VEGETATION CLASSIFICATION AND SURVEY

International Association for Vegetation Science (IAVS)

Vegetation Classification and Survey 7: 49–63
doi: 10.3897/VCS.176349

RESEARCH PAPER

SPECIAL COLLECTION ISLANDS AND ARCHIPELAGOS

A quantitative classification of chorotypes for the Socotra Archipelago's non-endemic plants

Dario La Montagna[1], Giuliano Fanelli[1], Michele De Sanctis[1], Alessio Farcomeni[2], Petr Maděra[3], Luca Malatesta[1], Kay Van Damme[3], Fabio Attorre[1]

1 *Botanical Garden of Rome, Department of Environmental Biology, Sapienza University of Rome, Rome, Italy*

2 *Department of Enterprise Engineering, University of Rome Tor Vergata, Rome, Italy*

3 *Department of Forest Botany, Dendrology and Geobiocoenology, Faculty of Forestry and Wood Technology, Mendel University in Brno, Brno, Czechia*

Corresponding author: Dario La Montagna (dario.lamontagna.6@gmail.com)

Academic editor: Riccardo Guarino ♦ **Linguistic editor:** Hallie Seiler
Received 31 October 2025 ♦ **Accepted** 13 March 2026 ♦ **Published** 20 May 2026

Abstract

Aims: To identify groups of species with similar geographic ranges in the non-endemic flora of the Socotra Archipelago, place them within a global biogeographic context, and examine how these chorotypes are reflected in Socotran vegetation types. **Study area:** The Socotra Archipelago (Yemen) and its surroundings, analysed against a 1° × 1° global grid to provide global context and a Socotra-centred regional focus. **Methods:** We assembled 4,493,566 GBIF records for 477 target taxa; 417 had usable occurrences and 389 met a ≥ 15-occurrence threshold for modelling. To obtain coarse-grained ranges, we fitted ensemble models and binarised predictions using conservative thresholds. Grid cells were grouped according to species turnover and geographic proximity using clustering methods, and statistical tests identified species associated with particular clusters. A second analysis focused on a Socotra-centred subset to resolve finer regional patterns. Species-level chorotypes were linked to synoptic vegetation tables to characterise vegetation types. **Results:** The global analysis identified ten spatially cohesive groups of species at the continent scale. The Socotra-centred analysis resolved finer regional structure. Socotra's chorological spectrum is dominated by the Xeric Tropical-Subtropical chorotype (13.8%), with contributions from the Paleoarabian-Somali-Masai chorotype (10.1%) and Temperate-Subtropical Belt chorotype (9.2%). Vegetation types differed markedly in their chorological signatures: endemic-rich woodlands and shrublands showed high specialisation, whereas grasslands and halophytic communities exhibited more mixed spectra. **Conclusions:** Turnover-based partitions recover major Afro-Arabian structures and highlight the Red Sea-Gulf of Aden rim as a potential biogeographic hinge. The dominance of xeric Afro-Arabian chorotypes, together with a secondary signal from mesic Afrotropical and Indo-Pacific groups, suggests a flora structured by arid corridors and monsoon- or fog-related refugia. Linking chorotypes with vegetation types reveals an ecological gradient from endemic-rich communities to more open assemblages with broader biogeographic affinities. Despite known limitations, the approach provides reproducible co-distribution units connecting classical chorology with quantitative vegetation analysis.

Keywords

Chorotype classification, non-endemic plant, plant bioregionalisation, quantitative analysis, Socotra Archipelago, species co-distribution

Introduction

Identifying groups of species that share coherent geographic ranges, usually referred to as "chorotypes", has long provided a compact way to summarise complex biogeographic structure and to generate hypotheses about shared history, dispersal barriers, and environmental filtering (Baroni-Urbani et al. 1978; Olivero et al. 2011; Fattorini

2015). The term was originally introduced to describe sets of species showing similar geographic distributions, and has since been widely used, particularly in European biogeography, to characterise the floristic composition and ecological affinities of plant communities (Baroni-Urbani et al. 1978; Passalacqua 2015). Conceptually, chorotypes can be understood either as sets of species whose overall ranges overlap (global chorotypes) or as sets whose within-region distributions are similar (regional chorotypes), distinctions that matter for interpreting processes operating across scales (Fattorini 2015, 2016). In practice, chorotypes serve as indicators of spatial co-occurrence patterns that may reflect common evolutionary legacies, long-standing climatic constraints, or contemporary connectivity and fragmentation (Olivero et al. 2011; Fattorini 2015; García-Carrasco et al. 2023). Whereas provinces and ecoregions typically partition space into non-overlapping units of communities or floras, chorotypes are species-set centric: they summarise recurrent co-distribution across space. They can legitimately span multiple regions or appear as partially overlapping belts (Fattorini 2015).

The idea of classifying geographical ranges has deep roots in European plant ecology. Classic "*Arealkunde*" framed the description and interpretation of species' areas, codifying range forms and their historical-environmental determinants (Walter and Straka 1970). This long effort, in particular by Central European authors, culminated in the monumental *Vergleichende Chorologie der Zentraleuropäischen Flora* that assembled synoptic range maps and typologies that remain as guidelines for diagnosing chorological patterns and for coining families of "areal types" (Meusel et al. 1965). However, a formal assignment of the entire flora of a European country to chorotypes is rare. Apart from the diagnosis of Meusel et al. (1965), probably the most important effort is the classification into geographic areas of the flora of Italy by Pignatti (1982). The formalisation of the "*Arealkunde*" approach yields interesting insights into ecology and can also be used to validate phytosociological classifications (Nimis and Bolognini 1993). Geographical distributions can also help derive ecological indicators (Berg et al. 2017).

Yet, despite this long history, explicit chorotype analysis has been applied unevenly outside Europe and remains underused for global syntheses. This is due in part to issues with data heterogeneity and sampling bias in occurrence datasets, but also because robust multivariate tools for detecting co-distribution have matured only recently, and that there has been some confusion on the definition of chorotypes (Boakes et al. 2010; Morrone 2014; Fattorini 2015; Vilhena and Antonelli 2015).

There is value in revisiting chorotypes with modern, data-driven methods at both macroecological and regional scales. Global-scale analyses can recover coarse-grained biogeographic signals linked to deep history and broad environmental gradients, while nested regional analyses can resolve finer-grain structure relevant to local endemism and stewardship (Wüest et al. 2020). Islands and continental fragments are particularly informative in this respect because they concentrate lineages under sharp environmental and geographic filters, producing mosaics of turnover and endemism that are ideal for testing chorotype logic (Warren et al. 2015).

The Socotra Archipelago (Yemen) is part of the Socotra Platform on the western Indian Ocean margin. Oceanic rifting in the Gulf of Aden began ~20–17 Ma, accompanied by substantial uplift, but biogeographically relevant land emergence in the area dates back at least to the late Eocene (38–34 Ma); subsequent sea-level oscillations repeatedly exposed parts of the platform through the Quaternary (Culek 2013). These dynamics set the stage for intermittent connectivity with Africa and Arabia. The archipelago is characterised by pronounced topographic and climatic gradients (granite Haggeher mountains and extensive karstic limestones) that structure habitats from fog-wet escarpments to arid plateaus. These settings have maintained both wet and dry refugia, which are repeatedly implicated in the archipelago's diversification and the present distribution of endemics (Banfield et al. 2011). Socotra hosts ~835 vascular plant species, ~308 of them endemic (~37%), with endemics concentrated disproportionately in wet refugia on sea-facing escarpments and highlands (Miller and Morri 2004).

The emergence of global biodiversity databases (e.g., Chytrý et al. 2016; GBIF 2025) together with today's computational power, has been a game-changer for both chorotype analysis and bioregionalisation, enabling quantitative patterns to be assessed with unprecedented resolution (e.g. Ball-Damerow et al. 2019; Heberling et al. 2021; Malatesta et al. 2023).

Methodologically, advances in similarity- and distance-based clustering, network community detection, and fuzzy set approaches now allow more objective identification of co-distributional units and better handling of transitional zones and sampling artefacts, challenges that historically complicated chorotype work (Olivero et al. 2011; Vilhena and Antonelli 2015). Combined with careful treatment of occurrence biases, these tools make it feasible to align classical chorological reasoning with modern macroecological data.

Recent global syntheses for plants underscore the value of quantitative delineation. A phylogeny-aware global regionalisation of vascular plants reveals a deep Laurasia-Gondwana split and clarifies cross-realm links (Carta et al. 2022). An updated map of floristic realms using distributions for > 12,000 angiosperm genera likewise identifies eight realms and 16 sub-realms and traces their assembly through time (Liu et al. 2023). These studies, together with network/community-detection approaches for delineating bioregions (e.g., Infomap Bioregions), highlight how co-distribution signal and phylogenetic turnover jointly inform regional boundaries and their temporal dynamics.

Against this background, our contribution is not to supplant the qualitative tradition, like the expert-driven regionalisations of Africa and the Middle East, with Takhtajan's floristic realms, regions and provinces (with emphasis on floristics and endemism), White's African

centres of endemism and transition mosaics, and Zohary's climatic belts and phytogeographic districts for the Middle East (Zohary 1973; White 1983; Takhtajan 1986). Instead, we aim for a complementary approach, quantitatively re-interrogating its boundaries and centres using modern co-distribution tools on harmonised global data. We use a global dataset of native non-endemic species of Socotra and a two-tier framework to: (1) identify emergent chorotypes at global extent from 1° grid data, emphasising *turnover*-based co-distribution patterns; (2) resolve global chorotypes in and around the Socotra Archipelago by re-clustering a geographically restricted subset of cells and species; and (3) link species-level chorotypes to vegetation types of Socotra in order to explore how biogeographic affinities are reflected in community composition. We then interpret these chorotypes in a biogeographic context, discussing their potential implications for environmental filters, regional connectivity, and historical legacies relevant to Socotra and surrounding regions.

Materials and methods

Study scope and species occurrence data

We targeted the native non-endemic plant species recorded for Socotra and characterised their distributions on a 1° × 1° global grid. Because chorotypes are defined as recurrent patterns of geographic co-distribution, strictly endemic species were excluded, as their ranges are restricted to the archipelago and therefore do not inform broader regional or global co-distribution patterns. Occurrence records for the 477 target species were downloaded from (GBIF 2025), retaining only those with valid geographic coordinates and excluding fossils, living specimens from cultivation (e.g. botanical gardens), records flagged as non-native, and observations with spatial uncertainty > 10 km or with missing uncertainty metadata. Non-native species were excluded from all analyses because their distributions reflect recent anthropogenic dispersal rather than historical biogeographical processes.

Records were then overlaid on the 1° grid following the approach of (Kreft and Jetz 2010), retaining at most one presence per species per cell (coded 1; absence coded 0). To obtain the final dataset used for species distribution modelling and subsequent chorotype analyses, we required a minimum of 15 occurrences per species as a threshold for model calibration (van Proosdij et al. 2016). All the analyses in this paper were performed in R (R Core Team 2025).

Environmental predictors and species distribution modelling (SDMs)

Climatic predictors were sourced from CHELSA (Karger et al. 2023) and resampled to the 1° analysis grid. Starting from an initial stack of 36 variables, we reduced multicollinearity using variance inflation factor (VIF) screening with a threshold of 5 using the R package 'usdm' (Naimi et al. 2014) and inspected pairwise correlations with a correlation plot. The final predictor set comprised Bio03 (isothermality), Bio07 (temperature annual range), Bio08 (mean temperature of wettest quarter), Bio15 (precipitation seasonality), Bio19 (precipitation of coldest quarter), humidity range (Hursrange), net primary productivity (NPP, as a coarse integrative proxy of ecosystem productivity), potential evapotranspiration range (PET range), and mean wind speed (Windmean), plus a longitude layer to help separate western and eastern hemispheres (acknowledging the circular nature of longitude).

We used species distribution models (SDMs) purely as a step to regularise GBIF occurrences and derive coarse-grained (1°) potential species ranges for subsequent chorotype analyses, not for ecological inference on predictors (Beck et al. 2014). Accordingly, standard model configuration was chosen to ensure consistency and scalability across a large number of species, rather than to maximise predictive accuracy for individual taxa, which would require species-specific tuning and extensive optimisation (Valavi et al. 2022). SDMs were fitted in 'biomod2' (Guéguen et al. 2025) using Gradient Boosting Machines (GBM), Random Forest (RF), and MaxEnt. For each species, we set a fixed seed (123), generated pseudo-absences with a disk strategy (3 replicates; 1,000 PAs per replicate; minimum/maximum distances 1,000–10,000 km), and split the data 80/20 for training/testing with three random cross-validation replicates. Single models were evaluated with ROC (AUC) and TSS, and ensembles retained only single models exceeding ROC ≥ 0.7 and TSS ≥ 0.6. Because we aimed to delineate broad ranges, we converted continuous suitability to binary presence/absence using the TSS-maximising threshold, prioritising sensitivity to limit commission errors and spurious spillover at 1° (Hellegers et al. 2025). Although threshold selection and binarisation may introduce uncertainty and information loss (Liu et al. 2013; Guillera-Arroita et al. 2015), especially for presence-only data, this step is a pragmatic compromise. Ensemble predictions were then projected globally and binarised per 1° cell to reconstruct the species × grid matrix of modelled presences. After removing cosmopolitan and very narrowly distributed species, this matrix formed the input for chorotype calculations.

Chorotype delineation at coarse global scale: ecological-geographical clustering

We aggregated species' binary ranges into a presence-absence matrix by grid cell. To avoid artefacts in between-cell compositional dissimilarity at coarse-grained scales, we filtered the species set by excluding cosmopolitan taxa known to inflate similarity across distant regions and removing very rare species occurring in ≤ 3 grid cells (to limit singletons).

Between-cell ecological dissimilarity was quantified with the Sørensen family of β-diversity using the turnover component (β_sim) from 'betapart' (Baselga and Orme

2012), following the guidelines from Kreft and Jetz (2010); we denote the resulting dissimilarity matrix as D_{ecol}. Geographical separation was computed as the great-circle (haversine) distance between cell centroids and min–max normalised to [0,1], yielding D_{geo}. We combined the two sources of dissimilarity with the convex mixing of ClustGeo (Chavent et al. 2018):

$$D_\alpha = (1-\alpha)D_{\text{ecol}} + \alpha D_{\text{geo}}$$

We scanned α in {0.0, 0.1, …, 0.9} and the number of clusters k in {5, …, 15}. For each (α, k), we partitioned cells by Partitioning Around Medoids (PAM) on D_α and computed within-cluster inertia curves as a diagnostic tool. Inspection of inertia curves across (α, k) and average silhouette width identified a parsimonious solution balancing ecological turnover and spatial cohesion; we retained $\alpha = 0.5$ and $k = 10$ and assigned final cluster memberships via PAM on $D_{0.5}$.

From clusters to species-chorotype relationships

To test whether a species is present in some clusters over others, we adopted a use-availability framework originally developed for habitat-selection studies in animals (Neu et al. 1974; Manly et al. 2004), transferring it to plant distributions across clusters. Although conceptually related to indicator-species approaches such as the framework of Dufrêne and Legendre (1997), our objective was not to identify species characteristic of single clusters, but to quantify their associations with one or multiple clusters and retain full compositional signatures across space. In our setting, "availability" is the spatial support of each cluster and "use" is the realised set of grid cells occupied by a species.

First, for each cluster k we computed its availability as the fraction of equal-area cells it covers, $p_k = N_k / N$ (with N_k cells in cluster k and N the total). For a given species, we then counted how many template cells it occupies in total (n) and how many of those fall in cluster k (x_k), which gives the species' observed proportion in that cluster, $\hat{p}_k^{\text{obs}} = x_k/n$. We measured the deviation from availability-based expectations using an availability-corrected enrichment, defined as a selection ratio (use/availability), conceptually equivalent to species fidelity indices used in vegetation ecology (Cáceres and Legendre 2009). For a given species and cluster k, enrichment was calculated as:

$$E_k = \frac{\hat{p}_k^{\text{obs}}}{p_k}, \log_2 E_k = \log_2\left(\frac{\hat{p}_k^{\text{obs}}}{p_k}\right)$$

Where values $E_k > 1$ mean the species appears in cluster k more often than expected from availability alone. For ease of interpretation, the ratio was $\log_2$-transformed, so that values of 0 indicate no deviation from expectation, positive values indicate higher-than-expected occupancy, and negative values indicate lower-than-expected occupancy. Statistically, we asked whether the count x_k could plausibly arise if occurrences were placed at random according to availability, by testing H_0:$x_k \sim \text{Binomial}(n, p_k)$ against the one-sided alternative $x_k > np_k$. We adjusted the resulting p-values across clusters within each species using Benjamini-Hochberg FDR. A species was deemed enriched in cluster k only if three conditions were all met: FDR-adjusted $q_k < 0.05$, effect size $\log_2 E_k \geq 0.8$ (i.e., at least 1.74× the null expectation), and a minimum support of $x_k \geq 5$ cells. This yields a binary species-by-cluster table flagging significant enrichments and for each species, a compositional signature p_k^{obs} describing how its occurrences are distributed across clusters.

Pattern mining on enriched associations

Our goal was not to redraw geographic regions, but to identify chorotypes, i.e., groups of species that share similar geographic distribution. To achieve this, we adopted a two-step procedure: first, grid cells were clustered into spatially coherent units based on species turnover and geographic distance; second, species were grouped into chorotypes according to their proportional occurrences across these spatial clusters. In this framework, a chorotype may occupy several geographic clusters identified through the cluster analysis, and in fact, it is defined by species composition, not by a single spatial cluster.

First, using the enrichment results, we treated each species as a set of geographic clusters in which it was enriched. To define chorotypes quantitatively, we compared species by their full compositional signatures across clusters. We built a species × cluster matrix of observed proportions (p_k^{obs}) and computed Jensen-Shannon distances between species. This metric quantifying the dissimilarity between probability distributions is symmetric and bounded, and is therefore appropriate for comparing compositional signatures of species across clusters, where values represent proportions rather than absolute counts (Lin 1991). Species were then clustered with Partitioning Around Medoids (PAM) on this distance matrix. We explored k = 2–40 and chose k by maximising mean silhouette width; for reporting and comparability with the coarse analysis, we retained a parsimonious k = 10 solution and used medoids as examples of each chorotype.

Regional refinement of chorotypes

To resolve the within-group structure that is blurred at the global extent and coarse-grained, we performed a restricted chorotype analysis centred on the region of interest (Socotra and surroundings). This step refines a subset of global macro-chorotypes into finer, regionally coherent groupings, which hereafter we call "meso-chorotypes", and reduces confounding from long-distance biogeographic breaks. These terms are purely operational labels to distinguish patterns at different spatial scales, and both correspond to global chorotypes *sensu* Fattorini (2015). In practice, they correspond to species with a broad and a narrow distribution range.

We first selected a set of global macro-chorotypes which exclude North and South America and Oceania. On the restricted grid, we recomputed between-cell ecological dissimilarity with the same distances as before. We scanned $\alpha \in \{0.0,0.1,\ldots,0.9\}$ and $k \in \{5,\ldots,30\}$, partitioned with PAM on D_α, and used within-cluster inertia to select a parsimonious solution. This yielded a locally finer partition (e.g., $\alpha = 0.5$, $k = 25$) used as the cluster raster for subsequent steps.

We repeated the availability-corrected enrichment test exactly as above to obtain a binary species × cluster matrix and per-species proportional signatures. We computed Jensen-Shannon distances among species' proportional signatures (ε-smoothing = 1×10^{-9}) and clustered species with PAM on this distance matrix. We explored $k = 2$–40 and chose k by maximum mean silhouette; the retained solution was $k = 12$, reflecting greater heterogeneity at the restricted extent.

Chorotype names were assigned after the analyses to aid interpretation and mapping. For each chorotype, we inspected its mapped extent and chose a geographic anchor based on the major area covered and the medoid cell, added habitat/biome qualifiers reflecting the prevailing environments within the group's footprint at 1°, and noted biogeographic affinities using standard terminology from regional floristic schemes.

Species that were not classified chorotypically by the main pipeline (e.g., excluded by data or occurrence thresholds) were not used in any statistics. For completeness, we provide a qualitative crosswalk in which each such species' known distribution was visually compared with the mapped chorotypes and annotated with the most compatible chorotype(s), following the common practice when chorotypes are already defined in the literature.

Chorological characterisation of vegetation types

To link chorological patterns with vegetation units, synoptic tables of vegetation types according to De Sanctis et al. (2013) were combined with species-level chorotype assignments. For each vegetation type, species frequencies were aggregated by chorotype and normalised to obtain relative chorotype contributions. The chorological signature of each vegetation type was summarised by identifying the dominant chorotype and the three most represented chorotypes. To quantify the degree of chorological specialisation versus mixing, we calculated Shannon diversity of chorotypes as a measure of chorological evenness.

Results

Data filtering, environmental predictors and SDMs summary

We retrieved GBIF occurrence records for Socotra's species (4,493,566 records, Suppl. material 1). Of the 477 target taxa, 417 had usable GBIF occurrences and were retained for analysis. Then we applied an ≥ 15-occurrence filter, which yielded 389 species to be modelled and excluded 28 species for insufficient data.

From the initial set of predictors, we retained ten variables for SDMs: Bio03 (isothermality), Bio07 (annual range of air temperature), Bio08 (mean daily mean air temperatures of the wettest quarter), Bio15 (precipitation seasonality), Bio19 (mean monthly precipitation amount of the coldest quarter), Hursrange (annual range of monthly near-surface relative humidity), NPP (net primary productivity), PET range (annual range of monthly potential evapotranspiration), Windmean (mean monthly near-surface wind speed), and longitude. 26 variables were removed for collinearity issues. We successfully fitted SDMs for 389 species. Average test performances across species are summarised in Table 1.

Looking at the model performance indicators, for the AUC on average, the models can be classified as excellent for all three algorithms. The TSS also offers good results for model performance, with values exceeding 0.8. The TSS had an average value of 0.826 for the ensemble models. This was the only performance metric we used to merge single models. From the binary maps of species resulting from the SDMs and the 28 species not used for modelling but useful for the next part of the analysis, we achieved 114,219 occurrences in the final matrix, covering 7,969 cells and a total of 417 species.

Global macro-chorotypes

Using ClustGeo mixing, scanning $\alpha \in [0, 0.9]$ and $k = 5$–15, the inertia surface identified a parsimonious solution at $\alpha = 0.5$, $k = 10$ (Suppl. material 1). The average silhouette of the retained PAM partition was 0.325.

The ten clusters form continent-scale, spatially coherent groups at 1°. In the Western Hemisphere, C5 and C10 cover North/Central America and South America, respectively. In the Eastern Hemisphere, Africa is split into two clusters, the Arabian Peninsula appears with extensions

Table 1. Performance of species distribution models used to derive 1° presence-absence ranges. Values are mean ± SD across modelled species for Area Under the ROC Curve (AUC) and True Skill Statistic (TSS) on training and testing data for Gradient Boosting Machines, Random Forest, and MaxEnt.

Model	AUC training (mean ± SD)	AUC testing (mean ± SD)	TSS training (mean ± SD)	TSS testing (mean ± SD)
GBM	0.993 ± 0.002	0.981 ± 0.030	0.947 ± 0.043	0.852 ± 0.040
Random Forest	0.989 ± 0.005	0.965 ± 0.010	0.986 ± 0.021	0.797 ± 0.025
MaxEnt	0.953 ± 0.040	0.928 ± 0.064	0.885 ± 0.082	0.829 ± 0.029
Ensemble	—	—	0.939 ± 0.048	0.826 ± 0.031

into the Horn of Africa and the Indian region, Australia forms a distinct cluster, and C2 spans the Afro-Mediterranean belt and much of the Levant region (Figure 1).

Availability-corrected tests (binomial one-sided with BH FDR) identified significant species-to-cluster enrichments (Suppl. material 2). Per cluster, enriched species counts were: C1 [45], C2 [85], C3 [65], C4 [4], C5 [8], C6 [230], C7 [195], C8 [179], C9 [19], C10 [8]. Since the total enriched counts are 838, it is clear that more than one species shares the same cluster.

In fact, Table 2 shows the combinations of clusters associated with each chorotype identified by the PAM clustering on compositional signatures, together with the number of species assigned. We ultimately retained a ten-group solution based on the highest mean silhouette (0.381), which forms the distinction of ten different chorotypes based on different cluster combinations. The cluster combination C6-C7-C8 includes the largest number of species, chiefly covering Africa and the Arabian Peninsula, whereas the least represented combination is C3-C6-C9, with only 16 species, spanning the Arabian Peninsula, India, Indochina, and Australia.

Table 2. Global chorotypes defined by clustering species' compositional signatures (proportions of occurrences across spatial clusters). The retained solution (k = 10) yields chorotypes M1–M10, each summarised by the combination of global clusters (C1–C10).

Chorotype ID	Cluster ID	Species Countsnumber
M1	C2-C6-C7-C8	36
M2	C3-C6-C8	38
M3	C3-C4-C6-C7-C9	22
M4	C6-C7-C8	79
M5	C3-C6-C7-C8-C9	66
M6	C5-C7-C8-C9-C10	24
M7	C2-C6-C8	37
M8	C3-C6-C9	16
M9	C6	56
M10	C1-C2-C3-C6-C9	44

Global meso-chorotypes

We retained global macro-chorotypes 1, 2, 4, 7, and 9 from the previous cluster analysis (which correspond to cluster 2, 3, 6, 7, and 8) to build a restricted dataset more focused on the surroundings of Socotra Island. After filtering, the analysis comprised 4,049 cells and 242 species.

Scanning $\alpha \in [0, 0.9]$ and k = 5–30, the selected partition was α = 0.5, k = 25, with a mean silhouette of 0.3438806 (Suppl. material 1). Here, the partition is finer, with many clusters that nevertheless delineate geographic areas clearly (Figure 2). For example, clusters 13 and 16 occupy the central-eastern belt of the Arabian Peninsula and its southern sector with extensions into the Horn of Africa, respectively. Across Africa, there is also a clear regionalisation, with several groups separating west from east and likewise north from south. Madagascar is, moreover, treated as a distinct cluster.

Availability-corrected tests (one-sided binomial with Benjamini-Hochberg FDR) showed a wide disparity in species enrichment among clusters: several have low or even zero counts (e.g. clusters 1, 2, 3, 5, and 10), whereas others are strongly enriched, with 174 species in cluster 16, 121 in cluster 19, and 119 in cluster 13. Nonetheless, the overall number of enriched species is again high (1,062), indicating that many species share the same clusters and thereby providing a robust basis for chorotype delineation (Suppl. material 3).

Proceeding to the next step, we examined combinations of clusters together with species' compositional signatures and obtained a 12-group solution, selected by a mean silhouette of 0.374. Given the larger number of groups

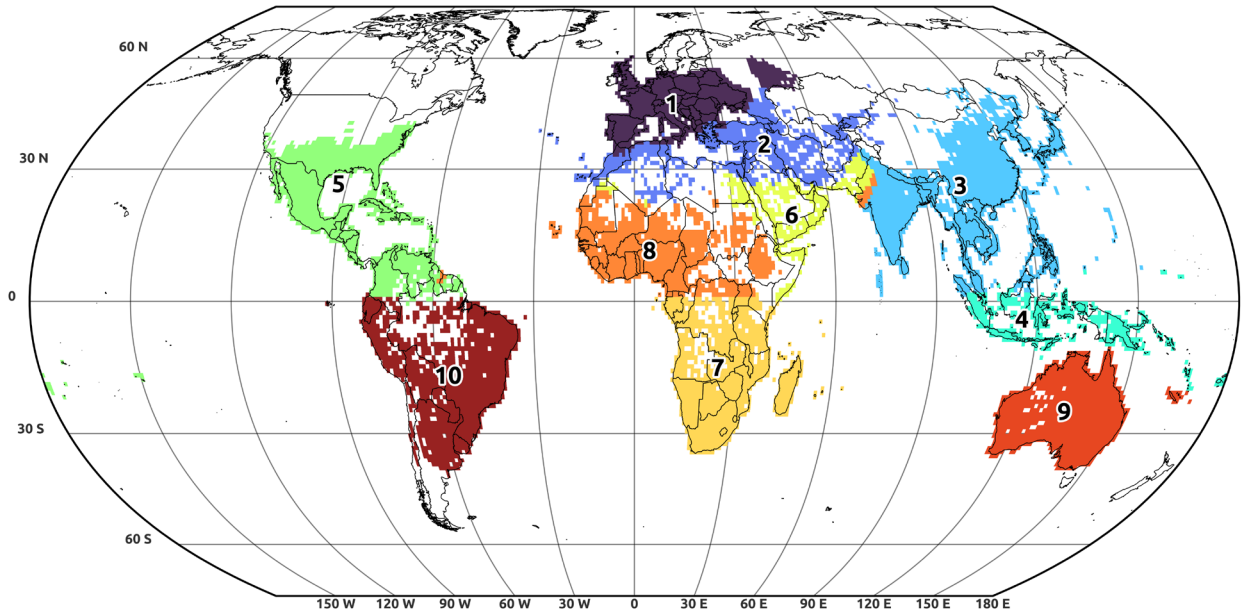

Figure 1. Global clusters found at 1° grid cells for the macro-chorotype analysis. Numbers (1–10) indicate the spatial clusters (C1–C10) highlighted by different colours.

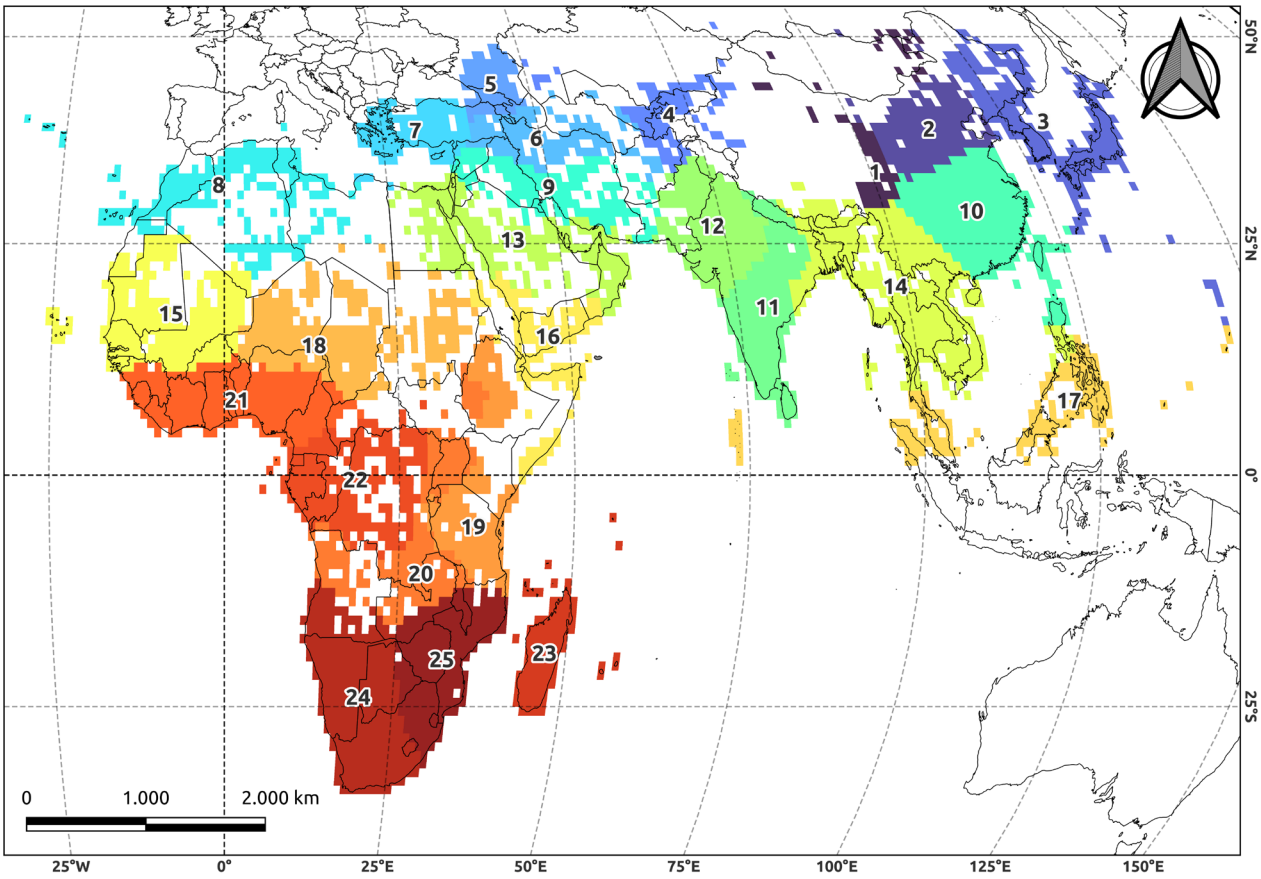

Figure 2. Global clusters found at 1° grid cells focused on the Socotra Archipelago surroundings for the meso-chorotype analysis. Numbers (1–25) indicate the spatial clusters (C1–C25) highlighted by different colours.

arising from the initial species-level clustering, the combinations available to define chorotypes are correspondingly richer and span many geographic areas (see Table 3). Even so, the number of species per chorotype is broadly stable across groups. Moreover, the clusters with low counts of species from the enrichment analysis are missing from the results and from the final cluster analysis.

Having established these results, we mapped both macro- and meso-chorotypes to provide a spatial depiction of their distributions, together with a detailed description (Suppl. material 4). By adding previously excluded species (e.g. cosmopolitans or taxa with too few usable occurrences) and assigning them to the most compatible chorotypes, we compiled a chorological spectrum for Socotra's non-endemic flora (Suppl. material 1). This spectrum is dominated by the Xeric Tropical-Subtropical chorotype (13.8%), followed by the Paleoarabian-Somali-Masai chorotype (10.1%) and the Temperate-Subtropical Belt (9.2%). The least represented chorotype is the Western Indian Ocean Humid Arc (0.8%) (Figure 3).

Table 3. Global chorotypes for the Socotra Archipelago surroundings defined by clustering species' compositional signatures (proportions of occurrences across spatial clusters). The retained solution (k = 12) yields chorotypes D1–D12, each summarised by the combination of global clusters (C1–C25).

Chorotype ID	Cluster ID	Species Countsnumber
D1	C7-C8-C9-C12-C13-C15-C16-C18-C19-C20-C21-C22-C24-C25	27
D2	C11-C12-C13-C15-C16-C18-C19-C21	22
D3	C11-C12-C14-C15-C18-C21	8
D4	C16-C19-C20	21
D5	C9-C12-C13-C16-C18-C19	16
D6	C6-C8-C9-C12-C13-C15-C16-C18	27
D7	C16	16
D8	C13-C16	23
D9	C18-C19-C20-C21-C22-C23-C24-C25	20
D10	C16-C19-C20-C22-C23-C24-C25	30
D11	C6-C9-C12-C13-C16	28
D12	C16-C19-C23	4

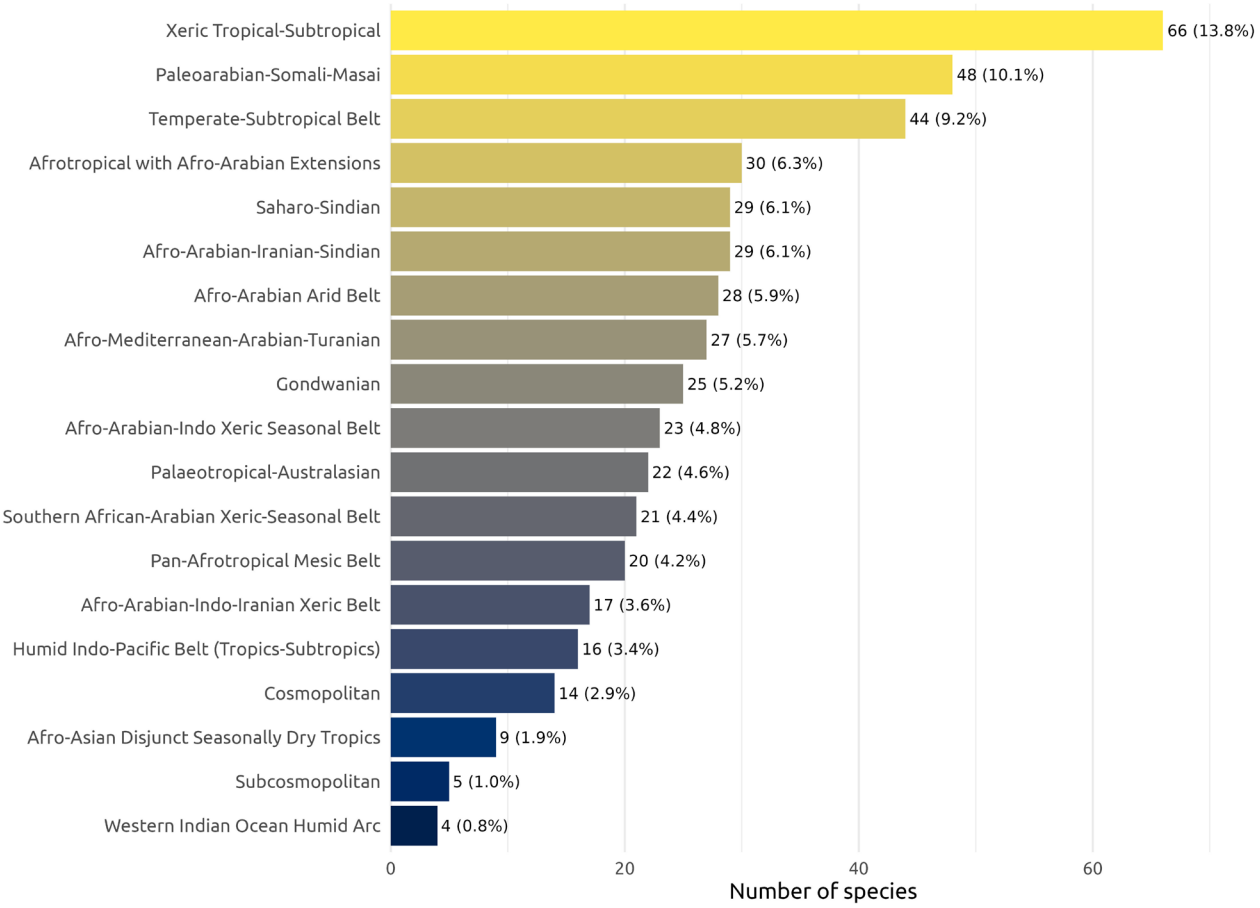

Figure 3. Chorological spectrum of non-endemic plant species occurring in the Socotra Archipelago, based on the resulting species-level chorotype assignments.

Chorological signatures of Socotran vegetation types

Across vegetation types, chorological signatures showed marked differences in both dominant biogeographic affinities and degree of mixing (Figure 4; Suppl. material 5). Most woodland and shrubland types were strongly dominated by endemic elements, with dominance values often exceeding 55–70% (e.g. BaW, ChHsS, CoW), suggesting high chorological specialisation. In contrast, several halophytic and grassland types showed lower dominance and higher Shannon diversity, reflecting a more mixed chorological composition (e.g. PaG, AaTaG). Shannon entropy revealed a gradient from highly specialised vegetation types, characterised by strong prevalence of a single chorotype, to more heterogeneous assemblages in which multiple chorotypes contribute substantially to community composition.

Discussion

Modern discourse on plant bioregionalisation in Africa and the Middle East draws on two intertwined traditions. One, largely European in origin, catalogued 'areal types' and chorotypes, emphasising floristic composition (presence/absence of diagnostic families, genera, species) and, at times, physiognomy (life forms, structure) to summarise recurring range patterns (e.g. Grisebach 1872; Schimper 1898). The other, broader bioregionalisation *sensu lato*, has combined floristics with explicit criteria of endemism and species turnover, and, more recently, quantitative clustering and phylogenetic evidence to delimit realms, regions and transition zones (Kreft and Jetz 2010; Daru et al. 2017; Liu et al. 2023). Landmark syntheses (e.g. Good 1974; Walter 1979; Takhtajan 1986) formalised hierarchical floristic units, while later African and Middle Eastern frameworks (e.g. Zohary 1973; White 1983) weighted centres of endemism, diagnostic taxa and environmental structuring to draw boundaries and recognise mosaics. Within Africa, White (1983) replaced the older, rigid hierarchy with an explicitly endemism-based framework: regional centres of endemism (thresholds > 1,000 endemic species and > 50% of species restricted), separated by transition zones and mosaics. This re-specified the continent in terms of centres and corridors of diversity and later informed links into Southwest Asia.

For the Middle East and Arabian Peninsula, Zohary (1973) argued that reliable regional boundaries should initially follow climatic zones, then be corroborated by diagnostic markers: high endemic richness and proportion, centres of diversification for particular clades, distinct geological-historical context, recipient/donor status of floras, and characteristic vegetation belts (including altitudinal

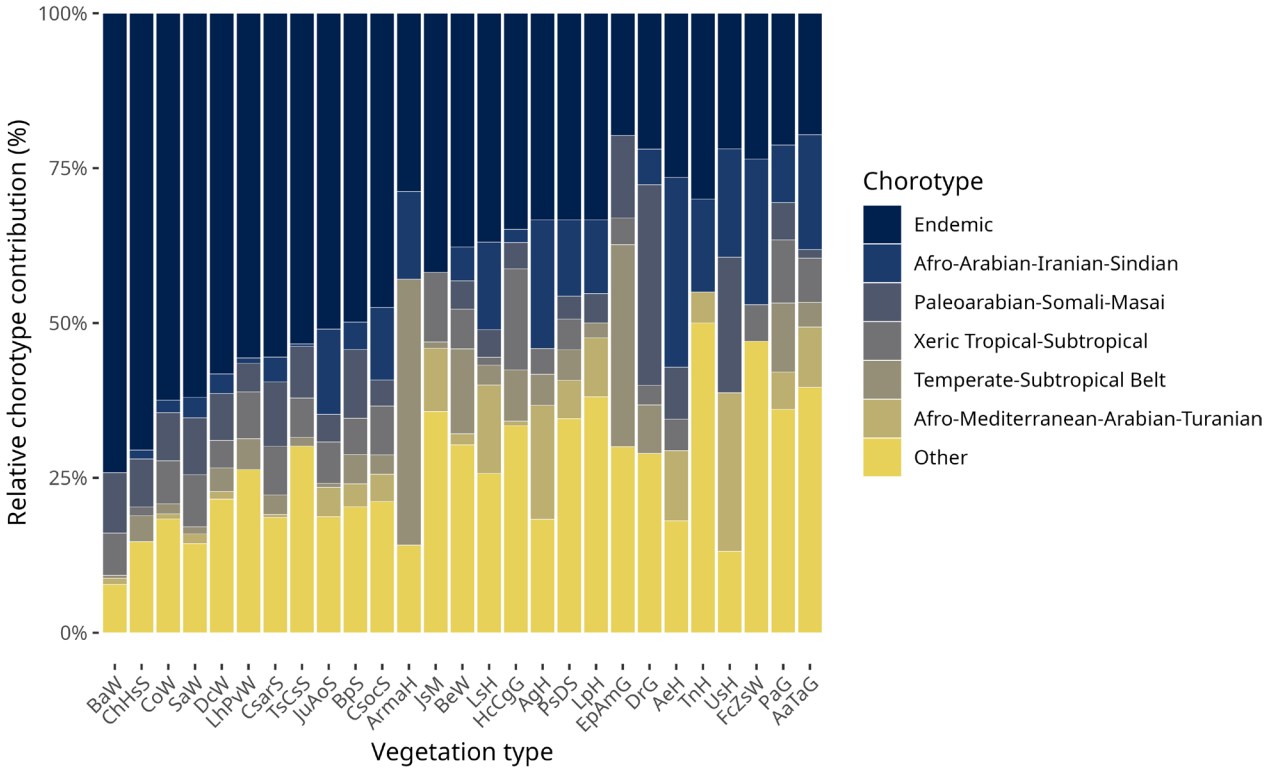

Figure 4. Chorological composition of Socotran vegetation types based on species-level chorotype assignments. Only the five most frequent chorotypes are shown explicitly; all remaining chorotypes are grouped as "Other". Vegetation types are ordered by decreasing dominance of the leading chorotype. Vegetation-type codes follow De Sanctis et al. (2013): BaW = *Boswellia ameero* woodland; ChHsS = *Coelocarpum haggierensis-Hypericum scopulorum* shrubland; CoW = *Commiphora ornifolia* woodland; SaW = *Sterculia africana* woodland; DcW = *Dracaena cinnabari* woodland; LhPvW = *Leucas haggh-ierensis-Pittosporum viridiflorum* woodland; CsarS = *Croton sarcocarpus* shrubland; TsCsS = *Trichodesma scottii-Cepha-locroton socotranus* shrubland; JuAoS = *Jatropha unicostata-Adenium obesum* shrubland; BpS = *Buxanthus pedicellatus* shrubland; CsocS = *Croton socotranus* shrubland; ArmaH = *Arthrocnemum macrostachyum* halophytic community; JsM = *Juncus socotranus* marshland; BeW = *Boswellia elongata* woodland; LsH = *Limonium sokotranum* halophytic community; HcCgG = *Heteropogon contortus-Chrysopogon serrulatus* grassland; AgH = *Atriplex griffithii* halophytic community; LpH = *Limonium paulayanum* halophytic community; PsDS = *Pulicaria stephanocarpa* dwarf shrubland; EpAmG = *Eragrostis papposa-Arthraxon micans* grassland; DrG = *Dactyloctenium robecchii* grassland; AeH = *Acacia edgeworthii* community; TnH = *Tamarix nilotica* community; UsH = *Urochondra setulosa* community; FcZsW = *Ficus cordata-Ziziphus spina-christi* woodland; PaG = *Panicum atrosanguineum* grassland; AaTaG = *Aristida adscensionis-Tephrosia apollinea* grassland.

zonation). This program acknowledged the porous nature of boundaries across transitional deserts and mountain arcs, anticipating that "hard lines" would be elusive.

These classic maps were, by necessity, synthetic and qualitative. They relied on expert assessment of floristic composition, endemism, and environmental context at a time when regional checklists were incomplete and quantitative data sparse; even late syntheses on Arabia explicitly noted the prematurity of quantitative delineation and the difficulty of drawing precise lines across transition zones.

Global signal and regional refinement

The chorotype analysis for Socotra suggests a split into two broad categories: macro-chorotypes with wide distributions, and meso-chorotypes with more restricted extents. Macro-chorotypes represent distributional patterns that span entire floristic realms and may reflect wide ecological tolerances or effective dispersal. Meso-chorotypes, by contrast, have narrower ranges and are typically tightly connected to the ecology and palaeoecological history of Africa and Eurasia. Socotra's meso-chorotypes often display relictual or disjunct characters. The relatively coarse 1° grid is nevertheless sufficient for the spatial scale presented here.

It is important not to confuse our analysis, based on present-day geographical patterns, with a phylogeographic analysis, which would consider the clade rather than the species as the unit of study and would therefore allow the placement of endemics within broader chorotypes. Unfortunately, this is not currently feasible because genetic knowledge of Socotran, and indeed African, endemic species remains too limited and patchy. While some emblematic lineages have been studied, multi-taxon phylogeographic datasets are still scarce (Miller and Morris 2004; Linder 2014). It should also be noted that the

phylogeographic analyses carried out so far, with some exceptions, have mostly investigated colonisation processes rather than characterising chorotypes (Sanbonmatsu and Spalink 2022; Španiel and Rešetnik 2022; Dantas-Queiroz et al. 2023). As our study shows, a chorotype framework allows one to test hypotheses not only about migration and dispersal, but also about how ranges contract and expand in response to ecological change through time.

Agreement and departures from classical regionalisations

The units we retrieved often align with classic qualitative frameworks. In particular, the clusters encompassing the Horn of Africa and southern Arabia echo Somalia-Masai and Eritreo-Arabian concepts, whereas arid belts across northern Arabia and the Sahara match Saharo-Arabian/Saharo-Sindian delimitations. Socotra's placement within an Indian Ocean-facing tropical unit, yet with strong links to the southern Arabian escarpments, is consistent with its historical treatment as part of the Eritreo-Arabian complex and, in some schemes, as a distinct Socotra Province. Where our units diverge from classic outlines, three design choices plausibly contribute: our global step uses turnover-only dissimilarity (emphasising replacements over richness gradients), we analyse non-endemic taxa at coarse grain, and we reflect contemporary ranges that integrate recent dynamics. In these places, the 'lines' drawn historically may better represent broad transition zones than clear-cut borders. Our results should therefore be seen as complementary to those frameworks, highlighting patterns of regional connectivity and a shared species pool rather than centres of endemism. Miller and Morris (2004) do not formally define chorotypes in a quantitative sense but remain the most comprehensive qualitative synthesis for interpreting the chorology of the Socotran flora. They identify strong affinities with southern Arabia and the Horn of Africa, connections with the Somali-Masai region, and several distant disjunctions. These patterns are broadly consistent with those retrieved by our analysis. The dominance of xeric tropical-subtropical and Paleoarabian-Somali-Masai chorotypes in our results mirrors the emphasis placed by Miller and Morris (2004) on Eritreo-Arabian and Somalia-Masai elements as the primary non-endemic components of the flora.

Our study differs in approach rather than in conclusions. Indeed, Miller and Morris (2004) rely on expert floristic knowledge and historical-ecological reasoning, and we recover similar biogeographic signals through a quantitative analysis of species co-distribution at global scale, formalising chorological relationships long recognised in the literature.

Environmental filters and historical legacies

The flora shows well-known disjunctions (Tethyan, African dry-corridor, and boreotropical) and mixed affinities with the Old-World tropics, northeast tropical Africa/Arabia, and the Indian subcontinent (Mies 1996; Banfield et al. 2011). The spatial coherence of macro-chorotypes can be interpreted as consistent with combined effects of environmental filtering (aridity and seasonal monsoon regimes, persistent coastal fog, elevation-driven moisture) and historical legacies that limit exchange across barriers (e.g., sea straits, desert interiors) while fostering diversification in refugial uplands. Within the restricted domain, it highlights nested assemblages along escarpments and montane belts, consistent with upslope range stacking and asymmetric source-sink dynamics from mesic refugia to surrounding xeric lowlands.

The chorological spectrum of Socotra's non-endemic flora is dominated by xeric, tropical-subtropical Eastern Hemisphere chorotypes, suggesting that the island sits within a thermo-xeric macro-network linking Northeast Africa, southern Arabia, and adjoining arid belts. This pattern is consistent with classical placements of southern Arabia within Somalo-Arabian/Eritreo-Arabian phytochoria and with a long history of aridity-adapted lineages moving along desert and semi-desert corridors (Zohary 1973; White 1983; Takhtajan 1986).

Several lineages suggest deeper antecedents to these xeric affinities. Parts of the Somali-Masai complex may retain a signature of Miocene expansion and fragmentation of seasonally dry tropical formations across northeast Africa and Arabia (Trauth et al. 2005, 2009), before their subsequent contraction during late Miocene to Pliocene aridification (deMenocal 2004). Under this view, high counts in the Paleoarabian-Somali-Masai chorotype (Figure 3) could represent composite outcomes of the Neogene assembly of dry biomes under progressive aridity and tectonic reorganisation (Linder 2014), including the uplift of East African and Arabian topography and the strengthening of rain-shadow and monsoon seasonality (Sepulchre et al. 2006). These features were later modulated by Quaternary sea-level oscillations that periodically opened and closed coastal corridors around the Red Sea and Gulf of Aden (Siddall 2004; Lambeck et al. 2014), with less distance between the Horn of Africa and Socotra Archipelago during lowered sea levels. While our data cannot date chorotypes, the geographic concordance, with strong occupancy along rift-parallel uplands and the Red Sea margins, fits expectations from this two-stage scenario. Although certain Afro-Arabian lineages are very old (e.g. *Dirachmaceae*), Neogene aridification profoundly reshaped the flora, obscuring much of its original signature (deMenocal 2004; Linder 2014). We therefore interpret the corridor signal not as exclusively late-Quaternary, but as a plausible scenario involving recurrent re-activation of older xeric linkages whose footprints were established during the Miocene and later modulated by glacial-interglacial sea-level lows.

A smaller but coherent contribution from Afrotropical and Indo-Pacific-aligned chorotypes accords with monsoon and fog influences on the south-facing escarpments and highlands. Conversely, occasional ties to intercontinental seasonally dry belts reflect long-range xeric linkages rather than recent exchange across humid barriers and appear as higher-order units that pair coastal-desert

clusters with steppe/plateau units. This mirrors what was already found by Banfield et al. (2011).

Interpreted through use/availability-corrected enrichments, Socotra appears as a gateway "sink/waystation" for thermo-xerophilous dispersers. Non-endemic species associated with coastal-desert chorotypes accumulate on the archipelago, whereas strictly mesic elements show weaker or patchier signals. As such, this conclusion reflects study design as much as biological reality; deeper insights are possible only with phylogeographic analysis. This asymmetry is consistent with the selective permeability of the corridor: efficient for dry-adapted lineages moving along coastal and plateau routes, restrictive for humid-adapted taxa constrained to fog-fed enclaves.

These interpretations are, for now, partly speculative because it is not possible to date chorotypes. However, this may become feasible in future by studying recurrent speciation patterns using molecular phylogenies. Although currently constrained by data scarcity and by the lack of statistical methods for comparing clades across different genera, such an approach would open a new phase in phylogeography, one in which the traditional chorotype framework is combined with molecular phylogenetics.

Taken together, these patterns point to a small number of dominant floristic assembly pathways shaping Socotran flora: a primary Afro-Arabian xeric corridor linking southern Arabia and the Horn of Africa, and a secondary, moisture-mediated pathway associated with monsoon-influenced escarpments and refugial uplands. As such, Socotra behaves as a continental-fragment island whose non-endemic flora largely reflects structured regional connectivity. Compared to the Arabian Peninsula and the Horn of Africa, Socotra acts less as a source area and more as a selective sink, where regional floristic signals are filtered, compressed, and preserved under insular conditions.

Interpreting enrichment and multi-cluster patterns

Availability-corrected enrichments identify species that occur in particular chorotypes more often than expected given each chorotype's spatial availability. This yields per-chorotype sets of indicator species that remain comparable even when chorotypes differ greatly in spatial extent. To compare species consistently across the full set of chorotypes, we then summarised for each species a compositional signature (the proportions of its occurrences in each chorotype) and measured between-species similarity using Jensen-Shannon distance, which is appropriate for compositional data and relatively insensitive to sparse zeros. Clustering species with PAM on this metric returns quantitative chorotypes with visually interpretable medoids and standard diagnostics of within-group cohesion (silhouettes). Taken together, availability-corrected enrichment and clustering of compositional signatures delineate reproducible co-distribution units, separating thermo-xeric coastal belts from more mesic montane complexes, and revealing where species share multi-chorotype affinities (e.g. coastal and plateau combinations around the Red Sea and Gulf of Aden). We emphasise that these are spatial co-occurrence patterns, not evidence of interaction; their value lies in defining operational units and testable hypotheses about environmental constraints and historical connectivity.

Chorotypes and vegetation classification

The chorological spectra observed across vegetation types provide a quantitative link between species biogeography and community composition. By linking species-level chorotypes to synoptic vegetation tables, our results suggest that chorological structure is not randomly distributed across vegetation types but instead reflects coherent biogeographic signatures of plant communities and an ecological gradient across vegetation types. Woodland and shrubland vegetation units are characterised by a strong dominance of endemic chorotypes, highlighting the importance of local evolutionary history in structuring these assemblages. Conversely, grasslands and halophytic communities show more heterogeneous chorological spectra, with a stronger contribution of Afro-Arabian and Temperate-Subtropical chorotypes, suggesting higher connectivity with regional species pools and higher turnover along environmental gradients and, in many cases, recent arrival from the continent. Overall, the chorotype composition reflects a gradient from locally specialised, refuge-like communities to more open, climatically driven assemblages with broader biogeographic affinities, a pattern also observed in other studies where endemic-rich communities correspond to more centralised and ecologically specialised vegetation types (Pignatti et al. 1991). For example, *Croton sarcocarpus* shrubland (CsarS), *Boswellia ameero* woodland (BaW), and *Buxanthus pedicellatus* shrubland (BpS) (see Suppl. material 5) show a higher proportion of the Paleoarabian-Somali-Masai chorotype than other woody communities, indicating a connection with ancient, relict vegetation types. Conversely, *Eragrostis papposa-Arthraxon micans* grassland (EpAmG) and *Arthrocnemum macrostachyum* halophytic communities (ArmaH) are dominated by the Temperate-Subtropical Belt and other chorotypes that suggest continued immigration from the mainland, despite the presence of a substantial endemic component.

This quantitative integration of chorotypes with vegetation types supports earlier qualitative interpretations of Socotra's vegetation mosaics (De Sanctis et al. 2013). Similar approaches have been applied in European vegetation studies where chorological spectra are used to characterise plant communities and interpret their ecological and biogeographic affinities (Nimis and Bolognini 1993; Berg et al. 2017).

Limitations and robustness

Our workflow was developed specifically to derive chorotypes for Socotra's non-endemic species from the data

currently available. It is still experimental and has several caveats that deserve explicit mention. First, GBIF sampling is uneven; we therefore have not used SDMs for ecological inference, and we binarised predictions conservatively to limit commission at the 1° grain. Because we relied on random (not spatial) cross-validation and pseudo-absences, absolute performance metrics may be optimistic. The downstream analyses are based on binary ranges and between-cell dissimilarity at a coarse spatial resolution and should therefore be interpreted primarily as broad-scale chorological patterns rather than precise distributional estimates. Second, the clustering solution depends on the choice of α (the mixing parameter) and k (the number of groups). Our selected values are supported by inertia and silhouette diagnostics, but they are not unique; importantly, the qualitative structure appeared stable under small changes in α, k, and in the β-diversity component used, although a formal sensitivity analysis was beyond the scope of this study. Third, for species-cluster enrichment, we controlled multiplicity within species using Benjamini-Hochberg FDR, and we report effect sizes ($\log_2 E$) to summarise the strength of associations across species rather than inflating claims with additional multiple testing. Finally, the scope is deliberately narrow: analyses are single-taxon and presence-based; we did not model phylogenetic relatedness, functional traits, or temporal dynamics, which remain important avenues for future work.

Implications and outlook

Our analysis differs from what is highlighted in studies of European chorotypes, where chorotypes are closely tied to macroclimatic patterns, as strongly argued by Meusel et al. (1965) and Berg et al. (2017). By contrast, the chorotypes we identify cannot be fully explained by ecology alone; they require a historical interpretation. Socotra's non-endemic species appear to retain signatures consistent with geological and climatic processes operating since the Neogene and Quaternary periods.

Chorotypes are routinely used in Europe to characterise plant communities ecologically and biogeographically (Pignatti et al. 1991; Dzwonko and Loster 2000; Košir et al. 2013; Mahmutaj et al. 2015). The gradient identified above allows the different vegetation communities of Socotra to be distinguished not only floristically, but also based on their chorological spectra, thus supporting the classification of De Sanctis et al. (2013). An interesting feature of our study is that, in the case of Socotra, it becomes possible not only to interpret the ecology of species and communities, but also their evolutionary history. This will not only refine their ecological characterisation but also help interpret community evolution. Evolutionary theory is strongly centred on population genetics; introducing chorotypes may shift the emphasis to the broader community context in which populations evolve through species interactions.

Conclusions

We combined turnover-based bioregionalisation at 1° with a restricted, equal-area refinement to delineate chorotypes from global to regional scales. The global macro-chorotypes recover traditionally recognised phytochoria across Africa and Southwest Asia, while the Socotra-centred analysis resolves nested sub-units and transition zones linked to orography and coastal climatic gradients. Availability-corrected enrichment provides interpretable indicator sets and recurrent range linkages, turning qualitative regional concepts into quantitative, reproducible units. Although our workflow inherits biases from occurrence data and modelling choices, key patterns are robust to reasonable parameter variation. The approach is modular and general: it can be reused for other taxa, extended with phylogenetic and functional information, and repeated through time to assess change. For Socotra and neighbouring regions, our results highlight representative cells for each chorotype and emphasise the importance of coastal and montane corridors for maintaining floristic connectivity. More broadly, chorotype-based analyses offer a potential bridge between classical phytosociology and modern vegetation ecology, linking community-level classifications with large-scale biogeographic patterns and ecological gradients, as long emphasised in European chorological traditions. In sum, modern co-distribution tools allow classical chorology to be quantified, mapped, and monitored, providing a practical scaffold for hypothesis-driven biogeography and conservation planning.

Data availability

All data supporting the findings of this study are provided in the Supplementary material.

Author contributions

Conceptualisation: DLM, GF. Methodology: DLM, AF, LM. Data curation: DLM. Formal analysis: DLM, AF. Writing — original draft: DLM, GF. Writing — review & editing: DLM, GF, MDS, AF, PM, LM, KVD, FA. Supervision: FA, PM, KVD.

Artificial Intelligence (AI) use

Regarding the use of AI in the preparation of this manuscript, the authors declare the following:

We used AI-assisted tools (e.g., ChatGPT) solely for English translation from Italian and grammar editing. The authors verified all content for accuracy and are fully responsible for the manuscript's scientific integrity; no AI was used to generate study design, analyses, results, or conclusions.

Acknowledgments

We sincerely thank the three anonymous reviewers for their thorough and constructive comments, which greatly improved the clarity and quality of the manuscript.

This study was partially supported by the Franklinia Foundation within the project "Conservation of the endangered endemic Boswellia trees on Socotra Island (Yemen)" Phase I (Grant No. 2020-03) and Phase II (Grant No. 2023-03).

References

Ball-Damerow JE, Brenskelle L, Barve N, Soltis PS, Sierwald P, Bieler R, LaFrance R, Ariño AH, Guralnick RP (2019) Research applications of primary biodiversity databases in the digital age. PLoS ONE 14: e0215794. https://doi.org/10.1371/journal.pone.0215794

Banfield LM, Van Damme K, Miller AG (2011) Evolution and biogeography of the flora of the Socotra archipelago (Yemen). In: Bramwell D, Caujapé-Castells J (Eds), The Biology of island floras. Cambridge University Press, Cambridge, GB, 197–225. https://doi.org/10.1017/CBO9780511844270.009

Baroni-Urbani C, Rufo S, Vigna-Taglianti A (1978) Materiali per una biogeografia Italiana fondata su alcuni generi di Coleotteri, Cicindelidi, Carabidi e Crisomelidi. Estratto della Memorie della Societa Entomologica Italiana 56: 35–92.

Baselga A, Orme CDL (2012) betapart: an R package for the study of beta diversity. Methods in Ecology and Evolution 3: 808–812. https://doi.org/10.1111/j.2041-210X.2012.00224.x

Beck J, Böller M, Erhardt A, Schwanghart W (2014) Spatial bias in the GBIF database and its effect on modeling species' geographic distributions. Ecological Informatics 19: 10–15. https://doi.org/10.1016/j.ecoinf.2013.11.002

Berg C, Welk E, Jäger EJ (2017) Revising Ellenberg's indicator values for continentality based on global vascular plant species distribution. Applied Vegetation Science 20: 482–493. https://doi.org/10.1111/avsc.12306

Boakes EH, McGowan PJK, Fuller RA, Chang-qing D, Clark NE, O'Connor K, Mace GM (2010) Distorted views of biodiversity: spatial and temporal bias in species occurrence data. PLoS Biology 8: e1000385. https://doi.org/10.1371/journal.pbio.1000385

Cáceres MD, Legendre P (2009) Associations between species and groups of sites: indices and statistical inference. Ecology 90: 3566–3574. https://doi.org/10.1890/08-1823.1

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: 1494–1504. https://doi.org/10.1111/nph.17844

Chavent M, Kuentz-Simonet V, Labenne A, Saracco J (2018) ClustGeo: an R package for hierarchical clustering with spatial constraints. Computational Statistics 33: 1799–1822. https://doi.org/10.1007/s00180-018-0791-1

Chytrý M, Hennekens SM, Jiménez-Alfaro B, Knollová I, Dengler J, Jansen F, Landucci F, Schaminée JHJ, Aćić S, … Yamalov S (2016) European Vegetation Archive (EVA): an integrated database of European vegetation plots. Applied Vegetation Science 19: 173–180. https://doi.org/10.1111/avsc.12191

Culek M (2013) Geological and morphological evolution of the Socotra Archipelago (Yemen) from the biogeographical view. Journal of Landscape Ecology 6: 84–108. https://doi.org/10.2478/jlecol-2014-0005

Dantas-Queiroz MV, Hurbath F, de Russo Godoy FM, Lanna FM, Versieux LM, Palma-Silva C (2023) Comparative phylogeography reveals the demographic patterns of neotropical ancient mountain species. Molecular Ecology 32: 3165–3181. https://doi.org/10.1111/mec.16929

Daru BH, Elliott TL, Park DS, Davies TJ (2017) Understanding the processes underpinning patterns of phylogenetic regionalization. Trends in Ecology & Evolution 32: 845–860. https://doi.org/10.1016/j.tree.2017.08.013

De Sanctis M, Adeeb A, Farcomeni A, Patriarca C, Saed A, Attorre F (2013) Classification and distribution patterns of plant communities on Socotra Island, Yemen. Applied Vegetation Science 16: 148–165. https://doi.org/10.1111/j.1654-109X.2012.01212.x

deMenocal PB (2004) African climate change and faunal evolution during the Pliocene–Pleistocene. Earth and Planetary Science Letters 220: 3–24. https://doi.org/10.1016/S0012-821X(04)00003-2

Dufrêne M, Legendre P (1997) Species assemblages and indicator species: the need for a flexible asymmetrical approach. Ecological Monographs 67: 345–366. https://doi.org/10.1890/0012-9615(1997)067[0345:SAAIST]2.0.CO;2

Dzwonko Z, Loster S (2000) Syntaxonomy and phytogeographical differentiation of the *Fagus* woods in the Southwest Balkan Peninsula. Journal of Vegetation Science 11: 667–678. https://doi.org/10.2307/3236574

Fattorini S (2015) On the concept of chorotype. Journal of Biogeography 42: 2246–2251. https://doi.org/10.1111/jbi.12589

Fattorini S (2016) A history of chorological categories. History and Philosophy of the Life Sciences 38: 12. https://doi.org/10.1007/s40656-016-0114-1

García-Carrasco J-M, Muñoz A-R, Olivero J, Figuerola J, Fa JE, Real R (2023) Gone (and spread) with the birds: Can chorotype analysis highlight the spread of West Nile virus within the Afro-Palaearctic flyway? One Health 17: 100585. https://doi.org/10.1016/j.onehlt.2023.100585

GBIF (2025) What is GBIF? GBIF: The Global Biodiversity Information Facility. [Available from:] https://www.gbif.org/what-is-gbif [accessed 04/07/2025]

Good R (1974) The geography of the flowering plants. 4th ed. Longman, London, GB.

Grisebach A (1872) Die Vegetation der Erde nach ihrer klimatischen Anordnung. Engelmann, Leipzig, DE, 603 pp.

Guéguen M, Blancheteau H, Thuiller W (2025) biomod2: Ensemble Platform for Species Distribution Modeling. [Available from:] https://biomodhub.github.io/biomod2/ [accessed 01/02/2025]

Guillera-Arroita G, Lahoz-Monfort JJ, Elith J, Gordon A, Kujala H, Lentini PE, McCarthy MA, Tingley R, Wintle BA (2015) Matching distribution models to applications. Global Ecology and Biogeography 24: 276–292. https://doi.org/10.1111/geb.12268

Heberling JM, Miller JT, Noesgaard D, Weingart SB, Schigel D (2021) Data integration enables global biodiversity synthesis. Proceedings of the National Academy of Sciences of the USA 118: e2018093118. https://doi.org/10.1073/pnas.2018093118

Hellegers M, Van Hinsberg A, Lenoir J, Dengler J, Huijbregts MAJ, Schipper AM (2025) Multiple threshold-selection methods are needed to binarise species distribution model predictions. Diversity and Distributions 31: e70019. https://doi.org/10.1111/ddi.70019

Karger DN, Lange S, Hari C, Reyer CPO, Conrad O, Zimmermann NE, Frieler K (2023) CHELSA-W5E5: daily 1 km meteorological forcing data for climate impact studies. Earth System Science Data 15: 2445–2464. https://doi.org/10.5194/essd-15-2445-2023

Košir P, Casavecchia S, Čarni A, Škvorc Ž, Zivkovic L, Biondi E (2013) Ecological and phytogeographical differentiation of oak-hornbeam forests in southeastern Europe. Plant Biosystems 147: 84–98. https://doi.org/10.1080/11263504.2012.717550

Kreft H, Jetz W (2010) A framework for delineating biogeographical regions based on species distributions. Journal of Biogeography 37: 2029–2053. https://doi.org/10.1111/j.1365-2699.2010.02375.x

Lambeck K, Rouby H, Purcell A, Sun Y, Sambridge M (2014) Sea level and global ice volumes from the Last Glacial Maximum to the Holocene. Proceedings of the National Academy of Sciences of the USA 111: 15296–15303. https://doi.org/10.1073/pnas.1411762111

Lin J (1991) Divergence measures based on the Shannon entropy. IEEE Transactions on Information Theory 37: 145–151. https://doi.org/10.1109/18.61115

Linder HP (2014) The evolution of African plant diversity. Frontiers in Ecology and Evolution 2: Article 38. https://doi.org/10.3389/fevo.2014.00038

Liu C, White M, Newell G (2013) Selecting thresholds for the prediction of species occurrence with presence-only data. Journal of Biogeography 40: 778–789. https://doi.org/10.1111/jbi.12058

Liu Y, Xu X, Dimitrov D, Pellissier L, Borregaard MK, Shrestha N, Su X, Luo A, Zimmermann NE, … Wang Z (2023) An updated floristic map of the world. Nature Communications 14: Article 2990. https://doi.org/10.1038/s41467-023-38375-y

Mahmutaj E, Shuka L, Xhulaj M, Hoda P, Mersinllari M (2015) Rare and endemic plants in the southern mountain ecosystems of Albania, their threats and diversity. Albanian Journal of Agricultural Sciences 14(1): 1–10.

Malatesta L, De Sanctis M, Ammann E, Attorre F, Buffi F, Cambria VE, Fratarcangeli C, Hoda P, Mahmutaj E, … Fanelli G (2023) Bioregionalization of Albania: Mismatch between the flora and the climate suggests that our models of Southern European bioregions are in need of a revision. Folia Geobotanica 58: 71–87. https://doi.org/10.1007/s12224-023-09432-7

Manly BFJ, MacDonald LL, Thomas DL, McDonald TL, Erickson WP (2004) Resource selection by animals: statistical design and analysis for field studies. 2nd ed. Kluwer Academic Publishers, Dordrecht, NL. https://doi.org/10.1007/0-306-48151-0

Meusel H, Eckehart J, Weinert E (1965) Vergleichende Chorologie der zentraleuropäischen Flora. Gustav Fischer Verlag, Jena, DE.

Mies B (1996) The phytogeography of Soqotra: evidence for disjunctive taxa, especially with Macaronesia. In: Dumont HJ (Ed.) Soqotra: Proceedings of the First International Symposium on Soqotra Island, Aden 1996: Present and Future. United Nations Development Programme, Conservation and Sustainable Use of Biodiversity of Soqotra Archipelago, Technical Series, Volume 1. New York, US, 83–105.

Miller AG, Morris M (2004) Ethnoflora of the Soqotra Archipelago. Royal Botanic Garden Edinburgh, Edinburgh, GB, 759 pp.

Morrone JJ (2014) On biotas and their names. Systematics and Biodiversity 12: 386–392. https://doi.org/10.1080/14772000.2014.942717

Naimi B, Hamm NAS, Groen TA, Skidmore AK, Toxopeus AG (2014) Where is positional uncertainty a problem for species distribution modelling? Ecography 37: 191–203. https://doi.org/10.1111/j.1600-0587.2013.00205.x

Neu CW, Byers CR, Peek JM (1974) A technique for analysis of utilization-availability data. The Journal of Wildlife Management 38: 541–545. https://doi.org/10.2307/3800887

Nimis PL, Bolognini G (1993) Quantitative phytogeography of the Italian beech forests. Vegetatio 109: 125–143. https://doi.org/10.1007/BF00044745

Olivero J, Real R, Márquez AL (2011) Fuzzy chorotypes as a conceptual tool to improve insight into biogeographic patterns. Systematic Biology 60: 645–660. https://doi.org/10.1093/sysbio/syr026

Passalacqua NG (2015) On the definition of element, chorotype and component in biogeography. Journal of Biogeography 42: 611–618. https://doi.org/10.1111/jbi.12473

Pignatti S (1982) Flora d'Italia. Edagricole, Bologna, IT.

Pignatti E, Pignatti S, Huang CC, Ding GQ, Huang ZL (1991) β-diversity and phytogeographical patterns in the Ding Hu Shan Reserve (Guangdong-South China) forest vegetation. Rendiconti Lincei Scienze Fisiche e Naturali 2: 79–85. https://doi.org/10.1007/BF03010415

van Proosdij ASJ, Sosef MSM, Wieringa JJ, Raes N (2016) Minimum required number of specimen records to develop accurate species distribution models. Ecography 39: 542–552. https://doi.org/10.1111/ecog.01509

R Core Team (2025) R: A language and environment for statistical computing. Version 4.5.2. R Foundation for Statistical Computing, Vienna, Austria. [Available from:] https://www.R-project.org/

Sanbonmatsu KK, Spalink D (2022) A global analysis of mosses reveals low phylogenetic endemism and highlights the importance of long-distance dispersal. Journal of Biogeography 49: 654–667. https://doi.org/10.1111/jbi.14333

Schimper AFW (1898) Pflanzengeographie auf physiologischer Grundlage. Gustav Fischer Verlag, Jena, DE, 1064 pp.

Sepulchre P, Ramstein G, Fluteau F, Schuster M, Tiercelin J-J, Brunet M (2006) Tectonic uplift and Eastern Africa aridification. Science 313: 1419–1423. https://doi.org/10.1126/science.1129158

Siddall M (2004) Understanding the Red Sea response to sea level. Earth and Planetary Science Letters 225: 421–434. https://doi.org/10.1016/j.epsl.2004.06.008

Španiel S, Rešetnik I (2022) Plant phylogeography of the Balkan Peninsula: spatiotemporal patterns and processes. Plant Systematics and Evolution 308: Article 38. https://doi.org/10.1007/s00606-022-01831-1

Takhtajan A (1986) Floristic Regions of the World. University of California Press, Berkeley, US.

Trauth MH, Larrasoaña JC, Mudelsee M (2009) Trends, rhythms and events in Plio-Pleistocene African climate. Quaternary Science Reviews 28: 399–411. https://doi.org/10.1016/j.quascirev.2008.11.003

Trauth MH, Maslin MA, Deino A, Strecker MR (2005) Late Cenozoic moisture history of East Africa. Science 309: 2051–2053. https://doi.org/10.1126/science.1112964

Valavi R, Guillera-Arroita G, Lahoz-Monfort JJ, Elith J (2022) Predictive performance of presence-only species distribution models: a benchmark study with reproducible code. Ecological Monographs 92: e01486. https://doi.org/10.1002/ecm.1486

Vilhena DA, Antonelli A (2015) A network approach for identifying and delimiting biogeographical regions. Nature Communications 6: Article 6848. https://doi.org/10.1038/ncomms7848

Walter H (1979) Vegetation of the Earth and Ecological Systems of the Geo-biosphere. 2nd ed. Springer, New York, US.

Walter H, Straka H (1970) Arealkunde, Floristich-Historische Geobotanik. Ulmer-Verlag, Stuttgart, DE, 478 pp.

Warren BH, Simberloff D, Ricklefs RE, Aguilée R, Condamine FL, Gravel D, Morlon H, Mouquet N, Rosindell J, … Thébaud C (2015) Islands as model systems in ecology and evolution: prospects fifty years after MacArthur-Wilson. Ecology Letters 18: 200–217. https://doi.org/10.1111/ele.12398

White F (1983) The vegetation of Africa: a descriptive memoir to accompany the Unesco/AETFAT/UNSO vegetation map of Africa. UNESCO, Paris, FR, 356 pp.

Wüest RO, Zimmermann NE, Zurell D, Alexander JM, Fritz SA, Hof C, Kreft H, Normand S, Cabral JS, … Karger DN (2020) Macroecology in the age of Big Data – Where to go from here? Journal of Biogeography 47: 1–12. https://doi.org/10.1111/jbi.13633

Zohary M (1973) Geobotanical foundations of the Middle East. G. Fischer, Stuttgart, DE.

E-mail and ORCID

Dario La Montagna (Corresponding author, dario.lamontagna.6@gmail.com), ORCID: https://orcid.org/0000-0002-7124-493X
Giuliano Fanelli (giuliano.fanelli@gmail.com), ORCID: https://orcid.org/0000-0002-3143-1212
Michele De Sanctis (michele.desanctis@uniroma1.it), ORCID: https://orcid.org/0000-0002-7280-6199
Alessio Farcomeni (afarcome@gmail.com), ORCID: https://orcid.org/0000-0002-7104-5826
Petr Maděra (petr.madera@mendelu.cz), ORCID: https://orcid.org/0000-0001-5415-8290
Luca Malatesta (luca.malatesta@uniroma1.it), ORCID: https://orcid.org/0000-0003-1887-4163
Kay Van Damme (kay.vandamme@gmail.com), ORCID: https://orcid.org/0000-0002-8215-3398
Fabio Attorre (fabio.attorre@uniroma1.it), ORCID: https://orcid.org/0000-0002-7744-2195

Supplementary material

Supplementary material 1
Data sources and additional figures (*.pdf)
Link: https://doi.org/10.3897/VCS.176349.suppl1

Supplementary material 2
Species to cluster enrichment table for global analysis (*.ods)
Link: https://doi.org/10.3897/VCS.176349.suppl2

Supplementary material 3
Species to cluster enrichment table for regional analysis (*.ods)
Link: https://doi.org/10.3897/VCS.176349.suppl3

Supplementary material 4
Chorotype descriptions (*.pdf)
Link: https://doi.org/10.3897/VCS.176349.suppl4

Supplementary material 5
Chorological percentages for Socotran vegetation types (*.ods)
Link: https://doi.org/10.3897/VCS.176349.suppl5