Climatic stability and geological history shape global centers of neo- and paleoendemism in seed plants

Lirong Cai[a,1] (ID), Holger Kreft[a,b,c] (ID), Amanda Taylor[a], Julian Schrader[a,d] (ID), Wayne Dawson[e] (ID), Franz Essl[f] (ID), Mark van Kleunen[g,h] (ID), Jan Pergl[i], Petr Pyšek[i,j], Marten Winter[k] (ID), and Patrick Weigelt[a,b,c,1] (ID)

Assessing the distribution of geographically restricted and evolutionarily unique species and their underlying drivers is key to understanding biogeographical processes and critical for global conservation prioritization. Here, we quantified the geographic distribution and drivers of phylogenetic endemism for ~320,000 seed plants worldwide and identified centers and drivers of evolutionarily young (neoendemism) and evolutionarily old endemism (paleoendemism). Tropical and subtropical islands as well as tropical mountain regions displayed the world's highest phylogenetic endemism. Most tropical rainforest regions emerged as centers of paleoendemism, while most Mediterranean-climate regions showed high neoendemism. Centers where high neo- and paleoendemism coincide emerged on some oceanic and continental fragment islands, in Mediterranean-climate regions and parts of the Irano-Turanian floristic region. Global variation in phylogenetic endemism was well explained by a combination of past and present environmental factors (79.8 to 87.7% of variance explained) and most strongly related to environmental heterogeneity. Also, warm and wet climates, geographic isolation, and long-term climatic stability emerged as key drivers of phylogenetic endemism. Neo- and paleoendemism were jointly explained by climatic and geological history. Long-term climatic stability promoted the persistence of paleoendemics, while the isolation of oceanic islands and their unique geological histories promoted neoendemism. Mountainous regions promoted both neo- and paleoendemism, reflecting both diversification and persistence over time. Our study provides insights into the evolutionary underpinnings of biogeographical patterns in seed plants and identifies the areas on Earth with the highest evolutionary and biogeographical uniqueness—key information for setting global conservation priorities.

islands | mountains | past climate change | phylogenetic endemism | plant diversity

Plant species range sizes vary widely from being nearly cosmopolitan to extremely small, for example, being restricted to a single mountain or island (1). Understanding the global distribution of range-restricted or endemic species and the mechanisms that create centers of high endemism is a central question in biogeography (2) and is crucial for the preservation of biodiversity (3). Due to their restricted geographic ranges, endemic species are more vulnerable to extinction (4, 5), and, if simultaneously evolutionarily unique, their extinction may result in significant losses of evolutionary history (6–8). Evolutionarily unique endemics are also likely to be associated with irreplaceable ecological and functional characteristics (9, 10). It is therefore essential to account for the phylogenetic relatedness and evolutionary uniqueness of species when assessing endemism. Measures of phylogenetic endemism (PE) account for the phylogenetic uniqueness of range-restricted species (11) and allow for the differentiation between neo- and paleoendemism (12). Regions with high PE or paleoendemism harbor more evolutionarily unique lineages with restricted geographic distributions than regions with low PE. Assessing PE for seed plants, the different types of endemism, and their past and present environmental drivers is thus crucial for setting conservation priorities and for understanding the biogeographical mechanisms underpinning plant diversity.

Endemism can originate from multiple biogeographical and evolutionary processes, which promote the formation or persistence of range-restricted species (see *SI Appendix*, Table S1 for main hypotheses of PE determinants). For one, endemism may result from speciation and limited range expansion due to dispersal limitation, which is promoted by physical or ecological barriers such as oceans, mountain ranges, or climatic gradients (13–15). Isolated regions, like oceanic islands, are renowned for their high levels of endemism (3). Relatively recent speciation events on oceanic islands (16, 17) may have resulted in an accumulation of recently evolved lineages ("cradles of diversity") that are still restricted to their area of origin, leading to so-called neoendemism (12). Alternatively, endemism can be facilitated by the long-term persistence of range-restricted species and

Significance

Range-restricted and evolutionarily unique species are a crucial yet often overlooked facet of biodiversity. Understanding the distribution of neo- and paleoendemism, i.e., identifying centers of evolutionarily young or old endemics, helps to understand the processes that shaped today's distribution of biodiversity. Here, we uncovered global patterns and determinants of phylogenetic endemism and neo- versus paleoendemism for seed plants. Environmental heterogeneity, climate, and geographic isolation emerged as key drivers of phylogenetic endemism. Long-term climatic stability promotes paleoendemism, while isolation promotes neoendemism, jointly leading to oceanic and large continental islands, and mountain regions in the tropics and subtropics being global endemism centers. These results highlight the importance of climatic and geological history on diversification and persistence of biodiversity and aid conservation prioritization.

their accumulation over long time-scales, leading to paleoendemism ("museums of diversity") (18, 19). During periods of pronounced climate change (e.g., Quaternary glacial cycles), plant distributions shifted greatly, resulting in repeated range contractions followed by range expansions in more favorable periods (20, 21). Therefore, regions that were climatically stable over long time periods might have served as refugia (2, 22). Particularly, topographically heterogeneous regions allowed species to track climate change over only relatively short altitudinal distances reducing their extinction risk (23, 24). After periods of climatically unfavorable conditions, not all plants could reoccupy their former ranges (i.e., relictualization) (25, 26). The resulting paleoendemics that were once widespread and are now restricted to former refugia often represent evolutionarily old lineages (19, 27).

Factors favoring the formation or persistence of endemic species do not need to be mutually exclusive. However, the influence of these processes may vary across space and over geological time, leading to regional assemblages of more recently evolved endemics or those that diverged long ago, or both (28). Floras with high levels of neo- or paleoendemism have likely been shaped by different processes affecting species diversification and persistence, which jointly lead to high endemism. Assessing patterns and drivers of PE accounting for the evolutionary history of range-restricted species and distinguishing between neo- and paleoendemism thus provides insights into past and present determinants, including geological history, climatic changes, and evolutionary processes, that structure biodiversity (29) (see *SI Appendix*, Table S2 for main hypotheses of neo- and paleoendemism determinants). However, global tests of how plant PE and neo- versus paleoendemism are driven by climatic and biogeographical history are still lacking.

Here, we reveal global patterns and drivers of PE and neo- versus paleoendemism (12) for ~320,000 seed plant species by integrating the most comprehensive regional plant inventories across 912 geographic regions worldwide (30, 31) (*SI Appendix*, Fig. S1) with a broad seed plant phylogeny (32). Specifically, our aims are i) to reveal geographic patterns of PE for seed plants at the global scale; ii) to test hypotheses related to isolation, environmental heterogeneity, climate, and past climate change on global patterns of seed plant PE (*SI Appendix*, Table S1); iii) to identify centers of neoendemism and paleoendemism across the world; iv) and to assess how past climate change and geological history shaped the centers of neo- and paleoendemism (*SI Appendix*, Table S2).

Results

Global Patterns and Drivers of Phylogenetic Endemism. Phylogenetic endemism of seed plants varied greatly among regions, being highest on islands and in topographically heterogeneous tropical mainland regions (Fig. 1). These and all other main results are based on the global distribution of 212,525 seed plants excluding all species from 293 genera that contain apomictic species (33), to avoid biases introduced by the multitude of apomictic taxa in the temperate Northern Hemisphere (see *Materials and Methods* for more details and *SI Appendix* for results based on the datasets including apomictic taxa), and only retaining species that were originally included in the phylogeny (32, 34), if not stated otherwise. To test for potential biases introduced by incomplete phylogenetic knowledge (i.e., taxa missing from the phylogeny) (35), we repeated all analyses based on a phylogeny with unplaced species added to their congeners (including 267,105 species when excluding apomictic taxa; see *SI Appendix* for the results). Because PE is scale-dependent (36) and depends on reliable range size estimates, we calculated PE based on two different calculations of

species range sizes: i) the total area (PE.area) of regions a species occurs in and ii) the number of these regions (PE.count).

We found that PE.area was almost 17-fold higher on islands than in mainland regions (mean PE.area of islands and mainland regions: 0.50 vs. 0.03 Myr·km^{-2}). PE.area peaked on subtropical islands located in the Southern Hemisphere, with Lord Howe Island having the highest PE.area overall (30.80 Myr·km^{-2}), while the province of Pichincha in Ecuador showed the highest value among mainland regions (0.57 Myr·km^{-2}; Fig. 1 *A* and *B* and *SI Appendix*, Table S3). In contrast, PE.count peaked in the tropics both for islands (Madagascar: 91,364 Myr) and mainland regions (Peru: 82,911 Myr; Fig. 1 *C* and *D* and *SI Appendix*, Table S3).

The environmental factors we hypothesized to affect PE (i.e., isolation, environmental heterogeneity, climate, and past climate change; *SI Appendix*, Table S1) explained 79.8% of the variance in PE.area and 87.7% in PE.count (Fig. 2 and *SI Appendix*, Table S4). The effects of environmental factors on PE were largely similar regardless of how range size was quantified (differing most prominently for region area which had a positive effect on PE.count and a negative effect on PE.area) (Fig. 2*A*). PE was most strongly associated with environmental heterogeneity, increasing, as expected, with elevational range and number of soil types (Fig. 2*A*). Surrounding landmass proportion, a proxy for isolation, which is lowest for remote islands and highest for regions located in the centers of large continents (37), was negatively related to PE.area. This indicates that high PE occurred on islands and in mainland regions that are partly surrounded by water bodies such as coastal regions or peninsulas. When unplaced species were added to the phylogeny, surrounding landmass proportion turned out to be the most important driver of PE.area and also showed a significant negative effect on PE.count (*SI Appendix*, Fig. S2), which may be explained by many species from islands missing from the original phylogeny (*SI Appendix*, Fig. S3). Among climatic factors, energy and water availability had strong associations with PE, with increasing length of the growing season and mean annual temperature leading to higher PE (Fig. 2*A*). Temperature and precipitation seasonality, however, had no or only weak positive effects on PE. Relatively recent past climate change left prominent traces in PE, but this was not detectable for climatic changes in deeper time. PE increased with temperature stability since the Last Glacial Maximum (LGM; 21 Ka), while velocity of temperature change since the LGM had a negative effect. However, we found no significant relationship between PE and temperature anomaly since the mid-Pliocene warm period (~3.264 to 3.025 Ma). To test if the effects of environmental predictors on PE varied between isolated regions (e.g., islands) and less isolated regions (e.g., mainland regions), we included interactions between each predictor and surrounding landmass proportion in the models. We found that the positive effect of mean annual temperature on PE increased with decreasing surrounding landmass proportion (Fig. 2*B*). Nearly identical PE patterns and drivers were found across all datasets, regardless of the exclusion or inclusion of unplaced species (*SI Appendix*, Figs. S2 and S4 and Table S5) and apomictic taxa (*SI Appendix*, Figs. S5–S7 and Tables S6 and S7).

Global Centers and Determinants of Neo- and Paleoendemism. We uncovered centers of evolutionarily old and range-restricted species, centers of evolutionarily young and range-restricted species as well as centers of both using a categorical analysis of neo- and paleoendemism (CANAPE) (12). Regions identified as centers of neo- or paleoendemism occupied 27.4% (PE.area) and 31.4% (PE.count) of the global landmass area including mainland regions and islands, while regions that harbored both high neo- and paleoendemism (i.e., centers of mixed or superendemism) only
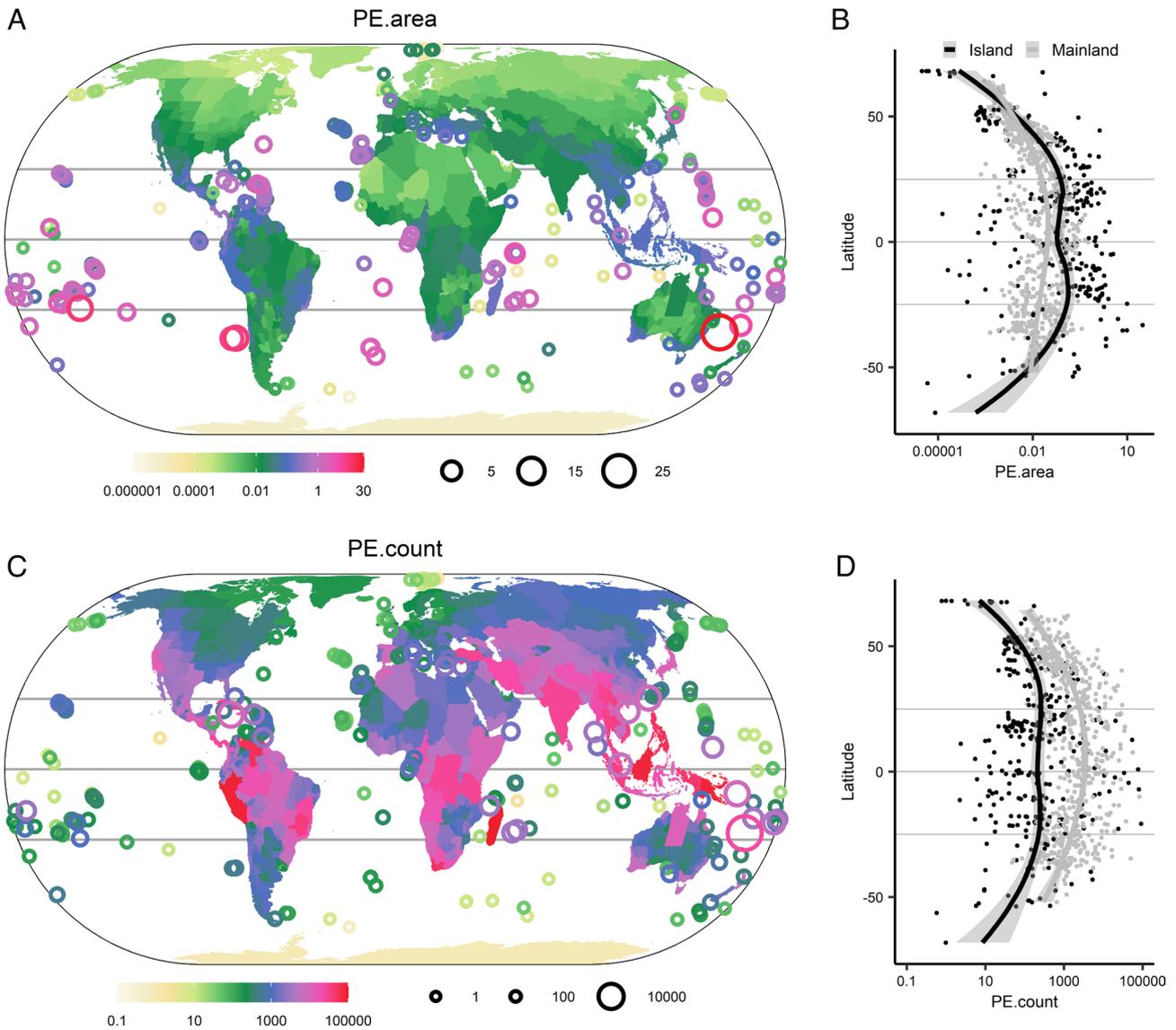
Fig. 1. Global patterns of phylogenetic endemism of seed plants and its distribution along latitude. In (*A* and *B*), phylogenetic endemism is calculated based on species range size measured as the total area of regions where a species occurs (PE.area); In (*C* and *D*), phylogenetic endemism is calculated based on species range size measured as the count of regions where a species occurs (PE.count). In (*B* and *D*), the fitted lines are lowess regressions, separately fitted for islands and mainland regions. Log_{10} scale is used for phylogenetic endemism in all panels and maps are shown in Eckert IV projection.

occupied 5.1% and 4.4%, respectively (Fig. 3*A* and *SI Appendix,* Fig. S8 *A* and *B*). Many remote islands (e.g., Mauritius, Juan Fernández Islands, and New Caledonia) emerged as centers of both neo- and paleoendemism (Fig. 3*A*). In contrast, some continental fragment islands, such as Madagascar, Cuba, and Hispaniola, and large continental islands in southeast Asia (e.g., New Guinea, Sumatra, and Java) were identified as centers of paleoendemism. When adding unplaced species to the phylogeny, Madagascar and Hispaniola turned out to be centers of superendemism, harboring both unusually high neo- and paleoendemism (Fig. 3*B* and *SI Appendix,* Fig. S8 *E* and *F*). Mainland regions characterized by tropical rainforests, such as Amazonia, Peru, western Colombia, central Africa, and large parts of Indochina, showed high paleoendemism (Fig. 3*A*). Mediterranean climate regions and large parts of the Irano-Turanian floristic region (38) stood out as extra-tropical hotspots of seed-plant PE among mainland regions. For example, south-western Australia, the Tibetan plateau, Afghanistan, Tajikistan, and Turkey were characterized by both high neo- and paleoendemism, while the Cape of South Africa, central Chile, California, and mainland Spain were centers of

neoendemism (Fig. 3*A*). When adding unplaced species to the phylogeny, regions from the Irano-Turanian floristic region tended to be centers of neoendemism (Fig. 3*B*). Moreover, some differences emerged depending on the measurement of species range size. For example, the Himalayas were a center of neo- and paleoendemism based on PE.count, while it did not emerge as an endemism center based on PE.area (*SI Appendix,* Fig. S8 *A* and *B*). In contrast, Iran tended to be a center of neoendemism based on PE.count, while it was a center of neo- and paleoendemism based on PE.area (*SI Appendix,* Fig. S8 *A* and *B*). Comparing patterns including and excluding apomictic taxa, the most prominent differences occurred in European countries that were identified as endemism centers when including apomictic species, due to high numbers of apomictic range-restricted species in genera like *Rubus* and *Hieracium* (*SI Appendix,* Figs. S8 and S9).

We assessed the impacts of geological history and past climate change on neo- and paleoendemism by modeling the standardized effect size of relative PE (see *Materials and Methods* for details) for regions that showed significantly high PE in response to past climatic and geological factors (*SI Appendix,* Table S2).
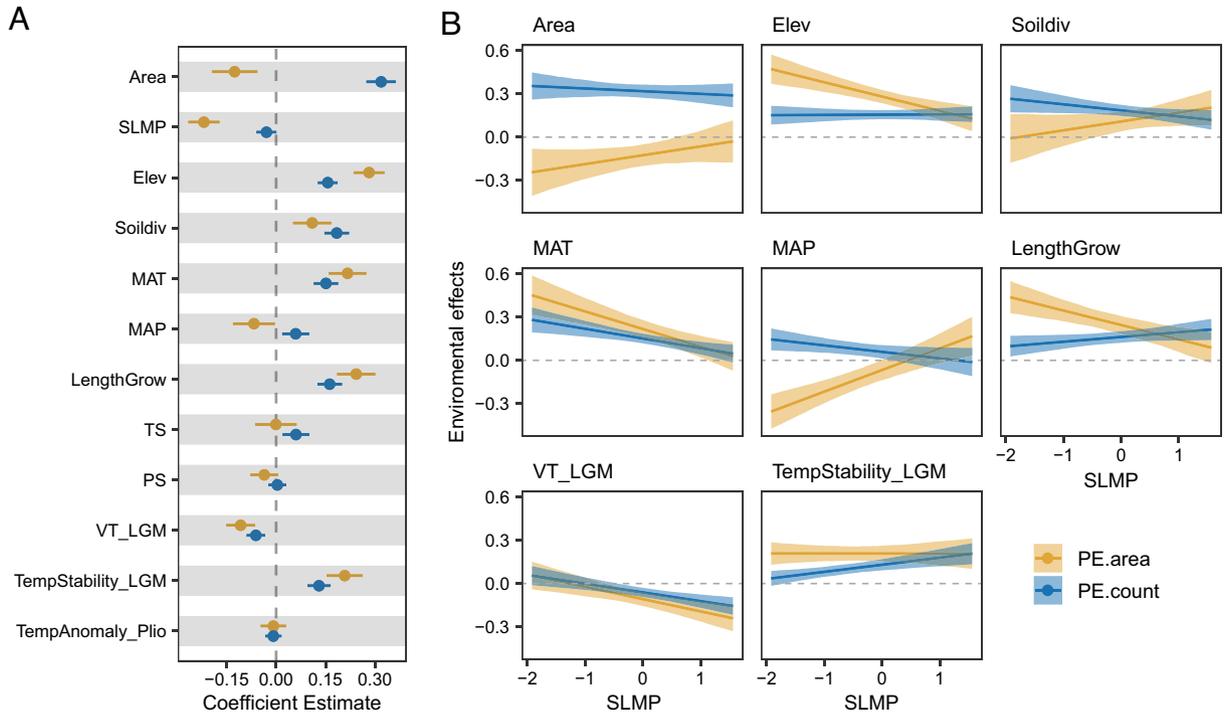
Fig. 2. Determinants of phylogenetic endemism in seed plants based on spatial models including environmental factors and interactions between each environmental factor and surrounding landmass proportion. (*A*) standardized regression coefficients of individual environmental factors. Bars around each point show the SE of the coefficient estimate. (*B*) significant interaction terms in the models visualized as effects of environmental factors on phylogenetic endemism (model coefficients on the *y* axis) with varying surrounding landmass proportion (*x* axis). Lines and shadings represent 95% CI. Results are shown for phylogenetic endemism based on two competing ways of measuring the range size of species. PE.area (yellow) indicates phylogenetic endemism calculated based on the range size of species as the area of regions where a species occurs, while PE.count (blue) is calculated based on the range size of species as the count of these regions. Area = region area; SLMP = surrounding landmass proportion; Elev = elevational range; Soildiv = number of soil types; MAT = mean annual temperature; MAP= mean annual precipitation; LengthGrow = length of the growing season; TS = temperature seasonality; PS = precipitation seasonality; VT_LGM = velocity of temperature change since the Last Glacial Maximum; TempStability_LGM = temperature stability since the Last Glacial Maximum; TempAnomaly_Plio = temperature anomaly between the mid-Pliocene warm period and present-day.

We included the geographic type of each region (distinguishing between mainland regions and continental shelf islands, continental fragments, and oceanic islands) and elevational range (distinguishing between mountainous and nonmountainous regions) to represent geological history. Oceanic islands showed significantly higher neoendemism than mainland endemism centers identified based on PE.count (Fig. 4 and *SI Appendix,* Table S8) and also based on PE.area when unplaced species were added to the phylogeny (*SI Appendix,* Fig. S10*A*). Continental islands did not show any significant difference to mainland regions. Elevational range had no significant effect on neo- versus paleoendemism without unplaced species added to the phylogeny (Fig. 4) and a positive effect on neoendemism when unplaced species were added (*SI Appendix,* Fig. S10*A*). However, when testing for differences of environmental factors among endemism types, centers of neo- and paleoendemism both consistently had significantly higher elevational ranges than regions with low PE (*SI Appendix,* Figs. S11 *A* and *B* and S12 *A* and *B*). Past climate change was a major driver of neo- versus paleoendemism, with increasing temperature stability since the LGM increasing paleoendemism. Besides climate since the LGM, we also found significant relationships between temperature anomaly since the mid-Pliocene warm period and neo- and paleoendemism. Specifically, regions of superendemism showed a significantly lower temperature anomaly since the mid-Pliocene than regions with low PE.area (*SI Appendix,* Fig. S11*G*). When unplaced species were added to the phylogeny, a significant difference also emerged between superendemism regions and other types of regions for both PE.area and PE.count (*SI Appendix,* Fig. S12

G and *H*). Comparing the results including and excluding apomictic taxa, the relationships between environmental variables and neo- versus paleoendemism were almost identical when unplaced species were added to the phylogenies (*SI Appendix,* Fig. S10*B*), while differences emerged without unplaced species added (*SI Appendix,* Fig. S10*C*). We found no difference or weakly increased neoendemism with increasing temperature stability since the LGM, while the significant difference in neo- versus paleoendemism for oceanic islands disappeared.

Discussion

Our study reveals islands and mountain regions in the tropics and subtropics as global centers of phylogenetic as well as neo- and paleoendemism of seed plants. Variation in the distribution of neo- and paleoendemism across the globe reflects the varied processes related to the generation and persistence of range-restricted species within a region, which jointly lead to high numbers of geographically restricted and evolutionarily unique lineages. We show that past climate change and geological history help to explain how diversification and relictualization (i.e., the persistence of species that went extinct elsewhere) shape the distribution of neo- and paleoendemism and simultaneously phylogenetic endemism worldwide. Understanding the drivers of different kinds of endemism and knowing particularly those regions with both high neo- and paleoendemism that act simultaneously as "museums" and "cradles" of biodiversity is of great importance for conservation prioritization, especially since global threat assessments for endemic plants are still incomplete (39).
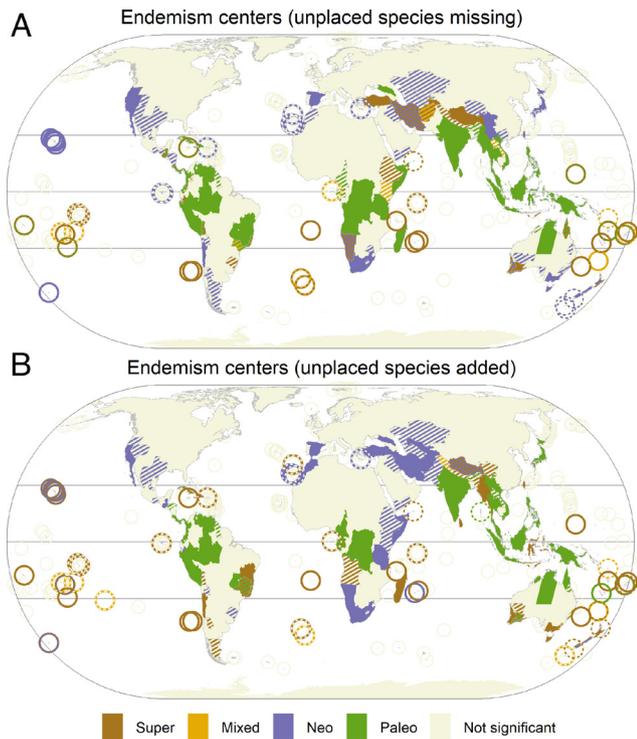
Fig. 3. Global centers of neo- and paleoendemism for seed plants. In (*A*) species unplaced in the phylogeny are missing while they are added in (*B*). Colored regions present different types of endemism centers according to a categorical analysis of neo- and paleoendemism (CANAPE): violet, neoendemism; green, paleoendemism; yellow, mixed-endemism (i.e., neo- and paleoendemism); and brown indicating superendemism (i.e., centers with both extremely high neo- and paleoendemism); beige, not significant. Patterns of neo- and paleoendemism have been calculated based on two competing ways of measuring species range size (i.e., as the area of regions where a species occurs versus as the count of these regions). Combinations of colors (hashed in mainland regions and dashed circles for islands) represent different types of endemism for a region based on these two metrics, while solid colors represent consistent endemism types. Islands that were not significant for both two metrics are represented by small and beige circles. See *SI Appendix*, Fig. S8 for endemism centers based on each metric separately. Maps are shown in Eckert IV projection.

Geographic isolation resulted in high neo- and paleoendemism and PE on islands, which may stem from in situ speciation in isolation and relictualization (26). While speciation events require time for island species to evolve into phylogenetically distinct species, endemic species may accumulate over shorter times through relictualization (18, 40), resulting from species extinctions on the mainland and other islands. Species from lineages that diversified on islands are often young and closely related, while relict lineages on islands may be old and only distantly related to other species on the same island. Sometimes ancient endemic species are even older than the formation of the island, such as the only member of the genus *Hillebrandia sandwicensis* on the Hawaiian Islands (41) and the only member of the oldest known angiosperm family (Amborellaceae) *Amborella trichopoda* on New Caledonia (42). High PE on islands may thus be a result of a combination of diversification leading to neoendemism and relictualization leading to paleoendemism. Furthermore, the diversification of island species is constrained by available resources and niches. For example, the probability of in situ speciation scales positively with island size (43). This may explain the stronger effects of some environmental factors, such as energy availability and elevational range, on island PE than on mainland PE.

When comparing islands of different geological origins, we found that oceanic islands are characterized by higher neoendemism than continental islands, which may be explained by their unique geological history (26). Oceanic islands have not been connected with continental landmasses in the past but emerged from the oceans due to volcanic or tectonic activity. Untapped resources and the lack of enemies and competitors allowed plant species that colonized oceanic islands to diversify (15, 16). Considering the relatively short geological lifespan of oceanic islands, the speciation on these islands happened comparatively recently, leading to neoendemism. However, some oceanic islands were identified as centers of superendemism (e.g., New Caledonia), where relictualization and diversification happened in concert. Continental fragments and continental shelf islands, in contrast, were once part of continents that became separated by tectonic drift or sea-level rise. These islands were originally inhabited by floras comparable to those of the continents they were connected to. The prolonged isolation (tens of millions of years) of continental fragments allowed for the accumulation of relict lineages as well as in situ speciation, which led to high neo- and paleoendemism on some of these islands. However, the origins of endemism on some large continental fragments are still debated (44). Apart from more recent colonization events, evolution after vicariance or early long-distance dispersal events may have led to particularly old endemic species. For example, the majority of endemics on Madagascar evolved from lineages that originated from Cenozoic dispersal events (45), while few groups (e.g., the genus *Takhtajania*) date back to a potential Gondwanan vicariance (46). Also, islands located in southeast Asia showed high paleoendemism, which is due to numerous relict lineages that have survived the last two mass extinctions (47). Consequently, our results reinforce the conservation urgency for islands which are often occupied by both neoendemics and paleoendemics that represent millions of years of unique evolutionary history (48).

Tropical mountain regions are well-known centers of taxonomic and phylogenetic plant diversity (e.g., ref. (49)). Due to their
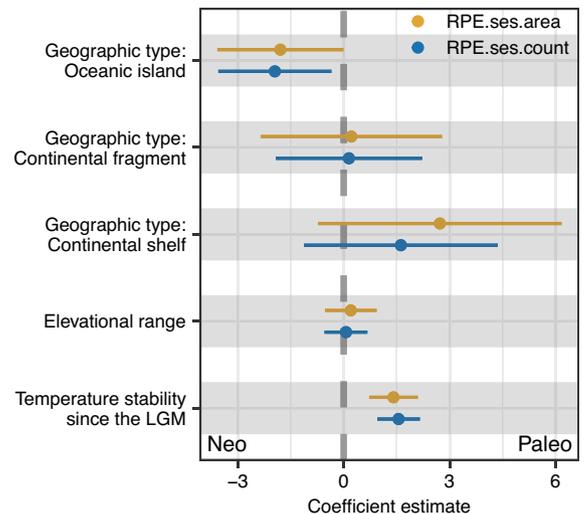
Fig. 4. Determinants of neo- and paleoendemism. Standardized regression coefficients of environmental factors are shown from spatial models of the standardized effect size of relative phylogenetic endemism of seed plants for regions with significantly high phylogenetic endemism. A positive effect of environmental factors represents higher paleoendemism at higher values of the environmental factor, while a negative effect represents higher neoendemism. RPE.ses.area (yellow) indicates the standardized effect size of relative phylogenetic endemism calculated based on the range size of species as the area of regions where a species occurs, while RPE.ses.count (blue) is calculated based on the range size of species as the count of these regions. The reference level of geographic type is mainland regions. LGM = Last Glacial Maximum.

complex topography and geological and climatic histories, they also hold exceptionally range-restricted species (50). In our study, mountain regions, especially in tropical regions, emerged not only as centers of PE but also of both neo- and paleoendemism in particular. On the one hand, mountain regions show remarkable diversification of their plant lineages and therefore foster high neoendemism, acting as "cradles" of biodiversity (51, 52). This diversification is the consequence of multiple mechanisms, including adaptation to diverse niches during long-term orogeny (52, 53), or divergence resulting from dynamic connectivity between habitats related to climatic fluctuation (54, 55). On the other hand, mountain regions support the persistence of ancient lineages over time, acting as "museums" for paleoendemics (50, 56). This results from steep environmental gradients with diverse microclimates in mountain regions, allowing species to track their climate niche through altitudinal range shifts during climate change periods (24).

Our results show how past climate has affected present patterns of neo- and paleoendemism, with climate stability since the Last Glacial Maximum promoting the accumulation of paleoendemics as well as PE in general. Cooler temperatures during glacial periods may have caused range contractions and selective extinctions of range-restricted species and thus likely removed or reduced their ranges in less stable regions (22). In contrast, some regions such as islands, coastal, or mountain regions have suffered less from past climate change because of the buffering effect of the oceans against climatic changes (27) and the topographically diverse microclimates in mountain regions (24). Also, the high concentration of both neo- and paleoendemism in regions with less climatic changes since the mid-Pliocene warm period emphasizes the vital role of long-term climatic stability on speciation and persistence of range-restricted species. Different from the climate during the LGM, the mid-Pliocene warm period, however, represented warmer climates compared to today. The positive effect of climatic stability on neo- and paleoendemism is therefore also in line with the general positive effect of energy and water availability on PE of plants. This may be linked to lower extinction risks for range-restricted plants under warm and wet climates by offering favorable environments and sufficient resources for larger populations in smaller areas (57). In addition, we found that Mediterranean-climate regions acted as extratropical hotspots of plant endemism, especially with high neoendemism. This may be attributed to the recent and rapid speciation in these regions, triggered by the unique climatic regime characterized by high seasonality and summer drought (13, 58).

Generally, larger regions host more endemics as well as wide-ranged species because of their overall higher plant diversity (49, 59). Here, we observed a negative association between region area and PE when species range sizes were measured as the total area of the occupied regions. Specifically, PE.area peaked on some small islands (e.g., Lord Howe Island) and showed lower values in large mainland regions. However, PE.area of large mainland regions was possibly underestimated because the range sizes of endemics that only occur in small suitable habitats within large regions were overestimated. In contrast, there was a positive association between region area and PE when we measured species range sizes as the count of occupied regions. This method ignores the variation of area across regions and disregards that endemics in small regions likely have smaller ranges than endemics in larger regions leading to an underestimation of PE for small regions. Area, therefore, acted as a covariate to control for biases in the estimation of range size in our two metrics of PE and the scale-dependency of endemism (36) rather than as an environmental predictor. The largely similar effects of environmental drivers on PE and neo- versus paleoendemism for the two ways

in which range size was quantified demonstrates the robustness of our results. Similarly, the results were largely the same with and without unplaced species added to the original phylogeny. Differences that did emerge, however, call for rigorous sensitivity analyses when dealing with species without phylogenetic information (34, 60, 61). For example, the emergence of Madagascar as a superendemism center when unplaced species were added may result from many species from genera endemic to Madagascar missing from the phylogeny, e.g., *Astiella* (38). Similarly, the lack of an effect of elevational range on neoendemism when unplaced species were not added suggests that particularly endemic species from recent diversification events may be missing phylogenetic data (35).

In conclusion, our study uncovers global patterns of phylogenetic endemism for seed plants and disentangles the complex joint effects of isolation, heterogeneity, climate, and long-term climatic stability on phylogenetic endemism. Integration of unprecedented phylogenetic information allowed us to distinguish global centers of neo- and paleoendemism, highlighting tropical mountains, oceanic, and large continental islands as well as Mediterranean-climate regions as hotspots of evolutionarily distinct endemic species. These regions have experienced unique climatic and geological histories, which have driven the interplay of important evolutionary and ecological processes of diversity generation and maintenance. Consequently, these regions are of crucial conservation value and need to be protected.

Materials and Methods

Species Distribution Data. We used regional species composition data for native seed plants from the Global Inventory of Floras and Traits (GIFT version 3.0: http://gift.uni-goettingen.de)(30, 62) and the World Checklist of Vascular Plants (WCVP, http://wcvp.science.kew.org/)(31). GIFT contains regional plant inventories from published floras and checklists for ~ 3,400 geographic regions worldwide representing islands, protected areas, biogeographical regions (e.g., botanical countries), and political units (e.g., countries, provinces). WCVP is a comprehensive taxonomic compilation of vascular plants and offers distribution information of species in WGSRPD Level-3 units (i.e., 369 botanical countries). We downloaded information for each nonsynonym species in WCVP (accessed 18 February 2022) using the function *pow_lookup* in the R package *taxize* (63) and extracted their distribution and biogeographical status across all botanical countries. We then combined all native seed plant occurrences from WCVP with all native seed plant checklists from GIFT available for the same regions. To obtain finer-grain distribution information for some large regions, we replaced botanical countries with smaller regions from GIFT where available (e.g., the individual departments of Bolivia instead of the entire country). We removed the larger regions only when smaller regions were nested within the larger regions and all nested regions completely covered the larger regions for mainland regions and replaced archipelagos with individual islands if the individual islands made up most of the archipelago. Because all nonhybrid species names in GIFT 3.0 were standardized and validated based on taxonomic information provided by WCVP, we were able to directly combine WCVP and GIFT data. We retained taxonomically unmatched species names because of the low percentage of these species per region (i.e., 99.7% of all species names were taxonomically matched on average across regions). We excluded regions with areas not permanently covered by ice smaller than 10 km^2. All small regions excluded were islands and only a few of them host endemic species (49 endemic species on 112 islands < 10 km^2 in GIFT). The final dataset included 317,985 seed plant species for 912 geographic regions covering all landmass worldwide with varying area sizes ranging from 10 to 3,069,766 km^2 (median: 23,192 km^2), consisting of 597 mainland regions and 315 islands or island groups (*SI Appendix*, Fig. S1).

Apomictic Taxa. Apomixis is a special case of uniparental reproduction via asexually formed seeds (64). Apomixis is tightly associated with hybridization and polyploidization and may promote reticulate evolution and the formation of a multitude of novel lineages (64). European brambles (*Rubus* subgen. *Rubus*, Rosaceae), for example, consist mostly of apomictic taxa (only 4 out of 748

accepted species are sexual) owing to speciation via reticulation and apomixis (65). However, taxonomic treatment of these complex groups of apomictic taxa and underlying species concepts are contentious. Additionally, regional floras and checklists differ in the level of detail at which these groups are included and taxonomically resolved. Consequently, the global distribution of apomictic taxa is geographically biased (particularly toward the well-sampled European flora), affecting the assessment of endemism, especially for regions with a high proportion of apomictic taxa. To account for the bias introduced by apomictic taxa, we repeated all analyses including and excluding all the species from 293 genera that contain apomictic species according to the Apomixis Database (http://www.apomixis.uni-goettingen.de) (33). The distribution dataset excluding species from apomictic genera included 273,838 species and was used for the main analyses. Results including apomictic species can be found in *SI Appendix*. The Apomixis Database has been constructed only for angiosperms. It has however been shown that apomixis is very rare in gymnosperms (64).

Phylogeny. To measure phylogenetic endemism, we linked the species from the distribution dataset to a large, dated species-level phylogeny of seed plants with 353,185 tips (32). A total of 212,525 species from the distribution dataset excluding apomictic taxa (77.6 % of the species), and 244,206 species from the dataset including apomictic taxa (76.8%), could be directly matched to the phylogeny. Species not present in the distribution dataset were excluded from the phylogeny (hereafter called matched phylogeny). Different ways to deal with species missing from phylogenies in biogeographic and macroecological analyses exist (34, 60, 66). Furthermore, it has been shown that range-restricted species are significantly less likely to have phylogenetic data (35), suggesting that excluding all species missing from the original phylogeny might systematically underestimate PE and neo- and paleoendemism. Therefore, to test whether removing unplaced species from the distribution dataset or adding them into the phylogeny affects patterns of PE and neo- and paleoendemism, we built an additional phylogeny and repeated all analyses for comparison (66). We bound the missing species to their congeners in the original phylogeny by replacing all species of a given genus by a polytomy using the function *congeneric.merge* in the R package *pez* (67). We then excluded species not present in the distribution dataset from the phylogeny (hereafter called merged phylogeny). The merged phylogeny included 267,105 out of 273,838 species (97.5 %) in the dataset excluding apomictic taxa and 311,250 species (97.9%) including apomictic taxa. Adding species as polytomies may introduce additional uncertainties when working with large phylogenies (49, 68). However, phylogenetic metrics based on phylogenies with higher numbers of polytomies have been shown to be highly correlated with metrics based on trees without or with fewer polytomies (69). We repeated all analyses using both the matched and merged phylogenies. PE derived from the matched phylogeny was highly correlated to PE based on the merged phylogeny (Pearson's r: 0.98 for PE.area and 0.99 for PE.count based on the dataset excluding species from apomictic genera). We, therefore, present results based on the matched phylogeny in the main text if not stated otherwise and discuss discrepancies between the different approaches critically. Results based on the merged phylogeny and excluding and including apomictic taxa can be found in *SI Appendix*.

Phylogenetic Endemism. To investigate the distribution of seed plant endemism worldwide, we calculated phylogenetic endemism for each region following ref. 11, as the sum of branch lengths connecting all species coexisting in a region, based on a phylogeny where each branch length is divided by the global range size of the species that descended from the branch. Because PE depends on reliable range size estimates and its pattern is sensitive to differences in grain size (36), we measured the range size of each species and of each branch in two different ways: i) as the number of regions a species occurs in (PE.count) and ii) as the total area (not permanently covered by ice) of these regions (PE.area). PE.count overestimates PE particularly for large regions, since it disregards that the ranges of species endemic to small regions are likely smaller than the ranges of species endemic to larger regions. In contrast, PE.area accounts for the varied areas of regions in our dataset but likely underestimates PE for large regions because their areas may be larger than the actual ranges of the species occurring inside. Despite the potential biases of both methods, the actual ranges and hence endemism fall within the range that is estimated by the two methods (*SI Appendix*, Fig. S13; see *Discussion* for more details). We, therefore, repeated all analyses based on both metrics, considered those results particularly robust that emerged for both metrics, and discussed differences critically.

Neo- and Paleoendemism. We used the categorical analysis of neo- and paleoendemism (CANAPE) (12) to distinguish between centers of neoendemism and paleoendemism. CANAPE is based on the assessment of the statistical significance of PE and relative phylogenetic endemism (RPE). RPE is the ratio of PE measured on the actual phylogenetic tree divided by PE measured on a comparison tree that retains the actual tree topology but with all branches having the same length (12). Therefore, RPE allowed us to examine the degree to which branch lengths and hence clade ages matter for the observed patterns of PE. We carried out the CANAPE analysis for PE.count and PE.area, respectively. To test the significance of the metrics, we ran 1,000 null model randomizations. In the null models, species occurrences across regions were randomly reassigned without replacement, keeping the species number in each region and the total number of regions occupied by each species constant (12). Distributions of null model values for each region were then used for nonparametric tests for significance of the observed values of the tested metrics and for calculating the standardized effect size of RPE. If the observed value of the tested metric fell into the highest 2.5% or lowest 2.5% of the null distribution for a region, it was identified as statistically significantly high or low, respectively (two-tailed test, $\alpha = 0.05$). This randomization-based significance test was carried out for PE measured on the actual tree (numerator of RPE), PE measured on the comparison tree with equal branch lengths (denominator of RPE), and RPE.

We then followed a two-step process to distinguish different centers of endemism following ref. 12. First, we identified regions with significantly high PE by testing whether PE measured on the actual tree (numerator of RPE), PE measured on the comparison tree with equal branch length (denominator of RPE), or both were significantly higher than expected (observed value > 95% of the randomization values; one-tailed test, $\alpha = 0.05$). Second, we divided regions with significantly high PE into four categories of centers of endemism (paleo-, neo-, mixed, and superendemism). If the RPE of a region was significantly high or low (two-tailed test, $\alpha = 0.05$), the region was defined as a center of paleoendemism or neoendemism, respectively. If the RPE was not significantly high or low, but both the numerator and denominator of RPE were significantly high ($\alpha = 0.05$), the region was defined as a center of mixed endemism. If a mixed endemism region had both a significantly high numerator and denominator of RPE at the $\alpha = 0.01$ level, the region was identified as a center of superendemism.

We also calculated the standardized effect size of relative phylogenetic endemism (RPE.ses) based on the null distributions of RPE obtained from the null model. RPE.ses was calculated as the difference between the observed values and the mean of the null distribution divided by the SD of the null distribution. In contrast to the nonparametric test in CANAPE, RPE.ses quantifies, for each region, the degree to which disproportionately young or old lineages (i.e., shorter or longer branches) are spatially restricted. When only considering regions with significantly high PE, lower values of RPE.ses represent more young or younger lineages that are spatially restricted, while higher values represent more old or older lineages than expected by chance. Thus, this metric offers an opportunity to model and explore the relationship between the historical drivers and the spatial patterns of neo- versus paleoendemism as a continuous variable. However, it should be noted that a region with both high neo- and paleoendemism may show a value close to zero.

Predictors of Phylogenetic Endemism. We hypothesized that phylogenetic endemism is shaped by biogeographical and evolutionary processes that promote the origin and maintenance of range-restricted lineages. We, therefore, identified a set of candidate predictor variables representing these processes and classified them into five categories: isolation, environmental heterogeneity, energy and water availability, climatic seasonality, and long-term climatic stability (*SI Appendix*, Table S1). These factors have been shown or hypothesized to contribute to geographic patterns of plant endemism in previous studies (3, 22, 56, 61). We measured geographic isolation as the sum of the proportions of landmass area in the surrounding of the target regions within buffer distances of 100 km, 1,000 km, and 10,000 km (37). Its value is lowest for remote islands and highest for regions located in the centers of large continents. We considered number of soil types (70) and elevational range (71) for each region as proxies for environmental heterogeneity. We also included five ecologically relevant climatic variables representing the main aspects of climate hypothesized to be important for plant endemism, namely mean annual temperature, mean annual precipitation, length of the growing season (i.e., number of days with temperatures exceeding a threshold of 0.9 °C, without snow cover, and with sufficient soil water), temperature seasonality (i.e., SD of mean monthly temperature

× 100), and precipitation seasonality (i.e., coefficient of variation in monthly precipitation) (72). Climatic variables were extracted as mean values per region from the input raster layers.

To determine the contribution of long-term climatic stability to PE, we calculated temperature stability since the LGM (21 Ka), velocity of temperature change since the LGM, and temperature anomaly since the mid-Pliocene warm period (~3.264 to 3.025 Ma). The LGM and the mid-Pliocene warm period represent cooler and warmer climates compared to the current climates, respectively. Temperature stability since the LGM was calculated using the *climateStablity* R package (73). It calculates temperature differences between 1,000-y time slices expressed as SD and averages the results across all time slices. Temperature stability is then calculated as the inverse of the mean SD rescaled to [0,1] (73). In addition, we calculated the velocity of temperature change since the LGM as the ratio between temporal change and contemporary spatial change in temperature, representing the speed with which a species would have to move its range to track analogous climatic conditions (56). Temperature anomaly since the mid-Pliocene warm period was calculated as the absolute difference in mean annual temperature between the mid-Pliocene warm period and present-day (74).

Predictors of Neo- and Paleoendemism. Neo- and paleoendemism are hypothesized to be driven primarily by the geological history of a region and by past climate change or stability (*SI Appendix*, Table S2). We, therefore, included the geographic type of each region (distinguishing between mainland regions and continental shelf islands, continental fragments, and oceanic islands) instead of surrounding landmass proportion, and elevational range (to distinguish between mountainous and non-mountainous regions) to represent geological history (30). We removed three islands with heterogeneous geological origin from further analyses on neo- and paleoendemism. To test for the impacts of past climate change on neo- and paleoendemism, we included the variables of long-term climatic stability introduced above.

Models of Phylogenetic Endemism. To assess the relationships between PE and environmental predictor variables, we fitted linear models with PE as a response variable. Beyond all predictor variables hypothesized to be important to PE (*SI Appendix*, Table S1), we included area size (km^2) to control for the over- and under-estimation of PE in large regions for PE.count and PE.area, respectively. We excluded regions with incomplete coverage of predictor variables, leading to a dataset including 818 regions (incl. 236 islands and 582 mainland regions; see *SI Appendix*, Fig. S14 for correlations between predictors). PE was $\log_{10}$-transformed before modeling. Some predictor variables (i.e., region area, elevational range, number of soil types, mean annual precipitation, temperature seasonality, precipitation seasonality, velocity in temperature change since the LGM, temperature stability since the LGM, and temperature anomaly since the mid-Pliocene warm period) were also $\log_{10}$-transformed to reduce the skewness of their distributions. All continuous predictor variables were then standardized to zero mean and unit variance to aid model fitting and make their parameter estimates comparable. To test whether the effects of environmental predictors on PE differ for isolated islands compared to less isolated mainland regions, we included the interaction between each predictor and surrounding landmass proportion. To test if including surrounding landmass proportion correctly encapsulated the effect of insularity, we updated the model by replacing surrounding landmass proportion with a categorical variable indicating whether a region is an island. Since these models performed worse than models including surrounding landmass proportion (*SI Appendix*, Table S9), we retained surrounding landmass proportion for all further analyses. We visualized the change in the coefficient of one variable in the interactions in dependence on the value of the other variable included using the function *interplot* in the R package *interplot* (75).

Species distributions, environmental predictors, and model residuals are often spatially autocorrelated, which may lead to biased parameter estimates and the violation of statistical assumptions (76). As spatial autocorrelation was detected in the model residuals (*SI Appendix*, Fig. S15 *A* and *B*), we included a spatial autocovariate that represents the spatial autocorrelation in the residuals of nonspatial models (residual autocovariate models, RAC) (77). This autocovariate term was implemented as a spatial weight matrix of nonspatial model residuals based on an optimized neighborhood structure. Because most of our regions

are political units with varying geometry and size, we used a sphere of influence to identify neighbors for each region (78). The sphere of influence for each focal region was defined as a circle around the centroid of a focal region within a radius equal to the distance to the centroid of the nearest neighboring region. When the sphere of influence of two regions overlapped, the two regions were considered neighbors. Overall, the RAC models successfully removed spatial autocorrelation from model residuals (*SI Appendix*, Fig. S15C).

Statistical Analyses of Neo- and Paleoendemism. To explore the potential drivers of spatial patterns of neo- and paleoendemism, we fitted ordinary linear models to explain the variation in RPE.ses only for regions that showed significantly high PE based on CANAPE (for CANAPE categories of each region, see doi: 10.6084/m9.figshare.21909822 (79); *SI Appendix*, Fig. S8). We removed velocity of temperature change since the LGM and temperature anomaly since the mid-Pliocene warm period because of their low explanatory power for RPE. ses based on AIC values. Predictors retained for modeling contained all three aspects (i.e., islands, mountains, and past climate change) hypothesized to affect neo- and paleoendemism. Likewise, we fitted spatial models by including a spatial autocovariate to remove spatial autocorrelation present in the residuals of the nonspatial models (*SI Appendix*, Fig. S16).

In addition, we compared the distribution of environmental factors for all regions (912 regions) across all CANAPE categories (i.e., neo, paleo, mixed, superendemic, and nonsignificant; *SI Appendix*, Figs. S11 and S12). Because the environmental factors were not normally distributed for each category separately, we used a nonparametric Kruskal–Wallis test followed by Wilcoxon pairwise comparisons (two-tailed tests with Holm's correction) to identify which categories were different from each other (50). We repeated all modeling procedures for two PE metrics (PE.area and PE.count) and the datasets with and without unplaced species added to the original phylogeny and excluding and including apomictic taxa.

Data, Materials, and Software Availability. Data and R codes needed to run the analyses are available at doi: 10.6084/m9.figshare.21909822 (79). An R package to retrieve the original plant checklists and floras from the Global Inventory of Floras and Traits (GIFT) (30) is available at https://CRAN.R-project.org/package=GIFT (62).

Author affiliations: [a]Biodiversity, Macroecology and Biogeography, University of Göttingen, Göttingen 37077, Germany; [b]Centre of Biodiversity and Sustainable Land Use, University of Göttingen, Göttingen 37077, Germany; [c]Campus-Institute Data Science, Göttingen 37077, Germany; [d]School of Natural Sciences, Macquarie University, Sydney, NSW 2109, Australia; [e]Department of Biosciences, Durham University, Durham DH1 3LE, United Kingdom; [f]Division of Bioinvasions, Global Change & Macroecology, University Vienna, Vienna 1030, Austria; [g]Ecology, Department of Biology, University of Konstanz, Konstanz 78464, Germany; [h]Zhejiang Provincial Key Laboratory of Plant Evolutionary Ecology and Conservation, Taizhou University, Taizhou 318000, China; [i]Czech Academy of Sciences, Institute of Botany, Department of Invasion Ecology, Průhonice 252 43, Czech Republic; [j]Department of Ecology, Faculty of Science, Charles University, Prague 128 44, Czech Republic; and [k]German Centre for Integrative Biodiversity Research Halle-Jena-Leipzig, Leipzig 04103, Germany

Author contributions: L.C., H.K., and P.W. designed research; L.C. performed research; L.C., H.K., A.T., J.S., W.D., F.E., M.v.K., J.P., P.P., M.W., and P.W. collected the data; and L.C., H.K., A.T., J.S., W.D., F.E., M.v.K., J.P., P.P., M.W., and P.W. wrote the paper.

1. S. N. Sheth, N. Morueta-Holme, A. L. Angert, Determinants of geographic range size in plants. *New Phytol.* **226**, 650–665 (2020).
2. B. J. Enquist *et al.*, The commonness of rarity: Global and future distribution of rarity across land plants. *Sci. Adv.* **5**, eaaz0414 (2019).
3. G. Kier *et al.*, A global assessment of endemism and species richness across island and mainland regions. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 9322–9327 (2009).
4. N. Myers, R. A. Mittermeier, C. G. Mittermeier, G. A. B. da Fonseca, J. Kent, Biodiversity hotspots for conservation priorities. *Nature* **403**, 853–858 (2000).
5. N. C. A. Pitman, P. M. Jørgensen, Estimating the size of the world's threatened flora. *Science* **298**, 989–989 (2002).
6. A. Purvis, P-M. Agapow, J. L. Gittleman, G. M. Mace, Nonrandom extinction and the loss of evolutionary history. *Science* **288**, 328–330 (2000).
7. G. M. Mace, J. L. Gittleman, A. Purvis, Preserving the tree of life. *Science* **300**, 1707–1709 (2003).
8. R. Gumbs *et al.*, The EDGE2 protocol: Advancing the prioritisation of evolutionarily distinct and globally endangered species for practical conservation action. *PLoS Biol.* **21**, e3001991 (2023).
9. D. P. Faith, Conservation evaluation and phylogenetic diversity. *Biol. Conserv.* **61**, 1–10 (1992).
10. S. Veron *et al.*, High evolutionary and functional distinctiveness of endemic monocots in world islands. *Biodivers. Conserv.* **30**, 3697–3715 (2021).
11. D. Rosauer, S. W. Laffan, M. D. Crisp, S. C. Donnellan, L. G. Cook, Phylogenetic endemism: A new approach for identifying geographical concentrations of evolutionary history. *Mol. Ecol.* **18**, 4061–4072 (2009).
12. B. D. Mishler *et al.*, Phylogenetic measures of biodiversity and neo- and paleo-endemism in Australian Acacia. *Nat. Commun.* **5**, 4473 (2014).
13. J. D. Thompson, S. Lavergne, L. Affre, M. Gaudeul, M. Debussche, Ecological differentiation of mediterranean endemic plants. *Taxon* **54**, 967–976 (2005).
14. C. E. Hughes, G. W. Atchison, The ubiquity of alpine plant radiations: From the andes to the hengduan mountains. *New Phytol.* **207**, 275–282 (2015).
15. M. Fernández-Mazuecos *et al.*, The radiation of Darwin's giant daisies in the Galápagos Islands. *Curr. Biol.* **30**, 4989–4998 (2020).
16. J. B. Losos, R. E. Ricklefs, Adaptation and diversification on islands. *Nature* **457**, 830–836 (2009).
17. P. Weigelt *et al.*, Global patterns and drivers of phylogenetic structure in island floras. *Sci. Rep.* **5**, 12213 (2015).
18. Q. C. B. Cronk, Relict floras of Atlantic islands: Patterns assessed. *Biol. J. Linn. Soc. Lond.* **46**, 91–103 (1992).
19. J. Fjeldså, J. C. Lovett, Geographical patterns of old and young species in African forest biota: The significance of specific montane areas as evolutionary centres. *Biodivers. Conserv.* **6**, 325–346 (1997).
20. M. Dynesius, R. Jansson, Evolutionary consequences of changes in species' geographical distributions driven by Milankovitch climate oscillations. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 9115–9120 (2000).
21. M. B. Davis, R. G. Shaw, Range shifts and adaptive responses to Quaternary climate change. *Science* **292**, 673–679 (2001).
22. R. Jansson, Global patterns in endemism explained by past climatic change. *Proc. R. Soc. Lond. B Biol. Sci.* **270**, 583–590 (2003).
23. K. D. Bennett, P. C. Tzedakis, K. J. Willis, Quaternary refugia of north European trees. *J. Biogeogr.* **18**, 103–115 (1991).
24. A. S. Jump, C. Mátyás, J. Peñuelas, The altitude-for-latitude disparity in the range retractions of woody species. *Trends Ecol. Evol.* **24**, 694–701 (2009).
25. M. D. Crisp, S. Laffan, H. P. Linder, A. Monro, Endemism in the Australian flora. *J. Biogeogr.* **28**, 183–198 (2001).
26. R. G. Gillespie, G. K. Roderick, Arthropods on islands: Colonization, speciation, and conservation. *Annu. Rev. Entomol.* **47**, 595–632 (2002).
27. Q. C. B. Cronk, Islands: Stability, diversity, conservation. *Biodivers. Conserv.* **6**, 477–493 (1997).
28. L.-M. Lu *et al.*, Evolutionary history of the angiosperm flora of China. *Nature* **554**, 234–238 (2018).
29. A. H. Thornhill *et al.*, Continental-scale spatial phylogenetics of Australian angiosperms provides insights into ecology, evolution and conservation. *J. Biogeogr.* **43**, 2085–2098 (2016).
30. P. Weigelt, C. König, H. Kreft, GIFT – A global inventory of floras and traits for macroecology and biogeography. *J. Biogeogr.* **47**, 16–43 (2020).
31. R. Govaerts, E. Nic Lughadha, N. Black, R. Turner, A. Paton, The world checklist of vascular plants, a continuously updated resource for exploring global plant diversity. *Sci. Data* **8**, 215 (2021).
32. S. A. Smith, J. W. Brown, Constructing a broadly inclusive seed plant phylogeny. *Am. J. Bot.* **105**, 302–314 (2018).
33. D. Hojsgaard, S. Klatt, R. Baier, J. G. Carman, E. Hörandl, Taxonomy and biogeography of apomixis in angiosperms and associated biodiversity characteristics. *CRC Crit. Rev. Plant Sci.* **33**, 414–427 (2014).
34. B. D. Mishler, Spatial phylogenetics. *J. Biogeogr.* **50**, jbi.14618 (2023).
35. A. V. Rudbeck *et al.*, The darwinian shortfall in plants: phylogenetic knowledge is driven by range size. *Ecography* **2022**, e06142 (2022).
36. B. H. Daru, H. Farooq, A. Antonelli, S. Faurby, Endemism patterns are scale dependent. *Nat. Commun.* **11**, 2115 (2020).
37. P. Weigelt, H. Kreft, Quantifying island isolation–insights from global patterns of insular plant species richness. *Ecography* **36**, 417–429 (2013).
38. A. L. Takhtajan, *Floristic Regions of the World* (University of California Press, 1986).
39. R. V. Gallagher *et al.*, Global shortfalls in threat assessments for endemic flora by country. *Plants People Planet* **5**, ppp3.10369 (2023).
40. Q. C. B. Cronk, The history of endemic flora of St Helena: A relictual series. *New Phytol.* **105**, 509–520 (1987).
41. W. L. Clement *et al.*, Phylogenetic position and biogeography of *Hillebrandia sandwicensis* (Begoniaceae): A rare Hawaiian relict. *Am. J. Bot.* **91**, 905–917 (2004).
42. L. Zhang *et al.*, The water lily genome and the early evolution of flowering plants. *Nature* **577**, 79–84 (2020).
43. Y. Kisel, T. G. Barraclough, Speciation has a spatial scale that depends on levels of gene flow. *Am. Nat.* **175**, 316–334 (2010).
44. A. Antonelli *et al.*, Madagascar's extraordinary biodiversity: Evolution, distribution, and use. *Science* **378**, eabf0869 (2022).
45. A. D. Yoder, M. D. Nowak, Has vicariance or dispersal been the predominant biogeographic force in Madagascar? Only time will tell. *Annu. Rev. Ecol. Evol. Syst.* **37**, 405–431 (2006).
46. S. Buerki, D. S. Devey, M. W. Callmander, P. B. Phillipson, F. Forest, Spatio-temporal history of the endemic genera of Madagascar. *Bot. J. Linn. Soc.* **171**, 304–329 (2013).
47. Ş Procheş, S. Ramdhani, S. J. Perera, J. R. Ali, S. Gairola, Global hotspots in the present-day distribution of ancient animal and plant lineages. *Sci. Rep.* **5**, 15457 (2015).
48. J. M. Fernández-Palacios *et al.*, Scientists' warning – The outstanding biodiversity of islands is in peril. *Glob. Ecol. Conserv.* **31**, e01847 (2021).
49. L. Cai *et al.*, Global models and predictions of plant diversity based on advanced machine learning techniques. *New Phytol.* **237**, 1432–1445 (2023).
50. L. M. J. Dagallier *et al.*, Cradles and museums of generic plant diversity across tropical Africa. *New Phytol.* **225**, 2196–2213 (2020).
51. V. S. F. T. Merckx *et al.*, Evolution of endemism on a young tropical mountain. *Nature* **524**, 347–350 (2015).
52. Y. Xing, R. H. Ree, Uplift-driven diversification in the Hengduan Mountains, a temperate biodiversity hotspot. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E3444–E3451 (2017).
53. A. Antonelli *et al.*, Geological and climatic influences on mountain biodiversity. *Nat. Geosci.* **11**, 718–725 (2018).
54. C. Rahbek *et al.*, Building mountain biodiversity: Geological and evolutionary processes. *Science* **365**, 1114–1119 (2019).
55. S. G. A. Flantua *et al.*, Snapshot isolation and isolation history challenge the analogy between mountains and islands used to understand endemism. *Global Ecol. Biogeogr.* **29**, 1651–1673 (2020).
56. B. Sandel *et al.*, The influence of late quaternary climate-change velocity on species endemism. *Science* **334**, 660–664 (2011).
57. K. L. Evans, P. H. Warren, K. J. Gaston, Species–energy relationships at the macroecological scale: A review of the mechanisms. *Biol. Rev.* **80**, 1–25 (2005).
58. L. M. Valente, V. Savolainen, P. Vargas, Unparalleled rates of species diversification in Europe. *Proc. R. Soc. B Biol. Sci.* **277**, 1489–1496 (2010).
59. D. Storch, P. Keil, W. Jetz, Universal species–area and endemics–area relationships at continental scales. *Nature* **488**, 78–81 (2012).
60. A. H. Thornhill *et al.*, Spatial phylogenetics of the native California flora. *BMC Biol* **15**, 96 (2017).
61. B. Sandel *et al.*, Current climate, isolation and history drive global patterns of tree phylogenetic endemism. *Glob. Ecol. Biogeogr.* **29**, 4–15 (2020).
62. P. Denelle, P. Weigelt, H. Kreft, GIFT – an R package to access the Global Inventory of Floras and Traits. bioRxiv [Preprint] (2023). https://doi.org/10.1101/2023.06.27.546704 (accessed 22 October 2022).
63. S. A. Chamberlain, E., Szöcs, taxize: Taxonomic search and retrieval in R. *F1000Res* **2**, 191 (2013).
64. L. Majeský, F. Krahulec, R. J. Vašut, How apomictic taxa are treated in current taxonomy: A review. *Taxon* **66**, 1017–1040 (2017).
65. M. Sochor, R. J. Vašut, T. F. Sharbel, B. Trávníček, How just a few makes a lot: Speciation via reticulation and apomixis on example of European brambles (Rubus subgen. Rubus, Rosaceae). *Mol. Phylogenet. Evol.* **89**, 13–27 (2015).
66. M. Tietje *et al.*, Global variation in diversification rate and species richness are unlinked in plants. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2120662119 (2022).
67. W. D. Pearse *et al.*, *pez*: Phylogenetics for the environmental sciences. *Bioinformatics* **31**, 2888–2890 (2015).
68. H. Qian, R. E. Ricklefs, W. Thuiller, Evolutionary assembly of flowering plants into sky islands. *Nat. Ecol. Evol.* **5**, 640–646 (2021).
69. H. Qian, Y. Jin, Are phylogenies resolved at the genus level appropriate for studies on phylogenetic structure of species assemblages? *Plant Divers.* **43**, 255–263 (2021).
70. T. Hengl *et al.*, SoilGrids250m: Global gridded soil information based on machine learning. *PLoS One* **12**, e0169748 (2017).
71. J. J. Danielson, D. B. Gesch Global multi-resolution terrain elevation data 2010 (GMTED2010) (2011). https://doi.org/10.3133/ofr20111073. Accessed 02 May 2017.
72. D. N. Karger *et al.*, Climatologies at high resolution for the earth's land surface areas. *Sci. Data* **4**, 170122 (2017).
73. H. L. Owens, R. Guralnick, climateStability: An R package to estimate climate stability from time-slice climatologies. *Biodivers. Inform.* **14**, 8–13 (2019).
74. J. L. Brown, D. J. Hill, A. M. Dolan, A. C. Carnaval, A. M. Haywood, PaleoClim, high spatial resolution paleoclimate surfaces for global land areas. *Sci. Data* **5**, 1–9 (2018).
75. F. Solt, Y., Hu, interplot: Plot the effects of variables in interaction terms (2015). https://CRAN.R-project.org/package=interplot. Accessed 19 May 2022.
76. C. Dormann *et al.*, Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography* **30**, 609–628 (2007).
77. B. Crase, A. C. Liedloff, B. A. Wintle, A new method for dealing with residual spatial autocorrelation in species distribution models. *Ecography* **35**, 879–888 (2012).
78. J. Y. Lim, J.-C. Svenning, B. Göldel, S. Faurby, W. D. Kissling, Frugivore-fruit size relationships between palms and mammals reveal past and future defaunation impacts. *Nat. Commun.* **11**, 4904 (2020).
79. L. Cai, Data and codes for exploring global centers of neo- and paleoendemism in seed plants. figshare. https://doi.org/10.6084/m9.figshare.21909822. Deposited 17 January 2023.