuentran endemismos amenazados y de distribución restringida.

References

- Aedo, C., Buira, A., Medina, L., & Fernández-Albert, M. (2017). The Iberian Vascular Flora: Richness, Endemicity and Distribution Patterns. In J. Loidi (Ed.), *The Vegetation of the Iberian Peninsula: Volume 1* (pp. 101–130). Springer International Publishing.
- Ali, J. R., & Heaney, L. R. (2021). Wallace's line, Wallacea, and associated divides and areas: history of a tortuous tangle of ideas and labels. *Biological Reviews*, 96(3), 922–942. <https://doi.org/10.1111/brv.12683>
- Anderson, R. P., Araújo, M., Guisan, A., Lobo, J. M., Martínez-Meyer, E., Townsend, A., & Soberon, J. (2016). Are species occurrence data in global online repositories fit for modeling species distributions? The case of the Global Biodiversity Information Facility (GBIF).
- Anderson, S. (2002). *Identifying important plant areas: a site selection manual for Europe and a basis for developing guidelines for other regions of the world*.
- Antonelli, A. (2017). Biogeography: Drivers of bioregionalization. *Nature Ecology & Evolution*, 1(4), 1–2. <https://doi.org/10.1038/s41559-017-0114>
- Antonelli, A., Kissling, W. D., Flantua, S. G. A., Bermúdez, M. A., Mulch, A., Muellner-Riehl, A. N., Kreft, H., Linder, H. P., Badgley, C., Fjeldså, J., Fritz, S. A., Rahbek, C., Herman, F., Hooghiemstra, H., & Hoorn, C. (2018). Geological and climatic influences on mountain biodiversity. *Nature Geoscience*, 11(10), 718–725. <https://doi.org/10.1038/s41561-018-0236-z>
- Araújo, M. B. (1999). Distribution patterns of biodiversity and the design of a representative reserve network in Portugal. *Diversity and Distributions*, 5(4), 151–163. <https://doi.org/10.1046/j.1472-4642.1999.00052.x>
- Araújo, M. B., Lobo, J. M., & Moreno, J. C. (2007). The Effectiveness of Iberian Protected Areas in Conserving Terrestrial Biodiversity. *Conservation Biology*, 21(6), 1423–1432. <https://doi.org/10.1111/j.1523-1739.2007.00827.x>
- Bachman, S. P., Brown, M. J. M., Leão, T. C. C., Lughadha, E. N., & Walker, B. E. (2023, August 31). Extinction risk predictions for the world's flowering plants to support their conservation. <https://doi.org/10.1101/2023.08.29.555324>

- Bañares, Á., Blanca, G., Moreno, J. C., & Ortiz, S. (Eds.). (2004). *Atlas y Libro Rojo de la Flora Vascular Amenazada de España*. Dirección General de Conservación de la Naturaleza.
- Baselga, A., Orme, D., Villeger, S., De Bortoli, J., Leprieur, F., & Logez, M. (2022). *betapart: Partitioning beta diversity into turnover and nestedness components* (manual).
- Birds Directive. Council Directive 79/409/EEC (1979, April 2).
- Blanca, G., Cueto, M., Martínez-Lirola, M. J., & Molero-Mesa, J. (1998). Threatened vascular flora of Sierra Nevada (Southern Spain). *Biological Conservation*, 85(3), 269–285. [https://doi.org/10.1016/S0006-3207\(97\)00169-9](https://doi.org/10.1016/S0006-3207(97)00169-9)
- Bolós, O. d. (1985). Le territoire subméditerranéen et le territoire carpetano-atlantique dans la Péninsule Ibérique. *Botanica Helvetica*, 95(1), 13–17.
- Bradshaw, P. L., Colville, J. F., & Linder, H. P. (2015). Optimising Regionalisation Techniques: Identifying Centres of Endemism in the Extraordinarily Endemic-Rich Cape Floristic Region. *PLOS ONE*, 10(7), e0132538. <https://doi.org/10.1371/journal.pone.0132538>
- Brooks, T. M., Mittermeier, R. A., da Fonseca, G. A. B., Gerlach, J., Hoffmann, M., Lamoreux, J. F., Mittermeier, C. G., Pilgrim, J. D., & Rodrigues, A. S. L. (2006). Global Biodiversity Conservation Priorities. *Science*, 313(5783), 58–61. <https://doi.org/10.1126/science.1127609>
- Buffon, G.-L. L., Comte de. (1761). *Histoire Naturelle, Generalle et Particuliere* (Vol. 9). Imprimerie Royale, Paris.
- Buira, A., Aedo, C., & Medina, L. (2017). Spatial patterns of the Iberian and Balearic endemic vascular flora. *Biodiversity and Conservation*, 26(2), 479–508. <https://doi.org/10.1007/s10531-016-1254-z>
- Buira, A., Cabezas, F., & Aedo, C. (2020). Disentangling ecological traits related to plant endemism, rarity and conservation status in the Iberian Peninsula. *Biodiversity and Conservation*, 29(6), 1937–1958. <https://doi.org/10.1007/s10531-020-01957-z>
- Buira, A., Fernández-Mazuecos, M., Aedo, C., & Molina-Venegas, R. (2021). The contribution of the edaphic factor as a driver of recent plant diversification in a Mediterranean biodiversity hotspot. *Journal of Ecology*, 109(2), 987–999. <https://doi.org/10.1111/1365-2745.13527>
- Cai, L., Kreft, H., Taylor, A., Schrader, J., Dawson, W., Essl, F., van Kleunen, M., Pergl, J., Pyšek, P., Winter, M., & Weigelt, P. (2023). Climatic stability and geological history shape global centers of neo- and paleoendemism in seed plants. *Proceedings of the National Academy of Sciences*, 120(30), e2300981120. <https://doi.org/10.1073/pnas.2300981120>
- Calatayud, J., Rodríguez, M. Á., Molina-Venegas, R., Leo, M., Horreo, J. L., & Hortal, J. (2019). Pleistocene climate change and the formation of regional species pools. *Proceedings of the Royal Society B: Biological Sciences*, 286(1905), 20190291. <https://doi.org/10.1098/rspb.2019.0291>
- Carapeto, A., Francisco, A., Pereira, P., & Porto, M. (Eds.). (2020). *Lista Vermelha da Flora Vascular de Portugal Continental*. Imprensa Nacional, Lisboa.

- Carrillo, E., Mercadé, A., Ninot, J., Carreras, J., Ferré, A., & Castell, X. (2008). *Check-list i Llista Vermella de la Flora d'Andorra*.
- Carta, A., Gargano, D., Rossi, G., Bacchetta, G., Fenu, G., Montagnani, C., Abeli, T., Peruzzi, L., & Orsenigo, S. (2019). Phylogenetically informed spatial planning as a tool to prioritise areas for threatened plant conservation within a Mediterranean biodiversity hotspot. *Science of The Total Environment*, 665, 1046–1052. <https://doi.org/10.1016/j.scitotenv.2019.02.127>
- Carta, A., Peruzzi, L., & Ramírez-Barahona, S. (2022). A global phylogenetic regionalization of vascular plants reveals a deep split between Gondwanan and Laurasian biotas. *New Phytologist*, 233(3), 1494–1504. <https://doi.org/10.1111/nph.17844>
- Castro Parga, I., Moreno Saiz, J. C., Humphries, C. J., & Williams, P. H. (1996). Strengthening the Natural and National Park system of Iberia to conserve vascular plants. *Botanical Journal of the Linnean Society*, 121(3), 189–206. <https://doi.org/10.1111/j.1095-8339.1996.tb00753.x>
- Castroviejo, S. (1986). *Flora iberica*. Real Jardín Botánico, CSIC, Madrid.
- Cavender-Bares, J., Kozak, K. H., Fine, P. V. A., & Kembel, S. W. (2009). The merging of community ecology and phylogenetic biology. *Ecology Letters*, 12(7), 693–715. <https://doi.org/10.1111/j.1461-0248.2009.01314.x>
- Chamberlain, S., Barve, V., Desmet, P., Geffert, L., Mcglinn, D., Oldoni, D., & Karthik, R. (2019). *rgbif: Interface to the Global Biodiversity Information Facility API* (Version 1.3.0).
- Chase, M. W., Christenhusz, M. J. M., Fay, M. F., Byng, J. W., Judd, W. S., Soltis, D. E., Mabberley, D. J., Sennikov, A. N., Soltis, P. S., & Stevens, P. F. (2016). An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Botanical Journal of the Linnean Society*, 181(1), 1–20. <https://doi.org/10.1111/boj.12385>
- Colwell, R. K., & Coddington, J. A. (1994). Estimating Terrestrial Biodiversity through Extrapolation. *Philosophical Transactions: Biological Sciences*, 345(1311), 101–118.
- COP. (2003). *Decision VI/9. Global Strategy for Plant Conservation*.
- Crisci, J. V., Katinas, L., Apodaca, M. J., & Hoch, P. C. (2020). The End of Botany. *Trends in Plant Science*, 25(12), 1173–1176. <https://doi.org/10.1016/j.tplants.2020.09.012>
- Croizat, L. (1958). *Panbiogeography* (Vol. 1, 2). Published by the Author, Caracas.
- Daru, B. H., Elliott, T. L., Park, D. S., & Davies, T. J. (2017a). Understanding the Processes Underpinning Patterns of Phylogenetic Regionalization. *Trends in Ecology & Evolution*, 32(11), 845–860. <https://doi.org/10.1016/j.tree.2017.08.013>
- Daru, B. H., Holt, B. G., Lessard, J.-P., Yessoufou, K., & Davies, T. J. (2017b). Phylogenetic regionalization of marine plants reveals close evolutionary affinities among disjunct temperate assemblages. *Biological Conservation*, 213, 351–356. <https://doi.org/10.1016/j.biocon.2016.08.022>

- Daru, B. H., Karunarathne, P., & Schliep, K. (2020). phyloregion: R package for biogeographical regionalization and macroecology. *Methods in Ecology and Evolution*, 11(11), 1483–1491. <https://doi.org/10.1111/2041-210X.13478>
- Daru, B. H., van der Bank, M., Maurin, O., Yessoufou, K., Schaefer, H., Slingsby, J. A., & Davies, T. J. (2016). A novel phylogenetic regionalization of phytogeographical zones of southern Africa reveals their hidden evolutionary affinities. *Journal of Biogeography*, 43(1), 155–166. <https://doi.org/10.1111/jbi.12619>
- Darwin, C. (1859). *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. John Murray, London.
- Davies, T. J. (2021). Ecophylogenetics redux. *Ecology Letters*, 24(5), 1073–1088. <https://doi.org/10.1111/ele.13682>
- Davies, T. J., Kraft, N. J. B., Salamin, N., & Wolkovich, E. M. (2012). Incompletely resolved phylogenetic trees inflate estimates of phylogenetic conservatism. *Ecology*, 93(2), 242–247. <https://doi.org/10.1890/11-1360.1>
- Davis, P., Miller, R., & Tan, K. (1988). *Flora of Turkey and the East Aegean Islands* (Edinburgh University Press, Vol. 10).
- de Candolle, A. L. P. P. (1855). *Géographie botanique raisonnée, ou exposition des faits principaux et des lois concertant la distribution géographique des plantes de l'époque actuelle*. Masson, Paris.
- Diniz-Filho, J. A. F., Loyola, R. D., Raia, P., Mooers, A. O., & Bini, L. M. (2013). Darwinian shortfalls in biodiversity conservation. *Trends in Ecology & Evolution*, 28(12), 689–695. <https://doi.org/10.1016/j.tree.2013.09.003>
- Domínguez Lozano, F., Galicia Herbada, D., Moreno Rivero, L., Moreno Saiz, J. C., & Sainz Ollero, H. (2000). Areas of high floristic endemism in Iberia and the Balearic islands : an approach to biodiversity conservation using narrow endemics. *Belgian Journal of Entomology*, 2, 171–185.
- Domínguez Lozano, F., Moreno Saiz, J. C., & Sainz Ollero, H. (2003). Rarity and threat relationships in the conservation planning of Iberian flora. *Biodiversity & Conservation*, 12(9), 1861–1882. <https://doi.org/10.1023/A:1024110925948>
- Droissart, V., Dauby, G., Hardy, O. J., Deblauwe, V., Harris, D. J., Janssens, S., Mackinder, B. A., Blach-Overgaard, A., Sonké, B., Sosef, M. S. M., Stévart, T., Svenning, J.-C., Wieringa, J. J., & Couvreur, T. L. P. (2018). Beyond trees: Biogeographical regionalization of tropical Africa. *Journal of Biogeography*, 45(5), 1153–1167. <https://doi.org/10.1111/jbi.13190>
- Escalante, T. (2009). Un ensayo sobre regionalización biogeográfica. *Revista mexicana de biodiversidad*, 80(2), 551–560.
- Fady, B., & Concord, C. (2010). Macroecological patterns of species and genetic diversity in vascular plants of the Mediterranean basin. *Diversity and Distributions*, 16(1), 53–64. <https://doi.org/10.1111/j.1472-4642.2009.00621.x>

- Faith, D. P. (1992). Conservation evaluation and phylogenetic diversity. *Biological Conservation*, 61(1), 1–10. [https://doi.org/10.1016/0006-3207\(92\)91201-3](https://doi.org/10.1016/0006-3207(92)91201-3)
- FitzJohn, R. G., Maddison, W. P., & Otto, S. P. (2009). Estimating Trait-Dependent Speciation and Extinction Rates from Incompletely Resolved Phylogenies. *Systematic Biology*, 58(6), 595–611. <https://doi.org/10.1093/sysbio/syp067>
- Forest, F. (2023). Species-level phylogenetic trees of all angiosperm species (100 trees). <https://doi.org/https://zenodo.org/doi/10.5281/zenodo.7600340>
- Forest, F., Crandall, K. A., Chase, M. W., & Faith, D. P. (2015). Phylogeny, extinction and conservation: embracing uncertainties in a time of urgency. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370(1662), 20140002. <https://doi.org/10.1098/rstb.2014.0002>
- Forest, F., Moat, J., Baloch, E., Brummitt, N. A., Bachman, S. P., Ickert-Bond, S., Hollingsworth, P. M., Liston, A., Little, D. P., Mathews, S., Rai, H., Rydin, C., Stevenson, D. W., Thomas, P., & Buerki, S. (2018). Gymnosperms on the EDGE. *Scientific Reports*, 8(1), 6053. <https://doi.org/10.1038/s41598-018-24365-4>
- Forster, J. R. (1778). *Observations made during a voyage round the world, on Physical Geography, Natural History and Ethic Philosophy*. G. Robinson, London.
- Gaiji, S., Chavan, V., Ariño, A. H., Otegui, J., Hobern, D., Sood, R., & Robles, E. (2013). Content assessment of the primary biodiversity data published through GBIF network: Status, challenges and potentials. *Biodiversity Informatics*, 8(2). <https://doi.org/10.17161/bi.v8i2.4124>
- Galicia, D., Hervás, J., Martínez, R., Seoane, J., & Hidalgo, R. (2010). Ambientes terrestres de España, una propuesta de clasificación territorial dirigida a facilitar el análisis y la gestión de la biodiversidad: *Ecosistemas*, 19(3).
- Gastauer, M., & Meira-Neto, J. A. A. (2013). Avoiding inaccuracies in tree calibration and phylogenetic community analysis using Phylocom 4.2. *Ecological Informatics*, 15, 85–90. <https://doi.org/10.1016/j.ecoinf.2013.03.005>
- Gaston, K. J., & Rodrigues, A. S. L. (2003). Reserve Selection in Regions with Poor Biological Data. *Conservation Biology*, 17(1), 188–195. <https://doi.org/10.1046/j.1523-1739.2003.01268.x>
- Gillings, S., Balmer, D. E., Caffrey, B. J., Downie, I. S., Gibbons, D. W., Lack, P. C., Reid, J. B., Sharrock, J. T. R., Swann, R. L., & Fuller, R. J. (2019). Breeding and wintering bird distributions in Britain and Ireland from citizen science bird atlases. *Global Ecology and Biogeography*, 28(7), 866–874. <https://doi.org/10.1111/geb.12906>
- González-Orozco, C. E., Laffan, S. W., Knerr, N., & Miller, J. T. (2013). A biogeographical regionalization of Australian Acacia species. *Journal of Biogeography*, 40(11), 2156–2166. <https://doi.org/10.1111/jbi.12153>

- Govaerts, R., Nic Lughadha, E., Black, N., Turner, R., & Paton, A. (2021). The World Checklist of Vascular Plants, a continuously updated resource for exploring global plant diversity. *Scientific Data*, 8(1), 215. <https://doi.org/10.1038/s41597-021-00997-6>
- Graham, C. H., & Fine, P. V. A. (2008). Phylogenetic beta diversity: linking ecological and evolutionary processes across space in time. *Ecology Letters*, 11(12), 1265–1277. <https://doi.org/10.1111/j.1461-0248.2008.01256.x>
- Guardiola, M., & Sáez, L. (2023). Are Mediterranean Island Mountains Hotspots of Taxonomic and Phylogenetic Biodiversity? The Case of the Endemic Flora of the Balearic Islands. *Plants*, 12(14), 2640. <https://doi.org/10.3390/plants12142640>
- Gumbs, R., Chaudhary, A., Daru, B. H., Faith, D. P., Forest, F., Gray, C. L., Kowalska, A., Lee, W.-S., Pellens, R., Pipins, S., Pollock, L. J., Rosindell, J., Scherson, R. A., & Owen, N. R. (2023a). Indicators to monitor the status of the tree of life. *Conservation Biology*, 37(6), e14138. <https://doi.org/10.1111/cobi.14138>
- Gumbs, R., Gray, C. L., Böhm, M., Burfield, I. J., Couchman, O. R., Faith, D. P., Forest, F., Hoffmann, M., Isaac, N. J. B., Jetz, W., Mace, G. M., Mooers, A. O., Safi, K., Scott, O., Steel, M., Tucker, C. M., Pearse, W. D., Owen, N. R., & Rosindell, J. (2023b). The EDGE2 protocol: Advancing the prioritisation of Evolutionarily Distinct and Globally Endangered species for practical conservation action. *PLOS Biology*, 21(2), e3001991. <https://doi.org/10.1371/journal.pbio.3001991>
- Gumbs, R., Gray, C. L., Wearn, O. R., & Owen, N. R. (2018). Tetrapods on the EDGE: Overcoming data limitations to identify phylogenetic conservation priorities. *PLOS ONE*, 13(4), e0194680. <https://doi.org/10.1371/journal.pone.0194680>
- Habitat Directive. Council Directive 92/43/EEC (1992, May 21).
- Haelewaters, D., Matthews, T. J., Wayman, J. P., Cazabonne, J., Heyman, F., Quandt, C. A., & Martin, T. E. (2024). Biological knowledge shortfalls impede conservation efforts in poorly studied taxa—A case study of Laboulbeniomycetes. *Journal of Biogeography*, 51(1), 29–39. <https://doi.org/10.1111/jbi.14725>
- Hengeveld, R. (1992). *Dynamic biogeography*. Cambridge University Press, Cambridge.
- Hengl, T., Jesus, J. M. d., Heuvelink, G. B. M., Gonzalez, M. R., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M. N., Geng, X., Bauer-Marschallinger, B., Guevara, M. A., Vargas, R., MacMillan, R. A., Batjes, N. H., Leenaars, J. G. B., Ribeiro, E., Wheeler, I., Mantel, S., & Kempen, B. (2017). SoilGrids250m: Global gridded soil information based on machine learning. *PLOS ONE*, 12(2), e0169748. <https://doi.org/10.1371/journal.pone.0169748>
- Heywood, V. H. (2003). The jewel in the crown. *Plant Talk*, 32, 42–43.
- Hörndl, E., & Stuessy, T. F. (2010). Paraphyletic groups as natural units of biological classification. *TAXON*, 59(6), 1641–1653. <https://doi.org/10.1002/tax.596001>

- Hortal, J., de Bello, F., Diniz-Filho, J. A. F., Lewinsohn, T. M., Lobo, J. M., & Ladle, R. J. (2015). Seven Shortfalls that Beset Large-Scale Knowledge of Biodiversity. *Annual Review of Ecology, Evolution, and Systematics*, 46(1), 523–549. <https://doi.org/10.1146/annurev-ecolsys-112414-054400>
- Humboldt, A. v. (1807). *Voyage aux régions équinoxiales du Nouveau Continent, fait en 1799-1804 par Alexandre de Humboldt et Aimé Bonpland*. Librairie grecque-latine-allemande, Paris.
- Huson, D. H., & Scornavacca, C. (2012). Dendroscope 3: An Interactive Tool for Rooted Phylogenetic Trees and Networks. *Systematic Biology*, 61(6), 1061–1067. <https://doi.org/10.1093/sysbio/sys062>
- Huxley, T. H. (1868). On the classification and distribution of the Alectromorphae and Heteromorphae. *Proceedings of the Zoological Society of London*, 1868, 294–319.
- IBRA. (2000). *Revision of the Interim Biogeographic Regionalisation of Australia (IBRA) and the Development of Version 5.1 - Summary Report - DCCEEW*. Department of Environment and Heritage. Canberra.
- IPBES. (2019). *Global Assessment Report on Biodiversity and Ecosystem Services*. Retrieved January 10, 2024, from <https://www.ipbes.net/node/35274>
- Isaac, N. J. B., & Pearse, W. D. (2018). The Use of EDGE (Evolutionary Distinct Globally Endangered) and EDGE-Like Metrics to Evaluate Taxa for Conservation. In R. A. Scherson & D. P. Faith (Eds.), *Phylogenetic Diversity: Applications and Challenges in Biodiversity Science* (pp. 27–39). Springer International Publishing.
- Isaac, N. J. B., Redding, D. W., Meredith, H. M., & Safi, K. (2012). Phylogenetically-Informed Priorities for Amphibian Conservation. *PLOS ONE*, 7(8), e43912. <https://doi.org/10.1371/journal.pone.0043912>
- Isaac, N. J. B., Turvey, S. T., Collen, B., Waterman, C., & Baillie, J. E. M. (2007). Mammals on the EDGE: Conservation Priorities Based on Threat and Phylogeny. *PLOS ONE*, 2(3), e296. <https://doi.org/10.1371/journal.pone.0000296>
- IUCN. (2021). *The IUCN Red List of Threatened Species*. Retrieved March 1, 2021, from <https://www.iucnredlist.org/en>
- IUCN. (2023). *The IUCN Red List of Threatened Species*. Retrieved April 1, 2023, from <https://www.iucnredlist.org/en>
- IUCN. (2024). *The IUCN Red List of Threatened Species*. Retrieved January 26, 2024, from <https://www.iucnredlist.org/en>
- Janssens, S., Couvreur, T. L. P., Mertens, A., Dauby, G., Dagallier, L.-P., Abeele, S. V., Vandeloek, F., Mascarello, M., Beeckman, H., Sosef, M., Droissart, V., Bank, M. v. d., Maurin, O., Hawthorne, W., Marshall, C., Réjou-Méchain, M., Beina, D., Baya, F., Merckx, V., ... Hardy, O. (2020). A large-scale species level dated angiosperm phylogeny for evolutionary and ecological analyses. *Biodiversity Data Journal*, 8, e39677. <https://doi.org/10.3897/BDJ.8.e39677>
- Jetz, W., McPherson, J. M., & Guralnick, R. P. (2012). Integrating biodiversity distribution knowledge: toward a global map of life. *Trends in Ecology & Evolution*, 27(3), 151–159. <https://doi.org/10.1016/j.tree.2011.09.007>

- Jin, Y., & Qian, H. (2019). V.PhyloMaker: an R package that can generate very large phylogenies for vascular plants. *Ecography*, 42(8), 1353–1359. <https://doi.org/10.1111/ecog.04434>
- Jin, Y., & Qian, H. (2023). U.PhyloMaker: An R package that can generate large phylogenetic trees for plants and animals. *Plant Diversity*, 45(3), 347–352. <https://doi.org/10.1016/j.pld.2022.12.007>
- Karger, D. N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R. W., Zimmermann, N. E., Linder, H. P., & Kessler, M. (2017). Climatologies at high resolution for the earth's land surface areas. *Scientific Data*, 4(1), 170122. <https://doi.org/10.1038/sdata.2017.122>
- Kembel, S., Cowan, P., Helmus, M., Cornwell, W., Morlon, H., Ackerly, D., Blomberg, S., & Webb, C. (2010). Picante: R tools for integrating phylogenies and ecology. *Bioinformatics (Oxford, England)*, 26, 1463–1464.
- Kidron, G. J., & Lázaro, R. (2020). Are coastal deserts necessarily dew deserts? An example from the Tabernas Desert. *Journal of Hydrology and Hydromechanics*, 68(1), 19–27. <https://doi.org/10.2478/johh-2020-0002>
- Kreft, H., & Jetz, W. (2010). A framework for delineating biogeographical regions based on species distributions. *Journal of Biogeography*, 37(11), 2029–2053. <https://doi.org/10.1111/j.1365-2699.2010.02375.x>
- Laguna, E., Ballester, G., Ranz, J., Fabregat, C., Olivares, A., Serra, L., Deltoro, V., Pérez Rovira, P., & Pérez Botella, J. (2001). *La red de microrreservas de flora de la Comunidad Valenciana*.
- Lamsdell, J. C., & Congreve, C. R. (2021). Phylogenetic paleoecology: macroecology within an evolutionary framework. *Paleobiology*, 47(2), 171–177. <https://doi.org/10.1017/pab.2020.61>
- Lobo, J. M., & Araújo, M. B. (2003). La aplicación de datos faunísticos para el diseño de redes de reservas: el caso de los anfibios y reptiles de la Península Ibérica. *Graellsia*, 59(2), 399–408. <https://doi.org/10.3989/graeellsia.2003.v59.i2-3.255>
- Lobo, J. M., Castro, I., & Moreno, J. C. (2001). Spatial and environmental determinants of vascular plant species richness distribution in the Iberian Peninsula and Balearic Islands. *Biological Journal of the Linnean Society*, 73(2), 233–253. <https://doi.org/10.1111/j.1095-8312.2001.tb01360.x>
- Lomolino, M. V. (2004). Conservation biogeography. In *Frontiers of Biogeography: New Directions in the Geography of Nature*, ed. M. V. Lomolino, L.R. Heaney (pp. 293–296). Sinauer Associates.
- Lomolino, M. V., Riddle, B. R., Whittaker, R. J., & Brown, J. H. (2010). *Biogeography* (Fourth Edition). Sinauer Associates.
- Lydekker, R. (1896). *A Geographical History of Mammals*. Cambridge University Press, Cambridge.
- Mackey, B. G., Berry, S. L., & Brown, T. (2008). Reconciling approaches to biogeographical regionalization: a systematic and generic framework examined with a case study of the Australian continent. *Journal of Biogeography*, 35(2), 213–229. <https://doi.org/10.1111/j.1365-2699.2007.01822.x>
- Maldonado, C., Molina, C. I., Zizka, A., Persson, C., Taylor, C. M., Albán, J., Chilquillo, E., Rønsted, N., & Antonelli, A. (2015). Estimating species diversity and distribution in the era of Big Data:

- to what extent can we trust public databases? *Global Ecology and Biogeography*, 24(8), 973–984. <https://doi.org/10.1111/geb.12326>
- Malécot, V., & Nickrent, D. L. (2008). Molecular Phylogenetic Relationships of Olacaceae and Related Santalales. *Systematic Botany*, 33(1), 97–106. <https://doi.org/10.1600/036364408783887384>
- Marchese, C. (2015). Biodiversity hotspots: A shortcut for a more complicated concept. *Global Ecology and Conservation*, 3, 297–309. <https://doi.org/10.1016/j.gecco.2014.12.008>
- Márquez, A. L., Real, R., & Vargas, J. M. (2001). Methods for comparison of biotic regionalizations: the case of pteridophytes in the Iberian Peninsula. *Ecography*, 24(6), 659–670. <https://doi.org/10.1111/j.1600-0587.2001.tb00528.x>
- Martins, W. S., Carmo, W. C., Longo, H. J., Rosa, T. C., & Rangel, T. F. (2013). SUNPLIN: Simulation with Uncertainty for Phylogenetic Investigations. *BMC Bioinformatics*, 14(1), 324. <https://doi.org/10.1186/1471-2105-14-324>
- May, R. M. (1988). How Many Species Are There on Earth? *Science*, 241(4872), 1441–1449. <https://doi.org/10.1126/science.241.4872.1441>
- Mayer-Schönberger, V., & Cukier, K. (2013). *Big Data: A Revolution that Will Transform how We Live, Work, and Think*. Houghton Mifflin Harcourt.
- Médail, F. (2022). Plant Biogeography and Vegetation Patterns of the Mediterranean Islands. *The Botanical Review*, 88(1), 63–129. <https://doi.org/10.1007/s12229-021-09245-3>
- Médail, F., & Diadema, K. (2009). Glacial refugia influence plant diversity patterns in the Mediterranean Basin. *Journal of Biogeography*, 36(7), 1333–1345. <https://doi.org/10.1111/j.1365-2699.2008.02051.x>
- Médail, F., & Quézel, P. (1997). Hot-Spots Analysis for Conservation of Plant Biodiversity in the Mediterranean Basin. *Annals of the Missouri Botanical Garden*, 84(1), 112–127. <https://doi.org/10.2307/2399957>
- Médail, F., & Quézel, P. (1999). Biodiversity Hotspots in the Mediterranean Basin: Setting Global Conservation Priorities. *Conservation Biology*, 13(6), 1510–1513.
- Mishler, B. D. (2023). Spatial phylogenetics. *Journal of Biogeography*, 50(8), 1454–1463. <https://doi.org/10.1111/jbi.14618>
- Mishler, B. D., Knerr, N., González-Orozco, C. E., Thornhill, A. H., Laffan, S. W., & Miller, J. T. (2014). Phylogenetic measures of biodiversity and neo- and paleo-endemism in Australian Acacia. *Nature Communications*, 5(1), 4473. <https://doi.org/10.1038/ncomms5473>
- Molina-Venegas, R., Aparicio, A., Lavergne, S., & Arroyo, J. (2015). The building of a biodiversity hotspot across a land-bridge in the Mediterranean. *Proceedings of the Royal Society B: Biological Sciences*, 282(1813), 20151116. <https://doi.org/10.1098/rspb.2015.1116>
- Molina-Venegas, R., Ramos-Gutiérrez, I., & Moreno-Saiz, J. C. (2020). Phylogenetic Patterns of Extinction Risk in the Endemic Flora of a Mediterranean Hotspot as a Guiding Tool for Preemptive Conservation Actions. *Frontiers in Ecology and Evolution*, 8.

- Molina-Venegas, R., Rodríguez, M. Á., Pardo-de-Santayana, M., Ronquillo, C., & Mabberley, D. J. (2021). Maximum levels of global phylogenetic diversity efficiently capture plant services for humankind. *Nature Ecology & Evolution*, 5(5), 583–588. <https://doi.org/10.1038/s41559-021-01414-2>
- Moreno Saiz, J. C., Castro Parga, I., & Sainz Ollero, H. (1998). Numerical analyses of distributions of Iberian and Balearic endemic monocotyledons. *Journal of Biogeography*, 25(1), 179–194. <https://doi.org/10.1046/j.1365-2699.1998.251118.x>
- Moreno Saiz, J. C. (2011). La diversidad florística vascular española. In *Biodiversidad. Aproximación a la diversidad botánica y zoológica de España* (pp. 75–108). Real Sociedad Española de Historia Natural.
- Moreno Saiz, J. C., Donato, M., Katinas, L., Crisci, J. V., & Posadas, P. (2013). New insights into the biogeography of south-western Europe: spatial patterns from vascular plants using cluster analysis and parsimony. *Journal of Biogeography*, 40(1), 90–104. <https://doi.org/10.1111/j.1365-2699.2012.02774.x>
- Moreno Saiz, J. C., Iriondo Alegría, J. M., Martínez García, F., Martínez Rodríguez, J., & Salazar Mendías, C. (Eds.). (2019). *Atlas y Libro Rojo de la Flora Vascular Amenazada de España. Adenda 2017*. Ministerio para la Transición Ecológica-Sociedad Española de Biología de la Conservación de Plantas.
- Moreno Saiz, J. C., & Lobo, J. M. (2008). Iberian–Balearic fern regions and their explanatory variables. *Plant Ecology*, 198(2), 149–167. <https://doi.org/10.1007/s11258-007-9392-8>
- Moreno Saiz, J. C., & Martínez García, F. (2023). La Sociedad Española de Biología de la Conservación de Plantas (SEBiCoP): Veinte años trabajando en pro de la conservación vegetal en España. *Conservación Vegetal*, 27, 28–30.
- Moreno-Saiz, J. C., & Sainz-Ollero, H. (1992). *Atlas corológico de las monocotiledóneas endémicas de la Península Ibérica e islas Baleares*. ICONA.
- Mouquet, N., Devictor, V., Meynard, C. N., Munoz, F., Bersier, L.-F., Chave, J., Couteron, P., Dalecky, A., Fontaine, C., Gravel, D., Hardy, O. J., Jabot, F., Lavergne, S., Leibold, M., Mouillot, D., Münkemüller, T., Pavone, S., Prinzing, A., Rodrigues, A. S., ... Thuiller, W. (2012). Ecophylogenetics: advances and perspectives. *Biological Reviews*, 87(4), 769–785. <https://doi.org/10.1111/j.1469-185X.2012.00224.x>
- R
Muñoz-Rodríguez, P., Draper Munt, D., & Moreno Saiz, J. C. (2016). Global strategy for plant conservation: inadequate in situ conservation of threatened flora in Spain. *Israel Journal of Plant Sciences*, 63(4), 297–308. <https://doi.org/10.1080/07929978.2016.1257105>
- Myers, N., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A. B., & Kent, J. (2000). Biodiversity hotspots for conservation priorities. *Nature*, 403(6772), 853–858. <https://doi.org/10.1038/35002501>
- Nee, S., & May, R. M. (1997). Extinction and the Loss of Evolutionary History. *Science*, 278(5338), 692–694. <https://doi.org/10.1126/science.278.5338.692>
- Nic Lughadha, E., Bachman, S. P., Leão, T. C. C., Forest, F., Halley, J. M., Moat, J., Acedo, C., Bacon, K. L., Brewer, R. F. A., Gâteblé, G., Gonçalves, S. C., Govaerts, R., Hollingsworth, P. M., Krisai-Greilhuber, I., de Lirio, E. J., Moore, P. G. P., Negrão, R., Onana, J. M., Rajaovelona, L. R., ... Walker,

- B. E. (2020). Extinction risk and threats to plants and fungi. *Plants, People, Planet*, 2(5), 389–408. <https://doi.org/10.1002/ppp3.10146>
- Nieto Feliner, G. (2014). Patterns and processes in plant phylogeography in the Mediterranean Basin. A review. *Perspectives in Plant Ecology, Evolution and Systematics*, 16(5), 265–278. <https://doi.org/10.1016/j.ppees.2014.07.002>
- Ninot, J. M., Carrillo, E., & Ferré, A. (2017). The Pyrenees. In J. Loidi (Ed.), *The Vegetation of the Iberian Peninsula: Volume 1* (pp. 323–366). Springer International Publishing.
- Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H., Szoecs, E., & Wagner, H. (2019). *vegan: Community ecology package* (Version 2.5-6).
- Pataro, L., Molina-Venegas, R., Calatayud, J., Moreno-Saiz, J. C., & Rodríguez, M. Á. (2021). An updated phylogenetic bioregionalization for the European fern flora. *Biodiversity and Conservation*, 30(1), 201–215. <https://doi.org/10.1007/s10531-020-02087-2>
- Pebesma, E., & Bivand, R. S. (2005). S Classes and Methods for Spatial Data: the sp Package.
- Pentinsaari, M., Ratnasingham, S., Miller, S. E., & Hebert, P. D. N. (2020). BOLD and GenBank revisited – Do identification errors arise in the lab or in the sequence libraries? *PLOS ONE*, 15(4), e0231814. <https://doi.org/10.1371/journal.pone.0231814>
- Pipins, S., Baillie, J. E., Bowmer, A., Pollock, L. J., Owen, N., & Gumbs, R. (2023, August 17). Advancing EDGE Zones: spatial priorities for the conservation of tetrapod evolutionary history.
- Plantlife. (2010). *Important Plant Area (IPA)*. Retrieved December 26, 2023, from <https://www.plantlife.org.uk/protecting-plants-fungi/important-plant-areas/>
- Posadas, P., Grossi, M. A., & Ortiz-Jaureguizar, E. (2013). Where is historical biogeography going? The evolution of the discipline in the first decade of the 21st century. *Progress in Physical Geography: Earth and Environment*, 37(3), 377–396. <https://doi.org/10.1177/0309133313478316>
- Rahbek, C., Borregaard, M. K., Antonelli, A., Colwell, R. K., Holt, B. G., Nogues-Bravo, D., Rasmussen, C. M. Ø., Richardson, K., Rosing, M. T., Whittaker, R. J., & Fjeldså, J. (2019). Building mountain biodiversity: Geological and evolutionary processes. *Science*, 365(6458), 1114–1119. <https://doi.org/10.1126/science.aax0151>
- Ramos-Gutiérrez, I., Lima, H., Pajarón, S., Romero-Zarco, C., Sáez, L., Pataro, L., Molina-Venegas, R., Rodríguez, M. Á., & Moreno-Saiz, J. C. (2021). Atlas of the vascular flora of the Iberian Peninsula biodiversity hotspot (AFLIBER). *Global Ecology and Biogeography*, 30(10), 1951–1957. <https://doi.org/10.1111/geb.13363>
- Ramos-Gutiérrez, I., Lima, H., Vilela, B., & Molina-Venegas, R. (2023). A generalized framework to expand incomplete phylogenies using non-molecular phylogenetic information. *Global Ecology and Biogeography*, 32(10), 1707–1716. <https://doi.org/10.1111/geb.13733>

- Rangel, T. F., Colwell, R. K., Graves, G. R., Fučíková, K., Rahbek, C., & Diniz-Filho, J. A. F. (2015). Phylogenetic uncertainty revisited: Implications for ecological analyses. *Evolution*, 69(5), 1301–1312. <https://doi.org/10.1111/evo.12644>
- Redding, D. W., & Mooers, A. Ø. (2006). Incorporating Evolutionary Measures into Conservation Prioritization. *Conservation Biology*, 20(6), 1670–1678. <https://doi.org/10.1111/j.1523-1739.2006.00555.x>
- Revell, L. J. (2012). phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, 3(2), 217–223. <https://doi.org/10.1111/j.2041-210X.2011.00169.x>
- Riddle, B. R., Ladle, R. J., Lourie, S. A., & Whittaker, R. J. (2011). Basic Biogeography: Estimating Biodiversity and Mapping Nature. In *Conservation Biogeography* (pp. 45–92). John Wiley & Sons, Ltd.
- Rivas Martínez, S. (2005). Avances en Geobotánica. *Discurso de Apertura del Curso Académico de la Real Academia Nacional de Farmacia del año 2005*.
- Rivas-Martínez, S., Fernández-González, F., Loidi, J., Lousa, M., & Penas, Á. (2002). Vascular plant communities of Spain and Portugal. *Itinera Geobotánica*, 15(1), 5–922.
- Robertson, M. P., Cumming, G. S., & Erasmus, B. F. N. (2010). Getting the most out of atlas data. *Diversity and Distributions*, 16(3), 363–375. <https://doi.org/10.1111/j.1472-4642.2010.00639.x>
- Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., & Müller, M. (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*, 12, 77.
- Rodríguez-Sánchez, F. (2023). *FloraIberica: Taxonomic and distribution data for the vascular plants of the Iberian Peninsula and Balearic Islands* (Version 0.0.3).
- Romero-Zarco, C. (2015). *Las gramíneas de la Península Ibérica e Islas Baleares: Claves ilustradas para la determinación de los géneros y catálogo preliminar de las especies*. Jolube Consultor Botánico y Editor.
- Ronquillo, C., Alves-Martins, F., Mazimpaka, V., Sobral-Souza, T., Vilela-Silva, B., García Medina, N., & Hortal, J. (2020). Assessing spatial and temporal biases and gaps in the publicly available distributional information of Iberian mosses. *Biodiversity Data Journal*, 8, e53474. <https://doi.org/10.3897/BDJ.8.e53474>
- Rosselló, J. A., & Castro, M. (2008). Karyological evolution of the angiosperm endemic flora of the Balearic Islands. *TAXON*, 57(1), 259–273. <https://doi.org/10.2307/25065967>
- Rouget, M., Richardson, D. M., & Cowling, R. M. (2003). The current configuration of protected areas in the Cape Floristic Region, South Africa—reservation bias and representation of biodiversity patterns and processes. *Biological Conservation*, 112(1), 129–145. [https://doi.org/10.1016/S0006-3207\(02\)00396-8](https://doi.org/10.1016/S0006-3207(02)00396-8)
- Rundel, P. W., Arroyo, M. T., Cowling, R. M., Keeley, J. E., Lamont, B. B., & Vargas, P. (2016). Mediterranean Biomes: Evolution of Their Vegetation, Floras, and Climate. *Annual Review of Ecology, Evolution, and Systematics*, 47(1), 383–407. <https://doi.org/10.1146/annurev-ecolsys-121415-032330>

- Sàbat, F., Gelabert, B., Rodríguez-Perea, A., & Giménez, J. (2011). Geological structure and evolution of Majorca: Implications for the origin of the Western Mediterranean. *Tectonophysics*, 510(1), 217–238. <https://doi.org/10.1016/j.tecto.2011.07.005>
- Safi, K., Armour-Marshall, K., Baillie, J. E. M., & Isaac, N. J. B. (2013). Global Patterns of Evolutionary Distinct and Globally Endangered Amphibians and Mammals. *PLOS ONE*, 8(5), e63582. <https://doi.org/10.1371/journal.pone.0063582>
- Sainz Ollero, H., & Hernández Bermejo, J. E. (1985). Sectorización fitogeográfica de la Península Ibérica e islas Baleares: la contribución de su endemoflora como criterio de semejanza. *Candollea*, 40, 485–508.
- Sainz Ollero, H., & Sánchez de Dios, R. (2011). La diversidad de los paisajes españoles. In *Biodiversidad. Aproximación a la diversidad botánica y zoológica de España* (pp. 109–156). Real Sociedad Española de Historia Natural.
- Sainz-Ollero, H., & Moreno-Saiz, J. C. (2002). Flora vascular endémica española. In *F. D. Pineda, J. M. De Miguel, M. A. Casado, & J. Montalvo (Eds.), La Diversidad Biológica de España* (pp. 175–195). CYTED. Prentice Hall.
- Sánchez de Dios, R., Benito-Garzón, M., & Sainz Ollero, H. (2009). Present and future extension of the Iberian submediterranean territories as determined from the distribution of marcescent oaks. *Plant Ecology*, 204(2), 189–205. <https://doi.org/10.1007/s11258-009-9584-5>
- Sánchez de Dios, R., Cabal Ruano, C., Domínguez Lozano, F., Sainz Ollero, H., & Moreno Saiz, J. C. (2017). The role of criteria in selecting important areas for conservation in biodiversity-rich territories. *Diversity and Distributions*, 23(4), 368–380. <https://doi.org/10.1111/ddi.12535>
- Sayers, E. W., Beck, J., Brister, J. R., Bolton, E. E., Canese, K., Comeau, D. C., Funk, K., Ketter, A., Kim, S., Kimchi, A., Kitts, P. A., Kuznetsov, A., Lathrop, S., Lu, Z., McGarvey, K., Madden, T. L., Murphy, T. D., O’Leary, N., Phan, L., ... Ostell, J. (2020). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research*, 48, D9–D16. <https://doi.org/10.1093/nar/gkz899>
- Sclater, P. L. (1858). On the general geographical distribution of the members of the class Aves. *Journal of the Proceedings of the Linnean Society: Zoology*, 2, 130–145.
- Sequeira, M., Espírito Santo, D., Aguiar, C., Capelo, J., & Honrado, J. J. (2011). *Checklist da Flora de Portugal Continental, Açores e Madeira*.
- Sérgio, C., Araújo, M., & Draper, D. (2000). Portuguese Bryophyte Diversity and Priority Areas for Conservation. *Lindbergia*, 25(2), 116–123.
- Sillero, N., Campos, J., Bonardi, A., Corti, C., Creemers, R., Crochet, P.-A., Isailović, J. C., Denoël, M., Ficetola, G. F., Gonçalves, J., Kuzmin, S., Lymberakis, P., Pous, P. d., Rodríguez, A., Sindaco, R., Speybroeck, J., Toxopeus, B., Vieites, D. R., & Vences, M. (2014). Updated distribution and biogeography of amphibians and reptiles of Europe. *Amphibia-Reptilia*, 35(1), 1–31. <https://doi.org/10.1163/15685381-00002935>

- Smith, S. A., & Brown, J. W. (2018). Constructing a broadly inclusive seed plant phylogeny. *American Journal of Botany*, 105(3), 302–314. <https://doi.org/10.1002/ajb2.1019>
- Smith, S. A., & Walker, J. F. (2019). PyPHLAWD: A python tool for phylogenetic dataset construction. *Methods in Ecology and Evolution*, 10(1), 104–108. <https://doi.org/10.1111/2041-210X.13096>
- Stebbins, G. L., & Major, J. (1965). Endemism and Speciation in the California Flora. *Ecological Monographs*, 35(1), 2–35. <https://doi.org/10.2307/1942216>
- Stein, R. W., Mull, C. G., Kuhn, T. S., Aschliman, N. C., Davidson, L. N. K., Joy, J. B., Smith, G. J., Dulvy, N. K., & Mooers, A. O. (2018). Global priorities for conserving the evolutionary history of sharks, rays and chimaeras. *Nature Ecology & Evolution*, 2(2), 288–298. <https://doi.org/10.1038/s41559-017-0448-4>
- Stork, N. E. (2018). How Many Species of Insects and Other Terrestrial Arthropods Are There on Earth? *Annual Review of Entomology*, 63(1), 31–45. <https://doi.org/10.1146/annurev-ento-020117-043348>
- Takhtajan, A. (1986). *Floristic regions of the world*. University of California Press.
- Thompson, J. D. (2020). *Plant Evolution in the Mediterranean: Insights for Conservation*. Oxford University Press.
- Upham, N. S., Esselstyn, J. A., & Jetz, W. (2019). Inferring the mammal tree: Species-level sets of phylogenies for questions in ecology, evolution, and conservation. *PLOS Biology*, 17(12), e3000494. <https://doi.org/10.1371/journal.pbio.3000494>
- Vane-Wright, R. I., Humphries, C. J., & Williams, P. H. (1991). What to protect?—Systematics and the agony of choice. *Biological Conservation*, 55(3), 235–254. [https://doi.org/10.1016/0006-3207\(91\)90030-D](https://doi.org/10.1016/0006-3207(91)90030-D)
- Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S* (4th ed.). Springer.
- Villar, L., Sesé, J. A., & Ferrández, J. V. (1997). *Atlas de la flora del Pirineo Aragonés*. Consejo de Protección de la Naturaleza de Aragón e Instituto de Estudios Altoaragoneses.
- Wallace, A. R. (1870). *Contributions to the Theory of Natural Selection*. London.
- Wallace, A. R. (1863). On the Physical Geography of the Malay Archipelago. *The Journal of the Royal Geographical Society of London*, 33, 217–234. <https://doi.org/10.2307/1798448>
- Wallace, A. R. (1876). *The Geographic Distribution of Animals*. Macmillan, London.
- Wang, Z., Du, S., Dayanandan, S., Wang, D., Zeng, Y., & Zhang, J. (2014). Phylogeny Reconstruction and Hybrid Analysis of Populus (Salicaceae) Based on Nucleotide Sequences of Multiple Single-Copy Nuclear Genes and Plastid Fragments. *PLOS ONE*, 9(8), e103645. <https://doi.org/10.1371/journal.pone.0103645>
- Webb, C. O., Ackerly, D. D., McPeek, M. A., & Donoghue, M. J. (2002). Phylogenies and Community Ecology. *Annual Review of Ecology and Systematics*, 33(1), 475–505. <https://doi.org/10.1146/annurev.ecolsys.33.010802.150448>
- Webb, C. O., & Donoghue, M. J. (2005). Phylomatic: tree assembly for applied phylogenetics. *Molecular Ecology Notes*, 5(1), 181–183. <https://doi.org/10.1111/j.1471-8286.2004.00829.x>

- Wegener, A. (1912). Die Entstehung der Kontinente. *Geologische Rundschau*, 3(4), 276–292. <https://doi.org/10.1007/BF02202896>
- Weitzman, M. L. (1992). On Diversity*. *The Quarterly Journal of Economics*, 107(2), 363–405. <https://doi.org/10.2307/2118476>
- Weitzman, M. L. (1998). The Noah's Ark Problem. *Econometrica*, 66(6), 1279–1298. <https://doi.org/10.2307/2999617>
- Whittaker, R. J., Araújo, M. B., Jepson, P., Ladle, R. J., Watson, J. E. M., & Willis, K. J. (2005). Conservation Biogeography: assessment and prospect. *Diversity and Distributions*, 11(1), 3–23. <https://doi.org/10.1111/j.1366-9516.2005.00143.x>
- Willis, K. J., & Whittaker, R. J. (2000). The Refugial Debate. *Science*, 287(5457), 1406–1407. <https://doi.org/10.1126/science.287.5457.1406>
- Winter, M., Devictor, V., & Schweiger, O. (2013). Phylogenetic diversity and nature conservation: where are we? *Trends in Ecology & Evolution*, 28(4), 199–204. <https://doi.org/10.1016/j.tree.2012.10.015>
- Woodland, D. W. (2007). Are botanists becoming the dinosaurs of biology in the 21st century? *South African Journal of Botany*, 73(3), 343–346. <https://doi.org/10.1016/j.sajb.2007.03.005>
- Ye, J.-F., Liu, Y., & Chen, Z.-D. (2020). Dramatic impact of metric choice on biogeographical regionalization. *Plant Diversity*, 42(2), 67–73. <https://doi.org/10.1016/j.pld.2019.12.003>
- Ye, J., Lu, L., Liu, B., Yang, T., Zhang, J., Hu, H., Li, R., Lu, A., Liu, H., Mao, L., & Chen, Z. (2019). Phylogenetic delineation of regional biota: A case study of the Chinese flora. *Molecular Phylogenetics and Evolution*, 135, 222–229. <https://doi.org/10.1016/j.ympev.2019.03.011>
- Zanne, A. E., Tank, D. C., Cornwell, W. K., Eastman, J. M., Smith, S. A., FitzJohn, R. G., McGlinn, D. J., O'Meara, B. C., Moles, A. T., Reich, P. B., Royer, D. L., Soltis, D. E., Stevens, P. F., Westoby, M., Wright, I. J., Aarssen, L., Bertin, R. I., Calaminus, A., Govaerts, R., ... Beaulieu, J. M. (2014). Three keys to the radiation of angiosperms into freezing environments. *American Journal of Botany*, 506, 89–92.

R

Appendices

A

Atlas of the vascular flora of the Iberian Peninsula biodiversity hotspot (AFLIBER)

Ignacio Ramos-Gutiérrez^{1,2}  | Herlander Lima³ | Santiago Pajarón⁴  |
 Carlos Romero-Zarco⁵  | Llorenç Sáez⁶  | Luciano Pataro¹ |
 Rafael Molina-Venegas³  | Miguel Á. Rodríguez³  | Juan Carlos Moreno-Saiz^{1,2} 

¹Department of Biology (Botany), Universidad Autónoma de Madrid, Madrid, Spain

²Research Center on Biodiversity and Global Change (CIBC-UAM), Universidad Autónoma de Madrid, Madrid, Spain

³GloCEE – Global Change Ecology & Evolution Group, Department of Life Sciences, University of Alcalá, Alcalá de Henares, Spain

⁴Departamento de Biodiversidad, Universidad Complutense de Madrid, Madrid, Spain

⁵Departamento de Biología Vegetal y Ecología, Universidad de Sevilla, Sevilla, Spain

⁶Departament of Animal Biology, Plant Biology and Ecology, Universitat Autònoma de Barcelona, Bellaterra, Spain

Correspondence

Ignacio Ramos-Gutiérrez, Department of Biology (Botany), Universidad Autónoma de Madrid, 28049, Madrid, Spain.

Email: ignacio.ramosgutierrez@uam.es

Funding information

Ministerio de Economía, Industria y Competitividad, Gobierno de España, Grant/Award Number: CGL2017-86926-P

Editor: Jonathan Lenoir

Abstract

Motivation: We accessed published and unpublished floristic sources to compile a comprehensive species list of the Iberian–Balearic terrestrial vascular flora and generate AFLIBER, an accurate floristic database of georeferenced plant occurrence records.

Main type of variable contained: Species distribution data totalling 1,824,549 plant occurrence records corresponding to 6,456 species and subspecies.

Spatial location and grain: The western Mediterranean, including inland territories of Spain, Portugal and Andorra and the adjacent archipelagos of Berlengas, Columbretes and the Balearic Islands, covered by 6,316 UTM quadrangular grid cells of 10 km resolution.

Time period: All distributional trustable records were considered to create the AFLIBER database, most of them dating from the 1960s onwards.

Major taxa and level of measurement: Terrestrial vascular plant species and subspecies.

Software format: Data are supplied as comma-separated text (csv) files.

KEY WORDS

Balearic Islands, biodiversity hotspot, geographical distribution, Portugal, Spain, vascular plants

1 | INTRODUCTION

With nearly 6,500 native vascular plant species and subspecies, of which 28% are endemics (Buira et al., 2017), the Iberian Peninsula (western Mediterranean) and its eastern Baetic offshore

prolongation (the Balearic archipelago; Sàbat et al., 2011) comprise about one-quarter of the total vascular taxa of the Mediterranean basin (c. 25,000 species), the world's third-richest plant biodiversity hotspot (Fady & Concord, 2010; Myers et al., 2000). This outstanding plant diversity prompted the *Flora iberica* (FI) project in the

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *Global Ecology and Biogeography* published by John Wiley & Sons Ltd.

mid-1980s, a long-lasting initiative aimed at synthesizing the knowledge of systematic botany of the Iberian–Balearic region. Since then, 22 volumes have been published describing c. 5,600 species and subspecies (Castroviejo, 1986–2019), providing valuable information on whether a taxon is endemic to the region and in which administrative provinces it occurs. Yet, FI is still a work in progress for two main reasons. First, the last two volumes devoted to the grass family Poaceae (c. 500 taxa in the region) are still underway. Second, there is a continuous publication of new plant occurrences and revisions of taxonomical treatments based on both classical and molecular systematics. Consequently, FI also tackles new botanical publications to collate newly described taxa and plant occurrences on its website (<http://www.floraiberica.es>).

Besides, the Spanish plant information system, Anthos (<http://www.anthos.es/>), was created soon after the FI project to host floristic occurrences, and it currently comprises c. 2 million plant records georeferenced according to the standard European UTM grid at 10 km resolution. A similar structure has been used for the equivalent database of Portugal, Flora.On (<https://flora-on.pt>), and many subnational biodiversity databases, atlases and/or red books on plant distributions across this region (e.g., Bañares et al., 2004; Villar et al., 1997–2001). Importantly, all plant records included in Anthos, plus c. 13 million extra records from other databases, have been uploaded to the Global Biodiversity Information Facility (GBIF; <https://www.gbif.org/>), the largest source of Iberian plant distributional data hitherto.

Although we acknowledge the extraordinary value of open-access biodiversity repositories, it should be noted that some records are often duplicated, imprecise or of dubious nature (Gaiji et al., 2013). Also, many records remain hardly accessible in non-digitized bibliographical sources (e.g., data from red books or floristic atlases) and restricted regional electronic repositories. Moreover, floristic information is often dated and riddled with errors (Anderson et al., 2016; Maldonado et al., 2015), which might have hampered fine-grained plant macroecological studies in the western Mediterranean. For instance, inflated omission error rates often force spatial analyses to be conducted at coarse grains (e.g., Buira et al., 2017; Lobo et al., 2001; Moreno-Saiz et al., 2013) that might not be appropriate to inform conservation planning and policy-making.

In this data paper, we provide AFLIBER, a comprehensive database connecting the up-to-date species list of the Iberian–Balearic vascular native flora with its matching occurrence records rescaled to a standard UTM 10 km grain resolution. The outstanding characteristics of this database are threefold: (1) it has been generated according to the latest knowledge in floristics and taxonomy published for the region; (2) it has reduced geographical biases by incorporating both subregional databases broadly absent from previous wide-ranging repositories and distributional information published in monographs, grey literature and floristic articles; and (3) it provides occurrence records that were filtered conservatively to minimize identification and georeferencing errors (see Methods) by means of a combined automatic and manual protocol undertaken by professional botanists specialized in the Iberian flora. Thus, this dataset represents an unprecedented effort to support

botanical, biogeographical and macroecological studies in the western Mediterranean by providing comprehensive and carefully curated plant distributional information across the Iberian Peninsula.

2 | METHODS

The study area includes the Iberian Peninsula (mainland Portugal, Spain and Andorra) and the adjacent archipelagos of Berlengas (0.99 km²), Columbretes (0.19 km²) and the Balearic Islands (4,992 km²), altogether comprising 6,316 terrestrial 10-km-side UTM grid cells (see Supporting Information Figure S2.1 in Supplement 2).

2.1 | Data compilation

First, we compiled all native vascular plant species and subspecies included in the published volumes of FI (Castroviejo, 1986–2019). Species in the Poaceae family (unpublished volumes) were obtained from different sources, including a comprehensive monograph by Romero-Zarco (2015), generic treatments published in recent years (see supplementary bibliography in Supporting Information Supplement 4a) and personal communications from experts involved in the preparation of the FI Poaceae volumes (i.e., C. Acedo, personal communication; J.A. Devesa, personal communication). Second, we conducted an exhaustive literature review to retrieve all newly published taxa (until July 2020) that were missing in the previous sources. Specifically, we revised the “new taxa” tab of the FI website (www.floraiberica.es/miscelania/nuevos_taxones.php), updated generic monographies and articles with newly recorded or described species in the study area (see Supporting Information Supplement 4a: Supplementary References). Taxa were considered endemic whenever their distribution was limited to the study area, including the northern slope of the Pyrenees (border with France), thus following previous biogeographical accounts of the Iberian flora (e.g., Buira et al., 2021; Moreno-Saiz & Sainz Ollero, 1992).

All distributional information at 10 km resolution was retrieved from high-quality, trustable databases available from electronic repositories and published research papers (Supporting Information Supplement 4b), including an unpublished database that was carried out over 30 years of data compilation from non-digitized sources (IberBal-Flora; see Supporting Information Supplement 4b). In order to complement the survey, GBIF plant occurrence records with a minimum resolution of 10 km were imported using the “rgbif” R package (Chamberlain et al., 2019). It is important to note that GBIF includes large numbers of unverified records that do not reflect the native distribution of species (e.g., corresponding to *ex situ* cultivation, botanical gardens). Thus, in order to tease apart doubtful and/or potentially erroneous information conservatively, we filtered GBIF records to include only those georeferenced within the limits of the administrative provinces of Spain, Portugal and Andorra where the species occur according to FI and other later reliable publications. Clearly outlying observations were removed using the *over* function of the

"sp" R package (Pebesma & Bivand, 2005). References on occurrence records retrieved from GBIF are shown in the Supporting Information (Supplement 4c). Furthermore, the resultant distributional maps were explored visually one by one by the professional botanists of the team (i.e., C.R.Z., I.R.G., J.C.M.S., L.S. and S.P.) in search for erroneous or doubtful occurrences and to amend potential information gaps manually on the basis of expert knowledge of the study region (Figure 1). A list including 29,600 disregarded records is provided in the Supporting Information (Supporting Information Table S3.3 in Supplement 3).

2.2 | Analysis of completeness

We used the nonparametric Chao2 estimator (Colwell & Coddington, 1994) to detect unequally sampled areas across the territory. To do so, the study area was divided into administrative provinces, following the FI scheme (Supporting Information Figure S2.2 in Supplement 2). Each province containing fewer than fifty 10 km grid cells was merged with its nearest neighbouring one to avoid extremely distinct sizes among them, making a total of 56 provinces (ranging from 4,500 to 21,700 km²). The analysis was performed within provinces and for the complete territory as a single unit, using the *specpool* function as implemented in the "vegan" R package (Oksanen et al., 2019).

3 | RESULTS

3.1 | Species list and distribution data

A total of 6,456 native species and subspecies were compiled (5,681 species after collapsing subspecies), of which 2,142 are endemic to

the study area (Table 1a). We gathered a total of 2,716,018 distributional records (Table 1b; Figure 2a) comprising 1,824,549 unique (i.e., non-duplicated) occurrences at the 10-km-side grid-cell resolution (Figure 2b). Plant occurrences were distributed across 6,303 grid cells (99.8% of total cells), of which 6,177 included species that are endemic to the study area (97.8% of the total; Figure 2c). Empty cells (0.2%) generally corresponded to coastal cells, only covering inland area in part, or to areas located between UTM transition zones.

3.2 | Database completeness

Floristic completeness for the entire study area was strikingly high (98.57%), suggesting that the flora in the Iberian Peninsula is already well known and well represented in our database. Indeed, when investigated separately throughout administrative provinces, completeness estimates remained high overall, with values ranging between 85 and 96% (cf. Figure 2d).

4 | DISCUSSION

The outstanding plant diversity and geographical extension of the Iberian Peninsula within the Mediterranean basin hotspot makes it an ideal setting to advance our understanding of the Mediterranean flora (Nieto Feliner, 2014), which has been greatly hampered by the lack of a comprehensive source of regional plant distribution. Given its size and exhaustive data curation, the AFLIBER database represents a major step towards filling in this central gap of floristic and distributional knowledge in the western Mediterranean, having the potential to become the new Iberian reference for future plant macroecological, biogeographical and conservation studies. Indeed, with

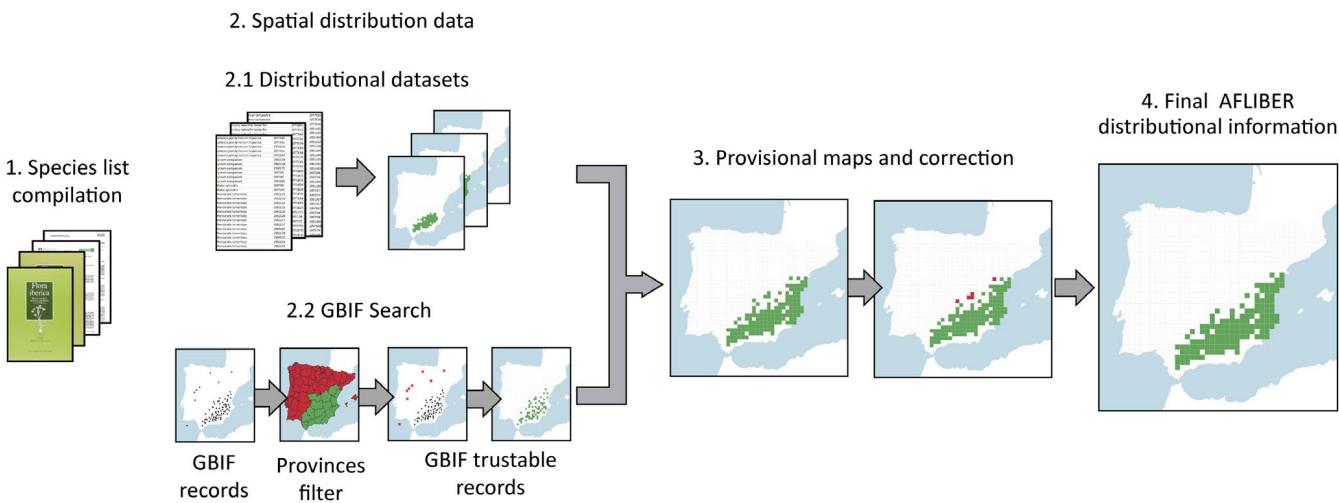


FIGURE 1 Workflow for creation of AFLIBER. (1) We compiled all species described in *Flora iberica* (FI) and later monographs and publications. (2) For each species in the list, we retrieved spatial distribution data in 10 km UTM cells, following a twofold procedure: (2.1) we retrieved information from reliable datasets (see Supporting Information Supplement 4b); and (2.2) the GBIF occurrence records were imported, disregarding occurrences with low coordinate resolution or that mapped away from the known distribution of the species (provinces filter). (3) These data were plotted as provisional maps that were visually explored and corrected by experts to achieve (4) the final result

Variable	Definition
(a) AFLIBER_Species_list.csv	
Taxon	Genus, species and subspecies (if applicable) epithets
Scientific_Name	Accepted scientific name
Endemic	Endemism of the Balearic Islands and/or the Iberian Peninsula (including French northern Pyrenean slope)
Genus	Adopted taxonomical category
Species	Adopted taxonomical category
Subspecies	Adopted taxonomical category
Class	Taxonomical category according to the National Center for Biotechnology Information (www.ncbi.nlm.nih.gov/guide/taxonomy/)
Order	Taxonomical category (according to NCBI)
Family	Taxonomical category (according to NCBI)
GBIF_id	Taxonomic numerical identifier in the Global Biodiversity Information Facility (https://www.gbif.org/)
POW_Name	Accepted taxonomic name in <i>Plants of the World online</i> (www.plantsoftheworldonline.org)
(b) AFLIBER_Distributions.csv	
Taxon	Genus, species and subspecies (if applicable) epithets
UTM.cell	UTM 10-km-side grid cell where the taxon is recorded
References	Sources from which the occurrence data were obtained. Numerical references correspond to those shown in the Supporting Information (Supplement 4b)

TABLE 1 Data files variable names and description (data can be found on Dryad at: <https://doi.org/10.5061/dryad.gmsbc2kv>)

its 6,456 species and subspecies, this database not only approaches the 6,500 taxa estimated by Buira et al., (2017), but also elevates the number of endemic Iberian taxa to 2,142 species and subspecies. These numbers posit a higher figure than earlier accounts of Iberian plant diversity (Buira et al., 2017, 2020; Sainz-Ollero & Moreno-Saiz, 2002), representing an increase in endemics above 30% and thus equalling Iberian plant endemism level to that of the Anatolian Peninsula in the Eastern Mediterranean (Davis et al., 1988). Interestingly, this pattern is much in line with earlier studies that recognized both ends of the Mediterranean basin as major centres of plant endemism (Médail & Quézel, 1997).

It is important to note that, despite all the efforts to detect mistakes and misidentifications, the AFLIBER database might include some erroneous records and taxonomic treatments that are still subject to debate. Yet, bearing in mind that our knowledge of the Iberian flora remains incomplete, AFLIBER is certainly the most exhaustive and refined Iberian plant database published hitherto, representing a major improvement over other popular repositories of distributional information widely used in the past. For instance, the latest update of Anthos, the main source of reference for Iberian plant distribution so far, dates back to 2016 (L. Medina, personal communication).

Although floristic completeness across the study area was acceptable (see Figure 2d), we note that some territories have been less prospected (Figure 2a) or lack distributional information in standard UTM grid cells (i.e., atlases), particularly towards western and southern areas. Such gaps of floristic knowledge might lead to

unbalanced accounts of distributional information, which could be addressed both by future prospection conducted in poorly sampled areas and by georeferencing extant information treasured in herbaria. Nevertheless, distinct levels of plant diversity across Iberian territories might also respond to natural processes. For example, plant species richness peaked in mountainous regions (see Figure 2b; Castro-Parga et al., 1996; Lobo et al., 2001), whereas human modification of natural ecosystems has led to depauperated levels of diversity over the centuries in certain areas, such as inner depressions and fertile river basins. A similar pattern can be observed for endemic species, which are particularly abundant in mountain regions, such as the Pyrenees, the Cantabrian range and, especially, in the Sierra Nevada (Baetic range), the hottest spot of floristic diversity in the Mediterranean (Blanca et al., 1998; Buira et al., 2021; Heywood, 2003).

The essence of the AFLIBER database lies in taxonomic botany, a discipline that is slowly disappearing from academic curricula (Crisci et al., 2020). This ongoing crisis in botany is attributable, in part, to the bibliometrics that are used to judge the impact of scientific research, which often undermine plant taxonomy despite it being the basis of many downstream disciplines, such as ecology, biogeography and conservation biology (Crisci et al., 2020; Woodland, 2007). Nonetheless, professional and amateur Iberian botanists continue to generate priceless floristic and taxonomic data, and the recently constituted Spanish Botanical Society will take up the challenge of maintaining taxonomic botany alive as a foundational task, including data curation of the

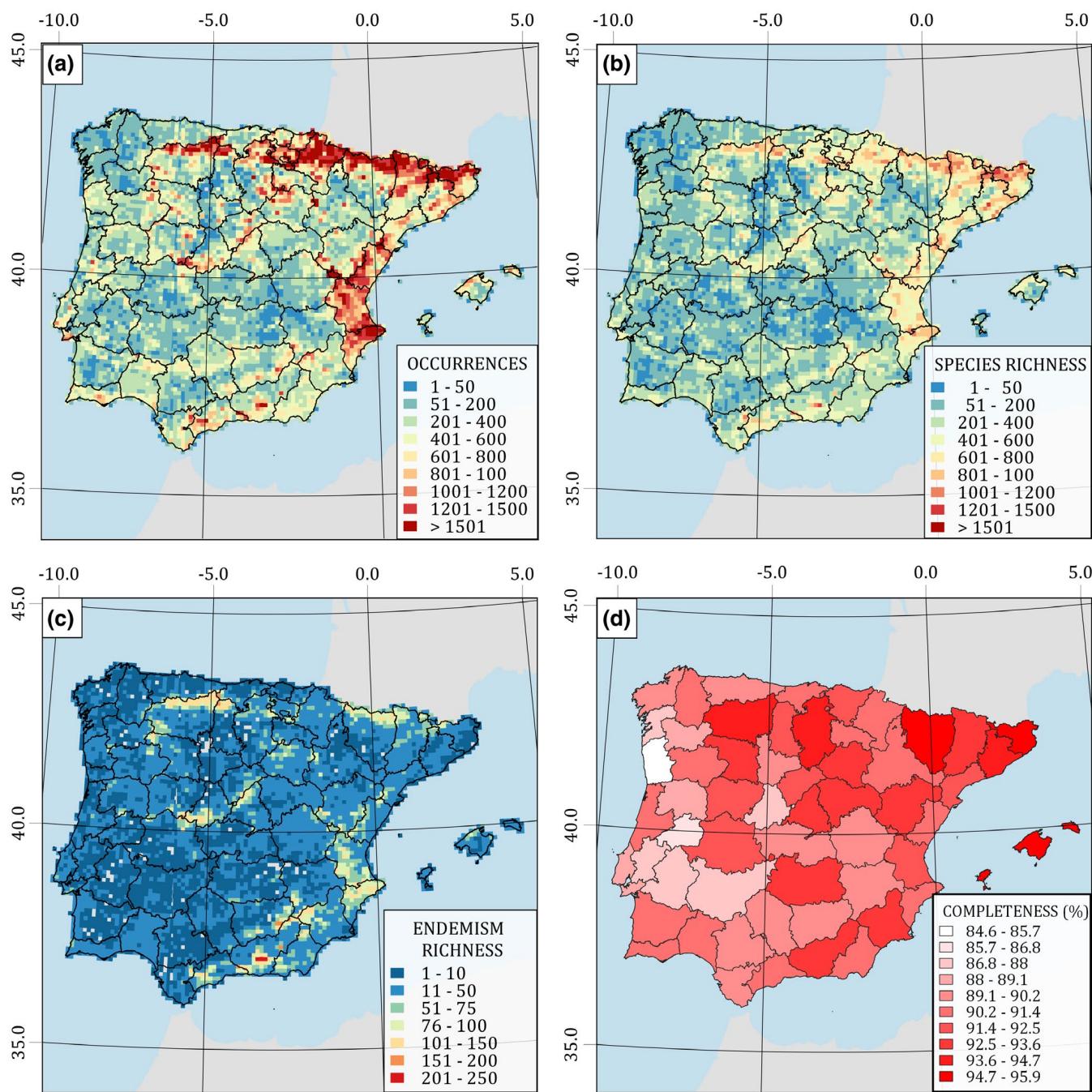


FIGURE 2 AFLIBER database number of (a) occurrences; (b) species richness (i.e., non-duplicated occurrences); (c) endemic species richness per 10 km UTM cell; and (d) percentage of regional floristic completeness

A

AFLIBER database and the incorporation of the two last volumes of *Flora iberica* (Poaceae) and other relevant essays in due course.

ACKNOWLEDGMENTS

We thank data curators who granted access to the restricted databases listed in the Supporting Information (Supplement 4b). We are also grateful to the many colleagues who revised the taxonomic treatment of some genera and contributed unpublished distributional data, especially Pedro Jiménez-Mejías, Carmen Acedo, Santiago Martín-Bravo and Juan Devesa. This work was funded by the Spanish Ministry of Science, Innovation and Universities through the Bioregions 2.0

project (CGL2017-86926-P) granted to M.A.R. and J.C.M.S. I.R.G. was supported by the Garantía Juvenil program (PEJ-2018-AI/AMB-9865) and R.M.-V. by the TALENTO program (2018-T2/AMB-10332), both from the Government of the Community of Madrid.

AUTHOR CONTRIBUTIONS

M.A.R., R.M.-V. and J.C.M.S. conceived the research; I.R.G., H.L., S.P., L.P., C.R.Z., L.S. and J.C.M.S. contributed to the dataset; I.R.G., S.P., C.R.Z., L.S. and J.C.M.S. revised and cleaned data; I.R.G. conducted analyses and drew the figures; R.M.-V., I.R.G. and J.C.M.S. led the writing with help from all the authors.

DATA AVAILABILITY STATEMENT

AFLIBER_Species_list.csv and AFLIBER_Distributions.csv data files are supplied as shown in Table 1 at Dryad Digital Repository (<https://doi.org/10.5061/dryad.gmsbcc2kv>). Updates on taxonomic and chorological information will be published at the AFLIBER database website: <https://iramosgutierrez.github.io/afliber>

ORCID

Ignacio Ramos-Gutiérrez  <https://orcid.org/0000-0002-8675-0114>

Santiago Pajaron  <https://orcid.org/0000-0003-2499-9341>

Carlos Romero-Zarco  <https://orcid.org/0000-0003-4178-2419>

Llorenç Sáez  <https://orcid.org/0000-0003-4551-2432>

Rafael Molina-Venegas  <https://orcid.org/0000-0001-5801-0736>

Miguel Á. Rodríguez  <https://orcid.org/0000-0002-4082-2995>

Juan Carlos Moreno-Saiz  <https://orcid.org/0000-0002-0793-9956>

REFERENCES

- Anderson, R. P., Araujo, M., Guisan, A., Lobo, J. M., Martinez-Meyer, E., Townsend, A., & Soberon, J. (2016). Are species occurrence data in global online repositories fit for modelling species distributions? In *The case of the Global Biodiversity Information Facility (GBIF), 2016: Final report of the task group on GBIF data fitness for use in distribution modelling*. Global Biodiversity Information Facility. <https://www.gbif.org/es/document/82612/report-of-the-task-group-on-gbif-data-fitness-for-use-in-distribution-modelling>
- Bañares, A., G. Blanca, J. Güemes, J. C. Moreno, & S. Ortiz (Eds.) (2004). *Atlas y Libro Rojo de la Flora Vascular Amenazada de España*. Dirección General de Conservación de la Naturaleza.
- Blanca, G., Cueto, M., Martínez-Lirola, M. J., & Molero-Mesa, J. (1998). Threatened vascular flora of Sierra Nevada (Southern Spain). *Biological Conservation*, 85, 269–285. [https://doi.org/10.1016/S0006-3207\(97\)00169-9](https://doi.org/10.1016/S0006-3207(97)00169-9)
- Buira, A., Aedo, C., & Medina, L. (2017). Spatial patterns of the Iberian and Balearic endemic vascular flora. *Biodiversity and Conservation*, 26, 479–508. <https://doi.org/10.1007/s10531-016-1254-z>
- Buira, A., Cabezas, F., & Aedo, C. (2020). Disentangling ecological traits related to plant endemism, rarity and conservation status in the Iberian Peninsula. *Biodiversity and Conservation*, 29, 1937–1958. <https://doi.org/10.1007/s10531-020-01957-z>
- Buira, A., Fernández-Mazuecos, M., Aedo, C., & Molina-Venegas, R. (2021). The contribution of the edaphic factor as a driver of recent plant diversification in a Mediterranean biodiversity hotspot. *Journal of Ecology*, 109, 987–999. <https://doi.org/10.1111/1365-2745.13527>
- Castro-Parga, I., Moreno-Saiz, J. C., Humphries, C. J., & Williams, P. H. (1996). Strengthening the Natural and National Park system of Iberia to conserve vascular plants. *Botanical Journal of the Linnean Society*, 121, 189–206. <https://doi.org/10.1111/j.1095-8339.1996.tb00753.x>
- Castroviejo, S., coord. (1986–2019). *Flora iberica 1–18, 20–21*. Real Jardín Botánico, CSIC.
- Chamberlain, S., Barve, V., Desmet, P., Geffert, L., Mcglinn, D., Oldoni, D., & Karthik, R (2019). *rgbif: Interface to the Global 'Biodiversity' Information Facility API*. R package version 1.3.0. <https://CRAN.R-project.org/package=rgbif>
- Colwell, R. K., & Coddington, J. A. (1994). Estimating terrestrial biodiversity through extrapolation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 345, 101–118.
- Crisci, J. V., Katinas, L., Apodaca, M. J., & Hoch, P. C. (2020). The end of botany. *Trends in Plant Science*, 25, 1173–1176. <https://doi.org/10.1016/j.tplants.2020.09.012>
- Davis, P. H., Mill, R. R., & Tan, K. (1988). *Flora of Turkey and the East Aegean Islands*, 10. Edinburgh University Press.
- Fady, B., & Concord, C. (2010). Macroecological patterns of species and genetic diversity in vascular plants of the Mediterranean basin. *Diversity and Distributions*, 16, 53–64. <https://doi.org/10.1111/j.1472-4642.2009.00621.x>
- Gaiji, S., Chavan, V., Ariño, A. H., Otegui, J., Hoborn, D., Sood, R., & Robles, E. (2013). Content assessment of the primary biodiversity data published through GBIF network: Status, challenges and potentials. *Biodiversity Informatics*, 8, 94–172.
- Heywood, V. H. (2003). The jewel in the crown. *Plant Talk*, 32, 42–43.
- Lobo, J. M., Castro, I., & Moreno, J. C. (2001). Spatial and environmental determinants of vascular plant species richness distribution in the Iberian Peninsula and Balearic Islands. *Biological Journal of the Linnean Society*, 73, 233–253. <https://doi.org/10.1111/j.1095-8312.2001.tb01360.x>
- Maldonado, C., Molina, C. I., Zizka, A., Persson, C., Taylor, C. M., Albán, J., Chilquillo, E., Rønsted, N., & Antonelli, A. (2015). Estimating species diversity and distribution in the era of big data: to what extent can we trust public databases? *Global Ecology and Biogeography*, 24, 973–984.
- Médail, F., & Quézel, P. (1997). Hot-spots analysis for conservation of plant biodiversity in the Mediterranean Basin. *Annals of the Missouri Botanical Garden*, 84, 112–127. <https://doi.org/10.2307/2399957>
- Moreno-Saiz, J. C., Donato, M., Katinas, L., Crisci, J. V., & Posadas, P. (2013). New insights into the biogeography of south-western Europe: Spatial patterns from vascular plants using cluster analysis and parsimony. *Journal of Biogeography*, 40, 90–104. <https://doi.org/10.1111/j.1365-2699.2012.02774.x>
- Moreno-Saiz, J. C., & Sainz-Ollero, H. (1992). *Atlas corológico de las monocotiledóneas endémicas de la Península Ibérica e islas Baleares*. ICONA.
- Myers, N., Mittermeier, R., Mittermeier, C., da Fonseca, G. A. B., & Kent, J. (2000). Biodiversity hotspots for conservation priorities. *Nature*, 403, 853–858. <https://doi.org/10.1038/35002501>
- Nieto Feliner, G. (2014). Patterns and processes in plant phylogeography in the Mediterranean Basin. A review. *Perspectives in Plant Ecology, Evolution and Systematics*, 16, 265–278. <https://doi.org/10.1016/j.ppees.2014.07.002>
- Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H., Szoecs, E., & Wagner, H. (2019). *vegan: Community ecology package*. R package version 2.5-6. <https://CRAN.R-project.org/package=vegan>
- Pebesma, E. J., & Bivand, R. S. (2005). Classes and methods for spatial data in R. *R News*, 5, 9–13.
- Romero-Zarco, C. (2015). Las gramíneas de la Península Ibérica e Islas Baleares. In Jolube Consultor Botánico y (Ed.). *Colección Monografías de Botánica Ibérica* (pp.15). Jaca.
- Sàbat, F., Gelabert, B., Rodríguez-Perea, A., & Giménez, J. (2011). Geological structure and evolution of Majorca: Implications for the origin of the Western Mediterranean. *Tectonophysics*, 510, 217–238. <https://doi.org/10.1016/j.tecto.2011.07.005>
- Sainz-Ollero, H., & Moreno-Saiz, J. C. (2002). Flora vascular endémica española. In F. D. Pineda, J. M. De Miguel, M. A. Casado, & J. Montalvo (Eds.), *La Diversidad Biológica de España* (pp. 175–195). CYTED. Prentice Hall.
- Villar, L., Sesé, J. A., & Ferrández, J. V. (1997–2001). *Atlas de la flora del Pirineo Aragonés*. Consejo de Protección de la Naturaleza de Aragón e Instituto de Estudios Altoaragoneses.
- Woodland, D. W. (2007). Are botanists becoming the dinosaurs of biology in the 21st century? *South African Journal of Botany*, 73, 343–346. <https://doi.org/10.1016/j.sajb.2007.03.005>

BIOSKETCH

Ignacio Ramos-Gutiérrez is a PhD student conducting a dissertation about biogeographical patterns in Iberian vascular plants. The research group in which his thesis is framed is focused on macroecological patterns at different resolutions and is currently working on the development of modern, up-to-date bioregionalizations. This dataset represents a cornerstone of Ignacio Ramos-Gutiérrez's doctoral dissertation, because it will be the data to use for his analyses.

SUPPORTING INFORMATION

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

How to cite this article: Ramos-Gutiérrez, I., Lima H., Pajarón S., Romero-Zarco C., Sáez L., Pataro L., Molina-Venegas R., Rodríguez M. Á., & Moreno-Saiz J. C. (2021). Atlas of the vascular flora of the Iberian Peninsula biodiversity hotspot (AFLIBER). *Global Ecology and Biogeography*, 00, 1–7. <https://doi.org/10.1111/geb.13363>

Extended Methods

Species list compilation

We extracted all native vascular plant species and subspecies described in the published volumes of *Flora iberica* so far (Castroviejo, 1986-2019), which cover all plant families but the Poaceae (two volumes, work in progress). To fill in this gap, we used the recent comprehensive monograph by Romero-Zarco (2015) on the Poaceae family, complemented with some generic treatments published in recent years (*cf.* Table S1.1) and personal communications from experts that were involved in the elaboration of the ongoing FI Poaceae volumes (i.e. Acedo, pers. com; Devesa, pers. com).

Given the long-lasting duration of the FI project (over 30 years), several taxa have subsequently been added to the Iberian-Balearic floristic catalogue since the publication of their respective FI volumes, either because they went unnoticed at the time or because they were taxa new to science. Therefore, we completed the initial species list (until July 2020) following several steps:

- A revision of the “new taxa” tab as recognized by the FI website (www.floraiberica.es/miscelania/nuevos_taxones.php).
- An exhaustive literature review of updated generic monographies and articles with newly recorded or described taxa for the study area (see Appendix 1a-Supplementary References).
- A revision of the official “Spanish standard list of vascular flora” published by the Ministry of the Environment (MITECO, 2016).
- A comparison with the recently published list of Iberian-Balearic endemics (Buira et al., 2020b) to seek for possible gaps in our search.

Table S1.1 Taxa number (and percentage) contributed by the sources used to build the AFLIBER species list.

Source	Number of taxa (%)
<i>Flora iberica</i>	5,764 (89.3%)
Poaceae monographs	465 (7.2%)
Recent revisions and taxonomical updates	166 (2.6%)
Newly described or discovered taxa	61 (0.9%)

The species listed in AFLIBER cover all accepted vascular plant species and subspecies known for the study area so far, identified by Linnean binomials or trinomials (subspecies, if applicable). The supra-generic taxonomic hierarchy for each taxon is also provided, including Family, Order and Class according to the National Center for Biotechnology Information taxonomic database (NCBI; www.ncbi.nlm.nih.gov/taxonomy). Besides, GBIF identification codes and accepted names according to Kew's Plants of the World Online taxonomy database (www.plantsoftheworldonline.org) are provided.

The endemic status of the taxa is provided (distribution limited to the Iberian Peninsula and/or the Balearic Islands, following FI criterion), including a few Iberian species whose distributions slightly overstepped into the Pyrenean French northern slope which were also considered as endemics (*cf.* Buira et al., 2020b; Moreno-Saiz & Sainz Ollero, 1992).

Spatial distribution data

Distributional information at 10-km resolution was compiled from non-digitized sources including atlases and research articles to an unpublished database carried out over 30 years (IberBal-Flora; see Table S3.3 in Supplement 3). Although distributional information at finer scales (i.e. 1-km) is available for some taxa (specially endemic and threatened species in certain mountain ranges, islands, and national parks), we are still very far from achieving such detailed spatial resolution for the entire flora of the territory. Therefore, occurrence data below the 10-km resolution were upscaled to the 10-km grid, which represents the most detailed scale possible for the vast majority of the species.

This initial bulk of data was enlarged with distributional information hosted in electronic repositories of both open and restricted access. For the latter, governmental or regional agencies granted us permits for accessing information pertaining threatened plants or species of regional interest, provided that we georeferenced them at coarse (10 km) resolution (for example, Andalusia, Castilla y León, Region of Murcia; see Appendix 1b) to safeguard the localities where threatened species thrive. Finally, GBIF records were imported using the function *occ_search* in 'rgbif' R package (Chamberlain & Boettiger, 2017) to complement the survey. We retrieved all available occurrences ranged between 35° and 45° latitude and -10° and 5° longitude, excluding records with low spatial

resolution (i.e. coordinates with 2 or less decimal digits). Regardless of the number of records available, only one specific quotation was downloaded per each 10-km cell. GBIF datasets considered in AFLIBER compilation were retrieved, and are referenced in Appendix 1c- GBIF References. Additionally, major data publishers of GBIF Iberian-Balearic plant records are acknowledged in Table S1.2 (Supplement 1).

GBIF includes many records that do not reflect the native distribution of species (e.g. georeferenced cultivations in parks or botanical gardens, herbarium sheets preserved in research institutions, etc.). Thus, in order to tease apart doubtful and/or potentially erroneous information, we filtered the records to exclude those that mapped clearly away from the known distribution of the species (i.e. administrative provinces of Spain, Portugal and Andorra according to FI and other subsequent reliable publications) using the *over* function of the ‘sp’ R package (Pebesma & Bivand, 2005). This conservative procedure reduces commission errors at the expense of increasing omission errors to some extent, a trade-off that nevertheless conferred AFLIBER with the robustness of being rooted on sound chorological information.

From all repositories, we imported only those records that exactly matched the taxonomic names accepted in the species list compilation. Afterwards, records referring to scientific names other than accepted were checked manually and incorporated into AFLIBER only if they met any of the following conditions:

- They referred to a clear synonym (i.e. different names for the same taxon),
- They referred to a different taxonomic rank that is directly transferable to an accepted taxon (e.g. records for *Nepeta ucranica* subsp. *hispanica* (Boiss. & Reut.) Bellot were assigned to *N. hispanica* Boiss. & Reut. in our list, and those for *Acer granatense* Boiss. were assigned to *A. opalus* subsp. *granatense* (Boiss.) Font Quer & Rothm.).
- They referred to a species for which we recognized several allopatric (i.e. non-overlapping) subspecies so that we could assign the former to the corresponding accepted subspecies.

Otherwise, we adopted the conservative criterion of rejecting the records to avoid mistakes in the distributional database.

Accepted records were plotted as provisional taxon-specific maps that were visually explored one by one to correct doubtful locations or questionable distribution gaps based

on expert knowledge. Doubtful records (e.g. isolated locations that mapped away from the known distribution of the species) were verified using authoritative electronic or bibliographical sources and otherwise removed from the database (see Table S3.3 in Supplement 3 for a list of disregarded records). We manually corrected distributional gaps due to spelling mistakes, nomenclatural disagreements or missed occurrences in local non-digitized journals (Figure 1).

Analysis of completeness

The non-parametric Chao2 estimator (Colwell & Coddington, 1994) was used to detect unequally sampled regions across the study area. To do so, the territory was divided into administrative provinces following the FI scheme (see Figure S2.2, Supplement 2). Each province containing less than fifty 10-km grid cells was merged with its nearest neighbouring one to avoid extreme variation in size, resulting in 56 provinces ranging from 21,700 to 4,500 km². The analysis was performed within provinces and for the complete territory as a single unit, using the *specpool* function as implemented in ‘vegan’ R package (Oksanen et al., 2019). For each individual test, AFLIBER database was trimmed in order to maintain only cells overlapping the target province, and used to create a presence-absence table from which we obtained the Chao2 estimator. After evaluating each region, these values were merged into the attribute table of administrative provinces shown in Figure 2d. The regional completeness map was then generated using QGIS 3.10 (QGIS Development Team, 2019).

Table S1.2. Largest plant datasets included in GBIF for the Iberian Peninsula and the Balearic Islands.

Dataset	Approx. number of occurrences
Localización de táxones botánicos de interés para la identificación de hábitats de la REDIAM. https://doi.org/10.15470/zatyqi	6,500,000
FloraCAT: Banco de datos de los cormófitos de Cataluña. https://doi.org/10.15470/fzwiiu	1,700,000
Sistema de Información de la vegetación Ibérica y Macaronésica. https://doi.org/10.15468/qyzfdt	1,600,000
Banco de Datos de la Biodiversidad de la Comunitat Valenciana. https://doi.org/10.15468/b4yqdy	1,600,000
CSIC-Real Jardín Botánico-Anthos. Sistema de Información de las Plantas de España. https://doi.org/10.15468/4wnutv	1,500,000
Pl@ntNet automatically identified occurrences. https://doi.org/10.15468/mma2ec	663,000
Tercer Inventario Forestal Nacional. Ministerio de Agricultura, Alimentación y Medio Ambiente (España). https://doi.org/10.15468/tkezjv	530,000
Cartografía de vegetación a escala de detalle 1:10.000 de la masa forestal de Andalucía. https://doi.org/10.15468/a7isx1	420,000
Flora-On: occurrence data of the flora of mainland Portugal. https://doi.org/10.15468/gi6aum	253,000
CSIC-Real Jardín Botánico-Colección de Plantas Vasculares (MA). https://doi.org/10.15468/mug7kr	195,000
RNF - Données de la Fédération des Réserves Catalanes. https://doi.org/10.15468/ufryrd	185,000
Colección de plantas vasculares del herbario de la Universitat de València (VAL). https://doi.org/10.15468/xmki52	163,000
Base de datos de plantas vasculares del País Vasco: ARAN-EH. https://doi.org/10.15468/m86bzk	150,000
CeDoc de Biodiversitat Vegetal: BCN-Cormophyta. https://doi.org/10.15468/x5ljvi	109,000
Catálogo Florístico Histórico de Navarra. https://doi.org/10.15468/nf54bd	103,000

New references

MITECO, Ministerio para la Transición Ecológica (2016). Retrieved from https://www.miteco.gob.es/en/biodiversidad/servicios/banco-datos-naturaleza/informacion-disponible/BDN_listas_patron.aspx

QGIS Development Team, 2019. QGIS Geographic Information System. Open Source Geospatial Foundation Project. <http://qgis.osgeo.org>

Supplementary figures

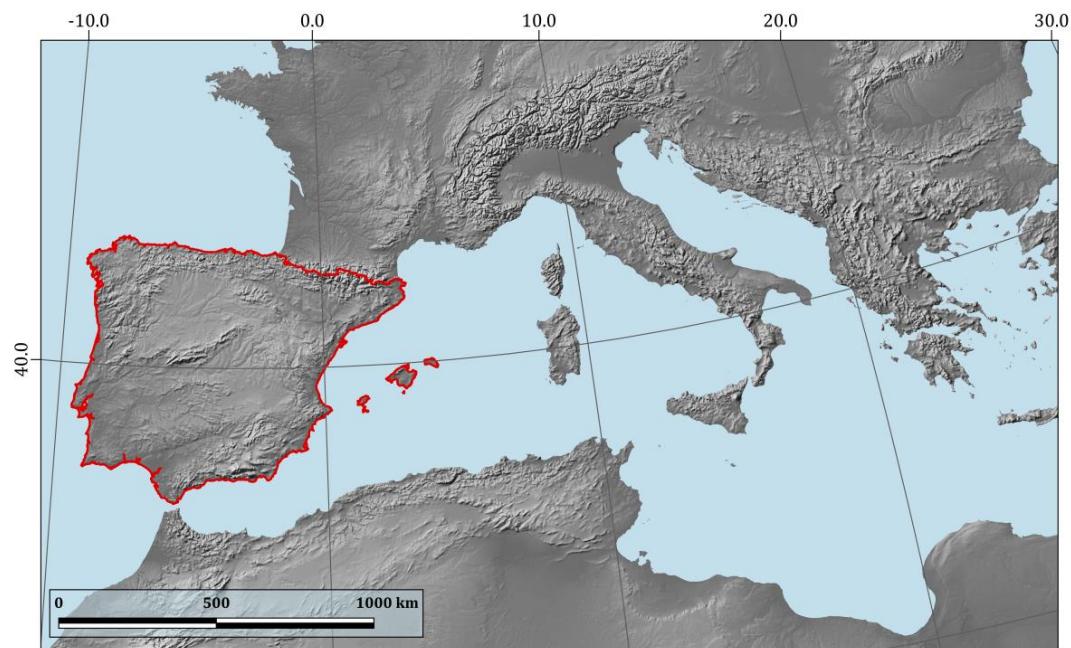


Figure S2.1. Study area (Iberian Peninsula and adjacent continental archipelagoes) on its western Mediterranean context.

A



Figure S2.2. Administrative regions in mainland Portugal, Spain and Andorra. Dotted lines represent the borders of the provinces that were merged for the analysis of completeness.

A

Supplementary Bibliography

Alarcón Cavero, M.L. (2008). *Biología reproductiva y evolución del género Erodium L'Her ex Aiton. Consecuencias y comparación con otras Geraniaceae*. Doctoral dissertation, Universidad Rey Juan Carlos, Madrid.

Alcaraz Ariza, F.J. (2015). *Helianthemum abelardi* sp. nov.: una especie malinterpretada de las zonas áridas del sureste de la Península Ibérica (España). *Flora Montiberica*, 60, 139-145.

Alonso, M.Á., Agulló, J.C., Villar, J.L., Juan, A. & Crespo, M.B. (2013). Taxonomic relationships in the *Hypericum ericoides* aggregate (*H. sect. Coridium*, Hypericaceae). *Annales Botanici Fennici*, 50(3), 195-207. <https://doi.org/10.5735/085.050.0312>

Amat, M.E., Vargas, P. & Gómez, J.M. (2013). Effects of human activity on the distribution and abundance of an endangered Mediterranean high-mountain plant (*Erysimum penyalarensis*). *Journal for Nature Conservation*, 21, 262-271. <https://doi.org/10.1016/j.jnc.2013.02.001>

Aparicio, A., Martín-Hernanz, S., Parejo-Farnés, C., Arroyo, J., Lavergne, S., Yeşilyurt, E.B., Zhang, M., Rubio, E. & Albaladejo, R.G. (2017). Phylogenetic reconstruction of the genus *Helianthemum* (Cistaceae) using plastid and nuclear DNA-sequences: Systematic and evolutionary inferences. *Taxon*, 66(4), 868-885. <https://doi.org/10.12705/664.5>

Aymerich, P., Kaplan, Z., Guardiola, M., Petit, A. & Schwarzer, U. (2012). *Potamogeton schweinfurthii* in the Iberian Peninsula. *Anales del Jardín Botánico de Madrid*, 69(2), 187-192. <https://doi.org/10.3989/ajbm.2297>

Aymerich, P., López-Alvarado, J. & Sáez, L. (2014). *Primula subpyrenaica* (Primulaceae) a new species from the Pyrenean range (south-western Europe). *Phytotaxa*, 163(2), 77-90. <http://dx.doi.org/10.11646/phytotaxa.163.2.2>

Aymerich, P. & Sáez, L. (2015). *Alchemilla cadinensis* (Rosaceae), a new species from the Pyrenees (SW Europe). *Willdenowia*, 45(3), 435-442. <https://www.jstor.org/stable/24753254>

Blanca, G., Cueto, M., Fuentes, J., Sáez, L. & Tarifa, R. (2018). *Linaria qartobensis* sp. nov. (Plantaginaceae) from the southern Iberian Peninsula. *Nordic Journal of Botany*, 36(8), e01914. <https://doi.org/10.1111/njb.01914>

Blanca, G., Cueto, M., Fuentes, J. & Tello, S. (2017). *Armeria villosa* subsp. *enritrianoi* subsp. nov. (Plumbaginaceae) en el sureste de España (provincia de Jaén). *Acta Botanica Malacitana*, 42(2), 287-291. <https://doi.org/10.24310/abm.v42i2.2882>

Bou, J & Bou i Tomàs, J. (2018). *Euphorbia duvalii* (Euphorbiaceae) in the Iberian Peninsula. *Anales del Jardín Botánico de Madrid*, 75(2), e073. <https://doi.org/10.3989/ajbm.2495>

Cabezudo, B., Casimiro-Soriguer Solanas, F., & Pérez Latorre, A.V. (2015). Sobre el género *Armeria* en la provincia de Málaga (Andalucía, España). *Acta Botanica Malacitana*, 40, 57-70. <https://doi.org/10.24310/abm.v40i0.2493>

Cano-Maqueda, J. & Talavera Lozano, S. (2011). A taxonomic revision of the *Campanula lusitanica* complex (Campanulaceae) in the Western Mediterranean region. *Anales del Jardín Botánico de Madrid*, 68(1), 15-47. <https://doi.org/10.3989/ajbm.2274>

Carlón, L., Gómez Casares, G., Laínz, M., Moreno Moral, G., Sánchez Pedraja, O. & Schneeweiss, G. M. (2008). Más, a propósito de algunas *Orobanche* L. y *Phelipanche* Pomel

(*Orobanchaceae*) del oeste del Paleártico. *Documentos del Jardín Botánico Atlántico (Gijón)*, 6, 1-128.

Carlón, L., Laínz, M., Moreno Moral, G. & Sánchez Pedraja, O. (2011). A new species (*Orobanche loscosii*), a priority name for *O. Icterica* (*O. Leptantha*) and a new member of the Spanish Flora (*O. Elatior*). *Flora Montiberica*, 48, 89-101.

Casimiro-Soriguer Solanas, F. & Cabezudo, B. (2018). *Aquilegia saxifraga* sp. nov. (Ranunculaceae) en la provincia de Málaga (España). *Acta Botanica Malacitana*, 43, 129-136. <http://dx.doi.org/10.24310/abm.v43i0.3512>

Casimiro-Soriguer Solanas, F. & Cabezudo, B. (2019). Dos nuevas subespecies de *Linaria amethystea* (Vent.) Hoffman. & Link en el sur de la Península Ibérica. *Acta Botanica Malacitana*, 44, 35-42. <https://doi.org/10.24310/abm.v44i0.6673>

Cires, E., Pinto-Cruz, C., Nava, H.S. & Fernández Prieto, J.A. (2020). A new species from the temporary ponds of southwest Portugal: *Helosciadium milfontinum*. *Phytotaxa*, 456(1), 49-62. <https://doi.org/10.11646/phytotaxa.456.1.3>

Coy Gómez, E., Hernández González, A. & Sánchez Gómez, P. (1997). *Diplotaxis tenuisiliqua* Delile, una especie nueva para la Península Ibérica. *Anales del Jardín Botánico de Madrid*, 55(2), 465.

Crespo, M.B., Güemes, J. & Mateo, G. (1992). Datos sobre algunos táxones iberolevantinos de *Biscutella* ser. *Laevigatae* Malinov. (Brassicaceae). *Anales del Jardín Botánico de Madrid* 50(1), 27-34.

Crespo, M.B., Martínez-Azorín, M. & Alonso-Vargas, M.Á. (2019). Morphological and molecular data support recognition of a new rupicolous species of *Pinguicula* (Lentibulariaceae) from the Iberian Peninsula. *Plant Biosystems-An International Journal Dealing with all Aspects of Plant Biology*, 153(1), 77-87. <https://doi.org/10.1080/11263504.2018.1461702>

Crespo, M.B., Martínez-Azorín, M., Alonso-Vargas, M.Á. & Sáez, L. (2020). Two new calcicolous species of *Pinguicula* sect. *Pinguicula* (Lentibulariaceae) growing on rocky habitats of the Iberian Peninsula. *Phytotaxa*, 456(3), 269-284. <https://doi.org/10.11646/phytotaxa.456.3.4>

Crespo, M.B. & Serra, L. (1993). A new Spanish species of the genus *Ononis* L. (Fabaceae). *Botanical Journal of the Linnean Society*, 111, 37-46. <https://doi.org/10.1111/j.1095-8339.1993.tb01888.x>

Cueto, M., Melendo, M., Giménez, E., Fuentes, J., López Carrique, E. & Blanca, G. (2018). First updated checklist of the vascular flora of Andalusia (S of Spain), one of the main biodiversity centres in the Mediterranean Basin. *Phytotaxa*, 339(1), 1-95. <https://doi.org/10.11646/phytotaxa.339.1.1>

Devesa, J.A., Catalán, P., Müller, J., Cebolla, C. & Ortúñez, E. (2013). Checklist de *Festuca* L. (Poaceae) en la Península Ibérica. *Lagascalia*, 33: 183-274.

Escobar García, P., Mascia, F. & Bacchetta, G. (2010). Typification of the name *Lavatera triloba* subsp. *pallescens* (Moris) Nyman and reassessment of *L. minoricensis* Cambess. (*L. triloba* subsp. *minoricensis* comb. nov.). *Anales del Jardín Botánico de Madrid*, 67(2), 79-86. <https://doi.org/10.3989/ajbm.2244>

Fernández-Mazuecos, M., Ferrer-Gallego, P. P., Miguel, M., Glover, B.J. & Sáez, L. (2018). A synopsis of the Iberian clade of *Linaria* subsect. *Versicolores* (Antirrhineae, Plantaginaceae)

based on integrative taxonomy. *Plant Systematics and Evolution*, 304, 871-884.
<https://doi.org/10.1007/s00606-018-1517-0>

Ferrer-Gallego, P.P., Sáez, L., Laguna, E. & Rosselló, J.A. (2018). Typification of the Linnaean name *Ononis crispa* and the related *O. zschackei* (Fabaceae). *Nordic Journal of Botany*, 36(9), e01920. <https://doi.org/10.1111/njb.01920>

Fraga-Arguimbau, P. (2014). Notes i contribucions al coneixement de la flora de Menorca (X). Notes florístiques. *Bulletí de la Societat d'Història Natural de les Balears*, 57, 161-189.

Fraga-Arguimbau, P., Mascaró-Sintes, C., Pallicer-Allès, X. & Carreras-Martí, D. (2018). Notes i contribucions al coneixement de la flora de Menorca (XIII). Notes florístiques. *Bulletí de la Societat d'Història Natural de les Balears*, 61, 183-198.

Fraga-Arguimbau, P. & Rosselló, J.A. (2011). *Euphorbia nurae* P. Fraga & Rosselló (Euphorbiaceae), a new species from Minorca (Balearic Islands). *Candollea*, 66, 181-190. <https://doi.org/10.15553/c2011v661a20>

Greuter, W. & Raus, T. (2012). Med-Checklist Notulae, 31. *Willdenowia*, 42, 287-295. <https://doi.org/10.3372/wi.42.42215>

Guardiola Bufí, M. (2019). *Herniaria alpina* (Caryophyllaceae) in the Iberian Peninsula. *Anales del Jardín Botánico de Madrid*, 76(2), e086. <https://doi.org/10.3989/ajbm.2526>

Güemes, J. & Mota, J.F. (2017). *Gadoria* (Antirrhineae, Plantaginaceae): A new genus, endemic from Sierra de Gádor, Almería, Spain. *Phytotaxa*, 298(3), 201-221. <https://doi.org/10.11646/phytotaxa.298.3.1>

Juan, A., Blanca, G., Cueto, M., Fuentes, J. & Sáez, L. (2018). *Linaria argillicola* (Plantaginaceae), a new species of *L. sect. Supinae* from the southern Iberian Peninsula. *Phytotaxa*, 343(2), 127-138. <https://doi.org/10.11646/phytotaxa.343.2.3>

Juan, A. & Crespo, M.B. (1999). Comportamiento fitosociológico de *Medicago citrina* (Font Quer) Greuter (Leguminosae), endemismo mediterráneo-iberolevantino. *Acta Botanica Malacitana*, 24, 221-229.

Le Driant, F. & Carlón, L. (2020). The Saharo-Arabian *Gymnocarpos sclerocephalus* (Caryophyllaceae) new to Europe in the semideserts of Almería, Spain. *Willdenowia*, 50, 187-194. <https://doi.org/10.3372/wi.50.50204>

Lihová, J., Marhold, K., & Neuffer, B. (2000). Taxonomy of *Cardamine amara* (Cruciferae) in the Iberian Peninsula. *Taxon*, 49(4), 747-763. <https://doi.org/10.2307/1223975>

Lizaur Sukia, X. (1998). *Atriplex glabriuscula* J.R. Edm. es planta ibérica. *Fontqueria*, 51, 1-2.

López, J., & Devesa, J.A. (2019). *Glyceria spicata* subsp. *onubensis* (Gramineae), nuevo taxón para la flora ibérica. *Acta Botanica Malacitana*, 44, 99-101. <https://doi.org/10.24310/abm.v44i0.6803>

López-Alvarado, J., Cobacho, I., Arán, V.J., Rosato, M., Rosselló, J.A. & Sáez, L. (2017). *Limonium ilergabonum* (Plumbaginaceae), a new species from northeastern Iberian Peninsula. *Phytotaxa*, 331(2), 199-212. <https://doi.org/10.11646/phytotaxa.331.2.4>

López-Alvarado, J., Sáez, L., Filigheddu, R., Guardiola, M. & Susanna, A. (2012). *Centaurea tripontina* (Compositae), a new species from the Pre-Pyrenean mountains, Spain. *Plant Biosystems-An International Journal Dealing with all Aspects of Plant Biology*, 146(sup1), 273-275. <https://doi.org/10.1080/11263504.2011.608736>

- López Udías, S., Mateo Sanz, G. & Crespo Villalba, M.B. (2004). Nuevo taxon del género *Galium* L. (sect. *Leptogalium* Lange) para el Sistema Ibérico. *Flora Montiberica*, 27, 47-53.
- Luceño, M., Fernández-Mazuecos, M. & Vargas, P. (2019). A new species of *Linaria* sect. *Supinae* from Sierra de Gredos (Sistema Central mountains, Iberian Peninsula). *Acta Botanica Malacitana*, 44, 43-49. <http://dx.doi.org/10.24310/abm.v44i0.6679>
- Maguilla, E. & Escudero, M. (2016). Cryptic species due to hybridization: a combined approach to describe a new species (*Carex*: Cyperaceae). *PloS one*, 11(12), e0166949. <https://doi.org/10.1371/journal.pone.0166949>
- Marques, I., Fuertes Aguilar, J., Martins-Louçao, M.A., Moharrek, F. & Nieto Feliner, G. (2017). A three-genome five-gene comprehensive phylogeny of the bulbous genus *Narcissus* (Amaryllidaceae) challenges current classifications and reveals multiple hybridization events. *Taxon*, 66(4), 832-854. <https://doi.org/10.12705/664.3>
- Martín-Bravo, S. & Jiménez-Mejías, P. (2009). Molecular data helps traditional taxonomy: Re-evaluation of *Reseda collina* (Resedaceae), and a new record for Europe. *Folia Geobotanica*, 44, 399-421. <https://doi.org/10.1007/s12224-009-9052-y>
- Martínez Fort, J. & Donat Torres, M.P. (2006). *Crassula alata* (Viv.) Berger subsp. *alata*: primera cita para la Península Ibérica. *Flora Montiberica*, 33, 41-44.
- Martínez-Sagarra, G. (2019). *Festuca devesae* (Poaceae), a new species from northern Spain. *Phytotaxa*, 424(2), 97-105. <https://doi.org/10.11646/phytotaxa.424.2.3>
- Montserrat, P. (1997). Una «Rosa inédita» de los Pirineos españoles: *Rosa jacetana* P. Monts., spec. nov. *Bulletin, Société pour l'Échange des Plantes Vasculaires de l'Europe Occidentale et du Bassin Méditerranéen*, 26, 107-113.
- Molino de Miguel, S., de la Fuente, P., Seral, A. & Gabriel y Galán, J.M. (2019). Nuevas localidades de tres helechos para la Península Ibérica. *Botanica Complutensis*, 43, 1-5. <https://doi.org/10.5209/BOCM.62875>
- Ortega-Olivencia, A. & Devesa, J.A. (2018). Updated checklist of Poa in the Iberian Peninsula and Balearic islands. *PhytoKeys*, 103, 27–60.
- Otero Gómez, A. (2019). *Evolution of Boraginaceae: the effect of homoplasy and plesiomorphy in systematics and biogeography of the subfamily Cynoglossoideae*. Doctoral dissertation, Universidad Rey Juan Carlos, Madrid.
- Peñas, J., Benito, B., Lorite, J., Morales, M.C. & Martínez-Ortega, M.M. (2008). *Astragalus peregrinus* subsp. *warionis* and *A. guttatus* (Fabaceae): New for the Iberian Peninsula. *Phyton*, 48(1), 155-167.
- Pujadas-Salvà, A. (2006). Nuevas combinaciones nomenclaturales y aportaciones florísticas en el género *Orobanche* L.(Orobanchaceae). *Lagascalia*, 26(1), 246-248.
- Pujadas-Salvá, A.J. & Gómez García, D. (2000). *Orobanche montserratii* A. Pujadas & D. Gómez (Orobanchaceae), especie nueva del Pirineo oscense. *Anales del Jardín Botánico de Madrid*, 57(2), 267-275. <https://doi.org/10.3989/ajbm.1999.v57.i2.202>
- Pyke, S., Sáez, L., Molero, J. & Garnatje, T. (2016). *Festuca dertosensis* (Poaceae), an overlooked fescue from the NE Iberian Peninsula. *Willdenowia*, 46, 367-377. <https://doi.org/10.3372/wi.46.46305>

- Ríos, S., Crespo, M.B. & Rivera, D. (2001). The west Mediterranean orophilous taxa of *Sideritis* L. (Lamiaceae): a new species of subsection *Hyssopifolia* from south-eastern Spain. *Botanical Journal of the Linnean Society*, 136, 247-254. <https://doi.org/10.1111/j.1095-8339.2001.tb00570.x>
- Romero-Zarco, C. (2019). *Medicago lesinsii* E. Small, (Leguminosae), nueva cita para la flora ibérica. *Acta Botanica Malacitana*, 44, 83-85. <https://doi.org/10.24310/abm.v44i0.5775>
- Romero-Zarco, C., García Castaño, J.L. & Berjano Pérez, R. (2019). Notas nomenclaturales sobre algunas gramíneas mediterráneas. *Acta Botanica Malacitana*, 44, 93-97. <https://doi.org/10.24310/abm.v44i0.5331>
- Rosselló, J.A., Mayol, M., Mus, M. & Torres, N. (1988). Taxonomia i nomenclatura de plantes baleàriques. *Bulletí de la Societat d'Història Natural de les Balears*, 32, 153-154.
- Rudner, M. & Deil, U. (2011). *Juncus maroccanus* Kirschner, a new record to the European flora. *Acta Botanica Malacitana*, 36, 221-223.
- Rufo, L., de la Fuente, V. & Sánchez-Mata, D. (2016). Sarcocornia plant communities of the Iberian Peninsula and the Balearic Islands. *Phytocoenologia*, 46(4), 383-396. <https://doi.org/10.1127/phyto/2016/0113>
- Sáez, L., Guasp, E., Ferrer-Gallego, P.P., López-Alvarado, J. & Rosselló, J.A. (2019). On the taxonomic identity and status of *Silene sericea* var. *balearica* (sect. *Dipterosperma*, Caryophyllaceae). *Mediterranean Botany*, 40(2), 193-202. <https://dx.doi.org/10.5209/mbot.64431>
- Sáez, L., López-Alvarado, J., Fraga, P., Berjano, R., Ortiz, M.A., & Romero-Zarco, C. (2020). Two New Species of *Aira* (Poaceae) from the Iberian Peninsula and the Balearic Islands. *Systematic Botany*, 45(1), 75-84. <https://doi.org/10.1600/036364420X15801369352324>
- Sáez, L. & Rosselló, J.A. (2000). A new species of *Agrostis* (Gramineae) in the *A. alpina* complex. *Botanical Journal of the Linnean Society*, 133, 359-370. <https://doi.org/10.1111/j.1095-8339.2000.tb01551.x>
- Sánchez-Gómez, P., Jiménez, J.F. & Vera, J.B. (2011, March). *Helianthemum motae* (Cistaceae), a new species from southeastern Spain. *Annales Botanici Fennici*, 48, 49-56. <https://doi.org/10.5735/085.048.0107>
- Schönswetter, P., Magauer, M. & Schneeweiss, G.M. (2015). *Androsace halleri* subsp. *nuria* Schönsw. & Schneew. (Primulaceae), a new taxon from the eastern Pyrenees (Spain, France). *Phytotaxa*, 201(3), 227-232. <https://doi.org/10.11646/phytotaxa.201.3.7>
- Senar, R., Ferrer-Gallego, P.P., Laguna, E., Rosselló, J.A., & Sáez, L. (2020). *Micromeria rodriguezii* (Lamiaceae) en la flora peninsular ibérica. *Collectanea Botanica*, 39, e006. <https://doi.org/10.3989/collectbot.2020.v39.006>
- Simón-Porcar, V.I., Pérez-Collazos, E. & Catalán, P. (2015). Phylogeny and systematics of the western Mediterranean *Vella pseudocytisus*-*V. aspera* complex (Brassicaceae). *Turkish Journal of Botany*, 39(3), 472-486. <https://doi.org/10.3906/bot-1406-87>
- Soler, J.X. (1995). Primera cita de *Boerhavia repens* L. (Nyctaginaceae) para la flora ibérica. *Anales del Jardín Botánico de Madrid*, 53(1), 123-125.
- Talavera Lozano, S., Sánchez-Gómez, P., López García, D., Jiménez Martínez, J.F., & Mota Poveda, J.F. (2010). A new species of *Astragalus* L. sect. *Sesamei* DC. (Leguminosae) from the

southeast of Spain: *Astragalus castroviejoi*. *Anales del Jardín Botánico de Madrid*, 67(1), 41-47. <https://doi.org/10.3989/ajbm.2248>

Verloove, F. & Mesterházy, A. (2013). *Cyperus glaber* L. (Cyperaceae), an enigmatic species “new” to Spain. *Webbia*, 68(1), 67-71. <https://doi.org/10.1080/00837792.2013.801110>

Warwick, S.I. & Al-Shehbaz, I.A. (1998). Generic evaluation of Boleum, Euzomodendron, and Vella (Brassicaceae). *Novon*, 8(3), 321-325. <https://doi.org/10.2307/3392030>

Zozomová-Lihová, J., Marhold, K. & Španiel, S. (2014). Taxonomy and evolutionary history of *Alyssum montanum* (Brassicaceae) and related taxa in southwestern Europe and Morocco: Diversification driven by polyploidy, geographic and ecological isolation. *Taxon*, 63(3), 562-591. <https://doi.org/10.12705/633.18>

AFLIBER compilation data sources.

AFLIBER data sources, sorted by decreasing number of occurrences in the database. Numeric references are referred to those shown on Table 1b-References, alongside with their name, citation, and the amount of information they contribute with in the database. This table is also supplied as a CSV file in Dryad (<https://doi.org/10.5061/dryad.gmsbcc2kv>).

NUMERIC REFERENCE	REFERENCE	CITATION	NUMBER OF OCCURRENCES
999	Global Biodiversity Information Facility	https://www.gbif.org/ . Last accessed January, 2020.	1,374,587
3	Base de Datos de Flora Vascular Silvestre de Castilla y León	http://jcyll.plumtic.es/ . Last accessed March, 2020.	236,357
6	Atlas de Flora de los Pirineos	http://www.atlasflorapyrenaea.eu/ . Last accessed March, 2020.	178,823
40	Flora-on	Flora-On: Flora de Portugal Interactiva. Sociedade Portuguesa de Botânica. www.flora-on.pt . Last accessed April, 2020.	151,385
114	IberBal-Flora	IberBal-Flora. <i>Database of the Iberian and Balearic vascular flora</i> . Department of Biology, Autonomous University of Madrid.	147,097
54	BDB-GVA	Banco de Datos de Biodiversidad Conselleria D'agricultura, Desenvolupament Rural, Emergència Climàtica i Transició Ecològica, Comunidad Valenciana. http://www.bdb.gva.es/ . Last accessed October, 2019.	140,547
55	Atlas de la Flora de Aragón	Atlas de la Flora de Aragón, Instituto Pirenaico de Ecología y Gobierno de Aragón (Departamento de Medio Ambiente) http://floragon.ipe.csic.es/ . Last accessed June, 2020.	99,087
12	Aranzadi Science Society	VVAA. ARAN Herbarium database.	57,075

	92	Cartografía de Vegetación de Andalucía	http://descargasrediam.cica.es/ . Last accessed March, 2020.	50,308
	25	Charco <i>et al.</i> , 2014	Charco, J., Becerra, M., Santa-Bárbara, C., Fernández, C., García, F.J., Triano, E., Baena, Vizoso, M.T. & Baena, L. (2014). <i>Árboles y arbustos autóctonos de Andalucía</i> . Centro de Investigaciones Ambientales del Mediterráneo, Ciudad Real.	40,580
	43	Database of Spanish Riparian Plants	VVAA. Compiled from Lara, F., Garilletti, R & Calleja, J.A. 2004. <i>La vegetación de ribera de la mitad norte española</i> . CEDEX, Madrid, and Garilletti, R., Calleja, J.A. & Lara, F. 2012. <i>Vegetación ribereña de los ríos y ramblas de la España meridional (península y archipiélagos)</i> . Ministerio de Agricultura, Alimentación y Medio Ambiente, Madrid.	36,272
A	4	Moreno-Saiz <i>et al.</i> , 2015	Moreno-Saiz, J. C., Pataro, L. & Pajarón Sotomayor, S. (2015). Atlas de los pteridófitos de la Península Ibérica e Islas Baleares. <i>Acta Botanica Malacitana</i> 40, 5-55. https://doi.org/10.24310/abm.v40i0.2540	33,211
	26	Charco <i>et al.</i> , 2008b	Charco, J., Fernández, F., García, R., Mateo, G. & Valdés, A. (2008). <i>Árboles y arbustos autóctonos de Castilla-La Mancha</i> . Centro de Investigaciones Ambientales del Mediterráneo, Ciudad Real.	27,901
	13	Base de Datos de Biodiversidad de la Región de Murcia	VVAA. 2013. <i>Base de Datos de Biodiversidad de la Región de Murcia</i> . Consejería de Medio Ambiente, Gobierno Regional de Murcia.	22,292
	83	Peralta <i>et al.</i> , 2015	Peralta, J. J.M. Olano & J.L. Remón. 2015. <i>Base de datos de los inventarios del Mapa de Vegetación Potencial de Navarra. Sección de Registros Agrarios</i> . Departamento de Desarrollo Rural, Medio Ambiente y Administración Local. Gobierno de Navarra, Pamplona.	16,919
	16	BioAtles	Servei de Protecció d'Espècies, Direcció General d'Espais Naturals i Biodiversitat, Govern de les Illes Balears. http://bioatles.caib.es/ . Last accessed April, 2020.	14,743
	36	Felicísimo <i>et al.</i> , 2011	Felicísimo, Á. M., Muñoz, J., Villalba, C.J. & Mateo, R. G. (2011). <i>Impactos, vulnerabilidad y adaptación al cambio climático de la biodiversidad española: Flora y vegetación: Proyecciones de las áreas de distribución potencial de la flora amenazada y las especies forestales de la España peninsular por efecto del cambio climático</i> . Oficina Española de Cambio Climático, Ministerio de Medio Ambiente y Medio Rural y Marino, Madrid.	11,928
	57	CLM-EXT Database	Trees and shrubs from Extremadura and Castilla-La Mancha compilation.	11,290

45	Inventario de Flora Vascular de La Rioja	VVAA. (1999). <i>Inventario de Flora Vascular de La Rioja 1997-1999</i> , Consejería de Medio Ambiente, Gobierno de La Rioja.	10,550
28	Charco <i>et al.</i> , 2015	Charco, J., Alcaraz, F., Carrillo, F.A. & Rivera, D. (2015). <i>Árboles y arbustos autóctonos de la Región de Murcia</i> . Centro de Investigaciones Ambientales del Mediterráneo, Ciudad Real.	9,486
10	Biocat	Font, X. (2020). Mòdul Flora i Vegetació. Banc de Dades de Biodiversitat de Catalunya. <i>Generalitat de Catalunya i Universitat de Barcelona</i> . http://biodiver.bio.ub.es/biocat/homepage.html . Last accessed June, 2020.	7,773
11	Moreno Saiz <i>et al.</i> , 2019	Moreno Saiz, J.C., Iriondo Alegria, J.M., Martínez García, F., Martínez Rodríguez, J. & Salazar Mendías, C. (2019). <i>Atlas y libro Rojo de la Flora Vascular Amenazada de España. Adenda 2017</i> . Ministerio para la Transición Ecológica-Sociedad Española de Biología de Conservación de Plantas. Madrid.	6,715
27	Charco <i>et al.</i> , 2008a	Charco, J., Devesa, J.A. & Ortega-Olivencia, A. (2008). <i>Árboles y arbustos autóctonos de Extremadura</i> . Centro de Investigaciones Ambientales del Mediterráneo, Ciudad Real.	6,708
82	UPNA-H Herbarium dataset	Peralta de Andrés J. (2020). Herbario de la Universidad Pública de Navarra, Pamplona: UPNA-H. Departamento de Ciencias, Universidad Pública de Navarra.	6,540
39	Alonso Felpete <i>et al.</i> , 2011	Alonso Felpete, J.I., González Robinson, S., Fernández Rodríguez, A., Sanzo Rodríguez, I., Mora Cabello de Alba, A., Bueno Sánchez, A. & Díaz González, T.E. (2011). <i>Catálogo florístico del Parque Nacional Picos de Europa</i> . Jardín Botánico Atlántico de Gijón, Ayuntamiento de Gijón. ISBN 978-84-615-5846-9.	5,304
35	Flora Amenazada y de Interés de Andalucía	http://descargasrediam.cica.es/ . Last accessed March, 2020.	4,989
2	Anthos	Anthos. Sistema de información de las plantas de España. Real Jardín Botánico, CSIC- Fundación Biodiversidad. http://www.anthos.es . Last accessed June, 2020.	4,009
44	Catálogo Florístico de la Provincia de Soria	Segura Zubizarreta, A., Mateo Sanz, G. & Benito Alonso, J.L. (2000). <i>Catálogo Florístico de la Provincia de Soria</i> . Excelentísima Diputación Provincial de Soria, Soria.	850

	60	Sáez, <i>Pers. Com.</i>		548
	17	Biodiversidade.eu	http://biodiversidade.eu/ . Last accessed October, 2019.	310
	56	Sánchez Pedraja (2016)	Sánchez Pedraja, Ó., Moreno Moral, G., Carlón, L., Piwowarczyk, R., Laínz, M. & Schneeweiss, G.M. (2016). <i>Index of Orobanchaceae</i> . http://www.farmalierganes.com/Otrospdf/publica/Orobanchaceae%20Index.htm . Liérganes, Cantabria, Spain. ISSN: 2386-9666 (accessed, October 2019).	253
A	30	Romero Zarco, <i>Pers. Com.</i>		219
	38	Fernández-Mazuecos <i>et al.</i> , 2018	Fernández-Mazuecos, M., Ferrer-Gallego, P.P., Miguel, M., Glover, B. J. & Sáez, L. (2018). A synopsis of the Iberian clade of <i>Linaria</i> subsect. <i>Versicolores</i> (Antirrhineae, Plantaginaceae) based on integrative taxonomy. <i>Plant Systematics and Evolution</i> , 304(7), 871-884. https://doi.org/10.1007/s00606-018-1517-0	148
	78	ORCA	Bolòs, O., Font, X., Pons, X. & Vigo, J. (1985-2010). <i>Atlas corològic de la flora vascular dels Països Catalans. Vols. 1-16</i> . Institut de Estudios Catalanes, Sección de Ciencias Biológicas, ORCA.	129
	5	Atlas de la flora vascular de Burgos	Alejandre, J.A., López, J.M.G. & Sanz, G.M. (2006). <i>Atlas de la flora vascular de Burgos</i> . Junta de Castilla y León, Burgos.	122
	51	Gramina Iberica	Romero-Zarco, C. (ed.). <i>Gramina Iberica. Genera graminum in opus "Flora iberica" intendentis</i> . https://sites.google.com/site/graminaiberica	106
	24	Catalán <i>et al.</i> , 2016	Catalán, P., López-Álvarez, D., Bellosta, C. & Villar, L. (2016). Updated taxonomic descriptions, iconography, and habitat preferences of <i>Brachypodium distachyon</i> , <i>B. stacei</i> , and <i>B. hybridum</i> (Poaceae). <i>Anales Del Jardín Botánico de Madrid</i> 73(1), e028. https://doi.org/10.3989/ajbm.2428	80
	96	Sáez <i>et al.</i> , 2020	Sáez, L., López-Alvarado, J., Fraga, P., Berjano, R., Ortiz, M. & Romero-Zarco, C. (2020). Two New Species of <i>Aira</i> (Poaceae) from the Iberian Peninsula and the Balearic Islands. <i>Systematic Botany</i> 45(1), 75-84. https://doi.org/10.1600/036364420X15801369352324	72

70	Mateo & Crespo, 2000	Mateo, G. & Crespo, M.B. (2000). Three new Spanish species of <i>Biscutella</i> L. (Brassicaceae) and remarks on <i>B. valentina</i> (L.) Heywood. <i>Botanical Journal of the Linnean Society</i> 132(1), 1-17. https://doi.org/10.1111/j.1095-8339.2000.tb01851.x	57
100	García Cardo <i>et al.</i> , 2020b	García Cardo, O., Martínez Labarga, J.M. & Mateo, G. (2020). Aportaciones a la flora de la provincia de Cuenca, VIII. <i>Flora Montiberica</i> 77, 70-76.	55
14	Baonza Díaz <i>et al.</i> , 2010	Baonza Díaz, J., Caparrós Callejo, R., García Medina, N., Martínez García, F. & Gómez Manzaneque, F. (2010). Flora vascular de los Quintos de Mora (Los Yébenes, Toledo). <i>Ecología</i> 23, 39-58.	54
127	Romero Zarco, 1990	Romero Zarco, C. (1990). <i>Avena</i> taxa of the group <i>barbata</i> in the Iberian peninsula and the Balearics. <i>Lagascalia</i> 16(2), 243-268.	41
104	Casimiro-Soriguer Solanas & Cabezudo-Artero, 2019	Casimiro-Soriguer Solanas, F. & Cabezudo-Artero, B. (2019). Dos nuevas subespecies de <i>Linaria amethystea</i> (Vent.) Hoffmanns. & Link (Plantaginaceae) en el sur de la Península Ibérica. <i>Acta Botanica Malacitana</i> 44, 35-42. https://doi.org/10.24310/abm.v44i0.6673	31
155	Fuentes <i>et al.</i> , 2019	Fuentes, J., Gutiérrez, L., Cueto, M. & Segura, J.M. (2019). Novedades corológicas destacables para la flora vascular de Andalucía (sur de España), III. <i>Anales de Biología</i> 41, 39-47. http://dx.doi.org/10.6018/analesbio.41.06	28
153	Fuentes <i>et al.</i> , 2020	Fuentes, J., Sánchez, R., Segura, J.M., Cueto, M., Ramírez, J. & Gutiérrez, L. (2020). Novedades corológicas destacables para la flora vascular de Andalucía, (sur de España) IV. <i>Anales de Biología</i> 42, 63-73. https://doi.org/10.6018/analesbio.42.08	23
32	de la Fuente <i>et al.</i> , 2015	de La Fuente, V., Rufo, L., Rodríguez, N., Sánchez-Mata, D., Franco, A. & Amils, R. (2015). A study of <i>Sarcocornia</i> AJ Scott (Chenopodiaceae) from Western Mediterranean Europe. <i>Plant Biosystems-An International Journal Dealing with all Aspects of Plant Biology</i> , 150(2), 343-356. https://doi.org/10.1080/11263504.2015.1022239	21
73	Molina <i>et al.</i> , 2008	Molina, A., Acedo, C. & Llamas, F. (2008). Taxonomy and new taxa in Eurasian <i>Carex</i> (section <i>Phaestoglochin</i> , Cyperaceae). <i>Systematic Botany</i> 33(2), 237-250. https://doi.org/10.1600/036364408784571563	21
85	Mateo, 2020	Mateo, G. (2020). De flora valentina, XX. <i>Flora Montiberica</i> 77, 98-103.	19

	103	Shiposha <i>et al.</i> , 2016	Shiposha, V., Catalán, P., Olonova, M. & Marques, I. (2016). Genetic structure and diversity of the selfing model grass <i>Brachypodium stacei</i> (Poaceae) in Western Mediterranean: out of the Iberian Peninsula and into the islands. <i>PeerJ</i> 4, e2407. https://doi.org/10.7717/peerj.2407	16
	97	Martínez-Sagarra & Devesa, 2019	Martínez-Sagarra, G. & Devesa, J.A. (2019). Two new taxa of <i>Festuca</i> (Poaceae) from the southeastern Iberian Peninsula. <i>Phytotaxa</i> 395(4), 251-264. https://doi.org/10.11646/phytotaxa.395.4.1	14
	93	Rivero-Guerra, 2011	Rivero-Guerra, A.O. (2011). Morphological variation within and between taxa of the <i>Santolina rosmarinifolia</i> L. (Asteraceae: Anthemideae) aggregate. <i>Systematic Botany</i> 36(1), 171-190. https://doi.org/10.1600/036364411X553261	14
A	154	Lázaro-Bello, 2020	Lázaro-Bello, J. A. (2020). Novedades agrostológicas en la provincia de Valladolid (España). <i>Anales de Biología</i> 42, 31-35. http://dx.doi.org/10.6018/analesbio.42.04	13
	94	Rubido-Bará <i>et al.</i> , 2010	Rubido-Bará, M., Horjales, M. & Villaverde, C. (2010). Dos nuevas subespecies del género <i>Jasione</i> L.(Campanulaceae) en el Noroeste de la Península Ibérica. <i>Nova Acta Científica Compostelana. Bioloxia</i> 19, 21-31.	12
	152	Sánchez-Gómez <i>et al.</i> , 2019	Sánchez-Gómez, P., Cánovas, J.L., Jiménez, J.F., Catalán, A. & López-Donate, J.A. (2019). Novedades florísticas para las provincias de Albacete, Almería y Granada (España). <i>Anales de Biología</i> 41, 73-76. http://dx.doi.org/10.6018/analesbio.41.10	12
	156	García Cardo <i>et al.</i> , 2020a	García Cardo, O., Domingo, L.M. & Lomas, L.M.F. (2020). Sobre la presencia de <i>Filago hispanica</i> (Degen & Hervier ex Pau) Chrtek & Holub en el Sistema Ibérico meridional. <i>Flora Montiberica</i> 77, 37-42.	10
	157	Ruiz de Gopegui & García 2020	Ruiz de Gopegui Valero, J.A. & Cayón, L.G. (2020). Adicciones y apuntes al catálogo de orquídeas de la provincia de Palencia. <i>Flora Montiberica</i> 77, 52-57.	10
	41	Aseginolaza Iparragirre <i>et al.</i> , 1984	Aseginolaza Iparragirre, C., Gómez García, D., Lizaur Sukia, X., Montserrat Martí, G., Morante Serrano, G., Salaverria Monfort, M.R. & Alejandre Sáenz, J.A. (1984). <i>Catálogo Florístico de Álava, Vizcaya y Guipúzcoa</i> . Gobierno Vasco, Vitoria-Gasteiz.	9
	29	Crespo Villalba & Mateo Sanz, 2010a	Crespo Villalba, M.B. & Mateo Sanz, G. (2010). Novedades taxonómicas y nomenclaturales para la flora valenciana, II. <i>Flora Montiberica</i> 45, 89-102.	9

8	Aymerich <i>et al.</i> , 2012	Aymerich, P., Kaplan, Z., Guardiola, M., Petit, A. & Schwarzer, U. (2012). <i>Potamogeton schweinfurthii</i> in the Iberian Peninsula. <i>Anales del Jardín Botánico de Madrid</i> 69(2), 187-192. http://dx.doi.org/10.3989/ajbm.2297	8
9	Barea-Azcón <i>et al.</i> , 2014	Barea-Azcón, J.M., Benito, B.M., Olivares, F.J., Ruiz, H., Martín, J., García, A.L. & López, R. (2014). Distribution and conservation of the relict interaction between the butterfly <i>Agriades zullichi</i> and its larval foodplant (<i>Androsace vitaliana nevadensis</i>). <i>Biodiversity and Conservation</i> 23(4), 927-944. https://doi.org/10.1007/s10531-014-0643-4	8
18	Blanca <i>et al.</i> , 2015	Blanca, G., Cueto, M. & Fuentes, J. (2015). <i>Sisymbrium isatidifolium</i> (Brassicaceae): a new species from southern Spain, and the identity of <i>S. hispanicum</i> Jacq. <i>Phytotaxa</i> 220(1), 043-053. http://dx.doi.org/10.11646/phytotaxa.220.1 .	8
122	Martín-Bravo & Jiménez-Mejías, 2009	Martín-Bravo, S. & Jiménez-Mejías, P. (2009). Molecular data helps traditional taxonomy: Re-evaluation of <i>Reseda collina</i> (Resedaceae), and a new record for Europe. <i>Folia Geobotanica</i> 44(4), 399. https://doi.org/10.1007/s12224-009-9052-y	8
90	Pyke <i>et al.</i> , 2016	Pyke, S., Sáez, L., Molero, J. & Garnatje, T. (2016). <i>Festuca dertosensis</i> (Poaceae), an overlooked fescue from the NE Iberian Peninsula. <i>Willdenowia</i> 46(3), 367-377. https://doi.org/10.3372/wi.46.46305	8
91	Quintanar <i>et al.</i> , 2007	Quintanar, A., Castroviejo, S. & Romero García, A.T. (2007). Apuntes taxonómicos sobre el género <i>Koeleria</i> Pers. (Gramineae) en Andalucía oriental. <i>Acta Botanica Malacitana</i> 32, 289-298.	7
126	Martín Marcos <i>et al.</i> , 2001	Martín Marcos, M.A., Sánchez Rodríguez, J.A. & Elías Rivas, M.J. (2001). Labiadas de la provincia de Salamanca. <i>Studia Botanica</i> 20, 13-41.	6
79	Ortiz & Rodríguez-Oubiña, 2005	Ortiz, S. & Rodríguez-Oubiña, J. (2005). Caracterización taxonómica das poboacións galegas de <i>Alyssum loiseleurii</i> P. Fourn. (Brassicaceae). <i>Nova acta científica compostelana. Bioloxía</i> 14, 51-56.	6
58	Lázaro <i>et al.</i> , 2016	Lázaro, A.B., Domínguez, R.D. & Ureña Plaza, J. (2016). Novedades en <i>Narcissus</i> L.(Amaryllidaceae). <i>Flora Montiberica</i> 63, 103-111.	5
67	Maguilla <i>et al.</i> , 2017	Maguilla, E., Escudero, M., Hipp, A.L. & Luceño, M. (2017). Allopatric speciation despite historical gene flow: Divergence and hybridization in <i>Carex furva</i> and <i>C. lucennoiberica</i> (Cyperaceae) inferred from plastid and nuclear RAD-seq data. <i>Molecular Ecology</i> 26(20), 5646-5662. https://doi.org/10.1111/mec.14253	5

69	Martínez-Sagarra, 2019	Martínez-Sagarra, G. (2019). <i>Festuca devesae</i> (Poaceae), a new species from northern Spain. <i>Phytotaxa</i> 424(2), 97-105. https://doi.org/10.11164/phytotaxa.424.2.3	5
145	Molina <i>et al.</i> , 2017	Molina, C., Montamarta, G. & Eugenio, M. (2017). Adiciones al Catálogo Florístico de Soria, V: nuevas aportaciones a la familia Orobanchaceae. <i>Flora Montiberica</i> 68, 18-26.	5
148	Sanchez-Gomez <i>et al.</i> , 2018	Sánchez-Gómez, P., López García, D., Cánovas, J. L., Jiménez, J. F., Vera, J.B. & Martínez López, P. (2018). Novedades florísticas para las provincias de Murcia, Albacete y Almería (España). <i>Anales de Biología</i> 40, 73-85. http://dx.doi.org/10.6018/analesbio.40.09	5
106	Talavera <i>et al.</i> , 2015a	Talavera, S., Talavera, M. & Sánchez, C. (2015). Los géneros <i>Thrincia</i> Roth y <i>Leontodon</i> L. (Compositae, Cichorieae) en Flora Iberica. <i>Acta Botanica Malacitana</i> 40, 344-364. https://doi.org/10.24310/abm.v40i0.2559	5
150	Acedo, <i>Pers. Com.</i>		5
31	de la Fuente <i>et al.</i> , 2011	de la Fuente, V., Nieto, L.R. & Sánchez-Mata, D. (2011). <i>Sarcocornia hispanica</i> (Chenopodiaceae), a new species from the Iberian Peninsula. <i>Lazaroa</i> 32, 9. https://doi.org/10.5209/rev_LAZA.2011.v32.37817	4
130	Fraga-Arguimbau <i>et al.</i> , 2019	Fraga-Arguimbau, P., Mascaró-Sintes, C., Pallicer-Allès, X. & Carreras-Martí, D. (2019). Notes i contribucions al coneixement de la flora de Menorca (XIII). Notes florístiques. <i>Bulletí de la Societat d'Història Natural de les Balears</i> 61, 183-198.	4
49	Gómez Navarro <i>et al.</i> , 2015	Gómez Navarro, J., Roselló Gimeno, R., Laguna Lumbreras, E., Ferrer Gallego, P.P., Peris Gisbert, J.B., Guillén Bas, A., Valdés Franzi, A. & Sanchis Duato, E. (2015). <i>Trinia castroviejoi</i> (Umbelliferae), especie nueva descrita para el sudeste de la Península Ibérica. <i>Sabuco: revista de estudios albacetenses</i> 11, 11-38.	4
81	Peñas <i>et al.</i> , 2008	Peñas, J., Benito, B., Lorite, J., Morales, M.C. & Martínez-Ortega, M.M. (2008). <i>Astragalus peregrinus</i> subsp. <i>warionis</i> and <i>A. guttatus</i> (Fabaceae): New for the Iberian Peninsula. <i>Phyton</i> 48(1), 155-167.	4
59	Sáez <i>et al.</i> , 2017	Sáez, L., Rosselló, J.A. & Fraga, P. (2017). <i>Llibre vermell de la flora vascular de les Illes Balears. Segona edició.</i> Conselleria de Medi Ambient, Agricultura i Pesca, Palma de Mallorca.	4

42	Senar Lluch <i>et al.</i> , 2020	Senar Lluch, R., Cardero Aguilera, S., Gumbau Vizcarro, P., Martínez Entonado, R. & Mesa Romeu, D. (2020). Aportaciones a la flora castellonense, XI. <i>Flora Montiberica</i> 77, 84-85.	4
46	Vázquez Pardo, 2013	Vázquez Pardo, F.M. (2013). Anotaciones a la diversidad de <i>Narcissus bulbocodium</i> L., (Amaryllidaceae) en Extremadura (España). <i>Folia Botanica Extremadurensis</i> 7, 13-21.	4
87	Jiménez-Mejías, <i>Pers. Com.</i>		4
132	Baonza & Martínez, 2013	Baonza, J. & Martínez, F. (2013). Notas florísticas de la sierra de Guadarrama (Madrid, Segovia). <i>Ecología</i> 25, 137-174.	3
34	Dorda & Fernández Casas, 1989	Dorda, E. & Fernández Casas, J. (1989). Estudios morfológicos en el género <i>Narcissus</i> L. Anatomía de hoja escapo III. <i>Fontqueria</i> 27, 103-162.	3
63	López Vélez, 1991	López Vélez, G. (1991). <i>Sarcocapnos baetica</i> (Boiss. & Reuter) Nyman subsp. <i>ardalii</i> G. López Vélez, subespecie nueva para la Península Ibérica. <i>Acta Botanica Malacitana</i> 16(2), 437-447.	3
64	Luceño <i>et al.</i> , 2019	Luceño, M., Fernández-Mazuecos, M. & Vargas, P. (2019). A new species of <i>Linaria</i> sect. <i>Supinae</i> from Sierra de Gredos (Sistema Central mountains, Iberian Peninsula). <i>Acta Botanica Malacitana</i> 44, 43-49. https://doi.org/10.24310/abm.v44i0.6679	3
137	Mateo Sanz & Arán Redó, 2000	Mateo Sanz, G. & Arán Redó, V.J. (2000). Nuevos datos sobre la flora de la provincia de Cuenca, XII. <i>Flora Montiberica</i> 16, 10-18.	3
131	Mayoral García-Berlanga & Gómez Serrano, 2006	Mayoral García-Berlanga, O. & Gómez Serrano, M.A. (2006). Distribución y ecología de <i>Rubus saxatilis</i> L. (Rosaceae) en el Sistema Ibérico. <i>Flora Montiberica</i> 33, 32-36.	3
105	Talavera Solís <i>et al.</i> , 2013	Talavera Solís, M., Sánchez Casimiro-Soriguer, C. & Talavera Lozano, S. (2013). <i>Crepis</i> sect. <i>Lepidoseris</i> sensu Babcock en la Península Ibérica y Baleares. <i>Acta Botanica Malacitana</i> 38, 231-240.	3

111	Vázquez, 2006	Vázquez, F.M. (2006). Anotaciones a la tribu Stipeae (Gramineae) para Andalucía oriental (España). <i>Acta Botanica Malacitana</i> 31, 73-80.	3
149	Alonso <i>et al.</i> , 2013	Alonso, M.Á., Agulló, J.C., Villar, J.L., Juan, A. & Crespo, M.B. (2013). Taxonomic relationships in the <i>Hypericum ericoides</i> aggregate (<i>H. sect. Coridium</i> , Hypericaceae). <i>Annales Botanici Fennici</i> 50(3), 195-207. https://doi.org/10.5735/085.050.0312	2
7	Aymerich & Sáez, 2015	Aymerich, P. & Sáez, L. (2015). <i>Alchemilla cadinensis</i> (Rosaceae), a new species from the Pyrenees (SW Europe). <i>Willdenowia</i> 45(3), 435-442. https://doi.org/10.3372/wi.45.45310	2
22	Cabezudo <i>et al.</i> , 2015	Cabezudo, B., Solanas, F.C.S. & Pérez-Latorre, A.V. (2015). Sobre el género <i>Armeria</i> en la provincia de Málaga (Andalucía, España). <i>Acta Botanica Malacitana</i> 40, 57-70. https://doi.org/10.24310/abm.v40i0.2493	2
147	Carlón <i>et al.</i> , 2008	Carlón, L., Gómez Casares, G., Laínz, M., Moreno Moral, G., Sánchez Pedraja, Ó. & Schneeweiss, G.M. (2008). Más, a propósito de algunas <i>Phelipanche</i> Pomel, Boulardia F. W. Schultz y Orobanche L. (Orobanchaceae) del oeste del Paleártico. <i>Documentos del Jardín Botánico Atlántico (Gijón)</i> 6, 1-128.	2
33	Castroviejo <i>et al.</i> , 2007, 2019	Castroviejo, S. (coord. gen.). 2007, 2019. <i>Flora iberica</i> vols 15, 16. Real Jardín Botánico, CSIC, Madrid.	2
143	Cires <i>et al.</i> , 2020	Cires, E., Pinto-Cruz, C., Nava, H.S. & Fernández Prieto, J.A. (2020). A new species from the temporary ponds of southwest Portugal: <i>Helosciadium milfontinum</i> . <i>Phytotaxa</i> 456(1), 49-62. https://doi.org/10.11646/phytotaxa.456.1.3	2
134	Cirujano, 1982	Cirujano, S. (1982). Aportaciones a la flora de los saladares castellanos. <i>Anales del Jardín Botánico de Madrid</i> 39(1), 167-173.	2
133	Fernández Prieto <i>et al.</i> , 2020	Fernández Prieto, J.A., Arjona Rodríguez, J.M., Sanna, M. & Cires, E. (2020): Phylogeographical patterns of <i>Campanula</i> gr. <i>arvatica</i> , an endemic group of the Cantabrian mountains (NW Iberian Peninsula), based on plastid and nuclear DNA polymorphisms. <i>Systematics and Biodiversity</i> 18(5), 434-446. https://doi.org/10.1080/14772000.2020.1783715	2
47	Fraga-Arguimbau <i>et al.</i> , 2018	Fraga-Arguimbau, P., Mascaró-Sintes, C., Allés, X.P. & Carreras-Martí, D. (2018). Notas y contribuciones al conocimiento de la flora de Menorca (XIII). Notas florísticas. <i>Boletín de la Sociedad de Historia Natural de Baleares</i> 61, 183-198.	2

151	Galán de Mera <i>et al.</i> , 2014	Galán-de-Mera, A., Martínez-Fresneda, M., Linares-Perea, E. & Vicente-Orellana, J.A. (2014). Taxonomic notes on <i>Gaudinia</i> (Poaceae) based upon morphological characters and the first record of <i>G. maroccana</i> for Portugal. <i>Plant Biosystems-An International Journal Dealing with all Aspects of Plant Biology</i> 148(5), 1056-1063. https://doi.org/10.1080/11263504.2013.809815	2
52	González Granados, 2011	González Granados, J. (2011). Descripción de una nueva subespecie de <i>Sedum aetnense</i> (Crassulaceae) del centro de España. <i>Boletín de la Real Sociedad Española de Historia Natural</i> 105 (1-4), 49-65.	2
72	Molero <i>et al.</i> , 1998	Molero, J., Sáez, L. & Villar Pérez, L. (1998). Interés florístico y geobotánico de la Sierra de Alcubierre (Monegros, Aragón). <i>Acta Botánica Barcinonensis</i> 45, 363-390.	2
88	Polatschek, 2014	Polatschek, A. (2014). Revision der gattung <i>Erysimum</i> (Cruciferae): Nachträge zu den bearbeitungen der Iberischen Halbinsel und Makaronesiens. <i>Annalen des Naturhistorischen Museums in Wien. Serie B für Botanik und Zoologie</i> 116, 87-105.	2
115	Romero-Zarco & Sáez, 2018	Romero-Zarco, C. & Sáez, L. (2018). Notas taxonómicas sobre el género <i>Avellinia</i> Parl. (Poaceae: Poeae: Aveninae). <i>Acta Botanica Malacitana</i> 43, 167-169. http://dx.doi.org/10.24310/abm.v43i0.5341	2
95	Sáez <i>et al.</i> , 2014	Sáez, L., Torres, N., Gil, L., Aymerich, P. & Fraga, P. (2014). Contributions to the knowledge of the flora of the Balearic Islands: Notes on the genus <i>Potamogeton</i> and <i>Stuckenia</i> (Potamogetonaceae). <i>Orsis: organismes i sistemes</i> 28, 81-96. http://doi.org/10.5565/rev/orsis.5	2
107	Talavera <i>et al.</i> , 2015b	Talavera, S., Ortíz, M.Á., Jiménez, F.J., Tremetsberger, K. & Talavera, M. (2015). Los géneros <i>Hypochaeris</i> L. y <i>Achyrophorus</i> Vaill. (Compositae, Cichorieae): nuevos taxones y combinaciones. <i>Acta Botanica Malacitana</i> 40, 332-343. https://doi.org/10.24310/abm.v40i0.2558	2
1	Galán de Mera, <i>Pers. Com.</i>		2
135	Aragón <i>et al.</i> , 1994	Aragón, G., Herrero, A., Izquierdo, J.L. & Martínez, I. (1994). Cartografía Corológica Ibérica. Aportaciones 70-72. <i>Botanica Complutensis</i> 19, 168-183.	1
141	Arán Redó & Mateo Sanz, 1999	Arán Redó, V.J. & Mateo Sanz, G. (1999). Nuevos datos sobre la flora de la provincia de Cuenca, X. <i>Flora Montiberica</i> 12, 33-39.	1

	125	Benítez-Benítez <i>et al.</i> , 2017	Benítez-Benítez, C., Míguez, M., Jiménez-Mejías, P. & Martín-Bravo, S. (2017). Molecular and morphological data resurrect the long neglected <i>Carex laxula</i> (Cyperaceae) and expand its range in the western Mediterranean. <i>Anales del Jardín Botánico de Madrid</i> 74(1), 1-12. http://dx.doi.org/10.3989/ajbm.2438	1
	15	Benito Ayuso, 2016	Benito Ayuso, J. (2016). <i>Eucera (Synhalonia) rufa</i> (Apioidea) polinizador efectivo de <i>Ophrys spectabilis</i> (Orchidaceae). <i>Flora Montiberica</i> 62, 67-73.	1
	19	Blanca <i>et al.</i> , 2016	Blanca, G., Cueto, M., Benavente, A. & Fuentes, J. (2016). <i>Rivasmartinezia cazorlana</i> sp. nov. (Apiaceae) from southern Spain. <i>Nordic Journal of Botany</i> 34(5), 517-521. http://dx.doi.org/10.1111/njb.01191	1
	20	Blanca <i>et al.</i> , 2017	Blanca, G., Cueto, M., Fuentes, J. & Tello, S. (2017). <i>Armeria villosa</i> subsp. <i>enritrianoi</i> subsp. nov.(Plumbaginaceae) en el sureste de España (provincia de Jaén). <i>Acta Botanica Malacitana</i> 42(2), 287-291. https://doi.org/10.24310/abm.v42i2.2882	1
A	21	Bou & Bou i Tomàs, 2018	Bou, J. & Bou i Tomàs, J. (2018). <i>Euphorbia duvalii</i> (Euphorbiaceae) in the Iberian Peninsula. <i>Anales del Jardín Botánico de Madrid</i> 75(2), e073. https://doi.org/10.3989/ajbm.2495	1
	119	Cabezudo <i>et al.</i> , 2013	Cabezudo, B., Latorre, A.V.P. & Solana, F.C.S. (2013). Vegetación gipsícola del sector antequerano (provincia bética, Andalucía). <i>Acta Botanica Malacitana</i> 38, 193-202.	1
	23	Carbajal <i>et al.</i> , 2017	Carbajal, R., Serrano, M., Ortiz, S. & Sáez, L. (2017). Two new combinations in Iberian <i>Santolina</i> (Compositae) based on morphology and molecular evidences. <i>Phytotaxa</i> 291(3), 217-223. https://doi.org/10.11646/phytotaxa.291.3.6	1
	140	Carlón <i>et al.</i> , 2010	Carlón, L., González del Valle, J.M., Laínz, M., Moreno Moral, G., Rodríguez Berdasco J.M. & Sánchez Pedraja, O. (2010). Contribuciones al conocimiento de la flora cantábrica, VIII. <i>Documentos del Jardín Botánico Atlántico (Gijón)</i> 7, 1-95.	1
	121	Carrasco <i>et al.</i> , 2005	Carrasco, M.A., García-Camacho, R. & Martín-Blanco, C.J. (2005). Notas de flora hispánica, V. <i>Botanica Complutensis</i> 29, 27-29.	1
	77	Casimiro-Soriguer Solanas & Cabezudo, 2018	Casimiro-Soriguer Solanas, F. & Cabezudo, B. (2018). <i>Aquilegia saxifraga</i> sp. nov.(Ranunculaceae) en la provincia de Málaga (España). <i>Acta Botanica Malacitana</i> 43, 129-136. https://doi.org/10.24310/abm.v43i0.4870	1

80	Crespo Villalba & Mateo Sanz, 2010b	Crespo Villalba, M. B. & Mateo Sanz, G. (2010). Un nuevo taxon de <i>Armeria</i> (Plumbaginaceae), propio de la Cordillera Ibérica central. <i>Flora Montiberica</i> 46, 49-55.	1
136	Cueto <i>et al.</i> , 1991	Cueto, M., González Rebollar, J.L. & Blanca, G. (1991). Fragmenta chorologica occidentalia. <i>Anales del Jardín Botánico de Madrid</i> 49(1), 121-123.	1
37	Fernández Casas, 1982	Fernández Casas, J. (1982). De Flora Occidentale. <i>Fontqueria</i> 1, 9-12.	1
89	Fernández Prieto & Cires, 2013	Fernández Prieto, J.A. & Cires, E. (2014). Phylogenetic placement of <i>Dethawia</i> , <i>Meum</i> , and <i>Rivasmartinezia</i> (Apioideae, Apiaceae): Evidence from nuclear and plastid DNA sequences. <i>Plant Biosystems-An International Journal Dealing with all Aspects of Plant Biology</i> 148(5), 975-987. https://doi.org/10.1080/11263504.2013.819818	1
116	Fraga-Arguimbau, 2015	Fraga-Arguimbau, P. (2015). Notes i contribucions al coneixement de la flora de Menorca (X). Notes florístiques. <i>Bulletí de la Societat d'Història Natural de les Balears</i> 57, 161-189.	1
139	Fuentes <i>et al.</i> , 2018	Fuentes, J., Cueto, M. & Segura, J.M. (2018). Novedades corológicas destacables para la flora vascular de Andalucía (sur de España) II. <i>Anales de Biología</i> 40: 1-7. https://doi.org/10.6018/analesbio.40.01	1
48	Galán de Mera & Vicente Orellana, 2010	Galán de Mera, A. & Vicente Orellana, J.A. (2010). <i>Taraxacum decastroi</i> and <i>T. lacianense</i> (Asteraceae), two new species from the Iberian Peninsula. <i>Annales Botanici Fennici</i> 47(4), 307-311. https://doi.org/10.5735/085.047.0408	1
50	Gonçalves <i>et al.</i> , 2018	Gonçalves, A.C., Castro, S., Paiva, J., Santos, C. & Silveira, P. (2018). Taxonomic revision of the genus <i>Calendula</i> (Asteraceae) in the Iberian Peninsula and the Balearic Islands. <i>Phytotaxa</i> 352(1), 1-91. https://doi.org/10.11646/phytotaxa.352.1.1	1
118	Guardiola & Romera, 2018	Guardiola, M. & Romera, X. (2018). Reaparició de <i>Reseda hookeri</i> (Resedaceae) a les platges del Maresme, una espècie considerada extingida a Catalunya. <i>Boletín de la Institución Catalana de Historia Natural</i> 82, 69-71.	1
53	Güemes <i>et al.</i> , 2016	Güemes, J., Marchal, F., Carrió, E. & Blasco, M.P. (2016). A new gypsophilous species of <i>Chaenorhinum</i> (Antirrhinaceae) from the south-east of the Iberian Peninsula. <i>Plant Biosystems-An International Journal Dealing with all Aspects of Plant Biology</i> , 150(1), 141-151. https://doi.org/10.1080/11263504.2014.987187	1

65	MA-611914	Herbarium specimen MA-611914	1
123	Jiménez-Mejías <i>et al.</i> , 2014	Jiménez-Mejías, P., Hilpold, A., Frajman, B., Puçsaş, M., Koopman, J., Mesterházy, A., Grulich, V., Lye, K. A. & Martín-Bravo, S. (2014). <i>Carex cespitosa</i> : reappraisal of its distribution in Europe. <i>Willdenowia</i> 44(3), 327-343. https://doi.org/10.3372/wi.44.44303	1
120	Juan <i>et al.</i> , 1995	Juan, A., Serra, L. & Crespo, M.B. (1995). Adiciones a la flora alicantina. <i>Acta Botanica Malacitana</i> 20, 284-290.	1
144	Le Driant & Carlón, 2020	Le Driant, F. & Carlón, L. (2020). The Saharo-Arabian <i>Gymnocarpos sclerocephalus</i> (Caryophyllaceae) new to Europe in the semideserts of Almería, Spain. <i>Willdenowia</i> 50(2), 187-194. https://doi.org/10.3372/wi.50.50204	1
62	López-Alvarado <i>et al.</i> , 2017	López-Alvarado, J., Cobacho, I., Arán, V. J., Rosato, M., Rossellò, J.A. & Sáez, L. (2017). <i>Limonium ilergabonum</i> (Plumbaginaceae), a new species from northeastern Iberian Peninsula. <i>Phytotaxa</i> 331(2), 199-212. https://doi.org/10.11646/phytotaxa.331.2.4	1
61	López-Alvarado <i>et al.</i> , 2012	López-Alvarado, J., Sáez, L., Filigheddu, R., Guardiola, M. & Susanna, A. (2012). <i>Centaurea tripontina</i> (Compositae), a new species from the Pre-Pyrenean mountains, Spain. <i>Plant Biosystems - An International Journal Dealing with all Aspects of Plant Biology</i> 146(sup1), 273-275. https://doi.org/10.1080/11263504.2011.608736	1
138	Lora González & Vivero Pol, 1998	Lora González, A. & Vivero Pol, J.L. (1998). Notas corológicas para la flora de Andalucía. <i>Acta Botanica Malacitana</i> 23, 243-247.	1
66	Maguilla & Escudero, 2016	Maguilla, E. & Escudero, M. (2016). Cryptic species due to hybridization: a combined approach to describe a new species (<i>Carex</i> : Cyperaceae). <i>PloS one</i> , 11(12), e0166949. https://doi.org/10.1371/journal.pone.0166949	1
68	Martínez Fort & Donat Torres, 2006	Martínez Fort, J. & Donat Torres, M.P. (2006). <i>Crassula alata</i> (Viv.) Berger subsp. <i>alata</i> : primera cita para la península Ibérica. <i>Flora Montiberica</i> 33, 41-44.	1
128	Martínez Sánchez & Herranz Sanz, 1992	Martínez-Sánchez, J.J. & Herranz Sanz, J.M. (1992). Notas corológicas albacetenses, I (España). <i>Anales de Biología</i> 18, 69-76.	1

101	Mateo Sanz & Crespo, 2008	Mateo Sanz, G. & Crespo, M.B. (2008). Novedades taxonómicas y nomenclaturales para la flora valenciana <i>Flora Montiberica</i> 40, 60-70.	1
71	Mejías, 2015	Mejías, J.A. (2015). <i>Cicerbita muralis</i> subsp. <i>gaditana</i> Mejías (Asteraceae, Cichorieae), nuevo taxón endémico de la Península Ibérica. <i>Acta Botanica Malacitana</i> 40, 323-331. https://doi.org/10.24310/abm.v40i0.2537	1
74	Molino <i>et al.</i> , 2019	Molino, S., de la Fuente, P., Seral, A. & Gabriel y Galán, J.M. (2019). Nuevas localidades de tres helechos para la Península Ibérica. <i>Botanica Complutensis</i> 40, 1-6. http://dx.doi.org/10.5209/BOCM.62875	1
75	Montserrat Recoder, 1996	Montserrat Recoder, P. (1996). Notas sobre violetas ibéricas. <i>Anales del Jardín Botánico de Madrid</i> 54, 234-240.	1
124	Morales-Molino & Martínez-Sagarra, 2015	Morales-Molino, C. & Martínez-Sagarra, G. (2015). <i>Carex extensa</i> Gooden. (Cyperaceae) en saladares continentales de montaña del Sistema Ibérico (provincia de Guadalajara, España). <i>Acta Botanica Malacitana</i> 40, 209-210. https://doi.org/10.24310/abm.v40i0.2539	1
76	Moreno <i>et al.</i> , 2016	Moreno, J., Terrones, A., Alonso, M.Á., Juan, A. & Crespo, M.B. (2016). <i>Limonium tobarrense</i> (Plumbaginaceae), a new species from the southeastern Iberian Peninsula. <i>Phytotaxa</i> 257(1), 61-70. https://doi.org/10.11646/phytotaxa.257.1.4	1
146	Pavon & Buisson, 2017	Pavon, D & Buisson, E. (2017). Sobre la presencia de <i>Orobanche pubescens</i> d'Urv. (Orobanchaceae) en la Península Ibérica. <i>Flora Montiberica</i> 68, 48-51.	1
84	Pino Pérez & Pino Pérez, 2018	Pino Pérez, R. & Pino Pérez, J.J. (2018). Asientos corológicos FBIGA, 2018. <i>Boletín BIGA</i> 16, 27-67.	1
86	Pino Pérez <i>et al.</i> , 2016	Pino Pérez, R., Silva-Pando, F.J. & Pino Pérez, J.J. (2016). Notas sobre <i>Limonium</i> (Plumbaginaceae) en el Noroeste Ibérico, I: <i>Limonium serpentinicum</i> , Nueva Especie. <i>Novon: A Journal for Botanical Nomenclature</i> 24(4), 380-388. https://doi.org/10.3417/2015026	1
142	Sáez <i>et al.</i> , 2015	Sáez, L., Galán de Mera, A., Pyke, S., Pié, G. & Carnicero, P. (2015). New data on vascular plants from Montseny massif (northeastern Iberian Peninsula). <i>Orsis</i> 29, 205-229.	1

98	Sales & Smith, 1990	Sales, F. & Smith, P. (1990). A new species in the genus <i>Bromus</i> . <i>Edinburgh Journal of Botany</i> 47(3), 361-366. https://doi.org/10.1017/S0960428600003498	1
99	Sanchez-Gómez <i>et al.</i> , 2011	Sánchez-Gómez, P., Jiménez, J.F. & Vera, J.B. (2011). <i>Helianthemum motae</i> (Cistaceae), a new species from southeastern Spain. <i>Annales Botanici Fennici</i> 48(1), 49-56. https://doi.org/10.5735/085.048.0107	1
102	Senar <i>et al.</i> , 2020	Senar, P., Ferrer-Gallego, P.P., Laguna, E., Rosselló, J.A., Sáez, L. (2020). <i>Micromeria rodriguezii</i> (Lamiaceae) en la flora peninsular ibérica. <i>Collectanea Botanica</i> 39, e006. https://doi.org/10.3989/collectbot.2020.v39.006	1
108	Talavera & Talavera, 2015	Talavera, S. & Talavera, M. (2015). El género <i>Andryala</i> L. (Compositae, Cichorieae) en la Península Ibérica y Baleares: una nueva especie e híbridos interespecíficos. <i>Acta Botanica Malacitana</i> 40, 296-310. https://doi.org/10.24310/abm.v40i0.2557	1
109	Ureña, 1994	Ureña, J.F. (1994). Nueva especie del género <i>Narcissus</i> L., sect. <i>Jonquillae</i> DC. de Ronda (Málaga). <i>Botanica Complutensis</i> 10, 83-88.	1
110	Vázquez & Barkworth, 2004	Vázquez, F.M. & Barkworth, M. E. (2004). Resurrection and emendation of <i>Macrochloa</i> (Gramineae: Stipeae). <i>Botanical Journal of the Linnean Society</i> 144(4), 483-495. https://doi.org/10.1111/j.1095-8339.2003.00256.x	1
112	Venhuis & Oostermeijer, 2011	Venhuis, C. & Oostermeijer, J.G.B. (2011). Distinguishing colour variants of <i>Serapias perez-chiscanoi</i> (Orchidaceae) from related taxa on the Iberian Peninsula. <i>Anales del Jardín Botánico de Madrid</i> 68(1), 49-59. https://doi.org/10.3989/ajbm.2269	1
113	Venhuis <i>et al.</i> , 2007	Venhuis, C., Venhuis, P., Oostermeijer, J.G.B. & van Tienderen, P.H. (2007). Morphological systematics of <i>Serapias</i> L. (Orchidaceae) in southwest Europe. <i>Plant Systematics and Evolution</i> 265(3-4), 165-177. https://doi.org/10.1007/s00606-007-0519-0	1
117	Verloove & Mesterházy, 2013	Verloove, F. & Mesterházy, A. (2013). <i>Cyperus glaber</i> L. (Cyperaceae), an enigmatic species “new” to Spain. <i>Webbia</i> 68(1), 67-71. https://doi.org/10.1080/00837792.2013.801110	1
129	Moreno Saiz, <i>Pers. Com</i>		1

GBIF References

- Academy of Natural Sciences. MAL. <https://doi.org/10.15468/xp1dhx>.
- Affouard A, Joly A, Lombardo J, Champ J, Goeau H, Bonnet P (2020). Pl@ntNet observations. Version 1.2. Pl@ntNet. <https://doi.org/10.15468/gtebaa>.
- Affouard A, Joly A, Lombardo J, Champ J, Goeau H, Bonnet P (2020). Pl@ntNet automatically identified occurrences. Version 1.2. Pl@ntNet. <https://doi.org/10.15468/mma2ec>.
- Agulló Villargona J, Aixart Sahun M (2021). Museu de Ciències Naturals de Barcelona: BGJBB. Version 1.9. Botanical Garden of Barcelona. <https://doi.org/10.15468/vvpgcy>.
- Agut A (2017). Banco de Germoplasma Vegetal del Jardín Botánico de Olarizu (Vitoria-Gasteiz)/Olarizuko Lorategi Botanikoaren Landare Germoplasma Bankua (Vitoria-Gasteiz). Jardín Botánico de Olarizu (Vitoria-Gasteiz)/Olarizuko Lorategi Botanikoa (Vitoria-Gasteiz). <https://doi.org/10.15468/kjdwh6>.
- Aizpuru I (2020). Aranzadi Zientzi Elkartea. Aranzadi Science Society. <https://doi.org/10.15468/jxs9rn>.
- Aizpuru Oiarbide I (2020). Base de datos de plantas vasculares del País Vasco: ARAN-EH. Version 1.7. Aranzadi Science Society. <https://doi.org/10.15468/m86bzk>.
- Alonso Felpete J I (2017). Banco de Germoplasma del Jardín Botánico Atlántico. Jardín Botánico Atlántico. <https://doi.org/10.15468/eaqv44>.
- Alonso Felpete J I (2018). Colección activa (BGVJBA) del Banco de germoplasma del Jardín Botánico Atlántico de Gijón. Version 1.1. Jardín Botánico Atlántico. <https://doi.org/10.15468/qs28g1>.
- Alonso Felpete J I (2018). Colección base (BGPVA) del Banco de germoplasma del Jardín Botánico Atlántico de Gijón. Version 1.3. Jardín Botánico Atlántico. <https://doi.org/10.15468/3fcy51>.
- Alonso Felpete J I (2020). Jardín Botánico Atlántico, Gijón: JBAG. Jardín Botánico Atlántico. <https://doi.org/10.15468/da9bjz>.
- Alonso Felpete J I (2020). Jardín Botánico Atlántico, Gijón: JBAG-Laínz. Jardín Botánico Atlántico. <https://doi.org/10.15468/6jpnna>.
- Alonso-Pastor del Coso F (2020). Catálogo Florístico Histórico de Navarra. Gobierno de Navarra. Directorate General of Environment and Water, Dept. of Rural Development and Environment, Government of Navarre. <https://doi.org/10.15468/nf54bd>.
- Anton M (2021). Data collected on citizen science web portal www.ornitho.cat. Version 2.21. ICO-Institut Català d'Ornitologia (Catalan Ornithological Institute). <https://doi.org/10.15470/dijq3a>.
- Appalachian State University (2020). Appalachian State University, I. W. Carpenter, Jr. Herbarium. <https://doi.org/10.15468/ivsxey>.
- Arcocha C (2018). EEIKO Base de datos sobre plantas invasoras. Version 2.7. EEIKO: a control of invasive alien flora species multi-platform application. <https://doi.org/10.15470/kmrohm>.

Arista Palmero M, Salgueiro González F J, Ariza Molina M J (2019). Herbario de la Universidad de Sevilla. Dep. of Plant Biology and Ecology, Univ. Seville. <https://doi.org/10.15468/zhmy7q>.

Arizona State University Biocollections (2021). Arizona State University Vascular Plant Herbarium. <https://doi.org/10.15468/a2o8vy>.

Arsénio P, Cunha A R, Paes P, Vasconcelos T, Figueira R (2021). Herbário João de Carvalho e Vasconcellos, I.S.A./U.L. Instituto Superior de Agronomia / Universidade de Lisboa. <https://doi.org/10.15468/olfpjv>.

Australia's Virtual Herbarium (2021). MEL AVH data. <https://doi.org/10.15468/rhzrxw>.

Bentley A, Morse C (2021). R. L. McGregor Herbarium Vascular Plants Collection. University of Kansas Biodiversity Institute. <https://doi.org/10.15468/htptzr>.

Berkeley Natural History Museums. UCBG TAPIR Provider. <https://doi.org/10.15468/loyiz6>.

Bernacci L C (2021). IAC - Herbário do Instituto Agronômico de Campinas. Version 1.76. Instituto Agronômico (IAC). <https://doi.org/10.15468/w48pii>.

Besson L, Maddi F A (2020). Collection botanique Jaubert et Saul (BOUM). Version 1.4. Muséum d'Histoire Naturelle de Bourges. <https://doi.org/10.15468/lcgħit>.

Besson L, Maddi F A (2020). Collection botanique Le Grand A. (BOUM). Version 13.5. Muséum d'Histoire Naturelle de Bourges. <https://doi.org/10.15468/ihclii>.

Besson L, Maddi F A (2020). Herbier Flora selecta exsiccata publiée par Ch. MAGNIER (BOUM). Version 3.4. Muséum d'Histoire Naturelle de Bourges. <https://doi.org/10.15468/0oij6l>.

Bijmoer R, Scherrenberg M, Creuwels J (2021). Naturalis Biodiversity Center (NL) - Botany. Naturalis Biodiversity Center. <https://doi.org/10.15468/ib5ypt>.

Biologiezentrum Linz Oberösterreich. Biologiezentrum Linz. <https://doi.org/10.15468/ynjblx>.

Bioversity International. EURISCO, The European Genetic Resources Search Catalogue. <https://doi.org/10.15468/a3lnmd>.

Bioversity International. The System-wide Information Network for Genetic Resources (SINGER). <https://doi.org/10.15468/uw1buh>.

Birk S (2017). BioFresh Pond Data. Version 2.2. BioFresh. <https://doi.org/10.13148/bf76>.

Boisset F (2019). Universitat de València, Colecciones de Criptogamas: VAL_Algae. Department of Botany, Faculty of Biological Sciences, University of Valencia. <https://doi.org/10.15468/odw9qj>.

Botanic Garden and Botanical Museum Berlin (2017). Herbarium Berolinense, Berlin (B). <https://doi.org/10.15468/dlwwhz>.

Botanic Garden and Botanical Museum Berlin (2019). BoBO - Botanic Garden and Botanical Museum Berlin Observations. <https://doi.org/10.15468/9ll2gz>.

Botanic Garden and Botanical Museum Berlin. Staatliches Museum für Naturkunde Stuttgart, Herbarium. <https://doi.org/10.15468/s1f4aw>.

Brouillet L, Sinou C (2021). Marie-Victorin Herbarium (MT) - Plantes vasculaires. Version 14.10. Université de Montréal Biodiversity Centre. <https://doi.org/10.5886/rzav8bu2>.

- Brun Christophe, Inventaire National du Patrimoine Naturel (2020). Données ONF faune-flore-fonge. Version 1.1. UMS PatriNat (OFB-CNRS-MNHN), Paris. <https://doi.org/10.15468/ykstli>.
- Bruno D, Solène R (2016). INPN - Données flore des CBN agrégées par la FCBN. Version 1.2. UMS PatriNat (OFB-CNRS-MNHN), Paris. <https://doi.org/10.15468/omae84>.
- Cabezudo Artero B, García Sánchez J (2018). MGC Herbarium of University of Malaga (Spain): MGC-Cormof dataset. University of Malaga. <https://doi.org/10.15468/2gfyxk>.
- California Department of Food and Agriculture (CDA) (2021). CDA - California Department of Food and Agriculture. <https://doi.org/10.15468/phc373>.
- Cameron E, Auckland Museum A M (2021). Auckland Museum Botany Collection. Version 1.59. Auckland War Memorial Museum. <https://doi.org/10.15468/mnjkvv>.
- Camilleri C (2014). INRA-Versailles Arabidopsis natural variants. Institut national de recherche pour l'agriculture, l'alimentation et l'environnement (INRAE). <https://doi.org/10.15468/epzraw>.
- Capers R (2014). CONN. University of Connecticut. <https://doi.org/10.15468/w35jmd>.
- Carapeto A, Porto M (2016). Flora-On: occurrence data of the flora of mainland Portugal. Version 4.1. Sociedade Portuguesa de Botânica. <https://doi.org/10.15468/gi6aum>.
- Caritg R, Martínez de la Riva S (2016). Flora d'Andorra. Centre d'estudis de la neu i de la muntanya d'Andorra (CENMA), Institut d'Estudis Andorrans. <https://doi.org/10.15468/pidttc>.
- Carnegie Museums (2021). Carnegie Museum of Natural History Herbarium. <https://doi.org/10.15468/d51v1f>.
- Carrera D (2018). DIBA-Parc de la Serralada de Marina. Version 2.4. Diputació de Barcelona. <https://doi.org/10.15470/nj86mt>.
- Carrera D (2018). DIBA-Parc del Montnegre i el Corredor. Version 1.3. Diputació de Barcelona. <https://doi.org/10.15470/dd5a-ce17>.
- Carrera D (2018). DIBA-Parc Natural del Montseny. Version 1.2. Diputació de Barcelona. <https://doi.org/10.15470/ug15ii>.
- Casassovici A, Brosens D (2021). Diveboard - Scuba diving citizen science observations. Version 54.36. Diveboard. <https://doi.org/10.15468/tnjrgy>.
- Caze Grégory, Hardy Franck, Inventaire National du Patrimoine Naturel (2020). Inventaire de la flore des Pyrénées-Atlantiques 2015. Version 1.1. UMS PatriNat (OFB-CNRS-MNHN), Paris. <https://doi.org/10.15468/yeqta3>.
- CeDoc of Plant Biodiversity (CeDocBIV), Univ. Barcelona (2009). Sistema de Información de la vegetación Ibérica y Macaronésica. <https://doi.org/10.15468/qyzfdt>.
- Cédric Elleboode, Guillaume Amirault, Marc Lherbette, Inventaire National du Patrimoine Naturel (2019). Données acquises dans le cadre de l'activité du bureau d'étude BIOTOPE (1994 - 2013). Version 1.1. UMS PatriNat (OFB-CNRS-MNHN), Paris. <https://doi.org/10.15468/d3mzxf>.
- Centre for Australian National Biodiversity Research (2021). CANB AVH data. <https://doi.org/10.15468/xycuk7>.
- Cezón K (2018). CeDoc de Biodiversitat Vegetal: BCN-Seeds. CeDoc of Plant Biodiversity (CeDocBIV), Univ. Barcelona. <https://doi.org/10.15468/kqz3uc>.

Charles University - Herbarium collections & Department of Botany. Charles University Prague - Herbarium PRC. <https://doi.org/10.15468/8xrt7r>.

Chen C (2020). Herbarium of Taiwan Forestry Research Institute. Version 1.7. Taiwan Forestry Research Institute. <https://doi.org/10.15468/vo3d2x>.

Citizen Science Observation Dataset B, Tiago P (2020). Biodiversity4all Research-Grade Observations. BioDiversity4All. <https://doi.org/10.15468/njmmp7>.

Colombet Laurent, Inventaire National du Patrimoine Naturel (2021). Observations naturalistes des Amis de BioObs. Version 1.1. UMS PatriNat (OFB-CNRS-MNHN), Paris. <https://doi.org/10.15468/lch7a>.

Colorado State University Herbarium (CSU). Colorado State University Herbarium. <https://doi.org/10.15468/zciacm>.

Conservatoire et Jardin botaniques de la Ville de Genève - G. Geneva Herbarium – De Candolle's Prodromus (G-DC). <https://doi.org/10.15468/s5auru>.

Conservatoire et Jardin botaniques de la Ville de Genève - G. Geneva Herbarium – General Collection (G). <https://doi.org/10.15468/rvjudu1>.

Crespo M B (2018). Herbario ABH (Universidad de Alicante). Depto. de Ciencias Ambientales y Recursos Naturales, Universidad de Alicante. <https://doi.org/10.15468/hs9hmr>.

Crop Wild Relatives Occurrence data consortia (2018). A global database for the distributions of crop wild relatives. Version 1.12. Centro Internacional de Agricultura Tropical - CIAT. <https://doi.org/10.15468/jyrthk>.

CSIC-Real Jardín Botánico (2021). CSIC-Real Jardín Botánico-Colección de Plantas Vasculares (MA). CSIC-Real Jardín Botánico. <https://doi.org/10.15468/mug7kr>.

Danis B, Coupremainne M (2019). Echinoderms occurrence in a strongly contaminated Bay (Portman, Spain), using ROV and Scuba-diving surveys. Version 1.5. Université Libre de Bruxelles (ULB). Sampling event dataset <https://doi.org/10.15468/wew7hf>.

de Montigny Jacky, Archipoff Gisèle, Inventaire National du Patrimoine Naturel (2020). Données Herbiers de l'Université de Strasbourg. Version 1.1. UMS PatriNat (OFB-CNRS-MNHN), Paris. <https://doi.org/10.15468/wmmuuuh>.

de Paz Canuria E (2018). Colección de plantas vasculares del Herbario "Jaime Andrés Rodríguez". LEB. Herbarium LEB Jaime Andrés Rodríguez. Facultad de Ciencias Biológicas y Ambientales. Universidad de León. <https://doi.org/10.15470/j3rugb>.

de Vries H, Lemmens M. Observation.org, Nature data from around the World. Observation.org. <https://doi.org/10.15468/5nilie>.

Devesa Alcaraz J A, Martínez Sagarra G (2021). Herbario COFC de la Universidad de Córdoba: colección general de plantas vasculares. Dept. of Botany, Ecology and Plant physiology, University of Cordoba. <https://doi.org/10.15468/fdzzal>.

Dipartimento di Biologia, Università di Pisa. Herbarium Horti Botanici Pisani (PI), Sistema Museale di Ateneo, Università di Pisa. <https://doi.org/10.15468/soyil7>.

Duprez Cédric, Inventaire National du Patrimoine Naturel (2019). Relevés floristiques du protocole de l'Inventaire National Forestier de l'IGN. Version 1.1. UMS PatriNat (OFB-CNRS-MNHN), Paris. <https://doi.org/10.15468/kqs0zw>.

Dusak François, Prat Daniel, Inventaire National du Patrimoine Naturel (2019). Atlas des Orchidées de France métropolitaine (2010). Version 1.1. UMS PatriNat (OFB-CNRS-MNHN), Paris. <https://doi.org/10.15468/ml5pe4>.

Edgar G J, Stuart-Smith R D (2016). Reef Life Survey: Global reef fish dataset. Version 2.1. Reef Life Survey. Sampling event dataset <https://doi.org/10.15468/qjgwba>.

Epicoco Cyril, Inventaire National du Patrimoine Naturel (2020). Données naturalistes de Cyril EPICOCO. Version 1.1. UMS PatriNat (OFB-CNRS-MNHN), Paris. <https://doi.org/10.15468/th8va6>.

Escuder Olivier, Inventaire National du Patrimoine Naturel (2020). Données naturalistes d'Olivier ESCUDER. Version 1.1. UMS PatriNat (OFB-CNRS-MNHN), Paris. <https://doi.org/10.15468/g5giua>.

Escuder Olivier, Inventaire National du Patrimoine Naturel (2020). Données naturalistes d'Olivier ESCUDER. Version 1.1. UMS PatriNat (OFB-CNRS-MNHN), Paris. <https://doi.org/10.15468/topt6v>.

Estrelles E (2018). Colección del Banco de Germoplasma de la Universitat de València. Version 1.2. Botanical Garden, University of Valencia. <https://doi.org/10.15470/dktqcx>.

European Nucleotide Archive (EMBL-EBI) (2019). Geographically tagged INSDC sequences. <https://doi.org/10.15468/cndomv>.

Fernández Cancio Á, de la Cruz Calleja A C, Montes Pita F, de la Cruz A C (2018). Fotografía digital para el estudio del impacto del cambio climático en la biodiversidad. Version 1.5. INIA-Centro de Investigación Forestal (CIFOR). <https://doi.org/10.15470/0zp7z0>.

Fernández Márquez M A (2018). Universidad de Oviedo. Departamento de Biología de Organismos y Sistemas: FCO. Department of Organisms and Systems Biology. University of Oviedo. <https://doi.org/10.15468/xtvrk2>.

Fernández Márquez M A (2020). Universidad de Oviedo. Departamento de Biología de Organismos y Sistemas: FCO-Briof. Department of Organisms and Systems Biology. University of Oviedo. <https://doi.org/10.15468/ciuxge>.

Fernández Rebollar I (2020). Institut Menorquí d'Estudis. Herbarium Generale Minoricae: HGM. Version 1.3. GBIF-Spain. <https://doi.org/10.15468/gww8x8>.

Fernández-Zamudio R, Rodriguez C (2018). Population dynamics of endangered plants in Doñana Natural Space. Biological Station of Doñana (CSIC). Sampling event dataset <https://doi.org/10.15470/tkydb2>.

Figueira R (2017). IICT Herbário LISC. Version 4.2. Instituto de Investigação Científica Tropical. <https://doi.org/10.15468/iinlqm>.

Font Castell X (2018). FloraCAT: Banco de datos de los cormófitos de Cataluña. Version 2.2. Banc de dades de biodiversitat de Catalunya. <https://doi.org/10.15470/fzwiiu>.

Font García J, Vilar Sais L (2020). Universitat de Girona: HGI-Cormophyta. Dep. Environmental Sciences, Faculty of Sciences, Univ. Girona. <https://doi.org/10.15468/i26r6z>.

Fuster Bejarano F, Rita Larrucea J (2020). Herbario de la Universitat de les Illes Balears. Version 1.2. Universitat de les Illes Balears. <https://doi.org/10.15470/qkd9dg>.

Gacia Passola E (2020). Centre d'Estudis Avançats de Blanes. Macròfits Pirineu. Centre for Advanced Studies of Blanes, CSIC. <https://doi.org/10.15468/pkekwl>.

Gaillard Thomas, Inventaire National du Patrimoine Naturel (2020). Données naturalistes de Thomas Gaillard. Version 1.1. UMS PatriNat (OFB-CNRS-MNHN), Paris. <https://doi.org/10.15468/hszqia>.

Gaisberger H, Endresen D (2019). Bioversity Collecting Mission Database. Version 1.10. Bioversity International. <https://doi.org/10.15468/ulk1iz>.

García San León D (2020). SANT Herbarium vascular plants collection. Herbario SANT, Universidade de Santiago de Compostela. <https://doi.org/10.15468/dgbpla>.

Garmendia Altuna J (2020). Gipuzkoako Foru Aldundiaren Landareen Germoplama Bankua / Banco de Germoplasma Vegetal de la Diputación Foral de Gipuzkoa. Version 5.4. Aranzadi Science Society. <https://doi.org/10.15468/bm2ulz>.

Gil Borrell P (2018). Herbario EMMA. Herbario de la Escuela Técnica Superior de Ingenieros de Montes. UPM. School of Forestry Engineering. Technical University of Madrid. <https://doi.org/10.15468/oasg9o>.

Gilg Olivier, Fons Roger, Inventaire National du Patrimoine Naturel (2019). RNF - Données de la Fédération des Réserves Catalanes. UMS PatriNat (OFB-CNRS-MNHN), Paris. <https://doi.org/10.15468/ufryrd>.

Gimenez Luque E, Yuste Salguero M (2020). Herbario de la Universidad de Almería. Dept. of Vegetal Biology and Ecology, Faculty of Experimental Science, University of Almeria. <https://doi.org/10.15468/d5is9n>.

Gourdain Philippe, Inventaire National du Patrimoine Naturel (2020). Données naturalistes de Philippe GOURDAIN. Version 1.1. UMS PatriNat (OFB-CNRS-MNHN), Paris. <https://doi.org/10.15468/x0oogl>.

Grant S, Niezgoda C (2020). Field Museum of Natural History (Botany) Seed Plant Collection. Version 11.12. Field Museum. <https://doi.org/10.15468/nxnqzf>.

Grant S, von Konrat M (2020). Field Museum of Natural History (Botany) Pteridophyte Collection. Version 2.9. Field Museum. <https://doi.org/10.15468/4nodxs>.

Gratacós Teixidor G, Feo Quer C (2019). Museu Darder de Banyoles. Version 2.1. Darder Natural History Museum of Banyoles. <https://doi.org/10.15470/wxaleu>.

Groom Q (2014). published Chenopodium vulvaria observations. Meise Botanic Garden. <https://doi.org/10.15468/oyorvb>.

Guardia R (2021). CeDoc de Biodiversitat Vegetal: BCN-Cormophyta. CeDoc of Plant Biodiversity (CeDocBIV), Univ. Barcelona. <https://doi.org/10.15468/x5ljvi>.

Guerrero E, Abelló P, Lombarte A, Villanueva R, Ramón M, Sabatés A, Santos R (2020). Biological Reference Collections ICM-CSIC. Version 1.28. Institute of Marine Sciences (ICM-CSIC). <https://doi.org/10.15470/qlqqdx>.

Gustafsson C, Holston K (2021). Herbarium GB, University of Gothenburg. Version 1.5. GBIF-Sweden. <https://doi.org/10.15468/asgd85>.

Hardy Franck, Inventaire National du Patrimoine Naturel (2020). Inventaire de la flore des Pyrénées-Atlantiques 2014. Version 1.1. UMS PatriNat (OFB-CNRS-MNHN), Paris. <https://doi.org/10.15468/ma6jus>.

Haus der Natur Salzburg, Museum für Natur und Technik. Biodiversitätsdatenbank Salzburg. <https://doi.org/10.15468/3pf855>.

Havström M, Källersjö M, Krüger Å (2021). Living Collection of the Gothenburg Botanical Garden. Version 1.5. GBIF-Sweden. <https://doi.org/10.15468/cchsv3>.

Hernandez García F J (2018). Herbario de Plantas Vasculares de la Universidad de Salamanca: SALA. Universidad de Salamanca. Nucleus. <https://doi.org/10.15468/ul946t>.

Hernández García F J, Martínez Ortega M M (2018). Biobanco de ADN Vegetal de la Universidad de Salamanca: SALA-DNA. Version 1.4. Universidad de Salamanca. Nucleus. <https://doi.org/10.15470/wzqivk>.

Herrera M, Torca M (2018). Universidad del País Vasco/EHU, Bilbao: Herbario BIO. Botany laboratory. Plant Biology and Ecology Department. University of the Basque Country. <https://doi.org/10.15468/xtjld4>.

Høiland K (2020). Photographs of higher plants, mosses, fungi, lichens, and algae. Version 1.13. UiO Department of Biosciences. <https://doi.org/10.15468/ft0c5f>.

Høiland K (2020). Photographs of higher plants, mosses, fungi, lichens, and algae. Version 1.13. University of Oslo. <https://doi.org/10.15468/ft0c5f>.

Hugueny Pierre, Inventaire National du Patrimoine Naturel (2020). Données naturalistes de Pierre HUGUENY. Version 1.1. UMS PatriNat (OFB-CNRS-MNHN), Paris. <https://doi.org/10.15468/w3wntg>.

Ibáñez Gastón R (2018). Universidad de Navarra, Herbarium: PAMP-Vascular Plants. University of Navarra – Department of Environmental Biology. <https://doi.org/10.15468/cerrll>.

Ibáñez R, Imas M, Báscones J C, Medrano L M (2018). PAMP-tracheo: Vascular plants in La Demanda and Cameros mountain ranges (La Rioja, Spain). PhD project, L.M. Medrano. Version 1.2. University of Navarra – Department of Environmental Biology. <https://doi.org/10.15470/rzrnsk>.

Ibáñez R, Imas M, Báscones J C, Medrano L M, García-Zamora C (2018). PAMP-tracheo: Vascular plants in Sierra del Mendaña (Navarra, Spain). Undergraduate thesis, C. García-Zamora. Version 1.2. University of Navarra – Department of Environmental Biology. <https://doi.org/10.15470/qdrvp3>.

Ibáñez R, Imas M, Báscones J C, Vicente D (2018). PAMP-tracheo: Vascular plants in Pamplona Basin (Navarra, Spain). Undergraduate thesis, D. Vicente. Version 2.2. University of Navarra – Department of Environmental Biology. <https://doi.org/10.15470/tglboy>.

Ibáñez R, Imas M, Cavero R Y, Akerreta S (2018). PAMP-tracheo: Vascular plants with pharmaceutical use in Navarra (Spain). PhD project, S. Akerreta. Version 2.3. University of Navarra – Department of Environmental Biology. <https://doi.org/10.15470/pycn5t>.

Ibáñez R, Imas M, Cavero R Y, Alberdi L (2018). PAMP-tracheo: Vascular plants recovery after a fire event in a mixed oak forest in Nazar (Navarra, Spain). PhD project, L. Alberdi. Version 1.5. University of Navarra – Department of Environmental Biology. <https://doi.org/10.15470/ltnq1o>.

Ibáñez R, Imas M, López M L, Garde L (2018). PAMP-tracheo: Vascular plants in Ribera Occidental de Navarra (Spain). PhD project, L. Garde. Version 2.6. University of Navarra – Department of Environmental Biology. <https://doi.org/10.15470/jxutdq>.

Ibáñez R, Imas M, López M L, Mayo R M (2018). PAMP-tracheo: Vascular plants in Pamplona and surroundings (Navarra, Spain). Undergraduate thesis, R. Mayo. Version 1.2. University of Navarra – Department of Environmental Biology. <https://doi.org/10.15470/zhbplm>.

Ibáñez R, Imas M, López-Fernández M L, García-Iribarren P (2019). PAMP-tracheo: Vascular plants in the municipality of Falces (Navarra, Spain). Undergraduate thesis, P. García-Iribarren. University of Navarra – Department of Environmental Biology. <https://doi.org/10.15470/sauxy7>.

Infante Sánchez M, Heras Pérez P (2020). VIT Herbarium - Vascular Plants (The Natural History Museum of Alava). Version 1.9. Natural History Museum of Alava (Museo de Ciencias Naturales de Álava). <https://doi.org/10.15468/tqcbr0>.

Inventaire National du Patrimoine Naturel (2019). Flore des Pyrénées. Version 1.1. UMS PatriNat (OFB-CNRS-MNHN), Paris. <https://doi.org/10.15468/vyh3vf>.

Inventaire National du Patrimoine Naturel (2020). Application INPN Espèces: Observations naturalistes, participatives et opportunistes, fondées sur des photographies. Version 1.1. UMS PatriNat (OFB-CNRS-MNHN), Paris. <https://doi.org/10.15468/2wgzwe>.

Inventaire National du Patrimoine Naturel (2020). Espèces Végétales Menacées de France métropolitaine. Version 1.1. UMS PatriNat (OFB-CNRS-MNHN), Paris. <https://doi.org/10.15468/uhrhsz>.

Inventaire National du Patrimoine Naturel (2020). Inventaire de la flore des Pyrénées-Atlantiques 2018. Version 1.1. UMS PatriNat (OFB-CNRS-MNHN), Paris. <https://doi.org/10.15468/rgqfrm>.

Japiot Xavier, Inventaire National du Patrimoine Naturel (2020). Données naturalistes de Xavier JAPIOT. Version 1.1. UMS PatriNat (OFB-CNRS-MNHN), Paris. <https://doi.org/10.15468/hswza8>.

Jennings L (2019). University of British Columbia Herbarium (UBC) - Vascular Plant Collection. Version 16.6. University of British Columbia. <https://doi.org/10.5886/rtt57cc9>.

Johannessen L E, Johnsen A (2021). NHMO DNA Bank Vascular plants collection. Version 1.64. University of Oslo. <https://doi.org/10.15468/nzszik>.

Kathryn Kalmbach Herbarium (Denver Botanic Gardens) (2021). Kathryn Kalmbach Herbarium. <https://doi.org/10.15468/axrelr>.

Kenneth S. Norris Center for Natural History (2021). UCSC - Kenneth S. Norris Center for Natural History, University of California Santa Cruz, Herbarium. <https://doi.org/10.15468/uavt0t>.

Kenneth S. Norris Center for Natural History (2021). UCSC Herbarium - Kenneth S. Norris Center for Natural History, University of California Santa Cruz. <https://doi.org/10.15468/uavt0t>.

Kobayashi M (2016). AIS Wildtype Populations of Arabidopsis. National Institute of Genetics, ROIS. <https://doi.org/10.15468/2hq9oo>.

Lahti T (2017). Hatikka Observation Database. Version 1.1. Finnish Biodiversity Information Facility. <https://doi.org/10.15468/te1t6l>.

Lapenu Pierre, Inventaire National du Patrimoine Naturel (2020). Observations occasionnelles Parc national des Pyrénées. Version 1.1. UMS PatriNat (OFB-CNRS-MNHN), Paris. <https://doi.org/10.15468/g3k5au>.

Lapenu Pierre, Inventaire National du Patrimoine Naturel (2020). Données 'Flore remarquable' du Parc national des Pyrénées. Version 1.1. UMS PatriNat (OFB-CNRS-MNHN), Paris. <https://doi.org/10.15468/tt1nfr>.

Leandro Brotto M (2021). MBM - Herbário do Museu Botânico Municipal. Version 1.76. Museu Botânico Municipal. <https://doi.org/10.15468/g6ppmt>.

Leibniz Institute of Plant Genetics and Crop Plant Research (IPK). Herbarium GAT. <https://doi.org/10.15468/hiiw6b>.

Loeuille Nicolas, Inventaire National du Patrimoine Naturel (2019). Données naturalistes de Nicolas LOEUILLE. Version 1.1. UMS PatriNat (OFB-CNRS-MNHN), Paris. <https://doi.org/10.15468/8frwda>.

López Martínez M, Martín-Consuegra E (2020). Jardín Botánico de Córdoba: Herbarium COA. Botanical Garden of Córdoba. <https://doi.org/10.15468/vjuvw1>.

Loskutov I, Korneev V (2015). Avena wild species collection. (N.I.Vavilov Institute of Plant Genetic Resources (VIR)). N. I. Vavilov Institute of Plant Genetic Resources (VIR). <https://doi.org/10.15468/xtcciv>.

Loskutov I, Korneev V (2015). Avena wild species VIR Herbarium. N.I.Vavilov Institute of Plant Genetic Resources (VIR). N. I. Vavilov Institute of Plant Genetic Resources (VIR). <https://doi.org/10.15468/cjzloe>.

MACB (2020). Herbario de la Facultad de Ciencias Biológicas de la Universidad Complutense de Madrid (MACB). Departamento de Biodiversidad, Ecología y Evolución, Unidad Docente de Botánica. Fac. de Ciencias Biológicas, Univ. Complutense de Madrid. <https://doi.org/10.15470/akznrb>.

Magnaghi E, Fong J (2021). CAS Botany (BOT). Version 157.282. California Academy of Sciences. <https://doi.org/10.15468/7gudyo>.

Marchante H, Morais M C, Marchante E (2021). Sightings Map of Invasive Plants in Portugal. Version 2.10. CFE - Centre for Functional Ecology, Department of Life Sciences, University of Coimbra. <https://doi.org/10.15468/ic8tid>.

Marín Solís F M (2019). Cartografía de vegetación a escala de detalle 1:10.000 de la masa forestal de Andalucía. Environmental Information Network of Andalusia (REDIAM), Regional Ministry for the Environment, Regional Government of Andalusia. <https://doi.org/10.15468/a7isx1>.

Masaryk University, Department of Botany and Zoology. Masaryk University - Herbarium BRNU. <https://doi.org/10.15468/soarvd>.

Masson Gaëtan, Inventaire National du Patrimoine Naturel (2020). Inventaire de la flore des ENS des Pyrénées-Atlantiques 2011-2012. Version 1.1. UMS PatriNat (OFB-CNRS-MNHN), Paris. <https://doi.org/10.15468/lkdtul>.

Mayfield-Meyer T, Zhuang V (2021). UTEP Plants (Arctos). Version 5.50. University of Texas at El Paso Biodiversity Collections. <https://doi.org/10.15468/yhb6ky>.

Mayfield-Meyer T, Zhuang V (2021). UTEP Plants (Arctos). Version 5.51. University of Texas at El Paso Biodiversity Collections. <https://doi.org/10.15468/yhb6ky>.

Medina Domingo L, Aedo Pérez C (2021). CSIC-Real Jardín Botánico-Anthos. Sistema de Información de las Plantas de España. Version 1.18. CSIC-Real Jardín Botánico. <https://doi.org/10.15468/4wnutv>.

Meise Botanic Garden (2021). Meise Botanic Garden Herbarium (BR). Version 1.19. Meise Botanic Garden. <https://doi.org/10.15468/wrthhx>.

Menting F (2020). Centre for Genetic Resources, the Netherlands, PGR passport data. Centre for Genetic Resources, The Netherlands. <https://doi.org/10.15468/mugslo>.

Millet Jérôme, Inventaire National du Patrimoine Naturel (2019). Inventaire Messicoles (FCBN). UMS PatriNat (OFB-CNRS-MNHN), Paris. <https://doi.org/10.15468/nxikmj>.

Millet Jérôme, Inventaire National du Patrimoine Naturel (2020). Livre rouge de la flore menacée de France (tomes I et II) - (FCBN). Version 1.1. UMS PatriNat (OFB-CNRS-MNHN), Paris. <https://doi.org/10.15468/sgoi2p>.

MNHN, Chagnoux S (2021). Herbarium specimens of Société des Sciences Naturelles et Mathématiques de Cherbourg (CHE). Version 70.204. Société des Sciences Naturelles et Mathématiques de Cherbourg. <https://doi.org/10.15468/lmznjw>.

MNHN, Chagnoux S (2021). Herbarium specimens of Université de Montpellier 2, Institut de Botanique (MPU)). Version 71.205. Herbarium of Université de Montpellier 2, Institut de Botanique. <https://doi.org/10.15468/gyvkrn>.

MNHN, Chagnoux S (2021). Herbarium specimens of Université de Montpellier 2, Institut de Botanique (MPU)). Version 71.206. Herbarium of Université de Montpellier 2, Institut de Botanique. <https://doi.org/10.15468/gyvkrn>.

MNHN, Chagnoux S (2021). The vascular plants collection (P) at the Herbarium of the Muséum national d'Histoire Naturelle (MNHN - Paris). Version 69.206. MNHN - Museum national d'Histoire naturelle. <https://doi.org/10.15468/nc6rxy>.

MNHN, Chagnoux S (2021). The vascular plants collection (P) at the Herbarium of the Muséum national d'Histoire Naturelle (MNHN - Paris). Version 69.207. MNHN - Museum national d'Histoire naturelle. <https://doi.org/10.15468/nc6rxy>.

MNHN, Chagnoux S (2021). The vascular plants collection (P) at the Herbarium of the Muséum national d'Histoire Naturelle (MNHN - Paris). Version 69.208. MNHN - Museum national d'Histoire naturelle. <https://doi.org/10.15468/nc6rxy>.

Mota P, Silva D, Santos M, Múrias T (2021). EDP Baixo Sabor: Construction and Reservoir Filling Phases – Vascular Flora and Habitats: species of conservation concern (RELAPE) - [2010-2014]. EDP - Energias de Portugal. Sampling event dataset <https://doi.org/10.15468/d8r8ck>.

Mota P, Silva D, Santos M, Múrias T (2021). EDP Baixo Sabor: Construction and Reservoir Filling Phases – Vascular Flora and Habitats: species from priority conservation habitats - [2010-2014]. EDP - Energias de Portugal. Sampling event dataset <https://doi.org/10.15468/f9755y>.

Museum für Naturkunde Berlin. Anymals+plants - Citizen Science Data. <https://doi.org/10.15468/ee6ps6>.

Museums Victoria (2021). Museums Victoria provider for OZCAM. <https://doi.org/10.15468/lp1ctu>.

Natural History Museum (2021). Natural History Museum (London) Collection Specimens. <https://doi.org/10.5519/0002965>.

Natural History Museum, Vienna - Herbarium W. Natural History Museum, Vienna - Herbarium W. <https://doi.org/10.15468/5sl7sh>.

Natural History Museum, Vienna. Natural History Museum, Vienna - Herbarium W. <https://doi.org/10.15468/5sl7sh>.

Naturgucker (2014). naturgucker.de. <https://doi.org/10.15468/uc1apo>.

Naturhistorisches Museum Mainz. Naturhistorisches Museum Mainz, Botanical Collection. <https://doi.org/10.15468/10wmu8>.

Ninot Sugrañes J M (2020). Universidad de Barcelona (Facultad de Biología): Plantas vasculares del valle de Son del Pi y áreas colindantes. Faculty of Biology, University of Barcelona. <https://doi.org/10.15468/10zm7v>.

Nordic Genetic Resource Center (NORDGEN). Nordic Genetic Resources. <https://doi.org/10.15468/3nyx9k>.

NTNU University Museum (2021). Vascular plant herbarium TRH, NTNU University Museum. Version 30.1505. <https://doi.org/10.15468/zrlqok>.

NTNU University Museum (2021). Vascular plant herbarium TRH, NTNU University Museum. Version 30.1512. <https://doi.org/10.15468/zrlqok>.

Nualart N (2018). Dataset of herbarium specimens of threatened vascular plants in Catalonia. Version 2.4. Herbarium specimens of threatened vascular plants in Catalonia. <https://doi.org/10.15470/sqanal>.

Nualart N, Muñiz D (2020). Institut Botanic de Barcelona (IBB-CSIC-ICUB), BC-Plantae. Botanical Institute of Barcelona (IBB-CSIC-ICUB). <https://doi.org/10.15468/pff0t6>.

Nyffeler R, Guggisberg A (2021). United Herbaria of the University and ETH Zurich. Version 1.106. United Herbaria of the University and ETH Zurich (Z+ZT). <https://doi.org/10.15468/tjvqci>.

Nyffeler R, Guggisberg A (2021). United Herbaria of the University and ETH Zurich. Version 1.107. United Herbaria of the University and ETH Zurich (Z+ZT). <https://doi.org/10.15468/tjvqci>.

Nyffeler R, Guggisberg A (2021). United Herbaria of the University and ETH Zurich. Version 1.108. United Herbaria of the University and ETH Zurich (Z+ZT). <https://doi.org/10.15468/tjvqci>.

Nyffeler R, Guggisberg A (2021). United Herbaria of the University and ETH Zurich. Version 1.114. United Herbaria of the University and ETH Zurich (Z+ZT). <https://doi.org/10.15468/tjvqci>.

Nyffeler R, Guggisberg A (2021). United Herbaria of the University and ETH Zurich. Version 1.115. United Herbaria of the University and ETH Zurich (Z+ZT). <https://doi.org/10.15468/tjvqci>.

Nyffeler R, Guggisberg A (2021). United Herbaria of the University and ETH Zurich. Version 1.116. United Herbaria of the University and ETH Zurich (Z+ZT). <https://doi.org/10.15468/tjvqci>.

Orrell T, Informatics Office (2021). NMNH Extant Specimen Records. Version 1.42. National Museum of Natural History, Smithsonian Institution. <https://doi.org/10.15468/hnhrg3>.

Orrell T, Informatics Office (2021). NMNH Extant Specimen Records. Version 1.43. National Museum of Natural History, Smithsonian Institution. <https://doi.org/10.15468/hnhrg3>.

Peralta de Andrés J (2018). Herbario de la Universidad Pública de Navarra, Pamplona: UPNA-H. Public University of Navarre Herbarium Service, Dept. of Sciences, Public University of Navarre. <https://doi.org/10.15468/vuqjrw>.

Pérez Luque A J (2018). Sinfonevada: Dataset of Floristic diversity in Sierra Nevada forest (SE Spain). Version 1.5. Sierra Nevada Global Change Observatory. Andalusian Environmental Center, University of Granada, Regional Government of Andalusia. <https://doi.org/10.15468/4gpr7e>.

Pérez Luque A J, Benito de Pando B M, Ramón P, Bonet García F J (2020). Dataset of Global Change, altitudinal range shift and colonization of degraded habitats in mediterranean mountains (MIGRAME). Version 1.9. Sierra Nevada Global Change Observatory. Andalusian Environmental Center, University of Granada, Regional Government of Andalusia. <https://doi.org/10.15470/orboj4>.

Perkins K D (2021). University of Florida Herbarium (FLAS). Version 11.1329. Florida Museum of Natural History. <https://doi.org/10.15468/v5wjn7>.

Perkins K D (2021). University of Florida Herbarium (FLAS). Version 11.1340. Florida Museum of Natural History. <https://doi.org/10.15468/v5wjn7>.

Pinheiro P, Aguiar F, Garcia C, Sérgio C, Múrias T (2020). EDP Foz-Tua: Macrophytes – Ecological Monitoring Program [2011-2015]. EDP - Energias de Portugal. Sampling event dataset <https://doi.org/10.15468/yyeq9d>.

Pino Pérez R (2018). Herbario de Plantas Vasculares de la Asociación BIGA para el estudio del patrimonio natural de Galicia: FBIGA. Version 1.4. Colecciones naturales de la Asociación BIGA para el estudio del patrimonio natural de Galicia: FBIGA y ABIGA. <https://doi.org/10.15468/3a8kec>.

Pizarro Domínguez J M (2020). Herbario de la Facultad de Farmacia de la Universidad Complutense de Madrid (MAF). Departamento de Farmacología, Farmacognosia & Botánica, Fac. de Farmacia, Univ. Complutense de Madrid. <https://doi.org/10.15468/ei2ijx>.

Porras Castillo I (2018). Banco de Germoplasma del Real Jardín Botánico Juan Carlos I. Version 1.2. Real Jardín Botánico Juan Carlos I. <https://doi.org/10.15468/sa5nby>.

Prosperi J (2014). The French Collection of *Medicago truncatula*. Institut national de recherche pour l'agriculture, l'alimentation et l'environnement (INRAE). <https://doi.org/10.15468/pqkvfyk>.

Pyle R (2016). Bernice P. Bishop Museum. Version 8.1. Bernice Pauahi Bishop Museum. <https://doi.org/10.15468/s6ctus>.

Questagame (2021). Earth Guardians Weekly Feed. <https://doi.org/10.15468/slqqt8>.

Ramirez J, Tulig M, Watson K, Thiers B (2021). The New York Botanical Garden Herbarium (NY). Version 1.32. The New York Botanical Garden. <https://doi.org/10.15468/6e8nje>.

Ramirez J, Tulig M, Watson K, Thiers B (2021). The New York Botanical Garden Herbarium (NY). Version 1.33. The New York Botanical Garden. <https://doi.org/10.15468/6e8nje>.

Ranz J (2017). Banco de Datos de la Biodiversidad de la Comunitat Valenciana. Biodiversity data bank of Generalitat Valenciana. <https://doi.org/10.15468/b4yqdy>.

Raz L, Agudelo H (2019). Herbario Nacional Colombiano (COL). Version 13.12. Universidad Nacional de Colombia. <https://doi.org/10.15472/ea8sek>.

REDIAM. Red de Información Ambiental de Andalucía (2019). Localización de táxones botánicos de interés para la identificación de hábitats de la REDIAM (Secretaría General de Medio Ambiente, Agua y Cambio Climático). Version 1.3. Environmental Information Network of Andalusia (REDIAM), Regional Ministry for the Environment, Regional Government of Andalusia. <https://doi.org/10.15470/zatyqi>.

Riera Vicent J (2020). Colección de plantas vasculares del herbario de la Universitat de València (VAL). Botanical Garden, University of Valencia. <https://doi.org/10.15468/xmki52>.

Rita Cabral Sales de Melo M (2021). PEUFR - Herbário Professor Vasconcelos Sobrinho. Version 1.73. Universidade Federal Rural de Pernambuco. <https://doi.org/10.15468/o6lt9x>.

Rob Cubey (2018). Royal Botanic Garden Edinburgh Herbarium (E). Royal Botanic Garden Edinburgh. <https://doi.org/10.15468/ypoair>.

Rob Cubey (2018). Royal Botanic Garden Edinburgh Living Plant Collections (E). Royal Botanic Garden Edinburgh. <https://doi.org/10.15468/bkzv11>.

Rodríguez Riaño T (2021). Herbarium of Vascular Plants Collection of the University of Extremadura (Spain). Dep. of Plant Biology, Ecology and Earth Sciences, Botany Area, Univ. Extremadura. <https://doi.org/10.15468/l8vpay>.

Rostanski A (2017). KTU Pteridophyta. University of Silesia, Laboratory of Botanical Documentation - Herbarium KTU. <https://doi.org/10.15468/ynpqsq>.

Rouveyrol Paul, Inventaire National du Patrimoine Naturel (2019). Données d'occurrence Espèces issues de la base Natura 2000 : espèces d'intérêt communautaire et autres espèces remarquables renseignées dans les FSD. Version 1.1. UMS PatriNat (OFB-CNRS-MNHN), Paris. <https://doi.org/10.15468/g2ptuw>.

Royal Botanic Gardens, Kew (2021). Royal Botanic Gardens, Kew - Herbarium Specimens. <https://doi.org/10.15468/ly60bx>.

Rubio López M Á (2020). Inventario de Flora y Vegetación del Municipio de Enguídanos (Cuenca), 2010. Asociación de Desarrollo Integral de la Manchuela Conquense (ADIMAN). Association for the Integral Development of the Manchuela Conquense area. <https://doi.org/10.15468/iotlat>.

Sales F, Santos J (2020). Herbarium of University of Coimbra (COI). Herbarium of University of Coimbra (COI). <https://doi.org/10.15468/7x9xtx>.

Sampoux J, Fourtier S (2012). The French National Collection of forage and turf species (grasses and legumes). BRC Forage and turf, INRA Lusignan. <https://doi.org/10.15468/zvm9gz>.

Sanabria A (2011). Herbarium areqvipense (HUSA). Universidad Nacional de San Agustín (Herbarium Areqvipense HUSA). <https://doi.org/10.15468/u6ypjm>.

Sánchez Durán S, Vega Elices E, García Sánchez A, Sánchez Sánchez J, Sánchez Reyes E, Miñambres Llamas C (2018). Banco de Germoplasma de la Universidad de Salamanca. Version 2.5. Banco de Germoplasma de la Universidad de Salamanca. <https://doi.org/10.15468/fkwhzo>.

Sánchez Gómez P (2020). Herbario de Universidad de Murcia: MUB. Department of Plant Biology. Faculty of Biological Sciences. Univ. Murcia. <https://doi.org/10.15468/qnv1aq>.

Santiago González A (2016). Banco de Germoplama del Jardín Botánico de Castilla-La Mancha. Jardín Botánico de Castilla-La Mancha. <https://doi.org/10.15468/xbgqyqj>.

Saraiva T, Canha P, Crespí A, Canteiro C, Almeida P, Carapeto A, Múrias T (2021). EDP Foz-Tua: Vascular Flora and Habitats – Ecological and Complementary Measures Monitoring Programs – Construction Phase II [2015-2016] and Flooding Phase [2016-2017]. EDP - Energias de Portugal. Sampling event dataset <https://doi.org/10.15468/gn73fg>.

Saraiva T, Canha P, Crespí A, Canteiro C, Almeida P, Carapeto A, Múrias T (2021). EDP Foz-Tua: Vascular Flora – Ecological and Complementary Measures Monitoring Programs – Construction Phase I [2012-2014]. Version 1.1. EDP - Energias de Portugal. Sampling event dataset <https://doi.org/10.15468/26agh>.

Saraiva T, Canha P, Crespí A, Canteiro C, Almeida P, Múrias T (2020). EDP Foz-Tua: Vascular Flora – Complementary Studies and Baseline Year of the Ecological Monitoring Program [2011]. EDP - Energias de Portugal. Sampling event dataset <https://doi.org/10.15468/pwqhpj>.

Schiavino D (2018). Herbario Gaspar Xuarez (BAA) - Facultad de Agronomía - Universidad de Buenos Aires. Herbario Gaspar Xuarez (BAA). <https://doi.org/10.15468/gjluak>.

Schuck A, Kraus D, Bebi P, Blaschke M, Bütler R, Flade M, Heintz W, Krumm F, Lachat T, Larrieu L, Lehnerova L, Levin M, Mergner U, Pach M, Paillet Y, Pyttel P, Rydkvist T, Santopuoli G, Sever K, Sturm K, Vandekerckhove K, Winter S, Witz M, Winnock M, Marcandella M, Roth I, Jakob A, Dickele R, Hofmann G, Ruis-Eckhardt D, Boschen S, Schölmrich U, Leder B, Guericke M, Merkel H, Löffler D, Schusser S, Runkel M, Steinmetz A, Marx K H, Mongelluzzi A, Wilshusen F, Boutteaux J, Duchamp L, Dericbourg N, Rouyer E, Csikos V, Sweeny K, Steichen D, Leytem M, Konczal S, Sterenczak K, Kazimirovic M, Šeben V, Baiges Zapater T, van der Maaten-Theunissen M, Pommer A, Nord-Larsen T, Fuhr M, Delebeque L, Navarro L M, Lasala D, Waez-Mousavi S, Sefidi K, Abellanas B, Zudin S (2021). Spatially explicit database of tree related microhabitats (TreMs). Version 1.17. Institut national de recherche pour l'agriculture, l'alimentation et l'environnement (INRAE). <https://doi.org/10.15468/ocof3v>.

Senckenberg. Herbarium Senckenbergianum (FR). <https://doi.org/10.15468/ucmdjy>.

Seregin A (2021). Moscow University Herbarium (MW). Version 1.177. Lomonosov Moscow State University. <https://doi.org/10.15468/cphbcc>.

Seregin A (2021). Moscow University Herbarium (MW). Version 1.178. Lomonosov Moscow State University. <https://doi.org/10.15468/cphbcc>.

Seregin A (2021). Moscow University Herbarium (MW). Version 1.179. Lomonosov Moscow State University. <https://doi.org/10.15468/cphbcc>.

Slieker F J A, van der Es H, Andeweg R, Langeveld B W (2021). Natural History Museum Rotterdam - Specimens. Version 1.24. Natural History Museum Rotterdam. <https://doi.org/10.15468/kwqaay>.

Slieker F J A, van der Es H, Andeweg R, Langeveld B W (2021). Natural History Museum Rotterdam - Specimens. Version 1.25. Natural History Museum Rotterdam. <https://doi.org/10.15468/kwqaay>.

Solomon J, Stimmel H (2021). Tropicos Specimen Data. Missouri Botanical Garden. <https://doi.org/10.15468/hja69f>.

Soudzilovskaia N A, Vaessen S, Barcelo M, He J, Rahimlou S, Abarenkov K, Brundrett M C, Gomes S, Merckx V, Martinez-Suz L, Tedersoo L. Taxon occurrence data for the FungalRoot database. PlutoF. <https://doi.org/10.15468/a7ujmj>.

Staatliche Naturwissenschaftliche Sammlungen Bayerns. Fungus Collections at Staatliches Museum für Naturkunde Karlsruhe (Herbarium KR). <https://doi.org/10.15468/0bhhip>.

Staatliche Naturwissenschaftliche Sammlungen Bayerns. The Collection of Lichenicolous Fungi at the Botanische Staatssammlung München. <https://doi.org/10.15468/ni5yho>.

Staatliche Naturwissenschaftliche Sammlungen Bayerns. The Erysiphales Collection at the Botanische Staatssammlung München. <https://doi.org/10.15468/sookye>.

Staatliche Naturwissenschaftliche Sammlungen Bayerns. The Fungal Collection at the Senckenberg Museum für Naturkunde Görlitz. <https://doi.org/10.15468/fhmwzv>.

Staatliche Naturwissenschaftliche Sammlungen Bayerns. The Vascular Plant Collection at the Herbarium Tubingense. <https://doi.org/10.15468/2wbxxh>.

Steiermärkisches Landesmuseum Joanneum - Herbarium GJO. Steiermärkisches Landesmuseum Joanneum - Herbarium GJO. <https://doi.org/10.15468/dmdck6>.

Tasmanian Museum and Art Gallery (2021). HO AVH data. <https://doi.org/10.15468/8wmoeb>.

Tela Botanica. Actualisation de la cartographie des Ptéridophytes de France et d'Europe occidentale. <https://doi.org/10.15468/xogtn8>.

Tela Botanica. Carnet en Ligne. <https://doi.org/10.15468/rydcn2>.

Tela Botanica. Cartographie des Leguminosae (Fabaceae) en France. <https://doi.org/10.15468/cjlhuy>.

Telenius A, Röhling Å (2017). Oskarshamn herbarium (OHN). GBIF-Sweden. <https://doi.org/10.15468/u5wjib>.

Telenius A, Shah M (2018). Botany (UPS). GBIF-Sweden. <https://doi.org/10.15468/ufmslw>.

Telenius A, Shah M (2020). Phanerogamic Botanical Collections (S). GBIF-Sweden. <https://doi.org/10.15468/yo3mmu>.

The International Barcode of Life Consortium (2016). International Barcode of Life project (iBOL). <https://doi.org/10.15468/inygc6>.

The Royal Botanic Gardens & Domain Trust (2021). NSW AVH data. <https://doi.org/10.15468/jf3yaе>.

The University of Vermont Pringle Herbarium (2020). University of Vermont, Pringle Herbarium. <https://doi.org/10.15468/zsgioig>.

Thomas Millon, Inventaire National du Patrimoine Naturel (2018). Données d'observation flore du réseau des CBN issues du système d'information de la FCBN. UMS PatriNat (OFB-CNRS-MNHN), Paris. <https://doi.org/10.15468/6klyyk>.

Tiroler Landesmuseum Ferdinandeum. Tiroler Landesmuseum Ferdinandeum. <https://doi.org/10.15468/eepzxm>.

Turpel A, Walisch T (2021). Collections and observation data National Museum of Natural History Luxembourg. Musée national d'histoire naturelle Luxembourg. <https://doi.org/10.15468/s2iu7d>.

- Ueda K (2021). iNaturalist Research-grade Observations. iNaturalist.org. <https://doi.org/10.15468/ab3s5x>.
- Uehbe de Oliveira M I (2021). ASE - Herbário da Universidade Federal de Sergipe. Version 1.82. Universidade Federal de Sergipe. <https://doi.org/10.15468/urlzfj>.
- UiO Natural History Museum (2021). Vascular Plant Herbarium, Oslo (O) UiO. Version 1.1473. <https://doi.org/10.15468/wtlymk>.
- UM Herbarium Data Group, IPT Admin L (2021). University of Michigan Herbarium. Version 1.24. University of Michigan Herbarium. <https://doi.org/10.15468/nl8bvi>.
- UMS PatriNat (OFB-CNRS-MNHN), Paris (2021). Données d'occurrences Espèces issues de l'inventaire des ZNIEFF. Version 1.1. <https://doi.org/10.15468/ikshke>.
- Universität Salzburg. Universität Salzburg. <https://doi.org/10.15468/ufqq8d>.
- University of Graz, Institute of Plant Sciences. University of Graz, Institute of Plant Sciences - Herbarium GZU. <https://doi.org/10.15468/axtkuz>.
- University of North Carolina at Chapel Hill Herbarium (NCU) (2020). University of North Carolina at Chapel Hill Herbarium. <https://doi.org/10.15468/63vxjd>.
- University of Vienna - Herbarium WU. University of Vienna, Institute for Botany - Herbarium WU. <https://doi.org/10.15468/tnj8wm>.
- Uribe F (2018). BioBlitz Barcelona 2010-14. Version 1.8. Museu de Ciències Naturals de Barcelona. <https://doi.org/10.15470/ssy7h3>.
- Uribe F, Agulló Villaronga J (2020). Museu de Ciències Naturals de Barcelona: MCNB-Malac. Museu de Ciències Naturals de Barcelona. <https://doi.org/10.15468/pnkuwh>.
- US National Plant Germplasm System. United States National Plant Germplasm System Collection. <https://doi.org/10.15468/ce7fox>.
- Utah State University (2021). Intermountain Herbarium (Vascular plants & algae). <https://doi.org/10.15468/t43wjj>.
- Vallejo Bombín R (2020). Tercer Inventario Forestal Nacional. Ministerio de Agricultura, Alimentación y Medio Ambiente (España/Spain). Version 1.4. Ministry of Agriculture, Food and Environment. <https://doi.org/10.15468/tkezjv>.
- Vázquez Pardo F M, Márquez García F (2020). Herbario HSS Finca La Orden-Valdesequera (CICYTEX). Gobierno de Extremadura. La Orden Estate, Dep. Forest Production and Grasslands, Agriculture and Environment Office, Regional Government of Extremadura. <https://doi.org/10.15468/siye1z>.
- Vicens M (2018). Banc de Germoplasma Vegetal de les Illes Balears-JBS. Version 7.3. Sóller Botanical Garden Foundation. <https://doi.org/10.15468/wjj1in>.
- Vicens M (2020). Hortus Botanicus Sollerensis Herbarium (FBonafè). Sóller Botanical Garden Foundation. <https://doi.org/10.15468/ghnxdi>.
- Villares J M (2018). Inventario Español de Especies Terrestres (MAGRAMA). Version 1.4. Ministry of Agriculture, Food and Environment. <https://doi.org/10.15468/f0qd41>.
- Villares Muyo J M, Ruiz Franco B (2020). Ministerio de Medio Ambiente, y Medio Rural y Marino. Dirección General de Medio Natural y Política Forestal. Inventario Nacional de

Biodiversidad 2007, Flora Vascular Amenazada. Ministry of Agriculture, Food and Environment. <https://doi.org/10.15468/htpk9n>.

Vizoso Paz M T (2020). Catalogue of type specimens of vascular plants deposited in the Herbarium of the University of Granada (Spain). Version 1.4. Herbario de la Universidad de Granada. <https://doi.org/10.15470/k97bjm>.

Wittzell H, Shah M (2021). Lund Botanical Museum (LD). Lund Botanical Museum (LD). <https://doi.org/10.15468/c4w4co>.

Zamora Rodríguez R J, Pérez-Luque A J (2020). Dataset of Phenology of flora of mediterranean high-mountains meadows (Sierra Nevada). Version 1.5. Sierra Nevada Global Change Observatory. Andalusian Environmental Center, University of Granada, Regional Government of Andalusia. <https://doi.org/10.15468/qhzub>.

A

A generalized framework to expand incomplete phylogenies using non-molecular phylogenetic information

Ignacio Ramos-Gutiérrez^{1,2}  | Herlander Lima³ | Bruno Vilela⁴ |
 Rafael Molina-Venegas^{2,5} 

¹Department of Biology (Botany), Universidad Autónoma de Madrid, Madrid, Spain

²Biodiversity and Global Change Research Center (CIBC-UAM), Universidad Autónoma de Madrid, Madrid, Spain

³GLOCEE—Global Change Ecology and Evolution Group, Department of Life Sciences, Universidad de Alcalá, Alcalá de Henares, Spain

⁴Instituto de Biología, Universidade Federal da Bahia, Salvador, Brazil

⁵Department of Ecology, Universidad Autónoma de Madrid, Madrid, Spain

Correspondence

Rafael Molina-Venegas and Ignacio Ramos-Gutiérrez, Biodiversity and Global Change Research Center (CIBC-UAM), Universidad Autónoma de Madrid, Madrid 28049, Spain.

Email: rafmolen@gmail.com and ignacio.ramosgutierrez@uam.es

Funding information

Ministry of Science and Innovation of Spain, Grant/Award Number: CGL2017-86926-P; Regional Government of Madrid, Spain, Grant/Award Number: CM/JIN/2019-005

Handling Editor: Franziska Schrot

Abstract

Aim: The increasing availability of molecular information has lifted our understanding of species evolutionary relationships to unprecedented levels. However, current estimates of the world's biodiversity suggest that about a fifth of all extant species are yet to be described, and we still lack molecular information for many of the known species. Hence, evolutionary biologists will have to tackle phylogenetic uncertainty for a long time to come. This prospect has urged the development of software to expand phylogenies based on non-molecular phylogenetic information, and while the available tools provide some valuable features, major drawbacks persist and some of the proposed solutions are hardly generalizable to any group of organisms.

Innovation: Here, we present a completely generalized and flexible framework to expand incomplete phylogenies. The framework is implemented in the R package “randtip”, a toolkit of functions that was designed to randomly bind phylogenetically uncertain taxa in backbone phylogenies through a fully customizable and automatic procedure that uses taxonomic ranks as a major source of phylogenetic information. Although randtip can generate fully operative phylogenies for any group of organisms using just a list of species and a backbone tree, we stress that the “blind” expansion of phylogenies using “quick-and-dirty” approaches often leads to suboptimal solutions. Thus, we discuss a variety of circumstances that may require customizing simulation parameters beyond default settings to optimally expand the trees, including a detailed step-by-step tutorial that was designed to provide guidelines to non-specialist users.

Main Conclusions: Phylogenetic uncertainty should be tackled with caution, assessing potential pitfalls and opportunities to optimize parameter space prior to launch any simulation. Used judiciously, our framework will help evolutionary biologists to efficiently expand incomplete phylogenies and thereby account for phylogenetic uncertainty in quantitative analyses.

KEY WORDS

backbone phylogeny, most derived consensus clade, most recent common ancestor, phylogenetic uncertainty, taxonomic rank

1 | INTRODUCTION

The past two decades have seen an explosive interest in incorporating evolutionary history into ecological analyses (Cavender-Bares et al., 2009; Mouquet et al., 2012; Webb et al., 2002), boosting several disciplines such as community ecology (Davies, 2021), macroecology (Lamsdell & Congreve, 2021) and conservation biology (Molina-Venegas et al., 2020). This eco-phylogenetic revolution was driven by the increased availability of molecular information (Sayers et al., 2020) and sophisticated tools for inferring phylogenetic trees (Smith & Walker, 2019), which have lifted our understanding of species evolutionary relationships to unprecedented levels. However, and despite the phylogeny of certain groups, such as mammals, is nearly completed (Upchurch et al., 2019), phylogenetic relationships remain vastly uncertain—particularly shallow ones (i.e., infra-family)—for many groups. For example, one of the largest global phylogenies of angiosperm plants published to date includes only ~12.5% of the species in the group (Janssens et al., 2020), and recent accounts of terrestrial arthropod biodiversity showed that up to 80% of insect species are yet to be discovered (Stork, 2018). These bleak figures suggest that evolutionary biologists will have to tackle phylogenetic uncertainty for a long time to come.

Conscious of the limited extent of molecular phylogenetic information, Rangel et al. (2015) developed a theoretical foundation to systematically account for phylogenetic uncertainty in quantitative analyses. Roughly, the procedure starts with the identification of *phylogenetically uncertain taxa* (PUTs), that is, taxonomic units (e.g., species, subspecies) that are well delineated in the continuum of biodiversity but remain missing from available phylogenies. Then, all acceptable taxonomic, morphological, or behavioural information on the PUTs is used to conservatively define their *most derived consensus clades* (MDCCs), that is, the less inclusive phylogenetic nodes that most certainly contain them. Finally, each PUT is assigned to a random point along one randomly selected branch of its corresponding MDCC, and the procedure is replicated a high number of times to obtain a distribution of possible trees that can be used in downstream analyses iteratively. While the “true” phylogenetic hypothesis will most certainly remain unsampled, the workflow allows exploring the parameter space, thereby quantifying the extent to which phylogenetic uncertainty has a significant impact in the analyses (e.g., Calatayud et al., 2019; Molina-Venegas et al., 2021). Rangel et al. (2015) accompanied their framework with the software SUNPLIN, a set of algorithms for randomly expanding phylogenies using the aforementioned procedure (Martins et al., 2013).

Although Rangel et al. (2015) suggested that the identification of MDCCs should be based on expert taxonomic evaluation, such knowledge is in practice beyond the reach of most researchers, particularly when dealing with very large phylogenies that often encompass a wide spectrum of taxonomic groups and thousands of species. In a valuable attempt to automatize the identification of MDCCs, Jin and Qian (2019) developed V.PhyloMaker, an R package that can generate large phylogenies of vascular plants (recently updated as U.PhyloMaker to include vertebrate animals; Jin & Qian, 2023).

PhyloMaker is based on the seminal idea of the classical software Phylomatic (Webb & Donoghue, 2005), which uses a taxonomically informed backbone mega-tree to automatically define MDCCs (in the case of PhyloMaker, genus or family nodes in case the former are not available) and bind the PUTs to the selected clades. Beyond covering features that were already implemented in Phylomatic, PhyloMaker provides an option to insert PUTs in randomly chosen nodes below the crown node of the corresponding MDCCs, so that a distribution of possible phylogenies can be generated with relatively little effort (Jin & Qian, 2019).

However, we note that current available tools for the insertion of PUTs, while valuable, have some important drawbacks. For example, PhyloMaker uses a pure node-based approach to insert PUTs, and thus the simulations often lead to the formation of polytomies even if a fully bifurcated backbone tree is used. In contrast, SUNPLIN allows the insertion of PUTs along randomly selected branches, but the user must manually set all the MDCCs for the simulations (Martins et al., 2013). PhyloMaker circumvents this limitation at the cost of requiring an “annotated” backbone mega-tree (a linkage between all the species represented in the backbone tree and their taxonomic genus and family) that is provided by the developers of the software, and thus the user is forced to use the backbone trees for which the software was implemented. Also, the definition of MDCCs on the basis of a few taxonomic ranks (e.g., PhyloMaker only considers genus or family nodes otherwise) might be excessively conservative and hence suboptimal under certain circumstances. For example, large taxonomic families often include taxonomic ranks between the family and genus level that may represent putative MDCCs (e.g., subfamilies, tribes and subtribes in the Asteraceae, Poaceae and Fabaceae plant families). Finally, there are shortcomings that are transversal to all available software for PUT binding, including the disregard of paraphyletic groups (Hörandl & Stuessy, 2010) and the impossibility to fully customize the space of phylogenetic edges for the insertion of PUTs among other issues.

Here, we present a completely generalized and flexible framework to expand incomplete phylogenies. The framework is implemented in the R package “randtip”, a toolkit of functions that was designed to randomly bind PUTs in backbone phylogenies through a fully customizable procedure that uses automatically retrieved and arranged taxonomic data as a major source of phylogenetic information. Although randtip can generate fully operative phylogenies for any group of organisms using just a list of species and a backbone tree, we discuss a variety of circumstances that may require customizing simulation parameters beyond default settings to optimally expand the trees, including a detailed step-by-step tutorial that was designed to provide guidelines to non-specialist users (see Supporting Information).

2 | GENERAL WORKFLOW

In this section, we describe the general workflow of randtip to expand phylogenies. Roughly, given a list of taxa (typically Linnean

binomials) for which a phylogeny is to be obtained and a backbone tree (provided by the user), the software identifies putative MDCCs for the PUTs in the list. MDCCs are defined based on taxonomic ranks, including genus, subtribe, tribe, subfamily, family, superfamily, order and class, and by default the software will select the less inclusive among the available. Once each PUT is assigned to a MDCC, randtip will automatically bind them to the backbone tree according to the parameters that are set for the simulations, and a phylogeny including all the taxa in the user's list is returned (Figure 1). The workflow can be customized using a variety of parameters that are either passed through the whole simulation or adjusted independently for each PUT (see Supporting Information for detailed step-by-step examples).

2.1 | Input files

The workflow of randtip is guided by a dataframe R object (hereafter “info”) and the instructions that are passed through the main function of the package (*rand_tip*). The dataframe *info* is a template with 21 columns –20 variables of type character or logical plus one integer variable for internal use—that must contain, as a minimum, all the taxa in the user's list (column 1) and their genus rank (column 2).

Optionally, the user may provide supra-generic taxonomic ranks and set parameter values specifically for individual PUTs. For simplicity, we will consider the most common scenario in the ecological literature where the operative taxa represent Linnean binomials (genus and species with or without subspecific epithets), although genus-level phylogenies are also supported. The *info* template can be created automatically using the auxiliary function *build_info*, which is fed with species names in a character vector or single-column dataframe. Besides, *build_info* can interact with a suite of taxonomic repositories—currently implemented for “ncbi” (default), “itis”, “gbif” and “bold” via the *classification* function of “taxize” R package (Chamberlain et al., 2020)—to automatically retrieve and arrange taxonomic information that will be used to identify putative supra-generic MDCCs for the PUTs (note that information to define genus-level MDCCs is intrinsically contained in the scientific names of the species). This can be done by setting the argument “find.ranks” of *build_info* to TRUE (default). We recommend providing at least one supra-generic rank (e.g., taxonomic family) for all the species in *info*, which will be used to define MDCCs whenever the genus of the PUTs is missing in the phylogeny (otherwise the PUTs will not be bound). Often the user will need to further edit *info* once the template is created (for example, to customize binding parameters for certain PUTs or to amend taxonomic mistakes in web repositories).

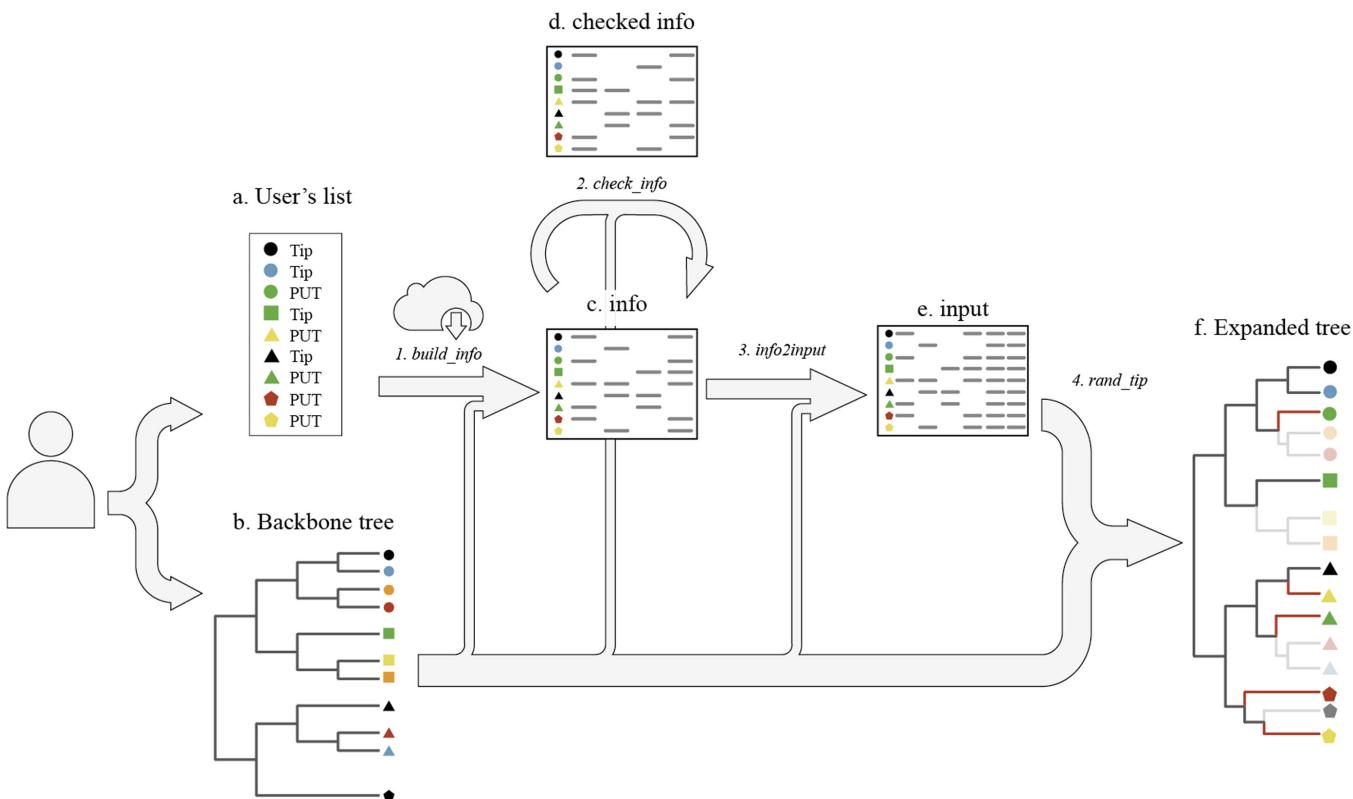


FIGURE 1 Schematic workflow of *randtip*. The user provides a backbone phylogeny and a list of taxa that are to be bound to the former (some of these are already placed in the tree while others represent *phylogenetically uncertain taxa* or PUTs). The function *build_info* creates the template *info* and retrieves taxonomic information for the listed taxa (and for those represented in the phylogeny if the “backbone” mode of *randtip* is set to TRUE) from web repositories. The resultant dataframe (*info*) can be evaluated with the function *check_info*. Once the user has edited *info* according to the particularities of each PUT, the dataframe is passed through *info2input* to create the *input* object for the *rand_tip* function, which in turn will expand the backbone phylogeny (in red, terminal branches subtending newly bound PUTs).

This can be done directly in R using the auxiliary function `edit_info` or exporting the dataframe as a spreadsheet (e.g., csv or xlsx) and importing it back into R once all the edits are completed.

The user must provide a backbone phylogeny as a `phylo` R object. Although `randtip` can identify MDCCs on the sole basis of taxonomic ranks of the species that are included both in the user's list and the backbone tree (hereafter "taxon list" mode), MDCCs can also be identified based on taxonomic ranks of all the species that are represented in the tree regardless of their presence in the user's list (hereafter "backbone" mode). Both approaches have pros and cons (see Section 3), but they will perform identically whenever the genus of the PUTs is represented by at least one species in the backbone tree. To use the "backbone" mode of `randtip`, the argument "mode" of `build_info` must be set to "backbone" (default) for the software to include all the species in the phylogeny as rows in the `info` dataframe (otherwise, only the species that appear both in the phylogeny and the user's list will be included), so that their taxonomic information can be automatically retrieved and arranged (if the argument "find.ranks" of `build_info` is set to TRUE).

Once the dataframe `info` is assembled, we strongly recommend checking the incidence of PUTs in the user's list and their putative MDCCs. This can be done with the auxiliary function `check_info`, which will inform on the PUT status of the species, the presence of possible spelling errors, putative MDCCs, and the phyletic nature of the set of species that are included in each MDCC and share taxonomic ranks (e.g., congenerics, contribrals, confamiliars) with the corresponding PUT—hereafter *phylogenetically placed and co-ranked* (PPCR) species. Also, the tip labels of the backbone tree are checked out for duplicates (e.g., *Ziziphora taurica taurica* and *Ziziphora taurica*), and the software evaluates if the tree is ultrametric or not. By default, `check_info` will make use of parallel processing to speed up the search for possible spelling errors and the identification of the phyletic nature of PPCR species, which is convenient for very large datasets. The auxiliary functions `get_clade` and `plot_clade` can in turn be used to extract and plot any subtree representing putative MDCCs, so that the user can visually explore them using the R graphic window (PPCR and non-PPCR species of the PUT are shown in contrasting colours, see Supporting Information for examples). Exploring MDCCs is particularly recommended to optimize PUT binding, and particularly when PPCR species form polyphyletic groups (see Section 2.2). Alternatively, subtrees can be exported in Newick format to visualize them using auxiliary software such as Dendroscope (Huson & Scornavacca, 2012), which may be convenient for very large clades. Once the MDCCs are defined and the user has optionally customized parameter values for individual PUTs, the wrapping function `info2input` is fed with the dataframe `info` and the backbone phylogeny to create a final dataset that will be passed through the `rand_tip` function to expand the tree. This final dataset ensures consistent structure for use in `rand_tip` and allows generating as many trees as desired without the need to search for putative MDCCs in `info` repeatedly. This is done by `info2input` just once, a computationally intense task that is, by default, expedited using parallel processing.

2.2 | Selecting MDCCs and binding PUTs

The binding of PUTs is conducted with the function `rand_tip`, which includes a variety of parameters that are passed through the whole simulation (Table S1). However, all the parameter arguments of `rand_tip` can be adjusted independently for each PUT by editing in the corresponding slots of `info`, which makes the framework completely flexible and customizable.

`Randtip` will always try to find the less inclusive MDCC of each PUT according to the taxonomic ranks that are provided in `info`, starting from genus level and up to class level until a MDCC is found. Regardless of the mode of `randtip` that is set by the user ("backbone" or "taxon list"), the software will always first attempt to define genus-level MDCCs as the *most recent common ancestor* (MRCA) of all the species in the backbone tree that are congeneric to the PUTs. However, MDCCs above the genus level may differ between the two modes of `randtip`. On "taxon list" mode, supra-generic MDCCs are defined as the MRCA of all the species in the user's list that are PPCR with the target PUT (e.g., contribrals, consubfamiliars, confamiliars). In contrast, the "backbone" mode (default) defines supra-generic MDCCs as the MRCA of all the species in the backbone phylogeny (regardless of their presence in the user's list) that are PPCR with the target PUT (see Figure 2 and Section 3 for an extended discussion).

By default, `rand_tip` will bind each PUT to a randomly selected branch below the crown node of the corresponding MDCC, the probability of being added along any branch being directly proportional to the length of the branch—if the argument "prob" is set to TRUE (default). Alternatively, branches can be selected on the basis of equal probability, and in either case the user can decide to add the stem branch of the clade to the pool of candidate branches—if the argument "use.stem" is set to TRUE (default is FALSE). The exact point to insert the PUT in the selected branch is sampled from a uniform distribution. Importantly, the extent to which the default behaviour of `rand_tip` to insert PUTs represents an optimal scenario may depend on the phyletic nature of their PPCR species. These can represent monophyletic (whenever the MDCC is exclusively shaped by species that are PPCR with the target PUT), singleton (terminal branch), paraphyletic (whenever the species that map within the MDCC but are not PPCR with the PUT form either a monophyletic or singleton group) or polyphyletic (set of PPCR species that does not fit any of the previous categories) groups (see Section 4.1 and Figure 3). The PPCR species of a given PUT could form a polyphyletic group simply because one of them maps clearly away from the main (monophyletic) cluster of PPCR species—for example, because the outlying PPCR species is labelled in error (Pentinsaari et al., 2020)—in which case the default behaviour of `rand_tip` to bind the PUT (i.e., any branch below the crown node of the largest monophyletic cluster) would be reasonable. However, the polyphyletic nature of the PPCR species could also be due to "intruder" species that map within an otherwise monophyletic cluster, in which case the default behaviour of `rand_tip` could be suboptimal because the evidence that the largest monophyletic cluster of the group includes the PUT is less conclusive (Figure 3).

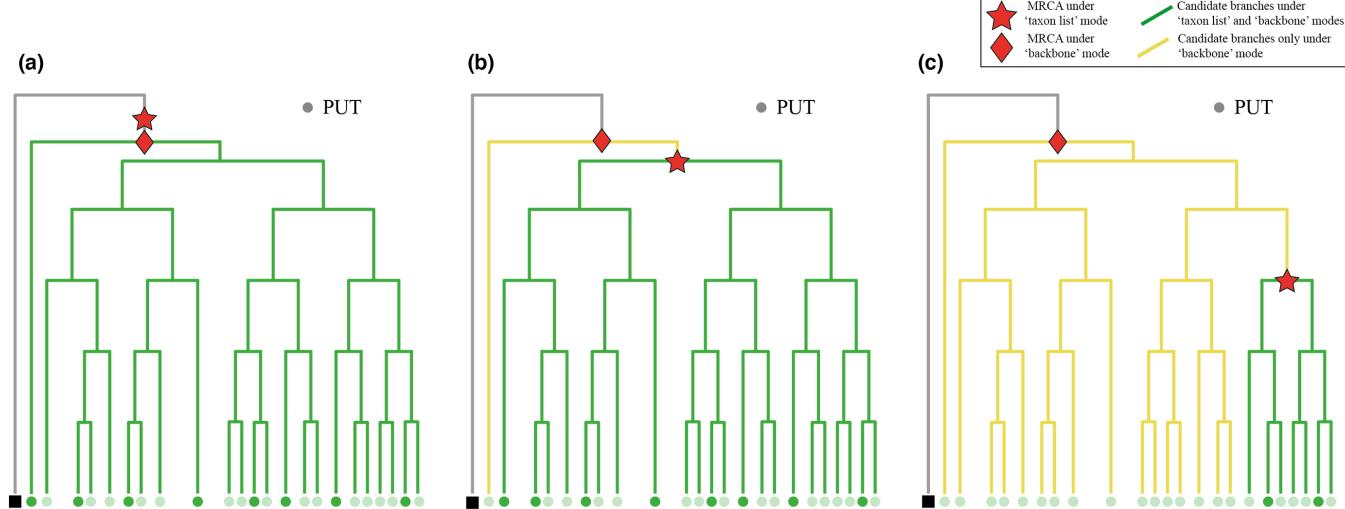


FIGURE 2 Scenarios of increasing divergence in the performance between the “taxon list” and “backbone” modes of *randtip*. The circle symbols on the phylogenetic tips represent *phylogenetically placed and co-ranked* (PPCR) species (e.g., *confamiliars*) of the PUT, and the highlighted ones are those included in the user’s list in each scenario. The diamond red symbol (hereafter “diamond node”) indicates the crown node of the *most derived consensus clade* (MDCC) that is identified for the PUT when taxonomic information is available for all the species in the backbone phylogeny (i.e., under “backbone” mode), and the star red symbol (hereafter “star node”) indicates the crown node of the MDCC that is identified when taxonomic information is available only for the species in the backbone phylogeny that are also included in the user’s list (i.e., under “taxon list” mode). In the first scenario (a), the diamond and star nodes are coincident, and thus both modes of *randtip* will use the same space of branch lengths (in green) to bind the PUT. In the second scenario (b), the *most recent common ancestor* (MRCA) of the subset of PPCR species that are represented in the user’s list includes all PPCR species but one, and therefore the branch subtending the latter (in yellow) will never be selected under “taxon list” mode. In the third scenario (c), a higher number of PPCR species are missing from the user’s list, resulting in a smaller space of branch lengths to bind the PUT under “taxon list” mode. Note that under “backbone” mode, both the green and yellow branches would be candidates to bind the PUT.

As we discuss in Section 4, *randtip* allows the user to optimize the binding of PUTs according to the specifics of each case.

It is important to note that the user can always decide to what extent they want to rely on the retrieved taxonomic ranks for the automatic identification of MDCCs. For example, if the taxonomic affiliation of a PUT to a given genus is controversial, the user may edit the dataframe *info* to change the genus-rank of the PUT into “NA”, in which case *randtip* will use the taxonomic rank immediately above to find a new MDCC.

3 | THE “TAXON LIST” AND “BACKBONE” MODES OF RANDTIP

The first decision the user will have to tackle is choosing between the “taxon list” and “backbone” modes of *randtip*. As we stated earlier, both approaches will perform identically as long as the genus of the PUTs is represented by at least one species in the backbone phylogeny, yet supra-generic MDCCs may differ between the two modes. For example, it might happen that some of the PPCR species of a given PUT (let us say *confamiliars*) are missing in the user’s list but are represented in the backbone phylogeny. Thus, in case these PPCR species were phylogenetically external to the *confamiliars* of the PUT that are included in the user’s list, the “backbone” mode of *randtip* would define an older MDCC than “taxon list” (Figure 2). It follows that the extent of the divergence in the functioning between

both modes (whenever a supra-generic MDCC is to be defined) depends on the phylogenetic placement of the PPCR species that are included in the user’s list. In sum, the “backbone” mode works based on the “true” supra-generic MDCCs (but note that these may neither represent the actual MDCCs as the backbone phylogenies are often not fully comprehensive) with the trade-off that it is a more time-consuming approach than “taxon list”. In contrast, the latter might define younger supra-generic MDCCs (meaning more restricted parameter space to bind PUTs) under some circumstances (Figure 2). We recommend considering the “backbone” mode as a first option (default) and use “taxon list” only when there is a low incidence of PUTs requiring supra-generic binding and/or low mismatch in the nodes defining supra-generic MDCCs between both approaches (see Figure 2 and Supporting Information for an extended discussion).

4 | NEWLY DESIGNED FEATURES FOR PUT BINDING

As discussed above, *rand_tip* will by default bind PUTs to randomly selected branches below the crown node of the corresponding MDCCs. However, this default behaviour can be modified using a variety of arguments that are implemented in *rand_tip*. For example, if the user is not interested in generating a distribution of possible phylogenies but one single tree without randomizing the PUTs, the argument “rand.

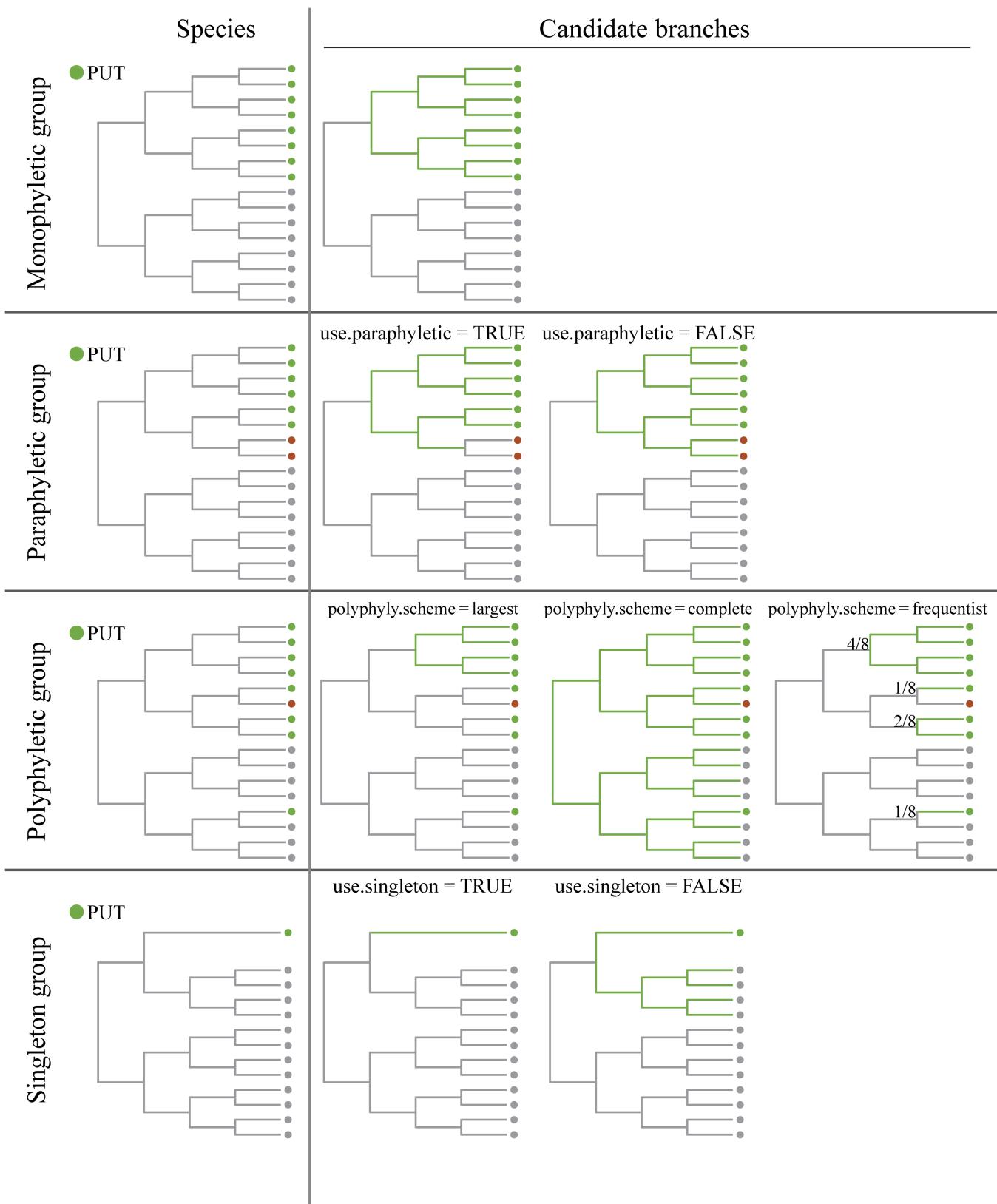


FIGURE 3 Types of phyletic groups formed by *phylogenetically placed and co-ranked* (PPCR) species (green circle symbols) and possible scenarios for PUT binding within each type. Non-PPCR species are in grey, and non-PPCR “intruder” species are in red. The candidate branches to bind the PUT in each scenario are in green (the vertical segments of the trees are purely aesthetic and were coloured to ease clade visualization). The fractions close to the phylogenetic nodes indicate the probability for the candidate clades to be selected under the scenario “frequentist”.

type" of `rand_tip` can be set to "polytomy" (default is "random") for the function to insert the PUTs as polytomies at the crown nodes of their corresponding MDCCs instead. This is the only binding option that was implemented in the seminal software Phylomatic (Webb & Donoghue, 2005), and it might still be convenient for extremely resource-consuming phylogenetic analyses where using a distribution of possible trees could be computationally prohibitive. Alternatively, the user may want to bind the PUTs following the default behaviour of `rand_tip` but still inserting some of them as polytomies in their corresponding MDCCs. To do so, the user can set the corresponding slots of the column "rand.type" of `info` to "polytomy" while keeping the argument "rand.type" of `rand_tip` to "random".

4.1 | Polyphyletic, paraphyletic and singleton groups of PPCR species

While PUT randomizations within monophyletic groups of PPCR species will always follow the same scheme (i.e., by default, randomly selected branches below the crown node of the corresponding MDCCs), the user must choose between different scenarios for polyphyletic, paraphyletic and singleton groups. In case the MDCC of a PUT is shaped by a polyphyletic group of PPCR species, the software allows the user to choose between three different binding scenarios using the "polyphyly.scheme" argument. If the default option "largest" is set, `rand_tip` will pick the largest monophyletic cluster of PPCR species among the available to insert the PUT (less conservative scenario; Figure 3). If the option "frequentist" is set, `rand_tip` will first pick one of the constituent clusters of PPCR species that conform the polyphyletic group, the probability of being selected being proportional to the size of the cluster, and then the PUT will be inserted in the selected cluster. If the option "complete" is set, `rand_tip` will bind the PUT to a randomly selected branch below the crown node of the MDCC (most conservative scenario).

In case the MDCC of a PUT is defined by a paraphyletic group of PPCR species, two different scenarios are eligible. If the argument "use.paraphyletic" is set to TRUE (default), the candidate branches are those that keep the paraphyletic nature of the group unchanged after the binding (Figure 3). Otherwise, the randomization will be conducted as if the MDCC were defined by a monophyletic group of species. Importantly, certain taxonomic groups such as the Olacaceae s.l. plant family are paraphyletic (Chase et al., 2016), and thus randomizing PUTs at any point below the crown node of this family (i.e., setting "use.paraphyletic" to FALSE) may result in an excessively conservative parameter space that would encompass almost the entire Santalales order (Malécot & Nickrent, 2008).

In case the MDCC of a PUT is defined by one single PPCR species (Figure 3), `rand_tip` will by default bind the PUT to the terminal branch subtending the only PPCR species, and whenever the MDCC is no longer singleton (because at least one PUT was already bound), `rand_tip` will consider the entire newly formed clade (same height as

the original singleton clade) to sample candidate branches. We will refer to this procedure as "bind-to-singleton" hereafter. However, if the argument "use.singleton" is set to FALSE (default is TRUE), the parent node of the singleton PPCR species will be defined as the MDCC of the PUT instead (Figure 3). Although the latter scheme is more conservative than the former, it may lead to suboptimal solutions under some circumstances. For example, the parameter space to randomize a PUT whose MDCC is shaped by one single species that is the only representative of a subfamily in the phylogeny can be drastically increased in case the subfamily is the sister group to the rest of the family. Note that all these parameters can be specifically set for individual PUTs by filling in the corresponding slots of `info`.

4.2 | Manual definition of MDCCs

Although `randtip` was conceived to automatize the definition of MDCCs based on taxonomic ranks, the user can manually define MDCCs for the PUTs. This can be done by filling in the corresponding slots of the columns "taxa1" and "taxa2" of `info`. As long as these slots are not set to "NA" (default), the MDCCs of the PUTs will be defined on the basis of this information instead. For example, if the slots "taxa1" and "taxa2" of a PUT are filled in with different species names, the PUT will be bound to a randomly selected branch below the MRCA of the two given species. If both slots are filled in with the same species name, `rand_tip` will follow the bind-to-singleton procedure to insert the PUT as sister to the so defined species, and in case the same genus is provided the PUT will be inserted as sister to the clade defined by the MRCA of all the species in that genus.

4.3 | Respecting monophyletic and paraphyletic groups

By default, `rand_tip` will never bind a PUT to a branch that results in breaking the monophyletic or paraphyletic nature of a group (of any taxonomic rank) unless the arguments "respect.mono" and "respect.para" are set to FALSE (default is TRUE). Thus, while previous software followed either approach (e.g., Phylomaker always respects monophyletic genera but SUNPLIN does not), `randtip` offers the user the possibility to choose between both options, either by setting the arguments of the `rand_tip` function or on a customized basis for individual PUTs by filling in the corresponding slots of `info`.

4.4 | Clumping PUTs

Some genera may not be represented in the phylogeny, and thus their representative species will likely form a polyphyletic group if they are to be bound randomly below the crown node of the corresponding supra-generic MDCC. However, the user could be certain in that a group of congeneric PUTs whose genus is missing in the phylogeny is monophyletic. Thus, if the argument "clump.puts" is set

to TRUE (default), `rand_tip` will first bind one of the congeneric PUTs, and then the rest will be bound following the bind-to-singleton procedure. Similarly, it may happen that supra-generic taxonomic groups are not represented in the phylogeny, in which case `rand_tip` will clump the PUTs as described above and following the taxonomic hierarchy so that the missing taxonomic groups will form monophyletic clusters once all the PUTs are bound. As any other randomization parameter of `randtip`, the user may decide the PUTs that will be clumped in this way by setting the “`clump.puts`” option individually in the corresponding slots of `info`.

Trinomials representing infra-specific taxa (e.g., subspecies) are also supported. If “`clump.puts`” is set to TRUE, `rand_tip` will clump PUTs with infra-specific information according to their specific epithets (i.e., second name in the trinomial). To do so, `rand_tip` will first check if any of the trinomial PUTs that share specific epithet are represented in the phylogeny. This search also takes into account the type subspecies of the species, which will be detected in either trinomial (e.g., *Ablepharus chernovi chernovi*) or binomial (e.g., *Ablepharus chernovi*) nomenclature. In case one or more PPCR subspecies are found in the backbone tree, `rand_tip` will define a MDCC for the infra-specific PUTs following the standard procedures described in Section 4.1. Finally, if none of the trinomials in the group are found, `rand_tip` will first bind any of them to the tree, and then all the others will be bound following the bind-to-singleton procedure.

We note that some available phylogenies use, likely in error, both the binomial and trinomial form of a species to label different tips. For example, the GBOTB.tre mega-tree (Smith & Brown, 2018) includes *Ziziphora taurica taurica* and *Ziziphora taurica* as two different tips, and the GBOTB.extended.tre mega-tree (Jin & Qian, 2019) includes both *Saxifraga serpyllifolia* and *Saxifraga serpyllifolia serpyllifolia*. In these cases, `rand_tip` will randomly select either tip as the actual type subspecies and ignore the other. Although the `check_info` function will warn the user about the existence of possible duplicate taxa in the backbone tree (see Supporting Information for an example), we strongly recommend the user to visually revise tip labelling before expanding any backbone tree.

4.5 | Non-ultrametric phylogenies

Previous software for PUT binding were conceived to be used with either ultrametric phylogenies (trees with branch lengths where all tips are equidistant from the root) or phylogenies without branch lengths. However, non-ultrametric trees where branch length is not proportional to time but character distance are also subject of ecological analyses (e.g., Mishler et al., 2014). The `check_info` function will warn the user in case the backbone phylogeny is non-ultrametric, and `rand_tip` will force non-ultrametric trees to be ultrametric—following the `nnls` method as implemented in “phytools” R package (Revell, 2012)—if the argument “`forceultrametric`” is set to TRUE (default is FALSE). It is important to note that forcing phylogenies to be ultrametric in this way should not be taken as a

formal statistical approach for inferring an ultrametric tree but a method to be deployed whenever a genuinely ultrametric phylogeny read from file fails due to issues related to numerical precision (Revell, 2012). Thus, we strongly recommend the user to visually explore phylogenetic trees that fail the ultrametricity test of `check_info` before assuming the failure is due to numerical precision of computer machinery.

If the backbone tree is non-ultrametric and the “`forceultrametric`” argument is set to FALSE, `rand_tip` will simulate the new branch lengths by sampling from a negative exponential distribution $EX(1/\lambda)$, where λ is the inverse of the mean terminal branch length in the backbone tree. In case a backbone phylogeny without branch lengths is provided, `rand_tip` will output a phylogeny without branch lengths as well (i.e., topological information only). Hence, the only condition for `rand_tip` to accept a phylogeny is that it is rooted.

4.6 | Customizing a subset of branches to randomize PUTs

The node-based workflow of `randtip` should suffice to cover most situations in PUT binding exercises. However, the distribution of possible branches for the simulation might not be drawn via MDCCs under some circumstances. For example, taxa of hybrid origin often appear as the sister species of either parent depending on the set of molecular markers that are used for the inference (Wang et al., 2014), in which case phylogenetic uncertainty may pertain to only two singleton putative MDCCs (assuming that the identity of the parents is known and both are represented in the backbone tree). Using the auxiliary function `custom_branch`, the user can customize specific subsets of branches to bind PUTs across any segment of the phylogeny.

5 | CONCLUDING REMARKS

`Randtip` is, to our knowledge, the only framework for PUT binding that is completely flexible and generalized, thus addressing several shortcomings of previous designs and offering new opportunities to optimize parameter space in tree expansion exercises. Although `randtip` can generate fully operative phylogenies using default settings, we stress that accounting for phylogenetic uncertainty should not be conceived as a “black box” procedure for the immediate generation of phylogenies. Indeed, previous studies have documented inaccuracies in the generation of such “quick-and-dirty” phylogenies due to the “blind” use of software packages (Gastauer & Meira-Neto, 2013). Phylogenetic uncertainty should always be tackled with caution and restraint, for there is a variety of circumstances that may require customizing simulation parameters for specific PUTs if we are to avoid suboptimal solutions. Beyond providing newly designed tools to expand phylogenetic trees, the framework presented here will help evolutionary biologists to get the most out of the evolutionary information that can be used to guide tree expansion exercises.

AUTHOR CONTRIBUTIONS

Rafael Molina-Venegas conceived the ideas with inputs from Ignacio Ramos-Gutiérrez; Ignacio Ramos-Gutiérrez developed the code with the help of Rafael Molina-Venegas, Herlander Lima and Bruno Vilela; and Rafael Molina-Venegas led the writing. All the authors read the manuscript and approved submission.

ACKNOWLEDGEMENTS

IR-G was primarily supported by the project CM/JIN/2019-005 entitled "Plant evolutionary history and human well-being in a changing world; assessing theoretical foundations using empirical evidence and new phylogenetic tools", granted to R.M.-V. (Regional Government of Madrid, Spain) and also by the project CGL2017-86926-P (Ministry of Science and Innovation of Spain).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

No data were used to produce this manuscript.

CODE AVAILABILITY STATEMENT

All the code can be sourced from the GitHub repository (<https://github.com/iramosgutierrez/randtip>) as explained in the supplementary material, and the software will be delivered as a formal R package in CRAN upon acceptance of the manuscript.

ORCID

Ignacio Ramos-Gutiérrez  <https://orcid.org/0000-0002-8675-0114>

Rafael Molina-Venegas  <https://orcid.org/0000-0001-5801-0736>

REFERENCES

- Calatayud, J., Rodríguez, M. Á., Molina-Venegas, R., Leo, M., Horreo, J. L., & Hortal, J. (2019). Pleistocene climate change and the formation of regional species pools. *Proceedings of the Royal Society B*, 286, 20190291. <https://doi.org/10.1098/rspb.2019.0291>
- Cavender-Bares, J., Kozak, K. H., Fine, P. V. A., & Kembel, S. W. (2009). The merging of community ecology and phylogenetic biology. *Ecology Letters*, 12, 693–715. <https://doi.org/10.1111/j.1461-0248.2009.01314.x>
- Chamberlain, S., Szoechs, E., Foster, Z., Arendsee, Z., Boettiger, C., Ram, K., Bartomeus, I., Baumgartner, J., O'Donnell, J., Oksanen, J., Greshake Tzovaras, B., Marchand, P., Tran, V., Salmon, M., Li, G., & Grenié, M. (2020). Taxize: Taxonomic information from around the web. R Package Version 0.9.98. <https://github.com/ropensci/taxize>
- Chase, M. W., Christenhusz, M. J. M., Fay, M. F., Byng, J. W., Judd, W. S., Soltis, D. E., Mabberley, D. J., Sennikov, A. N., Soltis, P. S., & Stevens, P. F. (2016). An update of the angiosperm phylogeny group classification for the orders and families of flowering plants: APG IV. *Botanical Journal of the Linnean Society*, 181, 1–20. <https://doi.org/10.1111/boj.12385>
- Davies, T. J. (2021). Ecophylogenetics redux. *Ecology Letters*, 24, 1073–1088. <https://doi.org/10.1111/ele.13682>
- Gastauer, M., & Meira-Neto, J. A. A. (2013). Avoiding inaccuracies in tree calibration and phylogenetic community analysis using Phylocom 4.2. *Ecological Informatics*, 15, 85–90. <https://doi.org/10.1016/j.ecoinf.2013.03.005>
- Hörndl, E., & Stuessy, T. F. (2010). Paraphyletic groups as natural units of biological classification. *Taxon*, 59, 1641–1653. <https://doi.org/10.1002/tax.596001>
- Huson, D. H., & Scornavacca, C. (2012). Dendroscope 3: An interactive tool for rooted phylogenetic trees and networks. *Systematic Biology*, 61, 1061–1067. <https://doi.org/10.1093/sysbio/sys062>
- Janssens, S., Couvreur, T. L. P., Mertens, A., Dauby, G., Dagallier, L.-P., Abeele, S. V., Vandelooy, F., Mascarello, M., Beeckman, H., Sosef, M., Droissart, V., van der Bank, M., Maurin, O., Hawthorne, W., Marshall, C., Réjou-Méchain, M., Beina, D., Baya, F., Merckx, V., ... Hardy, O. (2020). A large-scale species level dated angiosperm phylogeny for evolutionary and ecological analyses. *Biodiversity Data Journal*, 8, e39677. <https://doi.org/10.3897/BDJ.8.e39677>
- Jin, Y., & Qian, H. (2019). V.PhyloMaker: An R package that can generate very large phylogenies for vascular plants. *Ecography*, 42, 1353–1359. <https://doi.org/10.1111/ecog.04434>
- Jin, Y., & Qian, H. (2023). U.PhyloMaker: An R package that can generate large phylogenetic trees for plants and animals. *Plant Diversity*, 45, 347–352. <https://doi.org/10.1016/j.pld.2022.12.007>
- Lamsdell, J. C., & Congreve, C. R. (2021). Phylogenetic paleoecology: Macroecology within an evolutionary framework. *Paleobiology*, 47, 171–177. <https://doi.org/10.1017/pab.2020.61>
- Malécot, V., & Nickrent, D. L. (2008). Molecular phylogenetic relationships of Olacaceae and related Santalales. *Systematic Botany*, 33, 97–106. <https://doi.org/10.1600/036364408783887384>
- Martins, W. S., Carmo, W. C., Longo, H. J., Rosa, T. C., & Rangel, T. F. (2013). SUNPLIN: Simulation with uncertainty for phylogenetic investigations. *BMC Bioinformatics*, 14, 324. <https://doi.org/10.1186/1471-2105-14-324>
- Mishler, B. D., Knerr, N., González-Orozco, C. E., Thornhill, A. H., Laffan, S. W., & Miller, J. T. (2014). Phylogenetic measures of biodiversity and neo-and paleo-endemism in Australian Acacia. *Nature Communications*, 5, 4473. <https://doi.org/10.1038/ncomms5473>
- Molina-Venegas, R., Ramos-Gutiérrez, I., & Moreno-Saiz, J. C. (2020). Phylogenetic patterns of extinction risk in the endemic flora of a Mediterranean hotspot as a guiding tool for preemptive conservation actions. *Frontiers in Ecology and Evolution*, 8, 373. <https://doi.org/10.3389/fevo.2020.571587>
- Molina-Venegas, R., Rodríguez, M. Á., Pardo-de-Santayana, M., Ronquillo, C., & Mabberley, D. J. (2021). Maximum levels of global phylogenetic diversity efficiently capture plant services for humankind. *Nature Ecology & Evolution*, 5, 583–588. <https://doi.org/10.1038/s41559-021-01414-2>
- Mouquet, N., Devictor, V., Meynard, C. N., Muñoz, F., Bersier, L.-F., Chave, J., Couturon, P., Dalecky, A., Fontaine, C., Gravel, D., Hardy, O. J., Jabot, F., Lavergne, S., Leibold, M., Mouillot, D., Münkemüller, T., Pavoine, S., Prinzing, A., Rodrigues, A. S. L., ... Thuiller, W. (2012). Ecophylogenetics: Advances and perspectives. *Biological Reviews*, 87, 769–785. <https://doi.org/10.1111/j.1469-185X.2012.00224.x>
- Pentinsaari, M., Ratnasingham, S., Miller, S. E., & Hebert, P. D. (2020). BOLD and GenBank revisited—Do identification errors arise in the lab or in the sequence libraries? *PLoS One*, 15, e0231814.
- Rangel, T. F., Colwell, R. K., Graves, G. R., Fučíková, K., Rahbek, C., & Diniz-Filho, J. A. F. (2015). Phylogenetic uncertainty revisited: Implications for ecological analyses. *Evolution*, 69, 1301–1312. <https://doi.org/10.1111/evo.12644>
- Revell, L. J. (2012). phytools: An R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, 3, 217–223. <https://doi.org/10.1111/j.2041-210X.2011.00169.x>
- Sayers, E. W., Beck, J., Brister, J. R., Bolton, E. E., Canese, K., Comeau, D. C., Funk, K., Ketter, A., Kim, S., Kimchi, A., Kitts, P. A., Kuznetsov, A., Lathrop, S., Lu, Z., McGarvey, K., Madden, T. L., Murphy, T. D., O'Leary, N., Phan, L., ... Ostell, J. (2020). Database resources of the National Center for biotechnology information. *Nucleic Acids Research*, 48, D9–D16. <https://doi.org/10.1093/nar/gkz899>

- Smith, S. A., & Brown, J. W. (2018). Constructing a broadly inclusive seed plant phylogeny. *American Journal of Botany*, 105, 302–314. <https://doi.org/10.1002/ajb2.1019>
- Smith, S. A., & Walker, J. F. (2019). PyPHLAWD: A python tool for phylogenetic dataset construction. *Methods in Ecology and Evolution*, 10, 104–108. <https://doi.org/10.1111/2041-210X.13096>
- Stork, N. E. (2018). How many species of insects and other terrestrial arthropods are there on Earth? *Annual Review of Entomology*, 63, 31–45. <https://doi.org/10.1146/annurev-ento-020117-043348>
- Upham, N. S., Esselstyn, J. A., & Jetz, W. (2019). Inferring the mammal tree: Species-level sets of phylogenies for questions in ecology, evolution, and conservation. *PLoS Biology*, 17, e3000494. <https://doi.org/10.1371/journal.pbio.3000494>
- Wang, Z., Du, S., Dayanandan, S., Wang, D., Zeng, Y., & Zhang, J. (2014). Phylogeny reconstruction and hybrid analysis of *Populus* (Salicaceae) based on nucleotide sequences of multiple single-copy nuclear genes and plastid fragments. *PLoS One*, 9, e103645. <https://doi.org/10.1371/journal.pone.0103645>
- Webb, C. O., Ackerly, D. D., McPeek, M. A., & Donoghue, M. J. (2002). Phylogenies and community ecology. *Annual Review of Ecology and Systematics*, 33, 475–505. <https://doi.org/10.1146/annurev.ecolsys.33.010802.150448>
- Webb, C. O., & Donoghue, M. J. (2005). Phylomatic: Tree assembly for applied phylogenetics. *Molecular Ecology Notes*, 5, 181–183. <https://doi.org/10.1111/j.1471-8286.2004.00829.x>

BIOSKETCH

Ignacio Ramos-Gutiérrez is a PhD candidate interested in exploring floristic and biogeographic patterns of Iberian vascular plants, with a major focus on the phylogenetic dimension of biodiversity for conservation.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Ramos-Gutiérrez, I., Lima, H., Vilela, B., & Molina-Venegas, R. (2023). A generalized framework to expand incomplete phylogenies using non-molecular phylogenetic information. *Global Ecology and Biogeography*, 00, 1–10. <https://doi.org/10.1111/geb.13733>

Tutorial for randtip R package

Contents

Introduction	1
Package installation	2
Examples	2
Example 1 - The ‘backbone’ mode of randtip	2
Example 2 - The ‘taxon list’ mode of randtip	16
Manual definition of candidate branches	19
References	22

Introduction

This tutorial aims to provide a step-by-step user guide to expand incomplete phylogenies using the 'randtip' R package.

Specifically, the tutorial shows how to expand a hypothetical species-level backbone phylogeny with 56 tips using a list of 54 species. Some of the species in the list are already placed in the backbone tree but others are missing, the latter representing phylogenetically uncertain taxa (PUTs).

We focus on hypothetical rather than real world data to cover a wider array of PUT binding situations. All the datasets required to complete the tutorial are available as part of the 'randtip' R package.

Package installation

The 'randtip' R package is available as an open-source software hosted on GitHub at github.com/iramosgutierrez/randtip. The package can be installed (along with its dependencies) using the `install_github` function of the 'devtools' package (Wickham *et al.* 2021).

```
install.packages("devtools")
library(devtools)

devtools::install_github("iramosgutierrez/randtip")
```

Randtip will be delivered soon as a formal R package in CRAN, so that the user will be able to install the software using the utility function `install.packages` or the corresponding pathways of R interfaces (e.g. tab “packages” and then “install” if using *R Studio*). Once the installation is completed the package can be loaded.

```
library(randtip)
```

Examples

Example 1 - The ‘backbone’ mode of randtip

1. Data loading

Randtip requires the user to provide a backbone phylogeny (R object of class *phylo*) and a list of taxa (typically species and/or subspecies) for which a phylogenetic hypothesis is to be obtained.

The user must ensure that there are no duplicate `species` in the list, and that taxonomic criteria between the `species` in the list and the phylogenetic tip labels have been harmonized. Taxonomic harmonization can be conducted using several available resources that were conceived

to speed up this task, including the R packages ‘WorldFlora’ (Kindt 2020) for plants, and ‘U.Taxonstand’ (Zhang & Qian 2023) for plants and animals.

The phylogeny can be loaded into R using the function `read.tree` of 'ape' R package (Paradis & Schliep 2019), and the list of species can be loaded as a single column data frame or character vector.

```
sp.list <- read.table("/.../specieslist.txt")
back.tree <- ape::read.tree("/.../backbone.txt")
```

If randtip is already installed in the system, the hypothetical example files can be loaded into the working space.

```
sp.list <- mythology$sp.list
back.tree <- mythology$back.tree
```

2. Building the dataframe *info*

Once the backbone phylogeny and the list of species have been imported into the working space, the next step is to build the dataframe *info*. The most direct and safest way to assemble *info* is using the `build_info` function. Here, we will follow the 'backbone' mode of randtip (default) to bind the PUTs to the tree (otherwise the argument ‘mode’ of `build_info` must be set to “list”, see Example 2 below). Therefore, *info* will include not only the 54 species in the list, but also the species that are represented in the backbone phylogeny. By default, `build_info` will try to retrieve taxonomic information from the web for all the species in the list and the backbone phylogeny (calling to the ‘ncbi’ repository as default option). However, this procedure makes no sense in our hypothetical example, and thus the ‘find.ranks’ argument of `build_info` must be set to FALSE (default is TRUE).

Note that in case of unmatched or ambiguous queries (i.e. genera that may refer to different taxonomic groups) during taxonomic rank retrieval, NAs will be returned for the entire row (across all taxonomic ranks of the corresponding genus). Thus, we strongly recommend to visually explore the outputted dataframe and to manually

check if the empty rows (i.e. those filled with NAs in their entirety) correspond to either ambiguous or unmatched queries (browsing taxonomic repositories). Alternatively, the user can choose to solve ambiguities as they are found during the querying process by setting the argument ‘interactive’ to TRUE. However, this option will halt the function each time an ambiguous name is found, and therefore it may not be convenient for very large datasets that can take substantial time for completion. As a guideline, downloading taxonomic information for 10585 genera (~75000 species in total, but note that timing depends on the number of genera rather than species) took ~5 hours in a standard PC. Nonetheless, the user may consider using the taxonomic information we provide for vascular plants (Zanne *et al.* 2014; Smith & Brown 2018), fish (Rabosky *et al.* 2018), mammals (Upham *et al.* 2019), squamata (Pyron *et al.* 2013), amphibians (Jetz & Pyron 2018) and birds (Kimball *et al.* 2019) to assemble the *info* dataframe (data retrieved mostly from the NCBI Taxonomy Browser in January 2023). If the argument ‘prior.info’ is fed with a dataframe object (which must have the same structure as an *info* dataframe), `build_info` will first retrieve taxonomic information from this dataframe before starting to query the taxonomic repository for the genera that may still be missing.

```
my.info.noranks <- build_info(sp.list, tree = back.tree,
                                find.ranks = FALSE,
                                mode = "backbone")

my.info.noranks # print the data frame info on screen
```

Back to our hypothetical example, note that the only taxonomic information in the outputted dataframe (‘`my.info.noranks`’) is genus-level (second column, which has been filled in automatically with the genus of the species in the list and the backbone phylogeny), meaning that in case the genus of a PUT is missing in the backbone phylogeny, the PUT will not be bound. Thus, we recommend providing at least one supra-generic rank (e.g. taxonomic family) for all the species in *info*, which will be used to define supra-generic MDCCs whenever the genus of a PUT is missing in the phylogeny.

For the purpose of completing this tutorial, an alternative *info* dataframe including supra-generic taxonomic information for all the species in the list and the backbone

phylogeny can be loaded into the working space. This dataframe ('my.info') was filled in with class, order and family in all cases, and subfamily in some cases, as they would have been retrieved from a taxonomic repository if the species represented real taxa.

```
my.info <- mythology$info.backbone  
my.info # print the data frame info on screen
```

Note that the species that are represented in the backbone phylogeny but are missing in list show hyphens instead of 'NA' from columns 10th to 20th. This is because these columns are intended to customize simulation parameters for the species in the user's list representing PUTs. Thus, while we still do not know which species in the list represent PUTs, we can be certain that the species depicted in 'my.info' that are missing in the list (i.e. those showing hyphens from columns 10th to 20th) do not represent PUTs, for they are all included in the backbone phylogeny.

3. Checking the dataframe *info*

Once the *info* dataframe is created (and amended if necessary), it is very convenient to check for possible spelling errors. Otherwise, it may happen that species included both in the user's list and the backbone phylogeny are misidentified as PUTs simply because they were misspelled in either object.

Also, we strongly recommend the user to check the phyletic nature of the phylogenetically placed and co-ranked (PPCR) species that define putative most derived consensus clades (MDCCs) for the PUTs, so that informed decisions can be made in accordance with the particularities of each case (see section 4).

Finally, it is important to ensure that the backbone tree is ultrametric (in case the tree read from file is genuinely ultrametric) and does not include duplicate taxa. All these checks can be conducted using the function `check_info`, which will output a new dataframe with all the information. This can be a time-consuming task for very large datasets and, by default, the function will make use of parallel processing to expedite the analysis using all available cores minus one. Here, we will check the *info* dataframe

using two cores.

```
my.check <- check_info(my.info, back.tree, parallelize = T, ncores = 2)
```

Our example data set produces three **warning** messages.

```
## There may be misspelling errors in the species list or the
##   phylogenetic tips. Please, check the TYPO column in the
##   outputted data frame.

##

## Tips Yetis_abominabilis_abominabilis and Yetis_abominabilis may
##   represent the same taxon. Please consider removing one of
##   them.

##

## The backbone tree is not ultrametric.
```

First, the function warns about the possibility of misspelling errors. As such, the column “Typo.names” of the newly created dataframe ‘my.check’ reveals that the species *Gorgona medusi* (represented in the backbone phylogeny) was erroneously typed in the user’s list as *Gorgona medusii* (it may also happen the other way around, i.e. that a species is misspelled in the phylogeny but correctly typed in the user’s list). It is important that the user corrects misspelling errors in ‘my.info’ or the backbone tree, because otherwise the species will remain misidentified as PUTs (as shown in the second column of ‘my.check’). The *info* dataframe can be edited directly in R using the auxiliary function `edit_info` or exporting *info* as a spreadsheet (e.g. csv or xlsx) and importing it back into R once the edits are completed, and phylogenetic tip labels can be edited with the auxiliary function `edit_tree`.

Here, we will use the function `edit_info` for correcting the misspelling error. To do so, the user only needs to indicate the heading of the column that is to be edited, the corresponding species (as in the first column) and the new information.

```
my.info <- edit_info(my.info, taxa = "Gorgona medusii",
                      column = "taxon", edit = "Gorgona medusi")
```

Second, the function informs that two phylogenetic tips (*Yetis abdominabilis* and *Yetis abdominabilis abdominabilis*) may represent the same taxon, and thus the user may consider picking one of them and disregard the other. Otherwise, randtip will choose randomly between the two tips for the purpose of binding PUTs in case the taxon *Yetis abdominabilis* is defined as a MDCC. Phylogenetic tips can be easily pruned from the tree using the argument ‘remove.tip’ of the auxiliary function `edit_tree` (which also serves to edit tip labels, see Table S1), and the corresponding row of `info` should be removed as well. The latter amend can be conducted using the argument ‘remove.rows’ of `edit_info`.

```
back.tree <- edit_tree(back.tree,
                        tips="Yetis abdominabilis abdominabilis",
                        remove.tips=TRUE)

my.info <- edit_info(my.info,
                      taxa="Yetis abdominabilis abdominabilis",
                      remove.rows=TRUE)
```

Third, the function informs that the backbone tree is not ultrametric. However, we are certain that the tree read from file is genuinely ultrametric (it is simply detected as non-ultrametric due to numerical precision of computer machinery), and therefore we will force the tree to be ultrametric (see section 6).

Finally, `check_info` will by default evaluate the phyletic nature of the taxonomic ranks included in `info`, which is by default conducted with parallel processing (‘parallelize’ = TRUE). If the argument ‘find.phylicity’ is set to FALSE (default is TRUE), this evaluation will be skipped, but we strongly recommend the user to work in the light of this information, because the phyletic nature of taxonomic ranks can be instrumental for the optimal expansion of phylogenies (see main text).

4. Customizing simulation parameters

By default, `randtip` will bind the PUTs to the backbone tree using the parameters that are specified in the arguments of `rand_tip` (see section 5). However, using the same set of parameters to bind all the PUTs may lead to suboptimal solutions in many cases.

For example, consider the PUTs *Draco borealis*, *Draco troglodytes* and *Draco wiverny*, whose genus was identified as a polyphyletic group by `check_info`. We can take a closer look to the less inclusive clade that includes all the species in the genus *Draco* using the functions `get_clade` and `plot_clade` in tandem.

```
my.clade <- get_clade(my.info, back.tree, group = "Draco")
plot_clade(my.clade)
```

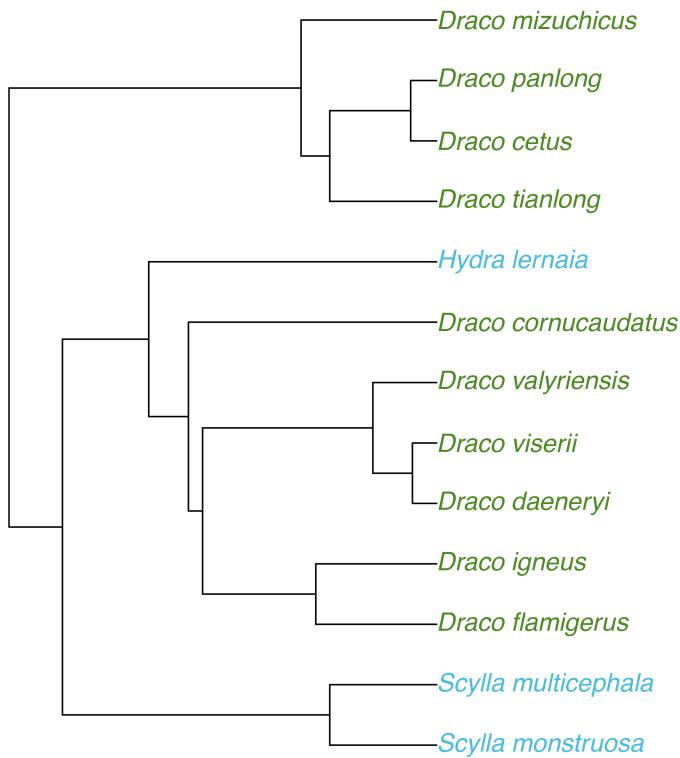


Figure 1: Backbone phylogeny pruned to the less inclusive clade that includes all the species in genus *Draco* (representative species in green)

Because the congenerics of these PUTs form two monophyletic clusters that are very similar in size (four and six species, respectively; Figure 1), the default behaviour

of `rand_tip` for binding *Draco borealis*, *Draco troglodytes* and *Draco wiverny* to this MDCC (largest monophyletic cluster) is risky, for the evidence that the largest cluster most likely include them is rather weak. Thus, a more conservative approach is desirable. For example, the user may use the “complete” scheme to bind these PUTs to a randomly selected branch below the crown node defining the most recent common ancestor (MRCA) of all the species in the genus instead (i.e. root node of the phylogeny displayed in Figure 1). Alternatively, the user may use the “frequentist” scheme to bind each of the PUTs to either of the two monophyletic clusters with a probability proportional to the size of the groups. Here, we will use the former scheme, and to do so, we can fill in the corresponding slots of `info` (column ‘polyphyly.scheme’) to set the “complete” scheme for these PUTs.

```
DracoPUTs <- c("Draco borealis", "Draco troglodytes",
              "Draco wiverny")

my.info <- edit_info(my.info,
                      taxa = DracoPUTs,
                      column = "polyphyly.scheme",
                      edit = "complete")
```

It may happen that the user is certain that the MDCC of a PUT does not correspond to any of the taxonomic groups considered by `randtip`. For example, the MDCC of the PUT *Draco balerion* could be infra-generic (e.g. a taxonomic section within the genus *Draco* including *Draco valyriensis*, *Draco viserii* and *Draco daeneryi*). The user may know that the MRCA of all the species constituting the target taxonomic section in the phylogeny is defined by *Draco valyriensis* and *Draco daeneryi* (the spanning path connecting both species in the tree traverses the MRCA of all the species in the section). Thus, we can fill in the slots “taxon1” and “taxon2” of the corresponding row of `info` with *Draco valyriensis* and *Draco daeneryi* to define an infra-generic MDCC for *Draco balerion*.

```

my.info <- edit_info(my.info,
                      taxa = "Draco balerion",
                      column = "taxon1",
                      edit = "Draco valyriensis")

my.info <- edit_info(my.info,
                      taxa = "Draco balerion",
                      column = "taxon2",
                      edit = "Draco daeneryi")

```

Now, consider the PUT *Lycanthropus albus*, whose genus also forms a polyphyletic group.

```

my.clade <- get_clade(my.info, back.tree, group = "Lycanthropus")
plot_clade(my.clade)

```

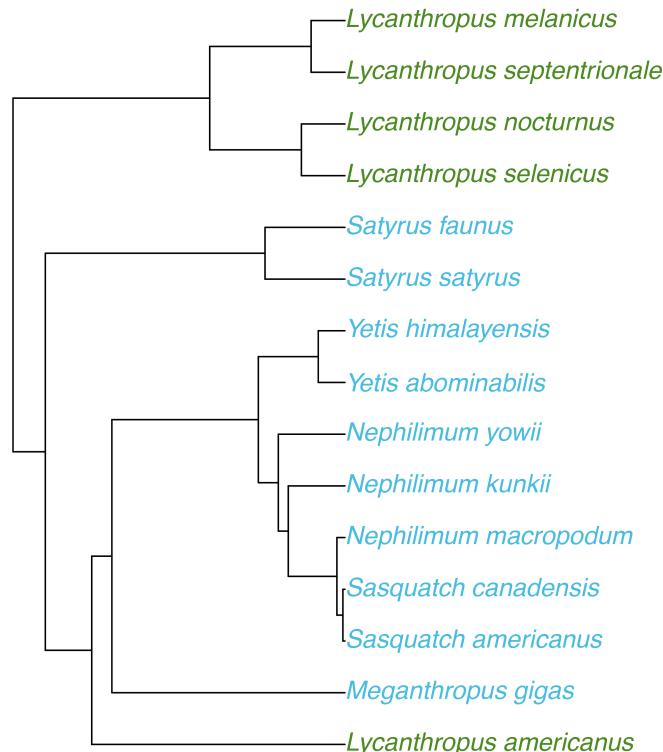


Figure 2: Backbone phylogeny pruned to the less inclusive clade that includes all the species in genus *Lycanthropus* (representative species in green).

In this case, the polyphyletic nature of the group is due to an outlying species (*Lycanthropus americanus*) that maps clearly away from the main cluster of *Lycanthropus* (Figure 2). Therefore, the default scheme “largest” seems adequate (i.e. it is highly likely that the largest cluster of the genus in the phylogeny actually includes *Lycanthropus americanus*) whereas the “complete” scheme could be excessively conservative.

5. Wrapping up

Once we have edited the data in *info* as we see fit (see above), the function `info2input` can be used to create the input object for the `rand_tip` function. This final dataset ensures consistent structure for use in `rand_tip` and allows generating as many expanded phylogenies as desired without the need to search for putative MDCCs in *info* repeatedly, which is a computationally intense task (this is done by `info2input` just once). By default, this function makes use of parallel processing.

In case `info2input` fails to find a MDCC for a PUT (which will only happen if the genus of the PUT is missing in the backbone tree and no supra-generic taxonomic information is provided), the function will return a `warning` message. Otherwise, `info2input` will select the less inclusive MDCCs of each PUT.

```
my.input.noranks <- info2input(my.info.noranks, back.tree)
```

```
## The following taxa were not assigned MDCC and will not be bound
## to the tree:
## Grindylowia_yorkii
## Harpia_feminicephala
## Leviathanus_cthulus
## Trolleolus_angmariensis
## Trolleolus_mordoriensis
```

```
my.input <- info2input(my.info, back.tree)
```

Note that the outputted dataframe 'my.input' is identical to 'my.info' except for the two newly added columns, namely 'MDCC' and 'MDCC.rank'. The column 'MDCC' shows the taxonomic groups defining the MDCCs to which the PUTs will be bound ('Tip' means that the species is already represented in the backbone phylogeny), and 'MDCC.rank' depicts the taxonomic rank of the groups.

6. PUT binding

The binding of PUTs in the selected MDCCs is carried out with the `rand_tip` function, which is fed with the output of `info2input` ('my.input'). Most arguments of `rand_tip` are used for defining simulation parameters (and thus they can be customized for individual PUTs via `info`) except for 'prune', 'forceultrametric' and 'verbose'.

By default, `rand_tip` will output a phylogenetic tree including only the species in the user's list unless 'prune' is set to FALSE, in which case the whole expanded backbone phylogeny will be outputted.

In case the tree read from file is detected as non-ultrametric despite being genuinely ultrametric (as in our hypothetical example, see [main text](#) for an extended discussion on this issue), the user can set the 'forceultrametric' argument to TRUE (default is FALSE) to force the tree to be ultrametric.

Lastly, the argument 'verbose' allows the user to print the progress of the function on screen (default is TRUE). Here, we will use the function with default settings except for (1) forcing the backbone phylogeny to be ultrametric, and (2) outputting the whole expanded tree. This will enable us to better visualize the MDCCs that were selected to bind the PUTs (see Figure 3):

```
A  
new.tree <- rand_tip(my.input, back.tree,  
                      forceultrametric = TRUE,  
                      prune = FALSE)
```

7. Tree visualization

We can visualize the result of the simulation using the `plot.phylo` function of ‘ape’ R package. To distinguish between phylogenetically placed species and PUTs, we can set the color pattern of phylogenetic tips using the auxiliary function `put_tip_col` before `plot.phylo`. Note that visualizing very large phylogenies may require auxiliary software such as Dendroscope (Huson & Scornavacca 2012).

```
my.tip.col <- put_tip_col(new.tree, back.tree,
                           placed.col = "dark grey",
                           put.col = "red")
plot.phylo(new.tree, tip.color = my.tip.col)
```

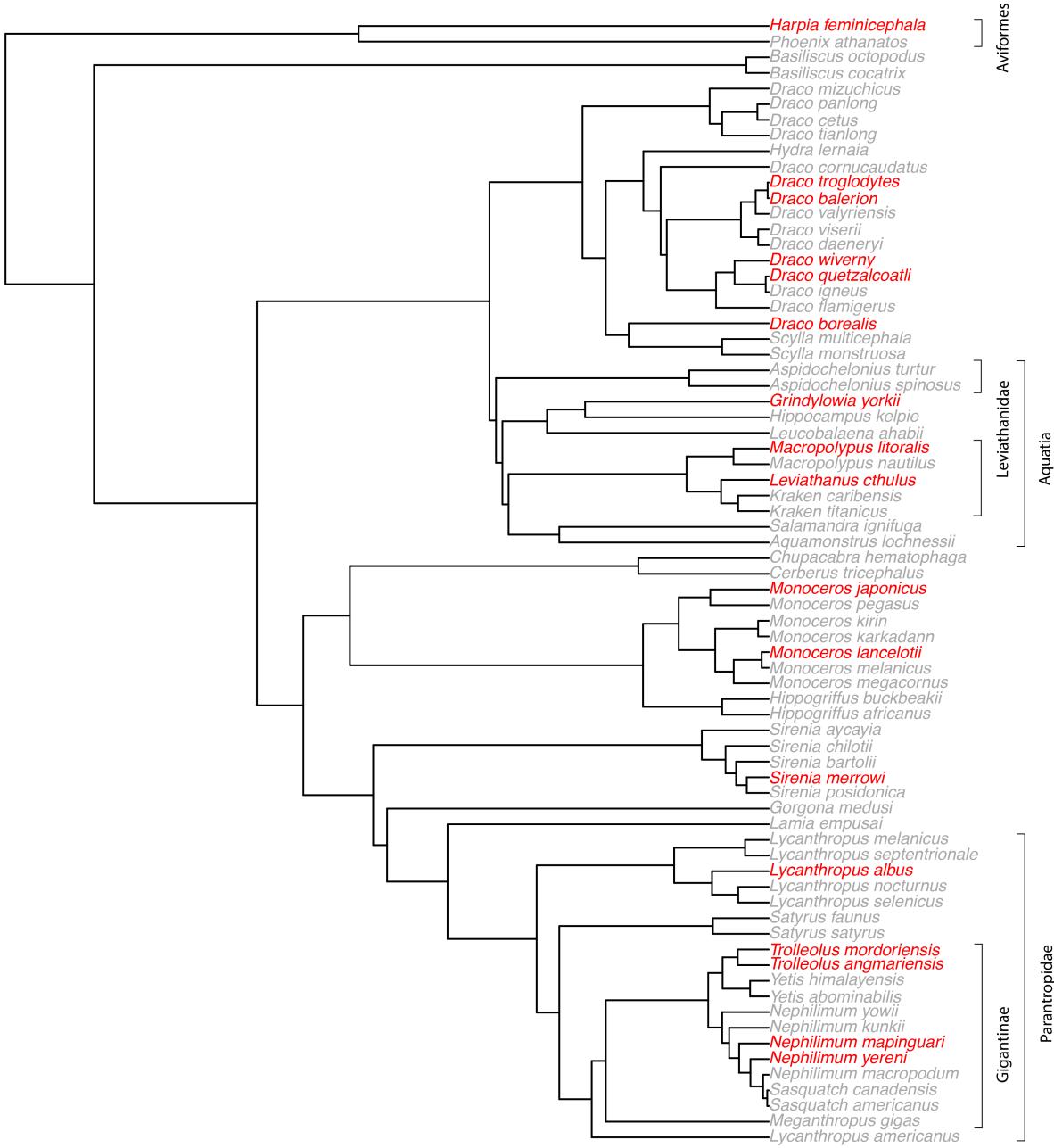


Figure 3: Expanded phylogenetic tree using the ‘backbone’ mode of randtip. Phylogenetic tips in red represent phylogenetically uncertain taxa (PUTs).

A
Most of the PUTs were bound to a randomly selected branch below the crown node of their corresponding genus-level MDCCs.

For example, the congenerics of the PUT *Sirenia merrowi* form a monophyletic group, and thus *Sirenia merrowi* is now placed below the crown node of the group (Figure 3).

In contrast, the congenerics of the PUT *Lycanthropus albus* form a polyphyletic group, and thus it was bound below the crown node of the largest cluster of the genus (default scheme “largest”).

However, the PUTs *Draco borealis*, *Draco troglodytes* and *Draco wiverny* may not necessarily appear bound below the crown node of the largest cluster of *Draco*, as we specifically set the polyphyletic scheme “complete” to bind these PUTs. Thus, they could have been bound to any branch below the node representing the MRCA of all the species in the genus. For example, in this specific simulation *Draco borealis* was bound as sister to *Scylla*.

The congenerics of the PUTs *Nephilimum yereni* and *Nephilimum mapinguari* form a paraphyletic group, and since the argument ‘use.paraphyletic’ of `rand_tip` was set to TRUE (default), none of the branches subtending the species in *Sasquatch* were considered as part of the parameter space for binding these PUTs. Otherwise, *Nephilimum yereni* and *Nephilimum mapinguari* could have been bound as sister to either species of *Sasquatch*, hence breaking the paraphyletic nature of *Nephilimum*.

In a few cases, the genera of the PUTs were missing in the backbone tree, and thus they were bound to supra-generic MDCCs instead. For example, the MDCC of the PUT *Leviathanus cthulus* was the family Leviathanidae, a polyphyletic group. Note that *Leviathanus cthulus* was bound to the largest cluster of Leviathanidae (default scheme) in such a way that the genera *Macropolypus* and *Kraken* remained monophyletic (Figure 3). This is because the argument ‘respect.mono’ of `rand_tip` was set to TRUE (default).

The less inclusive MDCC of the PUT *Harpia feminicephala* was order Aviformes, which is uniquely represented by the species *Phoenix athanatos* in the backbone phylogeny, and thus the PUT was bound as sister to the latter because the argument ‘use.singleton’ of `rand_tip` was set to TRUE (default). Otherwise, the parameter space to bind *Harpia feminicephala* would have been substantially larger (any branch where the insertion of the PUT would not compromise the monophyletic or paraphyletic nature of the taxonomic groups represented in the phylogeny).

The PUTs *Trolleolus mordoriensis* and *Trolleolus angmariensis* were bound below the crown node of the subfamily Gigantinae (which is less inclusive than Parantropidae), and because the argument ‘clump.puts’ of `rand_tip` was set to TRUE (default), the two PUTs appear clumped together forming a monophyletic group.

Finally, *Grindylowia yorkii* was bound below the crown node of the order Aquatia, its less inclusive MDCC in the backbone phylogeny. Again, note that the monophyletic status of the groups within Aquatia (*Aspidochelonius*, *Macropolyplus* and *Kraken*) was respected.

Note that the PUT *Draco balerion* was bound to a branch below the MRCA of the species *Draco valyriensis* and *Draco daenerysi*, as we specifically instructed the software to use an infra-generic MDCC to bind this PUT (Figure 3).

Example 2 - The ‘taxon list’ mode of randtip

Now that we are more familiar with the workflow of `randtip`, we will expand the backbone tree using the ‘taxon list’ mode. Under ‘taxon list’, `randtip` defines MDCCs on the sole basis of taxonomic information of the species provided in the user’s list. This implies shorter execution times, because backbone phylogenies often include thousands of species for which no taxonomic information needs to be retrieved under this mode. However, supra-generic MDCCs may differ between both approaches, which may or may not have an impact on the final tree. Remember that both modes of `randtip` will behave identically whenever the genera

of the PUTs are minimally represented in the backbone phylogeny (see Fig. 3 in main text).

As in the previous example, the first step is building the *info* dataframe.

```
my.info.noranks.list <- build_info(sp.list, tree = NULL,  
                                    find.ranks = FALSE,  
                                    mode = "list")  
  
my.info.noranks.list # print the data frame info on screen
```

The outputted dataframe ('my.info.noranks.list') is identical to that generated in the previous example ('my.info.noranks') with the exception that only the species included in the user's list are displayed. Again, we have instructed `build_info` not to retrieve taxonomic information from the web, and thus the only available information is that corresponding to genus rank. For the purpose of completing this tutorial, an alternative *info* dataframe including taxonomic information for all the species in the list can be loaded into the working space.

```
my.info.list <- mythology$info.list  
my.info.list # print the data frame info on screen
```

Now we can use `check_info`:

```
my.check.list <- check_info(my.info.list, back.tree)
```

Again, we get the same warnings as in the previous example except for that pertaining to the species that was duplicated in the backbone phylogeny, as we already pruned one of the duplicated tips earlier.

A closer look to the outputted dataframe ('my.check.list') reveals that the phyletic status of some groups has changed. For example, the family Leviathanidae is now identified as monophyletic instead of polyphyletic. This is because the dataframe 'my.info.list' only includes taxonomic information for the species in the user's list, and thus supra-generic taxonomic ranks for the species that are represented in the backbone phylogeny but missing in the user's list remain undisclosed.

In this example, the two species that conform the small phylogenetic cluster of Leviathanidae (*Aspidochelonius turtur* and *Aspidochelonius spinosus*) are not included in the user's list, the reason why the group is now identified as monophyletic (Figure 4).

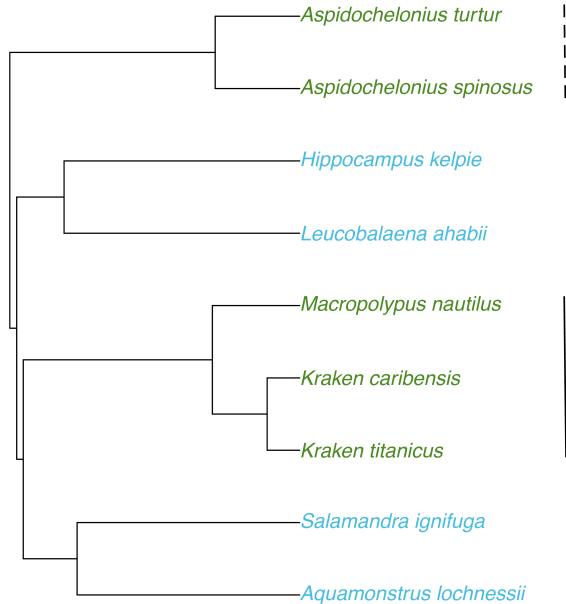


Figure 4: Backbone phylogeny pruned to the less inclusive clade that includes all the species in Leviathanidae (representative species in green). The species next to the solid vertical bar are included in the user's list, whereas those next to the dashed bar are not.

Whether or not the different functioning of the 'backbone' and 'taxon list' modes of randtip will have an impact in the expanded tree will depend on the specifics of each situation. For example, the PUT *Leviathanus cthulus* will always be bound below the crown node of the clade defined by *Kraken* and *Macropolyapus* regardless of the mode of randtip (assuming default settings). This is because under 'backbone' mode, *Leviathanus cthulus* will be bound to the largest cluster of Leviathanidae ($n = 3$ species; Figure 3), which is the only cluster of Leviathanidae that can be identified under 'taxon list' mode (because the two species in the small cluster of Leviathanidae are not included in the user's list; Figure 4). However, it may have happened that other Leviathanidae species not included in the user's list represented the largest cluster of the group in the backbone phylogeny (for example, if the genus *Aspidochelonius* would have been represented by four or more species in the tree), in which case

Leviathanus cthulus would be bound to different clades under ‘backbone’ and ‘taxon list’ modes, respectively.

Manual definition of candidate branches

The clade-based approach of `randtip` should cover most real world situations for PUT binding. Yet, the auxiliary function `custom_branch` allows the user to manually define any subset of candidate branches to bind PUTs.

For example, the phylogenetic place for the PUT of hybrid origin *Monoceros x alaricornus* could be any point along the branches subtending the parental species *Monoceros pegasus* and *Monoceros megacornus*, respectively, and such parameter space cannot be specified by one single clade. To solve this, the user can define the set of candidate branches as an `edges` dataframe. The dataframe `edges` must contain five columns, each row representing a different set of candidate branches for a given PUT. The first column must include the PUT to which the row refers to, the second and third columns are used to set the older node (MRCA of two given species), the fourth and fifth columns refer to the younger one, and all the branches traversed by the spanning path connecting the older and younger nodes are selected as candidate branches (the user can add any number of rows to the `edges` dataframe).

To define a terminal node (phylogenetic tip) as the younger node, the user must fill in the corresponding slots of the fourth and fifth columns with the corresponding species. Inserting the same species in the four slots will make the PUT to be bound as sister to this species. Finally, in case the same pair of species is set for columns 2-3 and 4-5 within a row, all branches below the MRCA of the two species will be defined as candidates.

In order to ensure that candidate branches have been correctly encoded in `edges`, the user can use the auxiliary function `plot_custom_branch` to visually explore the selected space of branch lengths.

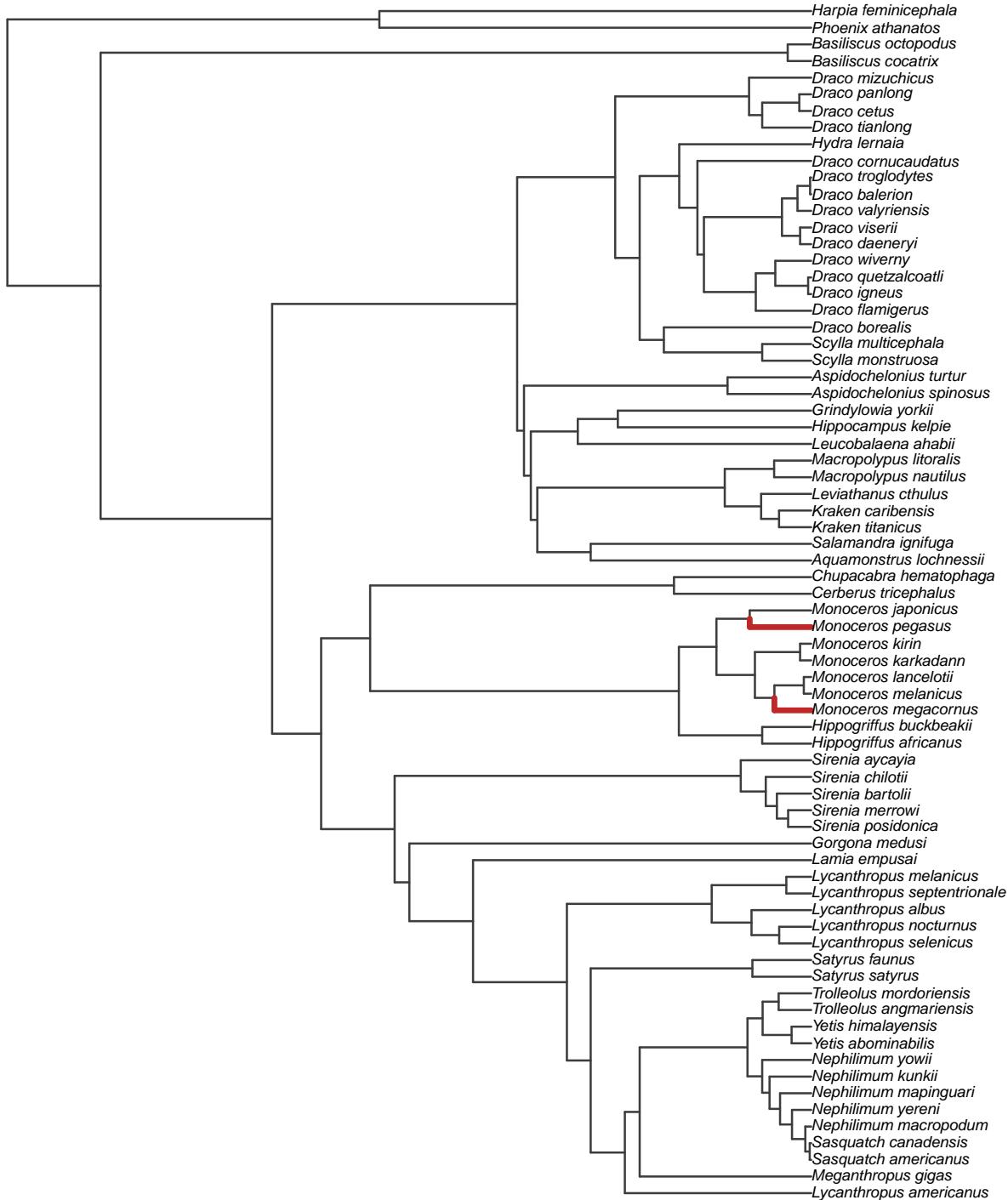
A

```

edges <- mythology$edges

plot_custom_branch(new.tree, edges, cex = 0.7,
candidate.lwd = 4, forbidden.lwd = 1.5)

```



A

Figure 5: Phylogeny with candidate branches for the insertion of *Monocerus x alaricor-nus* highlighted in red.

And now we can further expand the phylogeny that was generated earlier.

```
new.tree2 <- custom_branch(new.tree, edges, forceultrametric = TRUE)  
plot.phylo(new.tree2, cex = 0.7)
```

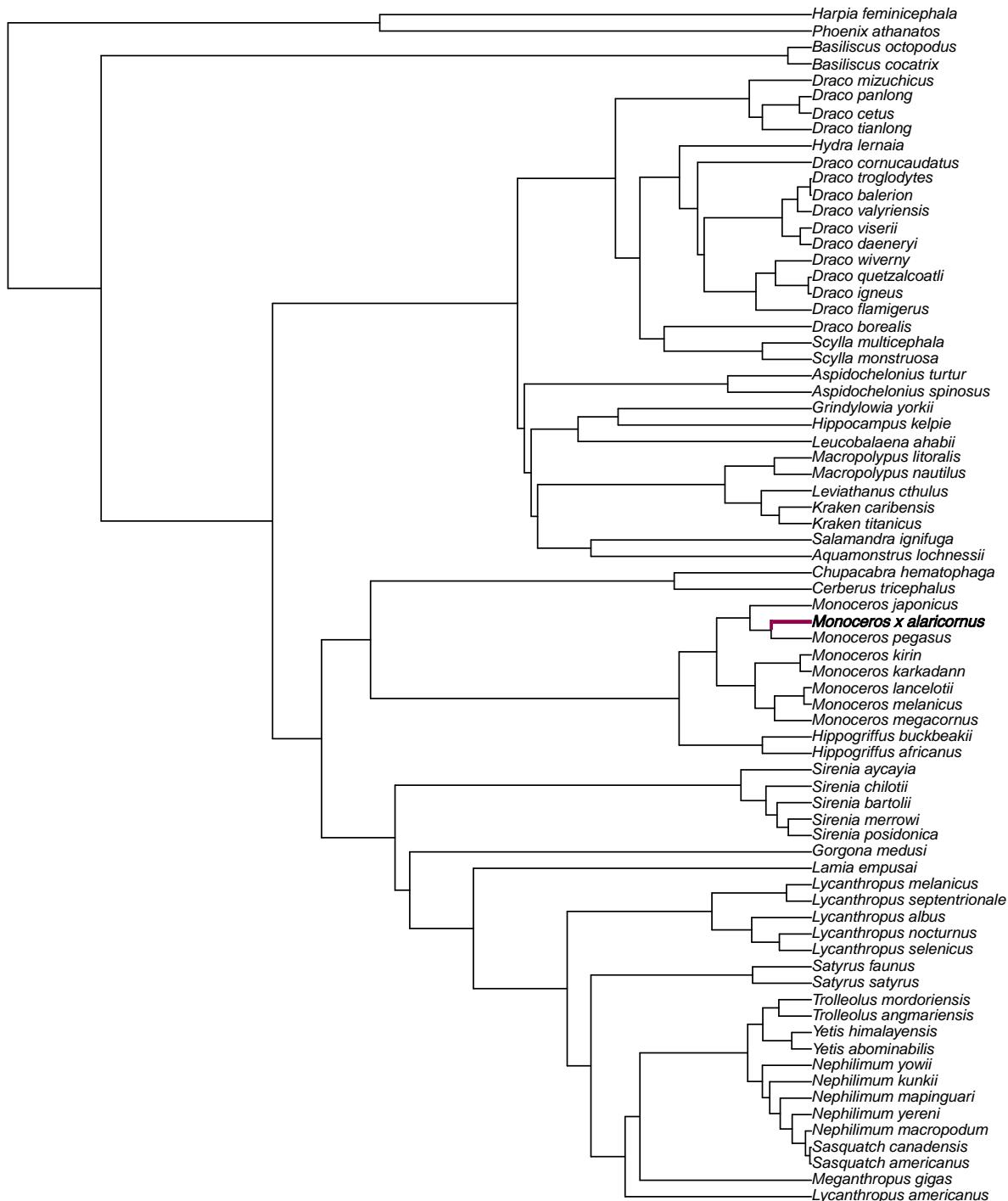


Figure 6: Phylogeny after the insertion of *Monoceros x alaricornus*.

References

- Rabosky, D.L., Chang, J., Title, P.O. et al. (2018). An inverse latitudinal gradient in speciation rate for marine fishes. *Nature*, **559**, 392–395.
- Huson, D.H. & Scornavacca, C. (2012). Dendroscope 3: An interactive tool for rooted phylogenetic trees and networks. *Systematic biology*, **61**, 1061–1067.
- Jetz, W. & Pyron, R.A. (2018). The interplay of past diversification and evolutionary isolation with present imperilment across the amphibian tree of life. *Nature Ecology & Evolution*, **2**, 850–858.
- Kimball, R.T., Oliveros, C.H., Wang, N. et al. (2019). A phylogenomic supertree of birds. *Diversity*, **11**, 109.
- Kindt, R. (2020). WorldFlora: An r package for exact and fuzzy matching of plant names against the world flora online taxonomic backbone database. *Applications in Plant Sciences*, **8**, e11388.
- Paradis, E. & Schliep, K. (2019). Ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, **35**, 526–528.
- Pyron, R.A., Burbrink, F.T. & Wiens, J.J. (2013). A phylogeny and revised classification of squamata, including 4161 species of lizards and snakes. *BMC Evolutionary Biology*, **13**, 93.
- Smith, S.A. & Brown, J.W. (2018). Constructing a broadly inclusive seed plant phylogeny. *American Journal of Botany*, **105**, 302–314.
- Upham, N.S., Esselstyn, J.A. & Jetz, W. (2019). Inferring the mammal tree: Species-level sets of phylogenies for questions in ecology, evolution, and conservation. *PLoS Biology*, **17**, e3000494.
- Wickham, H., Hester, J. & Chang, W. (2021). Devtools: Tools to make developing r packages easier. R package version 2.4.5
- Zanne, A.E., Tank, D.C., Cornwell, W.K. et al. (2014). Three keys to the radiation of angiosperms into freezing environments. *American Journal of Botany*, **506**, 89–92.

Zhang, J. & Qian, H. (2023). U.taxonstand: An r package for standardizing scientific names of plants and animals. *Plant Diversity*, **45**, 1–5

A

A

Table S3.1. Description of the functions implemented in the ‘randtip’ R package.

Function	Description	Fixed arguments	Variable arguments	Options (first as default)	Description of variable arguments
<i>build_info</i>	Creates the <i>info</i> template and retrieves taxonomic information from web repositories	<ul style="list-style-type: none"> • species = character vector or single-column dataframe 	tree	NULL, phylo object	A backbone tree of class <i>phylo</i> must be provided if using randtip in “backbone” mode
			find.ranks	TRUE, FALSE	Retrieve taxonomic information of the species in <i>info</i> that will be used to identify supra-generic MDCCs for the PUTs
			db	“ncbi”, “itis”, “gbif”, “bold”	Taxonomic database from which information is retrieved
			mode	“backbone”, “list”	Whether or not the species in the backbone tree are to be included in <i>info</i>
			interactive	FALSE, TRUE	By default, ambiguous names will make <i>info</i> to be filled in automatically with ‘NA’ when retrieving taxonomic information. Alternatively, ambiguities can be resolved manually by the user
			genus	FALSE, TRUE	Whether or not a genus-level backbone tree is to be expanded with PUTs representing genera
			prior.info	NULL, dataframe	A previously created <i>info</i> dataframe whose information will be used to create the final <i>info</i>
			verbose	TRUE, FALSE	Whether or not to print information on the flow of the function. If interactive is TRUE, this argument will be ignored
<i>check_info</i>	Informs on the PUT status of the species in <i>info</i> , spelling errors, putative MDCCs and the phyletic nature of groups of PPCR species	<ul style="list-style-type: none"> • info = dataframe <i>info</i> • tree = backbone phylogeny 	sim	numeric (0-1)	Similarity threshold to detect possible misspelling errors (default is 0.85)
			find.phylicity	TRUE, FALSE	Whether or not to assess the phyletic nature of PPCR species
			search.typos	TRUE, FALSE	Whether or not to assess misspelling errors
			verbose	TRUE, FALSE	Whether or not to print information on the flow of the function.
			parallelize	TRUE, FALSE	Whether or not to make use of parallel processing
			ncores	NULL	Number of cores for parallel processing. If NULL, all available cores minus one will be used
<i>edit_info</i>	Auxiliary function to easily edit the	• info = dataframe <i>info</i>	column	string (colname of <i>info</i>), NULL	Column of <i>info</i> that is to be edited

	<i>info</i> dataframe	<ul style="list-style-type: none"> • taxa = character vector with the PUTs to be edited or removed 	edit	NULL, string or logical	Any allowed value for the column of <i>info</i> that is to be edited
			remove.rows	FALSE, TRUE	Whether or not to remove the corresponding rows of <i>info</i>
<i>edit_tree</i>	Auxiliary function to easily edit the tip labels of the backbone tree	<ul style="list-style-type: none"> • tree = backbone phylogeny • tips = character vector with the tips to be edited or pruned 	edit	NULL, character vector	New labels for the specified phylogenetic tips
			remove.tips	FALSE, TRUE	Whether or not to prune the corresponding tips of the backbone tree
<i>get_clade</i>	Extracts any subtree representing a putative MDCC	<ul style="list-style-type: none"> • info = dataframe <i>info</i> • tree = backbone phylogeny 	group	string (taxonomic group)	MDCC to extract, defined as the MRCA of all the species in the given taxonomic group
<i>plot_clade</i>	Plots the clade extracted with <i>get_clade</i> and highlights PPCR species in the corresponding taxonomic group	<ul style="list-style-type: none"> • get.clade.out = output of <i>get_clade</i> function 	ppcr.col	string (color)	Color for tip labels corresponding to the species in the taxonomic group indicated in <i>get_clade</i> (default is “green”)
			nonppcr.col	string (color)	Color for tip labels corresponding to the species not included in the taxonomic group indicated in <i>get_clade</i> (default is “blue”)
			unknown.col	string (color)	Color for tip labels corresponding to the species whose taxonomic group is unknown or missing (default is “grey”)
			More plotting arguments passed to the <i>plot.phylo</i> function of ‘ape’ R package
<i>info2input</i>	Wrapping function to create the input dataframe for <i>rand_tip</i>	<ul style="list-style-type: none"> • info = dataframe <i>info</i> • tree = backbone phylogeny 	parallelize	TRUE, FALSE	Whether or not to make use of parallel processing
			ncores	NULL	Number of cores for parallel processing. If NULL, all available cores minus one will be used
			verbose	TRUE, FALSE	Whether or not to print information on the flow of the function

			rand.type	"random" or "r", "polytomy" or "p"	Defines whether PUTs will be bound to randomly selected branches below the crown node of their selected MDCCs or as polytomies instead
			polyphyly.scheme	"largest" or "l", "frequentist" or "f", "complete" or "c"	Defines the procedure to bind PUTs whenever their PPCR species form polyphyletic groups
			use.paraphyletic	TRUE, FALSE	If TRUE, branches constituting monophyletic or singleton groups of non-PPCR species within the MDCCs of the PUTs are prohibited
			use.singleton	TRUE, FALSE	If TRUE, PUTs are bound to terminal branches whenever their PPCR species form singleton clades in the phylogeny. Otherwise, they are bound to any branch below the parent node of the singleton PPCR species
			use.stem	FALSE, TRUE	Whether or not including the stem branch of the MDCCs as candidate branches
			respect.mono	TRUE, FALSE	Whether or not to respect monophyletic groups when binding PUTs above genus level
			respect.para	TRUE, FALSE	Whether or not to respect paraphyletic groups when binding PUTs above genus level
			clump.puts	TRUE, FALSE	Whether or not co-ranked PUTs (e.g. congeners, confamilials) should be clumped together in the phylogeny in case their taxonomic group (e.g. genus, family) is missing in the tree. Will also clump conspecific PUTs (subspecies) in case none of them are represented in the phylogeny
			prob	TRUE, FALSE	Whether branch selection probability must be proportional to branch length or equiprobable
			prune	TRUE, FALSE	Whether or not the newly expanded tree will include only the species in the user's list
			forceultrametric	FALSE, TRUE	Whether or not the backbone tree will be forced to be ultrametric following the nmls method as implemented in 'phytools' R package
			verbose	TRUE, FALSE	Whether or not to print information on the flow of the function

<i>put_tip_col</i>	Set color pattern to distinguish between phylogenetically placed taxa and PUTs when plotting trees with the <i>plot.phylo</i> function of ‘ape’ R package using the “tip.color” argument	<ul style="list-style-type: none"> • newtree = expanded phylogeny • oldtree = backbone phylogeny 	placed.col	string (color)	Color for tip labels representing taxa present in the phylogeny (default is “grey”)
			put.col	string (color)	Color for tip labels representing PUTs (default is “red”)
<i>custom_branch</i>	Binds PUTs across customized segments of the phylogeny	<ul style="list-style-type: none"> • tree = backbone phylogeny • edges = dataframe <i>edges</i> 	rand.type	“random” or “r”, “polytomy” or “p”	Defines whether the PUT will be bound to the randomly selected branch or as a polytomy in the parent node of the selected branch
			forceultrametric	FALSE, TRUE	Whether or not the backbone tree will be forced to be ultrametric following the <i>nmls</i> method as implemented in ‘phytools’ R package
			prob	TRUE, FALSE	Whether branch selection probability must be proportional to branch length or equiprobable
<i>plot_custom_branch</i>	Auxiliary function for <i>custom_branch</i> to visualize candidate branches	<ul style="list-style-type: none"> • tree = backbone phylogeny • edges = dataframe <i>edges</i> 	PUT	NULL, string	Only required for <i>edges</i> dataframes that include more than one PUT. Specifies which PUT’s candidate branches must be highlighted.
			candidate.col	string (color)	Colour of candidate branches (default is “green”)
			forbidden.col	string (color)	Colour of non-candidate branches (default is “grey”)
			candidate.lwd	numeric	Width of candidate branches (default is 2)
			forbidden.lwd	numeric	Width of non-candidate branches (default is 1)
			More plotting arguments passed to the <i>plot.phylo</i> function of ‘ape’ R package

Appendix 4a. Supplementary Tables

Table S4.1. Variables used for the Linear Descriptor Analyses calculated for each grid cell, description of the variable name and origin of the data.

Variable name	Description	Origin of the data
<i>alt.mn</i>	Mean altitude	EU-DEM (raster) - version 1.1, Apr. 2016
<i>alt.heter</i>	Maximum altitude – minimum altitude	EU-DEM (raster) - version 1.1, Apr. 2016
<i>ph.mn</i>	Mean pH value	Soilgrids database; https://doi.org/10.1371/journal.pone.0169748
<i>cont</i>	Continentiality index (bio5*0.1 - bio6*0.1)	CHELSA database; https://doi.org/10.1038/sdata.2017.122
<i>bio1.mn</i>	Annual Mean Temperature	CHELSA database; https://doi.org/10.1038/sdata.2017.122
<i>bio4.mn</i>	Temperature Seasonality (standard deviation)	CHELSA database; https://doi.org/10.1038/sdata.2017.122
<i>bio12.mn</i>	Annual Precipitation (mm/year)	CHELSA database; https://doi.org/10.1038/sdata.2017.122
<i>bio15.mn</i>	Precipitation Seasonality (coefficient of variation)	CHELSA database; https://doi.org/10.1038/sdata.2017.122
<i>isolation</i>	Minimum distance to mainland (only for islands)	<i>Ad-hoc</i> calculation

Table S4.2. Optimum number of clusters (k) found for each bioregion analysis, varying its species of study (endemic taxa *vs* whole flora), dissimilarity calculation index (taxonomic *vs* phylogenetic), and spatial resolution used (10, 20 or 50 km side UTM squares). An optimum number of clusters was not found for analyses using the whole flora at a 10 km resolution.

Spatial resolution	Similitude index	Taxa	k
10 km	Taxonomic	Endemic taxa	40
		All taxa	Not Found
	Phylogenetic	Endemic taxa	11
		All taxa	Not Found
20 km	Taxonomic	Endemic taxa	8
		All taxa	4
	Phylogenetic	Endemic taxa	10
		All taxa	11
50 km	Taxonomic	Endemic taxa	8
		All taxa	13
	Phylogenetic	Endemic taxa	13
		All taxa	14

Table S4.3. Results of the LDA analysis for each regionalization cluster split. Values shown in columns “AUC” and “Variable importance” are the averaged standardized values of 100 individual analyses. The names of the variables (as developed in Supplementary Table S4.1) stand for the following terms. *alt.heter*: altitude heterogeneity; *ph.mn*: mean pH value; *bio1.mn*: mean annual temperature; *bio4.mn*: mean temperature seasonality; *bio12.mn*: annual precipitation; *bio15.mn*: precipitation seasonality; *isolation*: distance of an island from mainland.

Regionalization	clusters	Mean AUC	Variable	Variable importance (mean± sd)
Taxonomic endemics	8 vs rest	0.987	isolation	-3.46±0.54
			bio4.mn	0.112±0.016
			bio12.mn	0.155±0.052
			bio1.mn	-0.06±0.015
			bio15.mn	0.041±0.014
			alt.heter	0.002±0.012
			ph.mn	0.059±0.044
	2,5,7 vs 1,3,4,6	0.909	bio12.mn	1.131±0.048
			bio15.mn	-0.936±0.034
			bio4.mn	-0.545±0.021
			alt.heter	0.32±0.027
			ph.mn	-0.078±0.041
			bio1.mn	0.1±0.049
			isolation	-
	1,3 vs 4,6	0.912	ph.mn	2.276±0.06
			alt.heter	0.642±0.036
			bio12.mn	0.541±0.068
			bio4.mn	-0.478±0.033
			bio15.mn	-0.467±0.055
			bio1.mn	-0.2±0.058
			isolation	-
	7 vs 2,5	0.938	bio4.mn	-1.123±0.107
			bio15.mn	0.649±0.065
			bio1.mn	-0.551±0.094
			alt.heter	-0.449±0.072
			bio12.mn	0.124±0.103
			ph.mn	0.152±0.107
			isolation	-

Table S4.3 (cont.)

Regionalization	clusters	Mean AUC	Variable	Variable importance (mean± sd)
Taxonomic whole flora	4 vs 1,2,3	0.732	isolation	-1.187±0.06
			bio1.mn	-0.689±0.047
			bio12.mn	0.623±0.042
			bio4.mn	0.56±0.031
			alt.heter	-0.514±0.041
			bio15.mn	0.382±0.032
			ph.mn	-0.029±0.029
	2 vs 1,3	0.894	ph.mn	0.996±0.043
			bio12.mn	-0.725±0.053
			bio1.mn	0.332±0.041
			bio4.mn	-0.135±0.029
			bio15.mn	0.105±0.032
			alt.heter	-0.049±0.028
			isolation	-
	1 vs 3	0.961	bio15.mn	-1.351±0.046
			ph.mn	1.212±0.05
			bio1.mn	-0.514±0.049
			bio4.mn	-0.446±0.036
			alt.heter	0.433±0.033
			bio12.mn	-0.243±0.054
			isolation	-

A

Table S4.3 (cont)

Regionalization	clusters	Mean AUC	Variable	Variable importance (mean± sd)
Phylogenetic endemics	10 vs rest	1.000	isolation	-3.155±0.453
			bio12.mn	-0.244±0.032
			bio4.mn	-0.101±0.023
			ph.mn	-0.168±0.041
			bio1.mn	0.052±0.015
			bio15.mn	-0.044±0.016
			alt.heter	-0.002±0.009
	1,4,2,5 vs 9,3,6	0.5	ph.mn	-1.185±0.073
			bio12.mn	-0.721±0.068
			bio1.mn	0.538±0.047
			bio4.mn	0.522±0.064
			isolation	-0.4±0.051
			bio15.mn	0.329±0.053
			alt.heter	-0.184±0.034
	1,4 vs 2,5	0.893	ph.mn	1.811±0.044
			bio15.mn	-1.099±0.036
			isolation	0.668±0.062
			alt.heter	0.62±0.029
			bio12.mn	0.635±0.051
			bio4.mn	-0.528±0.028
			bio1.mn	0.19±0.042
	1 vs 4	0.927	bio4.mn	1.513±0.045
			bio15.mn	-1.366±0.07
			bio12.mn	0.998±0.061
			ph.mn	-0.957±0.076
			alt.heter	-0.19±0.042
			bio1.mn	0.161±0.06
			isolation	-
	2 vs 5	0.903	bio12.mn	-0.718±0.087
			bio15.mn	0.766±0.045
			ph.mn	0.703±0.118
			bio4.mn	0.344±0.067
			alt.heter	-0.28±0.037
			bio1.mn	-0.19±0.063
			isolation	0.067±0.035

Table S4.3 (cont)

Regionalization	clusters	Mean AUC	Variable	Variable importance (mean± sd)
Phylogenetic whole flora	3,8,11,6,5,9 vs 10,7,4,1,2	0.908	bio1.mn	0.943±0.048
			bio12.mn	-0.721±0.043
			bio15.mn	0.57±0.042
			ph.mn	-0.502±0.052
			bio4.mn	-0.19±0.037
			isolation	0.153±0.01
			alt.heter	0.147±0.028
	3,8 vs 11,6,5,9	0.954	ph.mn	2.424±0.075
			bio4.mn	-0.444±0.052
			bio12.mn	0.213±0.058
			bio1.mn	-0.216±0.044
			bio15.mn	-0.12±0.04
			alt.heter	-0.016±0.03
			isolation	-
	1 vs 2	0.895	bio1.mn	1.837±0.12
			bio12.mn	-1.102±0.143
			bio15.mn	-1.308±0.104
			bio4.mn	-0.951±0.038
			alt.heter	0.877±0.074
			ph.mn	0.468±0.065
			isolation	-

A

Appendix 4b. Supplementary Figures

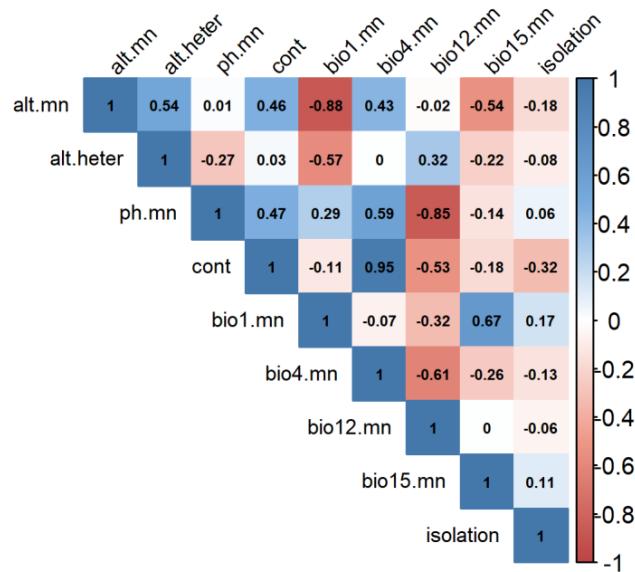


Figure S4.1. Correlation plot between topographical, biogeochemical and climatic variables. The names of the variables (as developed in Supplementary Table S4.1) stand for the following terms. *alt.mn*: average altitude of the grid cell; *alt.heter*: altitude heterogeneity; *ph.mn*: mean pH value; *cont*: continentality index; *bio1.mn*: mean annual temperature; *bio4.mn*: mean temperature seasonality; *bio12.mn*: annual precipitation; *bio15.mn*: precipitation seasonality; *isolation*: distance of an island from mainland.

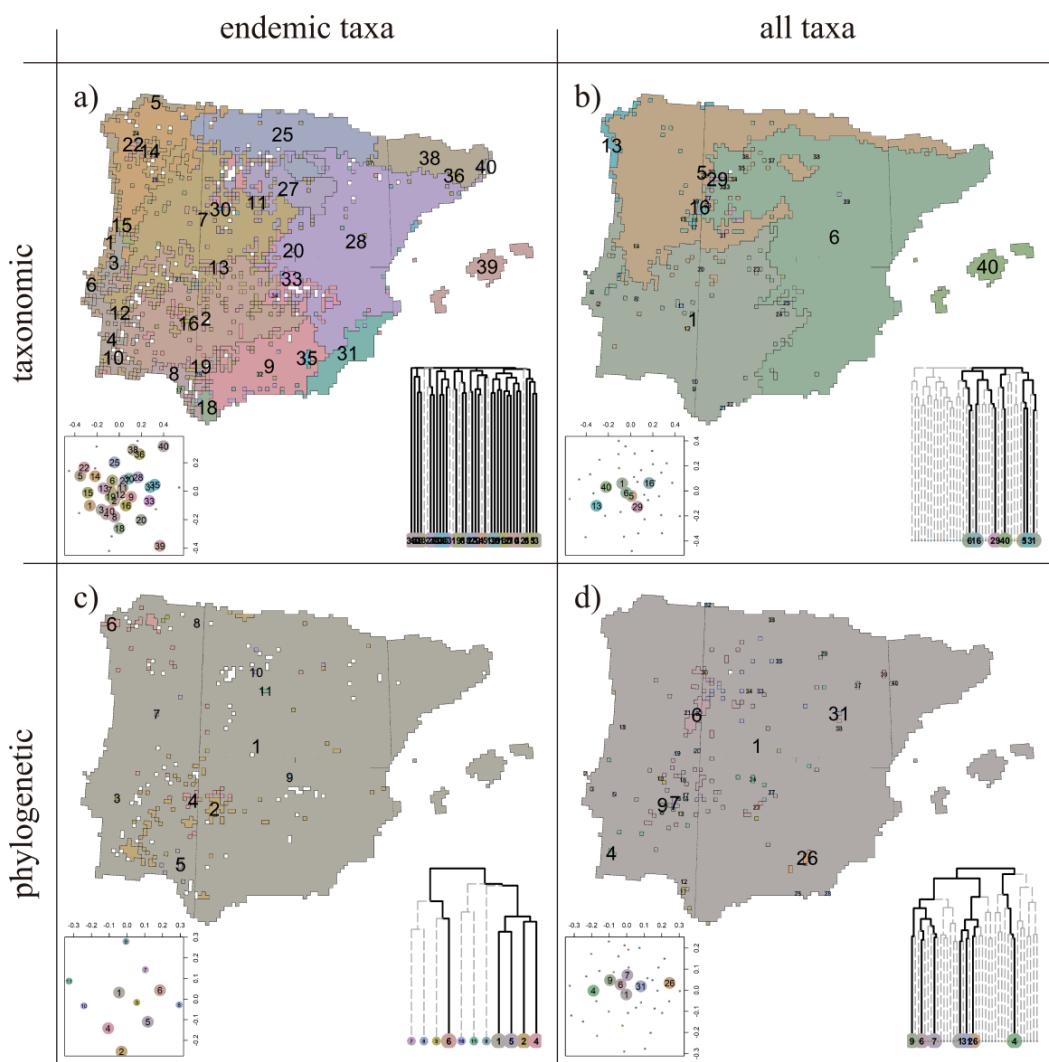


Figure S4.2. Delimitation of floristic bioregions in the Iberian Peninsula at a 10x10 resolution attending to their endemic species (subfigures a & c, on the left) and using the whole flora (b & d, on the right), performed using taxonomic (a & b, on the top) and phylogenetic (c & d, on the bottom) approaches. The optimal number of clusters found were 40 (a) and 11 (c). For figures (b) and (d) an optimum number was not found, and were analysed using $k=40$. Colours match within each panel in the map, NMDS plot and dendrogram, and are assigned using *phyloregion*'s package "NMDS" palette.

A

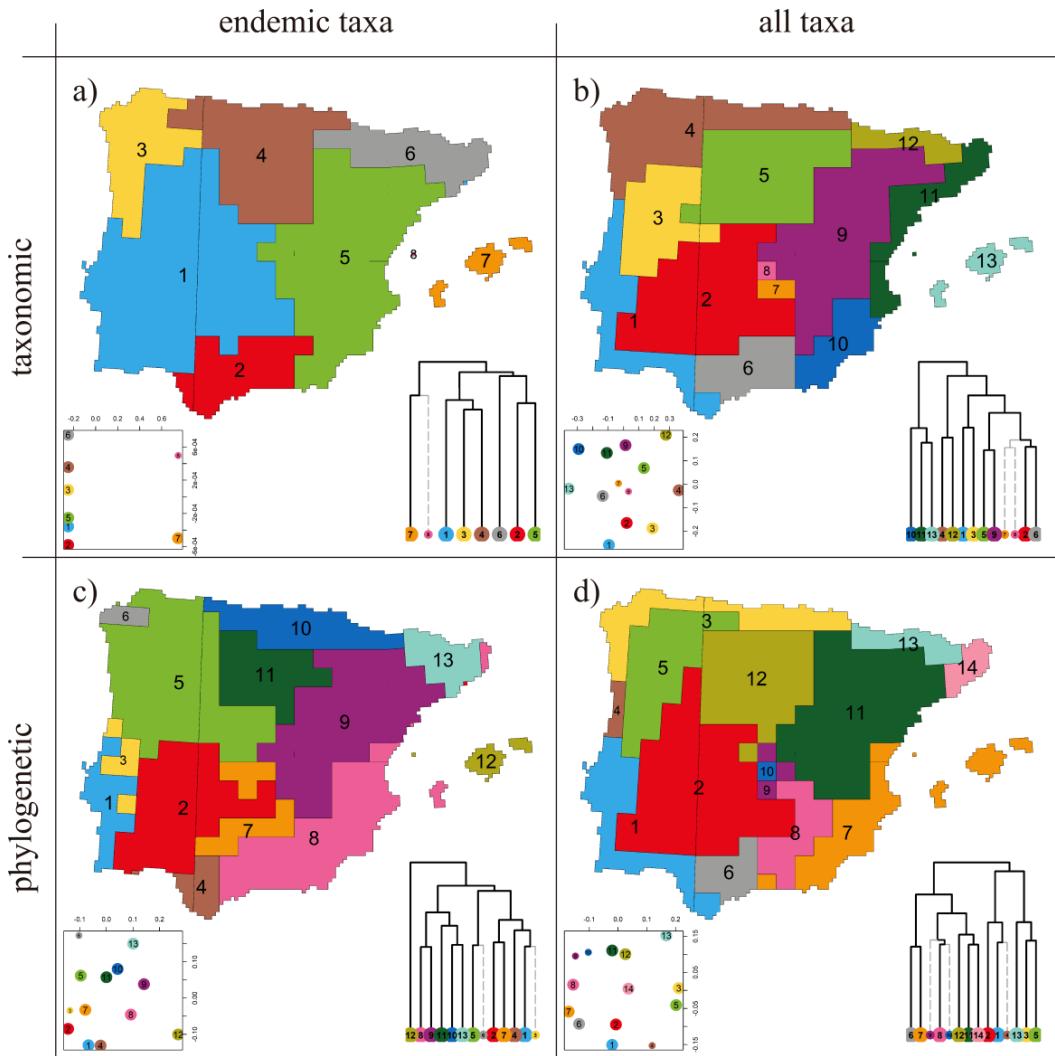


Figure S4.3. Delimitation of floristic bioregions in the Iberian Peninsula at a 50x50 resolution attending to their endemic species (subfigures a & c, on the left) and using the whole flora (b & d, on the right), performed using taxonomic (a & b, on the top) and phylogenetic (c & d, on the bottom) approaches. The optimal number of clusters found were 8 (a), 13 (b), 13 (c) and 14 (d). Colours match within each panel in the map, NMDS plot and dendrogram, and are assigned using *MetroBrewer's* package “Madrid” palette.

A

Appendix 4c. Supplementary Data

Specific parameters for PUT binding, following *randtip* package pipeline, can be found in the following dataset.

https://github.com/iramosgutierrez/PhD_code/blob/main/3.%20Iberian%20Bioregions/IberianChecklist_input.csv

Default values, whenever not specified else in the *input* table, were the following:

- `polyphyly.scheme = "largest"`
- `use.paraphyletic = TRUE`
- `use.singleton = TRUE`
- `use.stem = FALSE`
- `respect.mono = TRUE`
- `respect.para = TRUE`
- `clump.puts = TRUE`
- `prob = TRUE`

Appendix 5a. Supplementary Figures

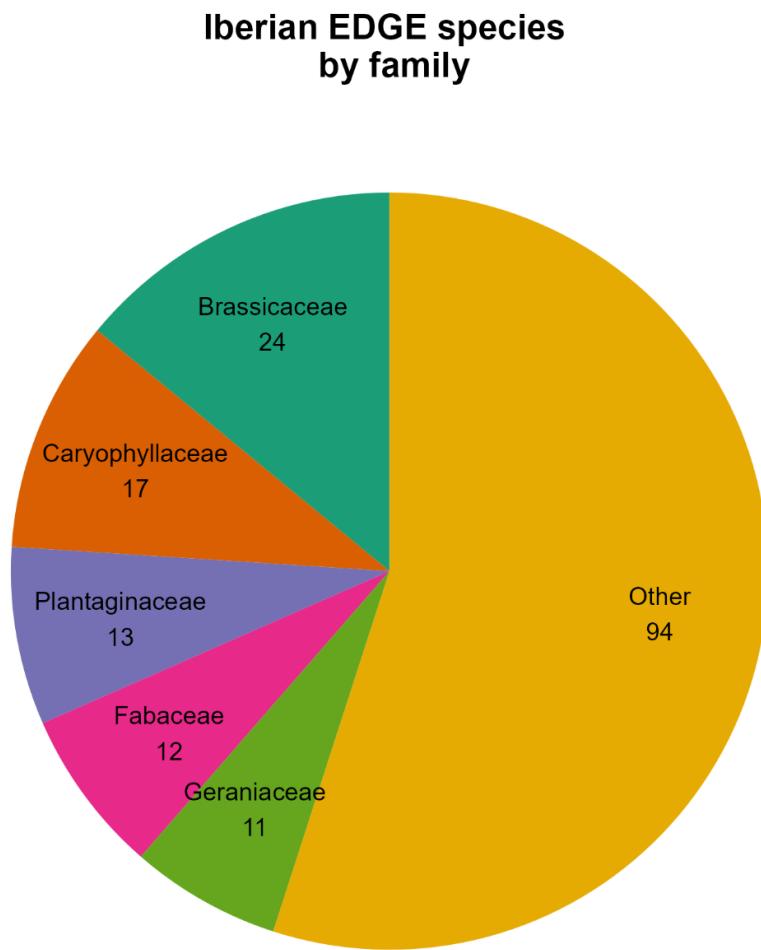


Figure S5.1. Proportion of Iberian EDGE species by family. Only the top 5 families (of a total number of 41) are shown, as they are the only ones containing more than 10 EDGE species.

A

Appendix 5b. Supplementary Tables

Table S5.1. List of EDGE species and family to which they belong.

Family	Species	Family	Species
Alismataceae	<i>Damasonium polyspermum</i>	Brassicaceae	<i>Noccaea nevadensis</i>
Amaranthaceae	<i>Patellifolia procumbens</i>	Brassicaceae	<i>Sisymbrium hispanicum</i>
Amaryllidaceae	<i>Acis valentina</i>	Brassicaceae	<i>Sisymbrium isatidifolium</i>
Apiaceae	<i>Eryngium galiooides</i>	Brassicaceae	<i>Vella bourgaeana</i>
Apiaceae	<i>Eryngium grossii</i>	Brassicaceae	<i>Vella castrilensis</i>
Apiaceae	<i>Eryngium huteri</i>	Brassicaceae	<i>Vella lucentina</i>
Apiaceae	<i>Eryngium tricuspidatum</i>	Brassicaceae	<i>Vella spinosa</i>
Apiaceae	<i>Eryngium viviparum</i>	Campanulaceae	<i>Campanula arvatica</i>
Apiaceae	<i>Ferulago brachyloba</i>	Caprifoliaceae	<i>Pseudoscabiosa saxatilis</i>
Apiaceae	<i>Naufraga balearica</i>	Caprifoliaceae	<i>Pterocephalus spathulatus</i>
Apiaceae	<i>Thapsia gymnesica</i>	Caprifoliaceae	<i>Succisella andreae-molinae</i>
Araceae	<i>Biarum carratracense</i>	Caprifoliaceae	<i>Succisella carvalhoana</i>
Asparagaceae	<i>Brimeura duvigneaudii</i>	Caprifoliaceae	<i>Succisella microcephala</i>
Asparagaceae	<i>Muscaria cazorlanum</i>	Caprifoliaceae	<i>Valeriana longiflora</i>
Asparagaceae	<i>Muscaria olivetorum</i>	Caryophyllaceae	<i>Arenaria fontqueri</i>
Asteraceae	<i>Centaurea lainzii</i>	Caryophyllaceae	<i>Arenaria funiculata</i>
Asteraceae	<i>Crepis triasii</i>	Caryophyllaceae	<i>Arenaria nevadensis</i>
Asteraceae	<i>Ptilostemon hispanicus</i>	Caryophyllaceae	<i>Arenaria tejedensis</i>
Asteraceae	<i>Scorzoneroides microcephala</i>	Caryophyllaceae	<i>Gypsophila bermejoi</i>
Boraginaceae	<i>Glandora nitida</i>	Caryophyllaceae	<i>Gypsophila tomentosa</i>
Boraginaceae	<i>Glandora oleifolia</i>	Caryophyllaceae	<i>Herniaria baetica</i>
Boraginaceae	<i>Gyrocarpum oppositifolium</i>	Caryophyllaceae	<i>Moehringia castellana</i>
Boraginaceae	<i>Solenanthus reverchonii</i>	Caryophyllaceae	<i>Petrocoptis crassifolia</i>
Brassicaceae	<i>Biscutella atropurpurea</i>	Caryophyllaceae	<i>Petrocoptis grandiflora</i>
Brassicaceae	<i>Biscutella bilbilitana</i>	Caryophyllaceae	<i>Petrocoptis guarensis</i>
Brassicaceae	<i>Biscutella conquensis</i>	Caryophyllaceae	<i>Petrocoptis montserratii</i>
Brassicaceae	<i>Biscutella dufourii</i>	Caryophyllaceae	<i>Petrocoptis pardoii</i>
Brassicaceae	<i>Biscutella fontqueri</i>	Caryophyllaceae	<i>Petrocoptis pseudoviscosa</i>
Brassicaceae	<i>Biscutella glacialis</i>	Caryophyllaceae	<i>Silene acutifolia</i>
Brassicaceae	<i>Biscutella marinae</i>	Caryophyllaceae	<i>Silene foetida</i>
Brassicaceae	<i>Biscutella segurae</i>	Caryophyllaceae	<i>Spergula viscosa</i>
Brassicaceae	<i>Coincyia longirostra</i>	Colchicaceae	<i>Colchicum androcymbioides</i>
Brassicaceae	<i>Coincyia monensis</i>	Convolvulaceae	<i>Convolvulus boissieri</i>
Brassicaceae	<i>Coincyia rupestris</i>	Crassulaceae	<i>Sedum lagascae</i>
Brassicaceae	<i>Draba dedeana</i>	Dioscoreaceae	<i>Dioscorea chouardii</i>
Brassicaceae	<i>Hormathophylla cadevalliana</i>	Fabaceae	<i>Adenocarpus hispanicus</i>
Brassicaceae	<i>Ionopsisidium aragonense</i>	Fabaceae	<i>Anthyllis ramburei</i>
Brassicaceae	<i>Lepidium navasii</i>	Fabaceae	<i>Galega cirujanoi</i>
Brassicaceae	<i>Moricandia foetida</i>	Fabaceae	<i>Genista ancistrocarpa</i>
Brassicaceae	<i>Moricandia moricandioides</i>	Fabaceae	<i>Genista carpetana</i>

Table S5.1. List of EDGE species and family to which they belong (cont).

Family	Species	Family	Species
Fabaceae	<i>Genista haensleri</i>	Plantaginaceae	<i>Lafuentea rotundifolia</i>
Fabaceae	<i>Genista legionensis</i>	Plantaginaceae	<i>Sibthorpia africana</i>
Fabaceae	<i>Genista monspessulana</i>	Plantaginaceae	<i>Veronica mampodrensis</i>
Fabaceae	<i>Lotus angustissimus</i>	Plumbaginaceae	<i>Limonium caesium</i>
Fabaceae	<i>Ononis azcaratei</i>	Poaceae	<i>Festuca rothmaleri</i>
Fabaceae	<i>Ononis broteroana</i>	Poaceae	<i>Macrochloa tenacissima</i>
Fabaceae	<i>Ononis varelae</i>	Polygonaceae	<i>Polygaloides vayredae</i>
Geraniaceae	<i>Erodium aguilellae</i>	Potamogetonaceae	<i>Zannichellia contorta</i>
Geraniaceae	<i>Erodium astragaloides</i>	Primulaceae	<i>Androsace rioxana</i>
Geraniaceae	<i>Erodium boissieri</i>	Primulaceae	<i>Soldanella villosa</i>
Geraniaceae	<i>Erodium carvifolium</i>	Ranunculaceae	<i>Ranunculus acetosellifolius</i>
Geraniaceae	<i>Erodium cazorlanum</i>	Resedaceae	<i>Sesamoides minor</i>
Geraniaceae	<i>Erodium daucooides</i>	Rosaceae	<i>Potentilla nevadensis</i>
Geraniaceae	<i>Erodium manescavi</i>	Rosaceae	<i>Potentilla reuteri</i>
Geraniaceae	<i>Erodium recoderi</i>	Rosaceae	<i>Rubus castroviejoi</i>
Geraniaceae	<i>Erodium rupicola</i>	Rosaceae	<i>Rubus galloecicus</i>
Geraniaceae	<i>Erodium sanguis-christi</i>	Rosaceae	<i>Rubus muricola</i>
Geraniaceae	<i>Geranium cataractarum</i>	Rosaceae	<i>Rubus urbionicus</i>
Hypericaceae	<i>Hypericum caprifolium</i>	Rosaceae	<i>Sanguisorba lateriflora</i>
Hypericaceae	<i>Hypericum ericoides</i>	Rutaceae	<i>Cneorum tricoccon</i>
Iridaceae	<i>Iris boissieri</i>	Rutaceae	<i>Haplophyllum linifolium</i>
Linaceae	<i>Linum appressum</i>	Saxifragaceae	<i>Saxifraga biternata</i>
Linaceae	<i>Linum maritimum</i>	Saxifragaceae	<i>Saxifraga campisii</i>
Malvaceae	<i>Malva maroccana</i>	Saxifragaceae	<i>Saxifraga canaliculata</i>
Malvaceae	<i>Malva oblongifolia</i>	Saxifragaceae	<i>Saxifraga conifera</i>
Orobanchaceae	<i>Orobanche haenseleri</i>	Saxifragaceae	<i>Saxifraga haenseleri</i>
Paeoniaceae	<i>Paeonia cambessedesii</i>	Saxifragaceae	<i>Saxifraga latepetiolata</i>
Papaveraceae	<i>Rupicapnos africana</i>	Saxifragaceae	<i>Saxifraga losae</i>
Papaveraceae	<i>Sarcocapnos baetica</i>	Saxifragaceae	<i>Saxifraga reuteriana</i>
Papaveraceae	<i>Sarcocapnos pulcherrima</i>	Scrophulariaceae	<i>Verbascum charidemi</i>
Plantaginaceae	<i>Antirrhinum charidemi</i>	Scrophulariaceae	<i>Verbascum fontqueri</i>
Plantaginaceae	<i>Chaenorhinum gamezii</i>	Scrophulariaceae	<i>Verbascum giganteum</i>
Plantaginaceae	<i>Chaenorhinum macropodum</i>	Scrophulariaceae	<i>Verbascum hervieri</i>
Plantaginaceae	<i>Chaenorhinum semiglabrum</i>	Scrophulariaceae	<i>Verbascum masguindalii</i>
Plantaginaceae	<i>Chaenorhinum tenellum</i>	Scrophulariaceae	<i>Verbascum prunellii</i>
Plantaginaceae	<i>Chaenorhinum villosum</i>	Scrophulariaceae	<i>Verbascum pseudocreticum</i>
Plantaginaceae	<i>Digitalis minor</i>	Solanaceae	<i>Withania somnifera</i>
Plantaginaceae	<i>Digitalis obscura</i>	Thymelaeaceae	<i>Daphne rodriguezii</i>
Plantaginaceae	<i>Digitalis parviflora</i>	Urticaceae	<i>Urtica bianorii</i>
Plantaginaceae	<i>Gadoria falukei</i>	Violaceae	<i>Viola jaubertiana</i>

A

Table S5.2. Top-EDGE selected UTM 10x10 grid cells for each iterative step of the complementarity analysis and number of adjacent cells bound to the EDGE zone.

EDGE zone	Selected grid cell	Number of adjacent grid cells
1	30SVG70	34
2	30TUN57	33
3	30STF70	16
4	31TDG38	98
5	31SDE80	55
6	30TTK56	1
7	30SYJ22	37
8	30SWG19	26
9	29TNG65	8
10	29SQA29	41
11	30TUK86	0
12	30SXG45	62
13	29TPH74	5
14	30TWK67	12
15	30TUK15	7
16	30TWN84	36
17	30SVH07	4
18	30SUF48	14
19	31SFE02	15
20	30TXL18	5
21	29SNC11	26

