ECOGRAPHY

Research Article

Plant diversity estimates of Mediterranean islands differ among biodiversity databases

Francesco Santi[1], Julian Schrader[2], Riccardo Testolin[1], Michele Di Musciano[1,3], Olivier Argagnon[4], Fabio Attorre[5], Irena Axmanová[6], Federico Bombardi[1], Gianmaria Bonari[7,8], Vanessa Bruzzaniti[1,9], Juan Antonio Campos[10], Johannes Foufopoulos[11], Emmanuel Garbolino[12], Riccardo Guarino[13], Behlül Güler[14], Borja Jiménez-Alfaro[15,16], Lea Klepka[17], Holger Kreft[18], Michele Lussu[1], Corrado Marcenò[19], Frédéric Médail[20], Maria Panitsa[21], Salvatore Pasta[22], Sofia Prandelli[1], Diletta Santovito[1], Željko Škvorc[23], Enrico Tordoni[24], Kiril Vassilev[25], Marlene Volz[26], Piero Zannini[1] and Alessandro Chiarucci[1]

[1]BIOME Lab, Department of Biological, Geological and Environmental Sciences, Alma Mater Studiorum – University of Bologna, Bologna, Italy
[2]School of Natural Sciences, Macquarie University, Sydney, New South Wales, Australia
[3]Department of Life, Health and Environmental Science, University of L'Aquila, L'Aquila, Italy
[4]Conservatoire Botanique National Méditerranéen, Hyères, France
[5]Department of Environmental Biology, Sapienza University of Rome, Italy
[6]Department of Botany and Zoology, Faculty of Science, Masaryk University, Brno, Czech Republic
[7]Department of Life Sciences, University of Siena, Siena, Italy
[8]NBFC, National Biodiversity Future Center, Palermo, Italy
[9]LifeWatch Research Infrastructure, Lecce, Italy
[10]Department of Plant Biology and Ecology, University of the Basque Country UPV/EHU, Bilbao, Spain
[11]School for Environment and Sustainability, University of Michigan, Ann Arbor, MI, USA
[12]Mines Paris PSL, ISIGE, Fontainebleau, France
[13]Department of Biological, Chemical and Pharmaceutical Sciences and Technologies (STEBICEF), University of Palermo, Italy
[14]Biology Education, Dokuz Eylül University, Buca, Izmir, Turkey
[15]Biodiversity Research Institute (IMIB), University of Oviedo-CSIC-Principality of Asturias, Mieres, Spain
[16]Department of Organismal and Systems Biology, University of Oviedo, Oviedo, Spain
[17]Conservation Biology, Philipps-University Marburg, Germany
[18]Biodiversity, Macroecology and Biogeography, University of Göttingen, Germany
[19]Department of Chemistry, Biology and Biotechnology, University of Perugia, Italy
[20]Institut Méditerranéen de Biodiversité et d'Ecologie (IMBE), Aix Marseille University, Avignon University, CNRS, IRD, Aix-en- Provence, France
[21]Division of Plant Biology, Department of Biology, University of Patras, Greece
[22]National Research Council (CNR), Institute of Biosciences and BioResources (IBBR), Unit of Palermo, Palermo, Italy
[23]University of Zagreb Faculty of Forestry and Wood Technology, Zagreb, Croatia
[24]Institute of Ecology and Earth Sciences, University of Tartu, Tartu, Estonia
[25]Institute of Biodiversity and Ecosystem Research, Bulgarian Academy of Science, Sofia, Bulgaria
[26]Department of Botany, University of Innsbruck, Austria

Correspondence: Francesco Santi (francesco.santi12@unibo.it, frncsc1697@gmail.com)

Ecography
2026: e08128
doi: 10.1002/ecog.08128

Subject Editor:
Manuel J. Steinbauer
Editor-in-Chief: Miguel Araújo
Accepted 26 January 2026

NORDIC SOCIETY OIKOS

www.ecography.org

Large-scale biodiversity databases encompass three main types of data for plants, namely single species point occurrences, co-occurrences in vegetation plots, and checklists for specific areas. Evidence shows that such data types exhibit specific biases, reporting different species assemblages at local scales. We used the Mediterranean Basin, a global biodiversity hotspot with more than 2200 islands larger than 0.01 km^2, to compare island vascular plant diversity patterns emerging from occurrence data (Global Biodiversity Information Facility; GBIF), vegetation plots (European Vegetation Archive; EVA), and species checklists (Global Inventory of Flora and Traits; GIFT). We aggregated plant data at the island level and compared geographic coverage, inventory completeness, and taxonomic coverage among these data sources. The combined databases accounted for 8702 species distributed on 790 islands (35.6% of the target islands). Data availability increased from small (26.8%) over medium (75.7%) to large islands (100.0%). Spatial coverage of databases on a 30 × 30 km grid was high for GBIF (52.8%) and EVA (45.4%), and low for GIFT (21.7%). GIFT provided higher native and alien species richness values for most of the islands, whereas GBIF and EVA consistently missed a considerable fraction of the expected species richness. Taking GIFT as reference, GBIF, and to a lesser extent EVA, showed a positive bias towards perennial species and an underrepresentation of annuals. Despite their lower taxonomic coverage, GBIF and EVA data can complement our knowledge on Mediterranean islands' plant diversity, providing data for islands lacking plant inventories. Moreover, GBIF and EVA's large datasets can be used for investigating other levels of ecological organisation and modelling single species (GBIF) or population (EVA) trends over space and time. Finally, our results advocate for a coordinated effort to fill the knowledge gaps through data collection and digitisation, possibly integrating data collected by experts by means of citizen science initiatives.

Keywords: Data gaps, European Vegetation Archive, Global Biodiversity Information Facility, Global Inventory of Flora and Traits, Mediterranean islands, Vascular plants

Introduction

Large-scale biodiversity databases have enabled unprecedented advances in plant ecology and biogeography (Wüest et al. 2020) by making biodiversity data widely accessible in standardized formats (Chytrý et al. 2016, Bruelheide et al. 2019, Weigelt et al. 2020, Govaerts et al. 2021, Heberling et al. 2021). This availability of data has provided new insights on how plant diversity is geographically distributed, thus unveiling global patterns of taxonomic diversity (Sabatini et al. 2022, Cai et al. 2023, Tordoni et al. 2024), trait distribution (Díaz et al. 2016, Bruelheide et al. 2018, Midolo et al. 2024), and endemism hotspots of island ecosystems (Schrader et al. 2024).

Species richness is considered one of the simplest and most direct measures to quantify plant diversity. The knowledge of how species richness is spatially distributed is particularly relevant for geographical units with discrete natural boundaries like islands (Warren et al. 2015, Whittaker et al. 2017, Schrader et al. 2025), where the number of species results from ecological and biogeographical processes (Schrader et al. 2024). Despite the great potential offered by large databases in documenting plant diversity, their outputs are inherently biased by the different nature of raw data (e.g. point occurrences, community data from vegetation plots, regional lists; König et al. 2019). Additionally, data collection often focuses on easily accessible locations (Ahrends et al. 2011) and varies widely across taxa (Troudet et al. 2017). Consequently, various data sources may provide contrasting patterns of species diversity, with differences in spatial data coverage (Nori et al. 2023), absolute species richness, or in the list of taxa reported (Wenk et al. 2024). Furthermore, they may be affected to different extents by omission errors (false negatives, i.e. species that occur in an area but are not recorded) and commission errors (false positives, i.e. species incorrectly reported to occur in an area) (Rondinini et al. 2006).

Here, we focus on the three major types of available plant diversity data: single species occurrences (e.g. the Global Biodiversity Information Facility: GBIF; GBIF.org 2026), site-specific co-occurrences in vegetation plots (e.g. the European Vegetation Archive: EVA; Chytrý et al. 2016), and species checklists (e.g. the Global Inventory of Flora and Traits: GIFT; Weigelt et al. 2020). Point occurrence data indicate the presence of a species at a given location, enabling us to map species distributions across different scales, e.g. with the use of species distribution models (Meyer et al. 2016, Heberling et al. 2021). The largest open-access provider for point occurrence data is the GBIF database, which collates data from regional and national plant occurrence databases and natural history collections (e.g. herbaria), and increasingly supplements them with occurrence records from citizen science projects (e.g. the iNaturalist platform). GBIF represents an increasingly important source of biodiversity data (Heberling et al. 2021), currently including over 520 million records for vascular plants worldwide (GBIF.org 2026).

Vegetation plot data consist of lists of species found in an area of fixed size at a given location, along with corresponding cover values. This type of data is primarily used to describe species co-occurrence patterns and habitat affinities, as well as population- and community-level dynamics (Garbolino et al. 2012, Hortal et al. 2015, Knollová et al. 2024). Among large-scale vegetation plot databases, EVA (Chytrý et al. 2016) is a continent-wide example for Europe and adjacent countries, including data from 108 individual databases and over 2 million plots.

Floras and checklists list all species found within a distinct geographical entity such as a country, region, mountain range, or an island. Therefore, they allow comparisons of biodiversity inventories among different areas. However, checklists can be heterogeneous (e.g. focusing only on selected groups of taxa) and may inflate diversity by including information collected from the same site at different times (Chiarucci et al. 2011). GIFT (Weigelt et al. 2020) is one of the most complete global checklist databases, providing plant data for 3400 distinct geographical regions worldwide, with a particular focus on islands (Schrader et al. 2024).

While checklists aim to report the overall species diversity of an area, lists obtained by aggregating point occurrences or vegetation plot data are usually missing a fraction of the overall local diversity (Garcillán and Ezcurra 2011, Alba et al. 2021, Wenk et al. 2024), as largely relying on opportunistic or unevenly distributed point records. Additionally, plant species with conspicuous traits (e.g. taller and with coloured flowers – also called 'charismatic' species) are usually overrepresented in citizen science data (Adamo et al. 2021, Mesaglio et al. 2023), as they are easily detectable in the field. In contrast, vegetation plot data typically reflect a preferential choice of sampling locations (Chiarucci 2007, Michalcová et al. 2011), which can result in good coverage of locally common species (e.g. Poaceae species) and a concurrent underrepresentation of rare or non-targeted species (Garcillán and Ezcurra, 2011).

The Mediterranean Basin is a global biodiversity hotspot (Médail and Myers 2004) and contains a large number of islands harbouring many endemic and range-restricted species (Médail 2022). This basin offers a valuable research opportunity to compare island plant diversity recorded by different data sources, given that plant data collection for scientific research has been ongoing for nearly two centuries on these islands (Chiarucci et al. 2017). This process is facilitated by the recent MEDIS database (Santi et al. 2024), which provides spatial and environmental data for the 2217 Mediterranean islands larger than 0.01 km^2 and enables a straightforward data retrieval from biodiversity databases.

Here, we test the comprehensiveness of vascular plant diversity data captured by three databases that focus on different types of plant diversity data, namely, EVA, GBIF, and GIFT for Mediterranean islands in MEDIS. We highlight the strengths and identify the gaps and biases of each database, by focusing on three key dimensions of data availability: 1) geographic coverage, 2) inventory completeness, and 3) taxonomic representation. We use GIFT as a reference to evaluate GBIF and EVA completeness for 2) and 3), given the current lack of a consolidated repository of Mediterranean islands' floras. Our approach offers crucial insights to guide future efforts in data collection and integration, and a foundational step towards the development of a comprehensive plant species database for the Mediterranean islands.

Material and methods

Data preparation

The number of islands present in the Mediterranean Basin has been reported to be up to 11 000 (Médail 2022) but many of them are really small in size. For this analysis, we included the 2217 islands with an area size larger than 0.01 km^2 that are reported in the MEDIS database (Santi et al. 2024). The island shapes from MEDIS were used to extract plant data from GBIF, EVA, and GIFT.

We retrieved occurrence data for vascular plants into R ver. 4.4.2 (www.r-project.org) from the GBIF portal (GBIF.org 2023), using the 'rgbif' R package (Chamberlain et al. 2022), filtering for Mediterranean islands records and obtaining 1 339 587 occurrences (GBIF.org 2024). We performed multiple data checks and ran validation procedures as detailed below, to further mitigate biases and errors in GBIF. In particular, we first harmonised the nomenclature using the Taxonomic Name Resolution Service (TNRS; Boyle et al. 2013), via the 'TNRS' R package, with World Checklist of Vascular Plants (WCVP; Govaerts 2025) as a taxonomic reference, and retrieved family information for each species. Taxa not identified nor matched at the species level were removed (2235 records). We also excluded records not directly referring to wild individuals ('living specimen'; 57 records), keeping only occurrence data recorded in the field ('human observation', 'occurrence', 'preserved specimen'; see Chamberlain et al. 2022 for detailed specification of GBIF syntax). As mislocated records can severely affect biodiversity patterns (Maldonado et al. 2015), we removed occurrences with location uncertainty > 500 m (313 660) or missing (369 177), in line with previous studies (de Araujo et al. 2022, Melis et al. 2025). We used the 'CoordinateCleaner' R package (Zizka et al. 2019) to exclude all records that were duplicates, or fell in proximity to country capitals and centroids (46 754). We also excluded marine species (*Cymodocea nodosa*, *Halophila stipulacea*, *Posidonia oceanica*, *Zostera marina* and *Z. noltii*; Marbà et al. 2013, Winters et al. 2020; 533 records) and hybrids (0 records). The final dataset encompassed 607 171 occurrences, accounting for 5046 species distributed across 464 islands, corresponding to 33 465 distinct island–species combinations.

From the European Vegetation Archive (Chytrý et al. 2016), we retrieved all plots falling inside the island polygons of MEDIS. We obtained 44 101 plots and excluded 40 of them as reporting non-vascular plant taxa only. We then associated each of the remaining 44 061 plots to its respective island, obtaining data for a total of 376 islands. We also digitised newly published and unpublished sources not already included in EVA, adding 1560 plots on 105 islands, with 70 new islands not previously covered by the EVA database. These new plots have been made available to EVA through the AMS-VegBank database (Alessi et al. 2022). The final dataset consisted of 45 621 plots. After nomenclatural harmonisation and removal of hybrids, unidentified taxa, and marine species, we obtained a dataset made up of 4575 species distributed across 446 islands, corresponding to 45 522 distinct island–species combinations.

We retrieved island vascular plant checklists from GIFT ver. 3.2, via the 'GIFT' R package (Denelle et al. 2023), selecting those belonging to single islands of MEDIS. We applied the specific filter '*suit_geo==1*' to retain checklists considered comprehensive in GIFT and exclude partial lists (e.g. accounting for endemic species only; see Denelle et al. 2023 and related tutorials for detailed use of the 'GIFT' R package). After taxonomic harmonisation and filtering out of hybrids, records not identified at the species level and marine species, we obtained a dataset made up of 7737 species distributed across 269 islands, corresponding to 83 366 distinct island–species combinations.

For each island, we distinguished between native and alien species in two ways. GIFT records already provided this information; we extended the GIFT attribution to the same species–island pairs occurring in GBIF and EVA. For other records, we retrieved information from the Global Naturalized Alien Flora database (van Kleunen et al. 2019) and checklists at the country level or, when available, from more localised checklists (e.g. those concerning one single archipelago). The long history of human presence in the Mediterranean region has significantly shaped the species assemblages on these islands (Pausas 1999, Vogiatzakis et al. 2016, Chiarucci et al. 2017, Médail and Pasta 2024), often complicating assessment of species status at the island level (Adams 1909, Hughes and Convey 2012). Therefore, in the alien subset, we only included neophytes, i.e. species introduced after 1500 C.E., considering archaeophytes (i.e. species introduced before 1500 C.E.) within the natives.

For each database, we then merged species records at the island level and retrieved checklists for each island. Following Santi et al. (2024), we categorized islands based on their surface area as 'small' (0.01–1.00 km^2), 'medium' (1–100 km^2), and 'large' (> 100 km^2). We then summarised and compared the number of islands covered by each database and the species richness for both native and alien taxa. We did not apply any exclusion filter based on island species richness to effectively assess the overall data availability.

Statistical analysis

To visualise the geographic coverage of the three databases, we aggregated the islands into squared grid cells with a resolution of 30 × 30 km, selecting the 443 cells intersecting with the island centroids. Each island was assigned to one grid cell only, based on the location of its centroid. For each database, we calculated the ratio of islands with at least one species occurrence compared to the total number of islands per cell in MEDIS. To highlight the geographic gaps, we calculated the same metric by pooling together all the databases.

To evaluate the inventory completeness, we calculated and compared the island species richness reported by each database. We then calculated the Island Species–Area Relationship (ISAR), by fitting the Arrhenius power function with the 'sars' R package (Matthews et al. 2019b) for both native and alien species. The power function is considered the most robust ISAR model across many island systems (Matthews et al. 2019a) and taxonomic groups (Dengler et al. 2020). Given the disproportionately higher number of alien species reported by GIFT for Corsica (1642) compared to recent checklist (533 alien taxa; Puddu et al. 2016), we computed the GIFT alien ISAR model both including and excluding Corsica, to check for any relevant difference between the model fits.

Due to the highly incomplete data for alien species in all the databases, we focused the following analyses on the native subset only.

As we selected comprehensive checklists from GIFT, we took it as a reference to evaluate the degree of completeness of GBIF and EVA species lists. We used the GIFT ISAR model to retrieve the expected values of species richness for the islands in MEDIS. We then calculated the negative or positive deviations of GBIF and EVA values from the expected ones as the difference between the observed and the expected value, divided by the expected one. We then applied two filters of sampling completeness to islands in GBIF and EVA (species richness respectively > 50% and > 75% of the expected value), to compare their distribution with the geographic coverage previously pointed out.

We analysed the taxonomic representation of GBIF and EVA in comparison with GIFT for the 63 islands the three databases had in common. We calculated species rank-frequency curves (Mouillot and Lepretre 2000), with frequency values reported as the number of islands of occurrence. Among the first 100-ranked species, we evaluated GBIF and EVA frequency patterns compared to GIFT, in terms of plant life strategies and phylogeny. For the first pattern, we categorised the species as 'annual' or 'perennial', retrieving data from the FloraVeg.EU portal (Dřevojan et al. 2023, Chytrý et al. 2024). For the phylogenetic comparison, we selected five of the species richest families of the Mediterranean flora, i.e. Asteraceae, Fabaceae, Lamiaceae, Orchidaceae, Poaceae, which also account for charismatic species (e.g. Orchidaceae) or common species often neglected by non-experts (e.g. Poaceae). We accounted for the first 100-ranked species only, to avoid a homogenisation of the frequency patterns due to the high number of species (GBIF: 4212; EVA: 3780; GIFT: 6066). Finally, we calculated and compared the ratios of unique species reported by pairs of databases for their shared islands, to visualise their relative contributions in reporting taxonomic diversity.

Results

GBIF, EVA and GIFT provided plant occurrence data for 464, 446, and 269 islands, respectively (Supporting information), covering a total of 8702 species of vascular plants distributed on 790 islands, i.e. 35.6% of the islands reported in MEDIS (Supporting information). The overall availability of data combining the three databases greatly increased from small (26.8%) over medium (75.7%) to large islands (100.0%). Similarly, the proportion of islands found in all the databases was 0.2% for small islands, 8.5% for the medium, and 65.3% for the large ones (Fig. 1).

Geographic coverage

The spatial distribution of islands with species data from the GBIF and EVA databases was largely overlapping (Fig. 2a–b), with most of the records located between southern France and central Italy, and within the Dalmatian, Ionian, and Aegean

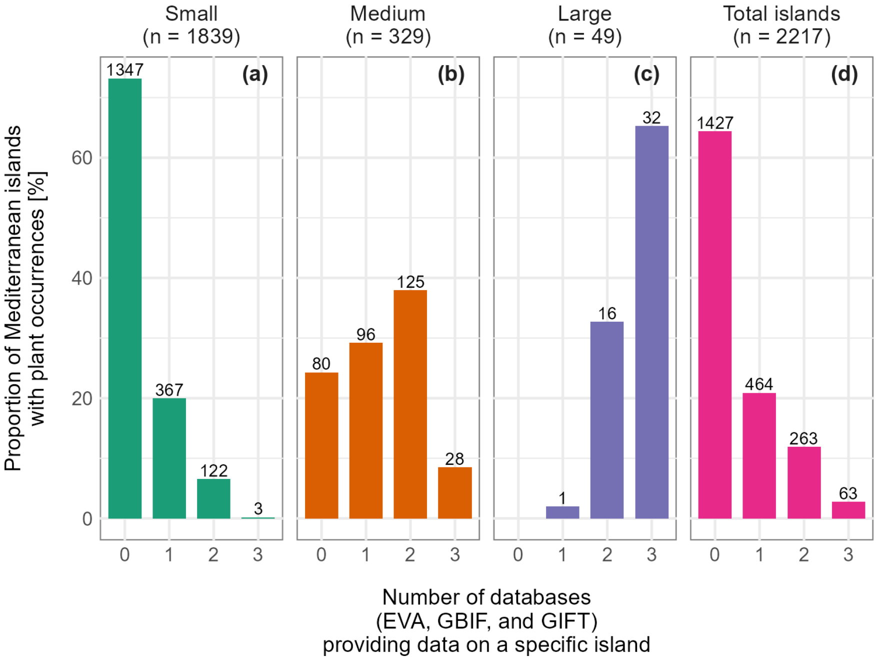

Figure 1. Bar chart showing the proportion of Mediterranean islands > 0.01 km^2 with plant occurrences in none (0), one (1), two (2), or all three (3) databases (EVA, GBIF, and GIFT). Islands are grouped on the basis of their size: a) small islands: 0.1–1.0 km^2, b) medium islands: 1–100 km^2, c) large islands: > 100 km^2, d) all the islands.

archipelagos. Whereas GBIF and EVA covered 234 (52.8%) and 201 (45.4%) grid cells, respectively, GIFT only covered 96 (21.7%) cells, with records mainly clustered within the Dalmatian, Ionian, and Aegean archipelagos (Fig. 2c). Of the 77 islands along the North African–Near East coast, 61 were not covered by any of the three databases (Fig. 2d), which provided data only for Algeria (4 islands out of 19; 21.1%) and Tunisia (12 islands out of 41; 29.3%, Supporting

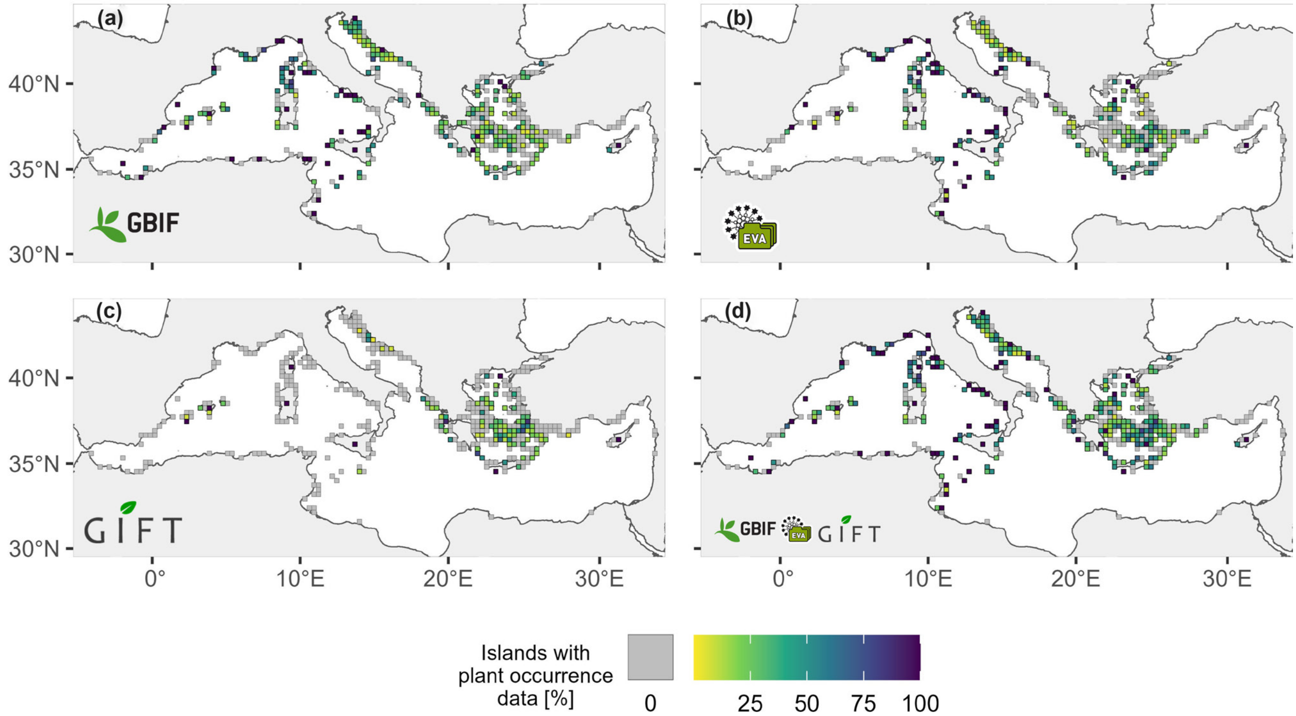

Figure 2. Geographic distribution of Mediterranean islands with plant occurrence data in GBIF (a), EVA (b), GIFT (c), and with cumulative data from the three databases (d). Islands are aggregated in grid cells with a 30 × 30 km resolution. The cell colour represents the proportion of islands covered by each database with respect to the total number of islands per cell.

information). Among the most island-rich countries, Turkey, Greece, and Croatia were represented by 12.5% (34 islands out of 271), 35.9% (312 islands out of 868), and 38.2% (240 islands out of 629), respectively. Moreover, France had a notably high data coverage, reaching 89.5% (68 islands out of 76), with Corsican islands almost totally represented (37 islands out of 42; Fig. 2a–b), compared to nearby Sardinian islands (14 islands out of 78; Fig. 2d). Apart from Sardinia, Italy was well covered by GBIF (62 islands; 33.9%) and EVA (81 islands; 44.3%), compared to GIFT (2 islands; 1.1%).

Inventory completeness

Native species richness per island had broader ranges in GIFT (5–2704 species), compared to GBIF (1–2171 species) and EVA (2–1761 species; Supporting information). Similarly, alien species richness per island had the broadest range in GIFT (1–1642 species; GBIF: 1–254 species; EVA: 1–61 species; Supporting information). Sicily was the island richest in native species both in EVA and GIFT, whereas Corsica was the richest for both native and alien species in GBIF, and the richest in alien species in GIFT; Crete was the richest in alien species in EVA.

Considering native species, GIFT checklists had most species in 152 (95.6%) of the 157 islands in common with EVA or GBIF (Supporting information). EVA reported more species than GIFT in 4 small islands (area range: 0.03–0.34 km^2), whereas GBIF had higher species richness than GIFT in one medium island (Nisida Kastos, Greece: 6.06 km^2; Supporting information). GBIF native richness was equal to or lower than 5 species in 192 islands (41.4%). When compared, GBIF and EVA showed opposite patterns for native and alien subsets (Supporting information), with EVA accounting for more native species than GBIF in 178 islands (77.1%), and GBIF reporting more alien species than EVA in 74 islands (71.8%).

When considering ISAR patterns (Table 1, Fig. 3), the models based on GIFT had the highest species estimates for unit area (natives: $c = 182.2$; aliens: $c = 23.7$), and the lowest values for the z parameter (natives: $z = 0.27$; aliens: $z = 0.30$), as well as the best fit for native species ($R^2_{adj} = 0.94$) and for alien species excluding Corsica ($R^2_{adj} = 0.92$; with Corsica included $R^2_{adj} = 0.27$).

The number of islands with respectively at least 50% and 75% of the species richness expected from the ISAR model based on GIFT was lower in GBIF (28 islands with more than 50% of expected species; 13 islands with more than 75%) than in EVA (111 islands with more than 50% and 53 with more than 75%; Supporting information).

Taxonomic coverage

The species rank-frequency distribution for the 63 islands shared by the three databases showed that GIFT systematically reported higher frequency values than GBIF and EVA and a longer tail of ranked species (Fig. 4a).

Within the first 100-ranked species in the three databases, GBIF and EVA had lower proportions of annual species than GIFT and higher proportions of perennial species (Fig. 4b).

Table 1. Summary of the parameters associated with the ISAR Arrhenius power function model fitted for the Mediterranean islands by using plant occurrence data obtained by EVA, GBIF, and GIFT. GIFT* indicates the ISAR model for GIFT alien species excluding Corsica.

	Database	No. of islands	c	z	R^2_{adj}
Native species	GBIF	460	20.8	0.43	0.74
	EVA	446	38.0	0.36	0.65
	GIFT	269	182.2	0.27	0.94
Alien species	GBIF	230	2.9	0.43	0.71
	EVA	153	1.1	0.34	0.31
	GIFT	39	23.7	0.30	0.27
	GIFT*	38	10.8	0.32	0.92

EVA and GIFT showed similar patterns for the frequency of the selected families (Fig. 4c), whereas GBIF exhibited marked differences, especially for Orchidaceae and Poaceae, with respectively higher and lower values than the other two databases.

GIFT reported the large majority of the islands' species diversity, but GBIF and EVA also provided a significant proportion of species not included in GIFT (Supporting information).

Discussion

This study presents the first comparison of plant species diversity as documented by three major large-scale plant diversity databases, each representing a different type of data format: species occurrences (GBIF), vegetation plots (EVA), and plant checklists (GIFT). We assessed their coverage across the full set of Mediterranean islands larger than 0.01 km^2, a testing ground for ecological and biogeographical analyses located within a global biodiversity hotspot (Médail and Myers 2004). Despite being one of the botanically most explored regions globally (Chiarucci et al. 2017, Médail 2022), we found that plant species data were available for 790 islands only, namely less than 40% of the islands larger than 0.01 km^2 occurring in the basin. Our findings therefore revealed substantial knowledge gaps (sensu Hortal et al. 2015) that persist even in such a relatively well-explored area. Our results also showed clear and consistent differences in the type and extent of data each of the three formats provides (Table 2).

Evaluating taxonomic completeness is challenging because it requires a reliable reference list. Among the three databases compared here, GIFT resulted in being the most comprehensive from the taxonomic point of view, as it specifically compiles checklists and floras designed to include all known species on a target island. In general, checklist data, as in GIFT, are considered the most authoritative source of plant occurrence information for a given area (Brummitt et al. 2021), as they are typically curated and verified by local expert botanists. Although some taxa may still be missing, this bias is likely small (König et al. 2019).

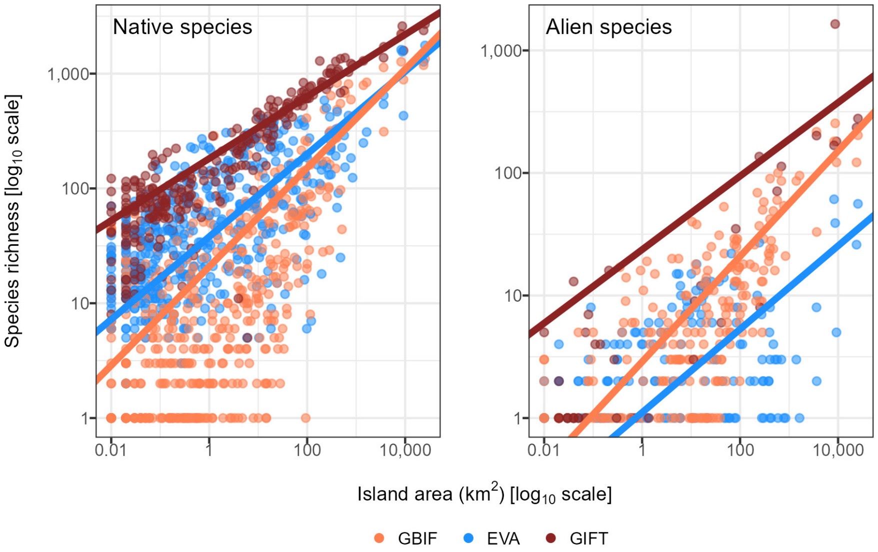

Figure 3. Island Species Area Relationship models for native (left) and alien species (right). The model is based on the Arrhenius power function: $S = c \times A^z$, where S is the species richness, A is the island area, c and z are two parameters respectively estimating the number of species per unit area and the rate of increase of species richness with increase of island area. The power model is graphically represented here in the log–log scale.

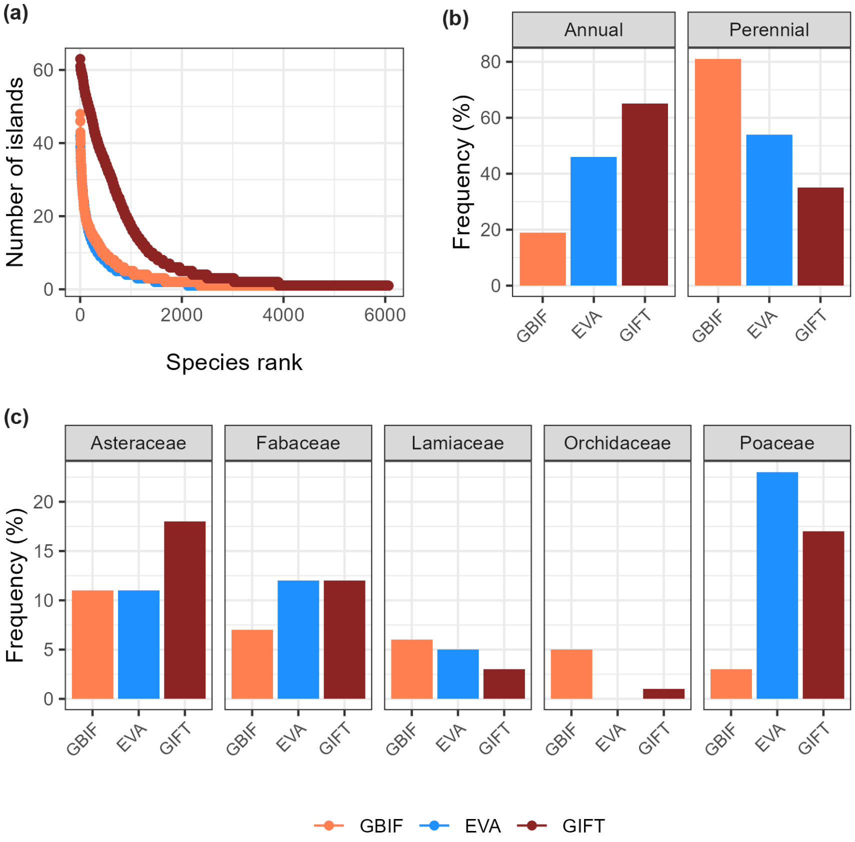

Figure 4. Taxonomic comparison of the species reported by GBIF, EVA, and GIFT for the 63 shared islands, in terms of frequency abundance values (a), frequency of annual and perennial species (b), and frequency of five species rich families in the Mediterranean region (c), i.e. Asteraceae, Fabaceae, Lamiaceae, Orchidaceae, and Poaceae. In (a) all the species are considered, whereas in (b) and (c) the first 100-ranked species only are considered.

Table 2. Summary of the main biases, limitations, and data opportunities related to GBIF, EVA, and GIFT for the Mediterranean islands in MEDIS. Cited references provide theoretical and technical context to the mentioned topics. GBIF = Global Biodiversity Information Facility; EVA = European Vegetation Archive; GIFT = Global Inventory of Flora and Traits.

	Bias	Limitations	Data quality	References
GBIF	**Taxonomic bias**: over-representation of perennial versus annual species and 'charismatic' taxa; under-representation of inconspicuous groups (e.g. Poaceae). **Geographic bias**: occurrences are often clustered in easily accessible areas or near research centres (opportunistic sampling bias)	**Spatial and taxonomic errors**: high incidence of false positives (commission errors), coordinate inaccuracies, or unresolved synonyms requiring intensive data cleaning. **Incompleteness:** frequently misses a significant fraction of expected species richness (false negatives) compared to complete checklists. Few species occurrences provided for large part of the islands	**High spatial resolution**: enables fine-scale Species Distribution Models (SDMs) for individual taxa. **Temporal analysis**: huge volume of data (including historical collections and citizen science) allows for the assessment of spatial and temporal trends. **Data for uninventoried areas**: provides preliminary biodiversity data for islands lacking formal checklists	Heberling et al. 2021 (GBIF overview), Troudet et al. 2017 (taxonomic bias), Adamo et al. 2021 (citizen science bias), Maldonado et al. 2015 (spatial errors)
EVA	**Preferential sampling**: plots are often located in 'typical' or well-preserved vegetation types (phytosociological bias), potentially ignoring marginal or degraded habitats. Potential avoidance of sites impacted by alien species. **Taxonomic bias**: tends to under-represent rare or non-target species not captured within standard plot sizes	**Irregular geographic coverage**: reflecting survey investigations and digitisation activities -> significant gaps in specific regions (e.g. Sardinian, North African islets) **Underestimation of regional flora**: captures only the diversity intercepted by plots, not the full species pool of an island/region	**Community data**: provide unique insights into species co-occurrence and habitat affinities. **Abundance data**: availability of cover-abundance scales allows for analysis of dominance patterns and population dynamics, which is impossible with presence-only data	Chytrý et al. 2016 (EVA overview), Michalcová et al. 2011 (plot sampling bias)
GIFT	**Source heterogeneity**: checklists vary in accuracy, taxonomic focus, or sampling effort depending on the original author and compilation date. **Temporal data aggregation**: cumulative checklists may include extinct or historical records, potentially overestimating current diversity	**Low spatial resolution**: data are aggregated at the island or political region level, preventing local-scale or micro-habitat analyses. **Limited coverage**: covers fewer islands compared to GBIF/EVA, particularly regarding very small islands. Low coverage on alien species as mainly targeting native species lists	**Taxonomic coverage**: comprehensive lists allow calculation of species composition changes between islands but also through time; baseline checklists for all sorts of other analyses. **Biogeographical analyses**: ideal for robust Island Species–Area Relationships (ISAR) and macroecological modelling at the regional scale	Weigelt et al. 2020 (GIFT overview), König et al. 2019, Schrader et al. 2024 (biodiversity data integration)

Using GIFT as a baseline, we found that checklists derived from GBIF and EVA consistently missed a fraction of the expected species diversity. This finding was further supported by the predominantly negative residuals in the Island Species–Area Relationship when GBIF- and EVA-derived checklists are compared to the GIFT-based ISAR. Indeed, GBIF and EVA are primarily designed for different research targets, such as analysing species distributions (GBIF) or abundance patterns (EVA), rather than providing exhaustive plant occurrence inventories.

However, the number of islands covered by GIFT was much lower than the number of islands with plant occurrence data available from GBIF and EVA, especially in the case of small islands. Therefore, GBIF and EVA data can still significantly complement our knowledge of plant diversity in Mediterranean islands, providing additional information on species occurrences and preliminary checklists for islands not covered by GIFT. In addition, we see potential that despite the aim of completeness, GIFT checklist data can still be supplemented by species recorded in GBIF or EVA (Alessi et al. 2023).

It is important to note that we did not apply any temporal filters to the record; therefore, our retrieved species lists represent the full set of species that have occurred on each island during the time periods of investigation, rather than a snapshot of current floras. This distinction is crucial, given the increasing recognition of the need for repeated surveys to monitor island biodiversity over time (Schrader et al. 2025) and the high temporal turnover dynamics of their communities (Panitsa et al. 2008, Chiarucci et al. 2017, Matthews and Triantis 2021, Schrader et al. 2023).

Geographic coverage

The limited number of islands with at least one species occurrence highlights the need for new survey campaigns and further data digitising to improve knowledge of plant diversity across Mediterranean islands, especially in the poorly surveyed areas, such as the southern and eastern margins of the Mediterranean Basin (Nieto Feliner 2014). Indeed, while most of the Algerian islands have been surveyed in recent decades (Hamimeche et al. 2024), these data have not yet been included in GIFT or EVA, whereas GBIF has data for only four of them. A similar pattern applies to Italian islands, for which floras (e.g. Pontine islands: Anzalone and Caputo 1976; Sardinian islands: Arrigoni and Bocchieri 1996; Tuscan Archipelago: Chiarucci et al. 2017) are available in the published literature but not yet integrated into GIFT. In contrast, the high coverage of vegetation plot data on 81 Italian islands (44.3%) is attributable to targeted digitisation projects (e.g. the small islands' focus within the AMS-VegBank database in EVA; Alessi et al. 2022).

The urgency of filling geographic data gaps was also amplified by the very low number of islands with considerable species richness in GBIF and EVA, which further reduced their geographic coverage. Digitising existing species occurrence data is therefore of paramount importance, also to improve data accessibility for research and conservation purposes (Maddock and Samways 2000, Pullin and Knight 2005).

On a long-term basis, establishing a comprehensive database of available floras would also facilitate the strategic planning of field surveys, e.g. pointing out both undersampled areas in need of initial data collection and sites that should be prioritised for long-term monitoring programmes.

Inventory completeness

The ISAR patterns estimated by different types of plant diversity data indicated that the choice of data source significantly influences both model fitting and parameter estimates, with GIFT yielding results that more closely align with those obtained from complete floras in several Mediterranean archipelagos (see Table 1 in Chiarucci et al. 2021b). Therefore, ISAR based on GIFT data was useful for comparison with the other two databases. However, the relatively low coverage on small and medium islands may affect the model parameters and may result in imprecise description of ISAR patterns for Mediterranean islands.

GBIF and EVA ISAR patterns mirrored those of the GIFT model, although the absolute values of c and z parameters differed, as already found by Chiarucci et al. (2021a). In particular, GBIF and EVA reported lower values for c and higher values of z compared to GIFT, in line with the patterns shown by remote archipelagos (Chiarucci et al. 2021b), which are known to be species-poor compared with archipelagos closer to the mainland (Matthews and Triantis 2021). Similarly, c and z values for the native GBIF model are highly influenced by the considerable number of islands with few species, resulting from opportunistic rather than systematic data collection (Adamo et al. 2021). Species richness data are indeed well known to be sensitive to changes in sampling effort, which depends on the duration and seasonality of surveys, the number of scientists involved, and other factors (Sastre and Lobo 2009, Petřík et al. 2010, Guarino et al. 2022, Klímová et al. 2025). This phenomenon affects many facets of biodiversity investigations, from the basic knowledge of distribution patterns (Petřík et al. 2010) to the study of global change effects (McDonald and Brown 1992, Arnillas et al. 2017) or the planning of conservation actions (Ahrends et al. 2011). This limitation has been conceptualised in different concepts, from the 'botanist effect', which predicts that places where scientists are more present tend to have more species occurrence data (Moerman and Estabrook 2006, Ahrends et al. 2011), to the Wallacean shortfall, which states that knowledge about the geographic distribution of most species is incomplete (Lomolino 2004, Hortal et al. 2015). The observed biodiversity patterns are also influenced by the long history of human presence on Mediterranean islands (Pausas 1999, Vogiatzakis et al. 2016, Chiarucci et al. 2017, Bradshaw et al. 2024, Médail and Pasta 2024), possibly resulting in different plant assemblages from prehistoric times (Carine and Menezes de Sequeira 2020, Orihuela-Rivero et al. 2025) due to alien species introduction and habitat modification.

Alien species data appeared to be more fragmented compared to natives, suggesting that botanists probably paid less attention to – or deliberately avoided – recording non-native plants. The recent increase of alien species introduction could also influence the observed lack of data (Seebens et al. 2017). EVA, in particular, lacked comprehensive information on the presence of alien species, as the recording scheme of vegetation plot data largely relies on preferential choice of sampling locations (Michalcová et al. 2011), which could have resulted in partial avoidance of sites impacted by alien species or strongly affected by anthropogenic disturbance (Axmanová et al. 2021). However, vegetation plot data provide valuable information on the abundance patterns of alien plants in local communities (Guarino et al. 2021) and could be used for a basin-scale assessment of alien invasion (Cao Pinna et al. 2024), given their broad coverage of Mediterranean islands. Moreover, the relative abundance of alien species records from GBIF could make them particularly useful to detect early stages of the naturalisation process of alien plants (Gallo and Waitt 2011, Johnson et al. 2020).

Taxonomic coverage

GBIF and EVA not only underestimate current island species richness, but also provide considerably lower frequency values for species, due to their sampling strategies which usually miss rare taxa (D'Antraccoli et al. 2020). Therefore, this potential bias should be taken into account when employing these data for models that strongly rely on spatially accurate information, like species distribution models (Lomba et al. 2010, Jeliazkov et al. 2022, Rocchini et al. 2023). Moreover, we found a strong prevalence of perennial or charismatic species among GBIF records, and a parallel neglect of common but easily overlooked species (e.g. grass species), in line with previous studies (Troudet et al. 2017, Adamo et al. 2021, Marcenò et al. 2021b). Systematic bias, more than random errors, can cause consistent distortions in diversity estimates, with more serious implications for conservation planning and macroecological inferences (Hortal et al. 2015, Meyer et al. 2016). For example, systematic underestimation of diversity in certain regions or taxonomic groups might make some areas appear less important for conservation, leading to an inefficient allocation of resources (Nori et al. 2023). Conversely, incorrect overestimation could divert attention away from truly critical areas. Despite our extensive data cleaning steps to filter out potentially erroneous GBIF records, we acknowledge that part of the included data could still contain errors, as it is normal in this type of extensive database. This potential issue is intrinsic to GBIF and could influence biodiversity assessments by overestimating the real species diversity (e.g. records from cultivated individuals treated as occurring in the wild; López-Guillén et al. 2024) or by reducing the variability (e.g. inability to identify distinct congeneric species belonging to critical groups; Popp et al. 2025). While our results suggested that a large part of GBIF diversity is nested within GIFT, we recommend a specific study assessing the validity of the GBIF species records not shared with GIFT.

Future perspectives

Our study gives insights to trace the future direction of data collection and integration. We recommend a coordinated initiative for the Mediterranean Basin to fill the identified data gaps. Such an achievement would encourage analogous actions for other biological groups (e.g. as done by Bonardi et al. 2022, Tonelli and Lobo 2025), eventually allowing multi-taxa investigations.

On a general point of view, promoting interoperability among major database formats (GBIF, EVA, GIFT) will streamline data sharing and usage. To enhance the quality and completeness of biodiversity data on island ecosystems worldwide and reduce systematic biases, a multi-pronged strategy is essential. First, digitising legacy data such as regional floras and herbarium records is crucial to fill existing geographic gaps. Ensuring regular updates to detect and report ongoing changes and processes (e.g. species turnover, extinction rate) should follow, especially for invasive or newly introduced alien species. Targeted field campaigns should prioritise under-sampled islands and overlooked taxa, while improvements in metadata quality will support better data integration and analysis.

Citizen science initiatives, guided by expert validation (Maldonado et al. 2015), can significantly boost data collection (Mason et al. 2025, Mesaglio et al. 2025, 2026), particularly for under-recorded species groups or alien species (Marcenò et al. 2021a). It is equally important to implement robust data curation processes to validate and clean records across platforms. Finally, long-term success depends on securing sustained funding and institutional support to maintain and expand biodiversity data infrastructure throughout the Mediterranean region.

Acknowledgements – We thank Mark Westoby for providing suggestions to perform statistical analyses. We are grateful to the thousands of people who collected or digitised vascular plant data, enabling us to carry out this study. Open access publishing was facilitated by Alma Mater Studiorum, University of Bologna, as part of the Wiley–CRUI-CARE agreement. Open access publishing facilitated by Universita degli Studi di Bologna, as part of the Wiley–CRUI-CARE agreement.

Funding – FM is financed by the CAIPIM project 'Know, Act, and Innovate to Protect Small Mediterranean Islands' supported by the PIM Initiative, the Conservatoire du littoral, Aix Marseille University/IMBE and whose main financier is the French Environment Fund worldwide (FFEM). GB, AC, and RT were funded under the National Recovery and Resilience Plan (NRRP), Mission 4 Component 2 Investment 1.4 - Call for tender no. 3138 of 16 December 2021, rectified by Decree no. 3175 of 18 December 2021 of Italian Ministry of University and Research funded by the European Union – NextGenerationEU; Award Number: Project code CN_00000033, Concession Decree no. 1034 of 17 June 2022 adopted by the Italian Ministry of University and Research, CUP B63C22000650007 and CUP J33C22001190001, Project title 'National Biodiversity Future Center - NBFC'. ET is currently supported by the Estonian Research Council (grant PSG1021). JAC has been supported by the Basque Government (IT1487-22). JF has been supported by PIM Initiative (France) as well as the Cycladic Preservation Fund (Greece) and the University of Michigan (USA). HK acknowledges funding by the German Research Foundation (DFG, grants 533271599 and 379417748).

Conflict of interest – The authors declare no conflict of interest.

Author contributions

Francesco Santi: Conceptualization (lead); Data curation (lead); Formal analysis (lead); Methodology (lead); Writing – original draft (lead); Writing – review and editing (equal). **Julian Schrader**: Conceptualization (equal); Methodology (equal); Writing – original draft (supporting); Writing – review and editing (equal). **Riccardo Testolin**: Conceptualization (equal); Methodology (equal); Supervision (equal); Writing – review and editing (equal). **Michele Di Musciano**: Conceptualization (equal); Methodology (equal); Writing – review and editing (equal). **Olivier Argagnon**: Data curation (equal); Writing – review and editing (equal). **Fabio Attorre**: Data curation (equal); Writing – review and editing (equal). **Irena Axmanová**: Data curation (equal); Writing – review and editing (equal). **Federico Bombardi**: Formal analysis (supporting); Writing – review and editing (equal). **Gianmaria Bonari**: Data curation (equal); Writing – review and editing (equal). **Vanessa Bruzzaniti**: Data curation (equal); Writing – review and editing (equal). **Juan Antonio Campos**: Data curation (equal); Writing – review and editing (equal). **Johannes Foufopoulos**: Data curation (equal); Writing – review and editing (equal). **Emmanuel**

Garbolino: Data curation (equal); Writing – review and editing (equal). **Riccardo Guarino**: Data curation (equal); Writing – review and editing (equal). **Behlül Güler**: Data curation (equal); Writing – review and editing (equal). **Borja Jiménez-Alfaro**: Data curation (equal); Writing – review and editing (equal). **Lea Klepka**: Data curation (equal); Writing – review and editing (equal). **Holger Kreft**: Writing – review and editing (equal). **Michele Lussu**: Conceptualization (supporting); Writing – review and editing (equal). **Corrado Marcenò**: Data curation (equal); Writing – review and editing (equal). **Frédéric Médail**: Data curation (equal); Writing – review and editing (equal). **Maria Panitsa**: Data curation (equal); Writing – review and editing (equal). **Salvatore Pasta**: Data curation (equal); Writing – review and editing (equal). **Sofia Prandelli**: Writing – review and editing (equal). **Diletta Santovito**: Methodology (equal); Writing – review and editing (equal). **Željko Škvorc**: Data curation (equal); Writing – review and editing (equal). **Enrico Tordoni**: Conceptualization (supporting); Writing – review and editing (equal). **Kiril Vassiliev**: Data curation (equal); Writing – review and editing (equal). **Marlene Volz**: Data curation (equal); Writing – review and editing (equal). **Piero Zannini**: Writing – review and editing (equal). **Alessandro Chiarucci**: Conceptualization (equal); Supervision (lead); Writing – review and editing (equal).

Transparent peer review

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/ecog.08128.

Data availability statement

Data and code supporting the manuscript can be accessed from the Zenodo Repository: https://doi.org/10.5281/zenodo.18339464 (Santi et al. 2026). Initial datasets were retrieved from the European Vegetation Archive (Chytrý et al. 2016, Project 168), the Global Biodiversity Information Facility (GBIF.org 2023, 2024), and the Global Inventory of Flora and Traits (Weigelt et al. 2020).

Supporting information

The Supporting information associated with this article is available with the online version.

References

Adamo, M., Chialva, M., Calevo, J., Bertoni, F., Dixon, K. and Mammola, S. 2021. Plant scientists' research attention is skewed towards colourful, conspicuous and broadly distributed flowers. – Nat. Plants 7: 574–578.

Adams, J. 1909. On the possibility of distinguishing between native and alien species of plants in Ireland. – Ir. Nat. 18: 123–132.

Ahrends, A., Rahbek, C., Bulling, M. T., Burgess, N. D., Platts, P. J., Lovett, J. C., Kindemba, V. W., Owen, N., Sallu, A. N., Marshall, A. R., Mhoro, B. E., Fanning, E. and Marchant, R. 2011. Conservation and the botanist effect. – Biol. Conserv. 144: 131–140.

Alba, C., Levy, R. and Hufft, R. 2021. Combining botanical collections and ecological data to better describe plant community diversity. – PLoS One 16: e0244982.

Alessi, N., Bruzzaniti, V., Buldrini, F., Centomo, E., Cervellini, M., Enea, M., Landi, S., Lelli, C., Montanari, I., Nascimbene, J., Pezzi, G., Virzí, G., Zannini, P. and Chiarucci, A. 2022. AMS-VegBank: a new database of vegetation plots for the Italian territory. – Veg. Classif. Surv. 3: 177–185.

Alessi, N., Bonari, G., Zannini, P., Jiménez-Alfaro, B., Agrillo, E., Attorre, F., Canullo, R., Casella, L., Cervellini, M., Chelli, S., Di Musciano, M., Guarino, R., Martellos, S., Massimi, M., Venanzoni, R., Zerbe, S. and Chiarucci, A. 2023. Probabilistic and preferential sampling approaches offer integrated perspectives of Italian forest diversity. – J. Veg. Sci. 34: e13175.

Anzalone, B. and Caputo, G. 1976. Flora e vegetazione delle Isole Ponziane (Golfo di Gaeta). – Delpinoa 16: 1–184.

Arnillas, C. A., Tovar, C., Cadotte, M. W. and Buytaert, W. 2017. From patches to richness: assessing the potential impact of landscape transformation on biodiversity. – Ecosphere 8: e02004.

Arrigoni, P. V. and Bocchieri, E. 1996. Caratteri fitogeografici della flora delle piccole isole circumsarde. – Biogeographia 18: 63–90.

Axmanová, I. et al. 2021. Neophyte invasions in European grasslands. – J. Veg. Sci. 32: e12994.

Bonardi, A., Ficetola, G. F., Razzetti, E., Canedoli, C., Falaschi, M., Lo Parrino, E., Rota, N., Padoa-Schioppa, E. and Sindaco, R. 2022. ReptIslands: Mediterranean islands and the distribution of their reptile fauna. – Global Ecol. Biogeogr. 31: 840–847.

Boyle, B., Hopkins, N., Lu, Z., Raygoza Garay, J. A., Mozzherin, D., Rees, T., Matasci, N., Narro, M. L., Piel, W. H., Mckay, S. J., Lowry, S., Freeland, C., Peet, R. K. and Enquist, B. J. 2013. The taxonomic name resolution service: an online tool for automated standardization of plant names. – BMC Bioinform. 14: 16.

Bradshaw, C. J. A., Reepmeyer, C., Saltré, F., Agapiou, A., Kassianidou, V., Demesticha, S., Zomeni, Z., Polidorou, M. and Moutsiou, T. 2024. Demographic models predict end-Pleistocene arrival and rapid expansion of pre-agropastoralist humans in Cyprus. – Proc. Natl Acad. Sci. USA 121: e2318293121.

Bruelheide, H. et al. 2018. Global trait–environment relationships of plant communities. – Nat. Ecol. Evol. 2: 1906–1917.

Bruelheide, H. et al. 2019. sPlot – a new tool for global vegetation analyses. – J. Veg. Sci. 30: 161–186.

Brummitt, N., Araújo, A. C. and Harris, T. 2021. Areas of plant diversity – what do we know? – Plants People Planet 3: 33–44.

Cai, L. et al. 2023. Global models and predictions of plant diversity based on advanced machine learning techniques. – New Phytol. 237: 1432–1445.

Cao Pinna, L., Gallien, L., Pollock, L. J., Axmanová, I., Chytrý, M., Malavasi, M., Acosta, A. T. R., Campos, J. A. and Carboni, M. 2024. Plant invasion in Mediterranean Europe: current hotspots and future scenarios. – Ecography 2024: e07085.

Carine, M. and Menezes de Sequeira, M. 2020. Sir Joseph Hooker on Insular Floras: human impact and the natural laboratory paradigm. – Rev. Sci. Insularum 3: 73–88.

Chamberlain, S., Oldoni, D. and Waller, J. 2022. rgbif: interface to the Global Biodiversity Information Facility API. – Zenodo, https://doi.org/10.5281/zenodo.1045299.

Chiarucci, A. 2007. To sample or not to sample? That is the question ... for the vegetation scientist. – Folia Geobot. 42: 209–216.

Chiarucci, A., Bacaro, G. and Scheiner, S. M. 2011. Old and new challenges in using species diversity for assessing biodiversity. – Philos. Trans. R. Soc. B 366: 2426–2437.

Chiarucci, A., Fattorini, S., Foggi, B., Landi, S., Lazzaro, L., Podani, J. and Simberloff, D. 2017. Plant recording across two centuries reveals dramatic changes in species diversity of a Mediterranean archipelago. – Sci. Rep. 7: 5415.

Chiarucci, A., Buldrini, F., Cervellini, M., Guarino, R., Caccianiga, M., Foggi, B., Viciani, D., Lazzaro, L., Casella, L., Angelini, P., Cerabolini, B. E. L., Pasta, S., Enea, M. and Zannini, P. 2021a. Habitat type and island identity as drivers of community assembly in an archipelago. – J. Veg. Sci. 32: jvs.12953.

Chiarucci, A., Guarino, R., Pasta, S., Rosa, A. L., Lo Cascio, P. L., Médail, F., Pavon, D., Fernández-Palacios, J. M. and Zannini, P. 2021b. Species–area relationship and small-island effect of vascular plant diversity in a young volcanic archipelago. – J. Biogeogr. 48: 2919–2931.

Chytrý, M. et al. 2016. European Vegetation Archive (EVA): an integrated database of European vegetation plots. – Appl. Veg. Sci. 19: 173–180.

Chytrý, M. et al. 2024. FloraVeg. EU – an online database of European vegetation, habitats and flora. – Appl. Veg. Sci. 27: e12798.

D'Antraccoli, M., Bacaro, G., Tordoni, E., Bedini, G. and Peruzzi, L. 2020. More species, less effort: designing and comparing sampling strategies to draft optimised floristic inventories. – Perspect. Plant Ecol. Evol. Syst. 45: 125547.

de Araujo, M. L., Quaresma, A. C. and Ramos, F. N. 2022. GBIF information is not enough: national database improves the inventory completeness of Amazonian epiphytes. – Biodivers. Conserv. 31: 2797–2815.

Denelle, P., Weigelt, P. and Kreft, H. 2023. GIFT – an R package to access the global inventory of floras and traits. – Methods Ecol. Evol. 14: 2738–2748.

Dengler, J. et al. 2020. Species–area relationships in continuous vegetation: evidence from Palaearctic grasslands. – J. Biogeogr. 47: 72–86.

Díaz, S. et al. 2016. The global spectrum of plant form and function. – Nature 529: 167–171.

Dřevojan, P., Čeplová, N., Štěpánková, P. and Axmanová, I. 2023. Life form. – www.FloraVeg.eu.

Gallo, T. and Waitt, D. 2011. Creating a successful citizen science model to detect and report invasive species. – BioScience 61: 459–465.

Garbolino, E., De Ruffray, P., Brisse, H. and Grandjouan, G. 2012. The phytosociological database SOPHY as the basis of plant socio-ecology and phytoclimatology in France. – Biodivers. Ecol. 4: 177–184.

Garcillán, P. P. and Ezcurra, E. 2011. Sampling procedures and species estimation: testing the effectiveness of herbarium data against vegetation sampling in an oceanic island. – J. Veg. Sci. 22: 273–280.

GBIF.org 2023. GBIF occurrence download (10 November 2023). – https://doi.org/10.15468/dl.4wk2wc.

GBIF.org 2024. Derived dataset GBIF.org (26 October 2024). Filtered export of GBIF occurrence data. – https://doi.org/10.15468/dd.9gpgjd.

GBIF.org 2026. GBIF home page. – https://www.gbif.org.

Govaerts, R. (ed.) 2025. WCVP: world checklist of vascular plants. – Facilitated by the Royal Botanic Gardens, Kew, http://sftp.kew.org/pub/data-repositories/WCVP.

Govaerts, R., Nic Lughadha, E., Black, N., Turner, R. and Paton, A. 2021. The World Checklist of Vascular Plants, a continuously updated resource for exploring global plant diversity. – Sci. Data 8: 215.

Guarino, R., Chytrý, M., Attorre, F., Landucci, F. and Marcenò, C. 2021. Alien plant invasions in Mediterranean habitats: an assessment for Sicily. – Biol. Invas. 23: 3091–3107.

Guarino, R., Guccione, M. and Gillet, F. 2022. Plant communities, synusiae and the arithmetic of a sustainable classification. – Veg. Classif. Surv. 3: 7–13.

Hamimeche, M., Véla, E., Gillet, F. and Moulaï, R. 2024. Biogeography and ecology of the Algerian island flora. – Plant Ecol. Evol. 157: Article 2.

Heberling, J. M., Miller, J. T., Noesgaard, D., Weingart, S. B. and Schigel, D. 2021. Data integration enables global biodiversity synthesis. – Proc. Natl Acad. Sci. USA 118: e2018093118.

Hortal, J., De Bello, F., Diniz-Filho, J. A. F., Lewinsohn, T. M., Lobo, J. M. and Ladle, R. J. 2015. Seven shortfalls that beset large-scale knowledge of biodiversity. – Annu. Rev. Ecol. Evol. Syst. 46: 523–549.

Hughes, K. A. and Convey, P. 2012. Determining the native/non-native status of newly discovered terrestrial and freshwater species in Antarctica –current knowledge, methodology and management action. – J. Environ. Manage. 93: 52–66.

Jeliazkov, A., Gavish, Y., Marsh, C. J., Geschke, J., Brummitt, N., Rocchini, D., Haase, P., Kunin, W. E. and Henle, K. 2022. Sampling and modelling rare species: conceptual guidelines for the neglected majority. – Global Change Biol. 28: 3754–3777.

Johnson, B. A., Mader, A. D., Dasgupta, R. and Kumar, P. 2020. Citizen science and invasive alien species: an analysis of citizen science initiatives using information and communications technology (ICT) to collect invasive alien species observations. – Global Ecol. Conserv. 21: e00812.

Klímová, B., Divíšek, J., Večeřa, M., Chytrý, M., Danihelka, J., Kaplan, Z., Novotný, P., Skokanová, H., Wild, J. and Axmanová, I. 2025. Spatial patterns of species richness in the Czech flora: effects of sampling intensity, environment and landscape history. – Preslia 97: 541–566.

Knollová, I. et al. 2024. ReSurveyEurope: a database of resurveyed vegetation plots in Europe. – J. Veg. Sci. 35: e13235.

König, C., Weigelt, P., Schrader, J., Taylor, A., Kattge, J. and Kreft, H. 2019. Biodiversity data integration — the significance of data resolution and domain. – PLoS Biol. 17: e3000183.

Lomba, A., Pellissier, L., Randin, C., Vicente, J., Moreira, F., Honrado, J. and Guisan, A. 2010. Overcoming the rare species modelling paradox: a novel hierarchical framework applied to an Iberian endemic plant. – Biol. Conserv. 143: 2647–2657.

Lomolino, M. V. 2004. Conservation biogeography. – In: Lomolino, M. V. and Heaney, L. R. (eds), Frontiers of biogeography: new directions in the geography of nature. Sinauer Assoc., pp. 293–296.

López-Guillén, E., Herrera, I., Bensid, B., Gómez-Bellver, C., Ibáñez, N., Jiménez-Mejías, P., Mairal, M., Mena-García, L., Nualart, N., Utjés-Mascó, M. and López-Pujol, J. 2024. Strengths and challenges of using iNaturalist in plant research with focus on data quality. – Diversity 16: 42.

Maddock, A. H. and Samways, M. J. 2000. Planning for biodiversity conservation based on the knowledge of biologists. – Biodivers. Conserv. 9: 1153–1169.

Maldonado, C., Molina, C. I., Zizka, A., Persson, C., Taylor, C. M., Albán, J., Chilquillo, E., Rønsted, N. and Antonelli, A. 2015. Estimating species diversity and distribution in the era of Big Data: to what extent can we trust public databases? – Global Ecol. Biogeogr. 24: 973–984.

Marbà, N., Krause-Jensen, D., Alcoverro, T., Birk, S., Pedersen, A., Neto, J. M., Orfanidis, S., Garmendia, J. M., Muxika, I., Borja, A., Dencheva, K. and Duarte, C. M. 2013. Diversity of European seagrass indicators: patterns within and across regions. – Hydrobiologia 704: 265–278.

Marcenò, C., Padullés Cubino, J., Chytrý, M., Genduso, E., Gristina, A. S., La Rosa, A., Salemi, D., Landucci, F., Pasta, S. and Guarino, R. 2021a. Plant hunting: exploring the behaviour of amateur botanists in the field. – Biodivers. Conserv. 30: 3265–3278.

Marcenò, C., Padullés Cubino, J., Chytrý, M., Genduso, E., Salemi, D., La Rosa, A., Gristina, A. S., Agrillo, E., Bonari, G., Giusso del Galdo, G., Ilardi, V., Landucci, F. and Guarino, R. 2021b. Facebook groups as citizen science tools for plant species monitoring. – J. Appl. Ecol. 58: 2018–2028.

Mason, B. M., Mesaglio, T., Barratt Heitmann, J., Chandler, M., Chowdhury, S., Gorta, S. B. Z., Grattarola, F., Groom, Q., Hitchcock, C., Hoskins, L., Lowe, S. K., Marquis, M., Pernat, N., Shirey, V., Baasanmunkh, S. and Callaghan, C. T. 2025. iNaturalist accelerates biodiversity research. – BioScience 75: 953–965.

Matthews, T. J. and Triantis, K. A. 2021. Island biogeography. – Curr. Biol. 31: R1201–R1207.

Matthews, T. J., Rigal, F., Triantis, K. A. and Whittaker, R. J. 2019a. A global model of island species–area relationships. – Proc. Natl Acad. Sci. USA 116: 12337–12342.

Matthews, T. J., Triantis, K. A., Whittaker, R. J. and Guilhaumon, F. 2019b. sars: an R package for fitting, evaluating and comparing species–area relationship models. – Ecography 42: 1446–1455.

McDonald, K. A. and Brown, J. H. 1992. Using montane mammals to model extinctions due to global change. – Conserv. Biol. 6: 409–415.

Médail, F. 2022. Plant biogeography and vegetation patterns of the Mediterranean islands. – Bot. Rev. 88: 63–129.

Médail, F. and Myers, N. 2004. Mediterranean Basin. – In: Mittermeier, R. A. and Robles Gil, P. (eds), Hotspots revisited: Earth's biologically richest and most endangered terrestrial ecoregions. CEMEX (Monterrey), Conservation International, pp. 144–147.

Médail, F. and Pasta, S. 2024. The intermittent history of exploitation of terrestrial biotic resources on the small islands of the western Mediterranean Basin. – Hum. Ecol. 52: 397–408.

Melis, R., Malavasi, M., Bazzichetto, M., Caria, M. C., Rivieccio, G., Denaro, A., Marchetto, E., Perrone, M., Livornese, M., Rocchini, D. and Bagella, S. 2025. Unravelling bias: a Sardinian perspective on taxonomic, spatial, and temporal biases in vascular plant biodiversity data from GBIF. – Ecol. Inform. 90: 103289.

Mesaglio, T., Sauquet, H., Coleman, D., Wenk, E. and Cornwell, W. K. 2023. Photographs as an essential biodiversity resource: drivers of gaps in the vascular plant photographic record. – New Phytol. 238: 1685–1694.

Mesaglio, T., Sauquet, H. and Cornwell, W. K. 2025. Citizen science records are fuelling exciting discoveries of new plant species. – Am. J. Bot. 112: e70048.

Mesaglio, T., Bean, A. R. and Bean, A. 2026. Rediscovery of a presumed extinct plant species, *Ptilotus senarius* (Amaranthaceae), through iNaturalist. – Aust. J. Bot. 74: BT25063.

Meyer, C., Weigelt, P. and Kreft, H. 2016. Multidimensional biases, gaps and uncertainties in global plant occurrence information. – Ecol. Lett. 19: 992–1006.

Michalcová, D., Lvončík, S., Chytrý, M. and Hájek, O. 2011. Bias in vegetation databases? A comparison of stratified-random and preferential sampling. – J. Veg. Sci. 22: 281–291.

Midolo, G. et al. 2024. Diversity and distribution of Raunkiær's life forms in European vegetation. – J. Veg. Sci. 35: e13229.

Moerman, D. E. and Estabrook, G. F. 2006. The botanist effect: counties with maximal species richness tend to be home to universities and botanists. – J. Biogeogr. 33: 1969–1974.

Mouillot, D. and Lepretre, A. 2000. Introduction of relative abundance distribution (RAD) indices, estimated from the rank-frequency diagrams (RFD), to assess changes in community diversity. – Environ. Monit. Assess. 63: 279–295.

2014. Patterns and processes in plant phylogeography in the Mediterranean Basin. A review. – Perspect. Plant Ecol. Evol. Syst. 16: 265–278.

Nori, J., Prieto-Torres, D. A., Villalobos, F., Loyola, R., Rojas-Soto, O., Parra, J. L., Lira-Noriega, A., Ortega-Andrade, H. M., Monjeau, A., Fuente, S. H. D. L., Martínez-Meyer, E. and Osorio-Olvera, L. 2023. Contrasting biogeographical patterns of threatened vertebrates on islands emerge from disparities between expert-derived maps and global biodiversity information facility data. – J. Biogeogr. 50: 418–427.

Orihuela-Rivero, R., Morente-López, J., Reyes-Betancort, J. A., Schaefer, H., Valido, A., Menezes de Sequeira, M., Romeiras, M. M., Góis-Marques, C. A., Salas-Pascual, M., Vanderpoorten, A., Fernández-Palacios, J. M. and Patiño, J. 2025. Geographic and biological drivers shape anthropogenic extinctions in the Macaronesian vascular flora. – Global Change Biol. 31: e70072.

Panitsa, M., Tzanoudakis, D. and Sfenthourakis, S. 2008. Turnover of plants on small islets of the eastern Aegean Sea within two decades. – J. Biogeogr. 35: 1049–1061.

Pausas, J. G. 1999. Mediterranean vegetation dynamics: modelling problems and functional types. – Plant Ecol. 140: 27–39.

Petřík, P., Pergl, J. and Wild, J. 2010. Recording effort biases the species richness cited in plant distribution atlases. – Perspect. Plant Ecol. Evol. Syst. 12: 57–65.

Popp, M. R., Zimmermann, N. E. and Brun, P. 2025. Evaluating the use of automated plant identification tools in biodiversity monitoring – a case study in Switzerland. – Ecol. Inform. 90: 103316.

Puddu, S., Podda, L., Mayoral, O., Delage, A., Hugot, L., Petit, Y. and Bacchetta, G. 2016. Comparative analysis of the alien vascular flora of Sardinia and Corsica. – Not. Bot. Horti Agrobot. Cluj-Napoca 44: 337–346.

Pullin, A. S. and Knight, T. M. 2005. Assessing conservation management's evidence base: a survey of management-plan compilers in the United Kingdom and Australia. – Conserv. Biol. 19: 1989–1996.

Rocchini, D. et al. 2023. A quixotic view of spatial bias in modelling the distribution of species and their diversity. – npj Biodiversity 2: 10.

Rondinini, C., Wilson, K. A., Boitani, L., Grantham, H. and Possingham, H. P. 2006. Tradeoffs of different types of species occurrence data for use in systematic conservation planning. – Ecol. Lett. 9: 1136–1145.

Sabatini, F. M. et al. 2022. Global patterns of vascular plant alpha diversity. – Nat. Commun. 13: 4683.

Santi, F. et al. 2024. MEDIS – a comprehensive spatial database on Mediterranean islands for biogeographical and evolutionary research. – Global Ecol. Biogeogr. 33: e13855.

Santi, F. et al. 2026. Data from: Plant diversity estimates of Mediterranean islands differ among biodiversity databases. – Zenodo Repository, https://doi.org/10.5281/zenodo.18339464.

Sastre, P. and Lobo, J. M. 2009. Taxonomist survey biases and the unveiling of biodiversity patterns. – Biol. Conserv. 142: 462–467.

Schrader, J., Wright, I. J., Kreft, H., Weigelt, P., Andrew, S. C., Abbott, I. and Westoby, M. 2023. ETIB-T: an equilibrium theory of island biogeography for plant traits. – J. Biogeogr. 50: 223–234.

Schrader, J., Weigelt, P., Cai, L., Westoby, M., Fernández-Palacios, J. M., Cabezas, F. J., Plunkett, G. M., Ranker, T. A., Triantis, K. A., Trigas, P., Kubota, Y. and Kreft, H. 2024. Islands are key for protecting the world's plant endemism. – Nature 634: 868–874.

Schrader, J. et al. 2025. A-islands: a vascular plant dataset for biodiversity research and species monitoring on Australian continental islands. – J. Veg. Sci. 36: e70019.

Seebens, H. et al. 2017. No saturation in the accumulation of alien species worldwide. – Nat. Commun. 8: 14435.

Tonelli, M. and Lobo, J. M. 2025. Human activities and dispersal capacity: key drivers of dung beetle colonization in Mediterranean and Macaronesian islands. – Biol. J. Linn. Soc. 144: blaf003.

Tordoni, E., Carmona, C. P., Toussaint, A., Tamme, R. and Pärtel, M. 2024. Global patterns and determinants of multiple facets of plant diversity. – Global Ecol. Biogeogr. 33: e13823.

Troudet, J., Grandcolas, P., Blin, A., Vignes-Lebbe, R. and Legendre, F. 2017. Taxonomic bias in biodiversity data and societal preferences. – Sci. Rep. 7: 9132.

van Kleunen, M. et al. 2019. The global naturalized alien Flora (GloNAF) database. – Ecology 100: e02542.

Vogiatzakis, I. N., Mannion, A. M. and Sarris, D. 2016. Mediterranean island biodiversity and climate change: the last 10 000 years and the future. – Biodivers. Conserv. 25: 2597–2627.

Warren, B. H. et al. 2015. Islands as model systems in ecology and evolution: prospects fifty years after MacArthur-Wilson. – Ecol. Lett. 18: 200–217.

Weigelt, P., König, C. and Kreft, H. 2020. GIFT – A Global Inventory of Floras and Traits for macroecology and biogeography. – J. Biogeogr. 47: 16–43.

Wenk, E., Mesaglio, T., Keith, D. and Cornwell, W. 2024. Curating protected area-level species lists in an era of diverse and dynamic data sources. – Ecol. Inform. 84: 102921.

Whittaker, R. J., Fernández-Palacios, J. M., Matthews, T. J., Borregaard, M. K. and Triantis, K. A. 2017. Island biogeography: taking the long view of nature's laboratories. – Science 357: eaam8326.

Winters, G., Beer, S., Willette, D. A., Viana, I. G., Chiquillo, K. L., Beca-Carretero, P., Villamayor, B., Azcárate-García, T., Shem-Tov, R., Mwabvu, B., Migliore, L., Rotini, A., Oscar, M. A., Belmaker, J., Gamliel, I., Alexandre, A., Engelen, A. H., Procaccini, G. and Rilov, G. 2020. The tropical seagrass *Halophila stipulacea*: reviewing what we know from its native and invasive habitats, alongside identifying knowledge gaps. – Front. Mar. Sci. 7: 300.

Wüest, R. O., Zimmermann, N. E., Zurell, D., Alexander, J. M., Fritz, S. A., Hof, C., Kreft, H., Normand, S., Cabral, J. S., Szekely, E., Thuiller, W., Wikelski, M. and Karger, D. N. 2020. Macroecology in the age of Big Data – where to go from here? – J. Biogeogr. 47: 1–12.

Zizka, A., Silvestro, D., Andermann, T., Azevedo, J., Duarte Ritter, C., Edler, D., Farooq, H., Herdean, A., Ariza, M., Scharn, R., Svantesson, S., Wengström, N., Zizka, V. and Antonelli, A. 2019. coordinateCleaner: standardized cleaning of occurrence records from biological collection databases. – Methods Ecol. Evol. 10: 744–751.