Research

Phytodiversity is associated with habitat heterogeneity from Eurasia to the Hengduan Mountains

Yaquan Chang[1,2,3] (iD), **Katrina Gelwick**[4] (iD), **Sean D. Willett**[4] (iD), **Xinwei Shen**[5] (iD), **Camille Albouy**[1,2] (iD), **Ao Luo**[6] (iD), **Zhiheng Wang**[6] (iD), **Niklaus E. Zimmermann**[3]* (iD) and **Loïc Pellissier**[1,2]* (iD)

[1]Ecosystems and Landscape Evolution, Department of Environmental Systems Science, ETH Zürich, Universitätstrasse 16, 8092, Zürich, Switzerland; [2]Ecosystems and Landscape Evolution, Land Change Science Research Unit, Swiss Federal Institute for Forest, Snow and Landscape Research (WSL), Zürcherstrasse 111, 8903, Birmensdorf, Switzerland; [3]Dynamic Macroecology, Land Change Science Research Unit, Swiss Federal Institute for Forest, Snow, and Landscape Research (WSL), Zürcherstrasse 111, 8903, Birmensdorf, Switzerland; [4]Earth Surface Dynamics, Department of Earth Sciences, ETH Zürich, Sonneggstrasse 5, 8092, Zürich, Switzerland; [5]Department of Mathematics, Seminar for Statistics, ETH Zürich, Rämistrasse 101, 8092, Zürich, Switzerland; [6]Institute of Ecology and Key Laboratory for Earth Surface Processes of the Ministry of Education, College of Urban and Environmental Sciences, Peking University, Beijing, 100871, China

Author for correspondence:
Yaquan Chang
Email: yaquanchang0623@gmail.com

Summary

- The geographic distribution of plant diversity matches the gradient of habitat heterogeneity from lowlands to mountain regions. However, little is known about how much this relationship is conserved across scales.
- Using the World Checklist of Vascular Plants and high-resolution biodiversity maps developed by species distribution models, we investigated the associations between species richness and habitat heterogeneity at the scales of Eurasia and the Hengduan Mountains (HDM) in China.
- Habitat heterogeneity explains seed plant species richness across Eurasia, but the plant species richness of 41/97 HDM families is even higher than expected from fitted statistical relationships. A habitat heterogeneity index combining growing degree days, site water balance, and bedrock type performs better than heterogeneity based on single variables in explaining species richness. In the HDM, the association between heterogeneity and species richness is stronger at larger scales.
- Our findings suggest that high environmental heterogeneity provides suitable conditions for the diversification of lineages in the HDM. Nevertheless, habitat heterogeneity alone cannot fully explain the distribution of species richness in the HDM, especially in the western HDM, and complementary mechanisms, such as the complex geological history of the region, may have contributed to shaping this exceptional biodiversity hotspot.

Introduction

Mountain regions cover only a quarter of the Earth's land surface but harbour a considerable proportion of its terrestrial biodiversity across different taxa (Körner, 2000; Rahbek *et al.*, 2019b). Different biogeographic principles have been proposed to explain the high biodiversity within mountains (Spehn & Körner, 2005), with habitat heterogeneity being one of the most prominent drivers. Mountain ranges are typically linked to high habitat heterogeneity (Rahbek *et al.*, 2019a), which offer diverse niches for species to occupy. Habitat heterogeneity can be measured in different dimensions that correspond to the biological requirements of the species, such as climate (Udy *et al.*, 2021), and soil and bedrock conditions (Jiménez-Alfaro *et al.*, 2021). Habitat dimensions jointly determine the niche space available for diversifying lineages (Ricklefs, 2010). Given the multidimensional niche axes of any given plant (Silvertown, 2004), habitat heterogeneity based on multiple environmental axes should be better at explaining species richness (SR) patterns than single-dimensional heterogeneity, but the generality of these relationships should be assessed across scales from continents to local mountain regions.

To study the drivers of biodiversity distribution, not only the inferred relationships but also their residuals can be informative (Rahbek *et al.*, 2019b). In the context of the relationship between habitat heterogeneity and species richness (hereafter given as heterogeneity–richness relationship), statistical residuals can help highlight regions or families that do not conform to the general patterns. Several mountain regions of Eurasia have been highlighted as showing an exceptionally high level of diversity (Rahbek *et al.*, 2019a), and among these mountain ranges, the Hengduan Mountains region (HDM) stands out. The HDM represents the main biodiversity hotspot outside of the tropics

*These authors contributed equally to this work.

(Rahbek *et al.*, 2019b), and hosts over 12 800 plant species (Sun *et al.*, 2017), compared with 4000 species in the European Alps (Chauvier *et al.*, 2021). Detailed knowledge of species distributions within this exceptional biodiversity hotspot, and the drivers that have shaped it, is limited, although some qualitative assessments exist (Wu, 1988). The complex geological and climate histories of the HDM (Mulch & Chamberlain, 2006; Favre *et al.*, 2015) have created diverse habitats, barriers, and crossroads within a relatively small area, with climates ranging from cold and partly humid in alpine areas to dry and hot in deep valleys (Yang *et al.*, 2020). Habitat heterogeneity is further increased in the HDM by widespread, variable tectonic uplift (Gourbet *et al.*, 2020). Together, the complex climate and geological histories of the HDM might have led to the emergence of the biodiversity hotspot by favouring *in situ* speciation and colonisation in response to the creation of novel ecological niches (Xing & Ree, 2017; Ding *et al.*, 2020). Furthermore, the region is characterised by pervasive tectonic deformation, with large displacements on transcurrent faults, leading to widespread habitat fragmentation and exposure of variable bedrock (Hartmann & Moosdorf, 2012). The relationships between habitat heterogeneity and biodiversity in this region remain largely unquantified, especially regarding how consistent the relationships are across plant families.

Understanding the interplay of the drivers of biodiversity distribution in mountain regions requires a broad comparison among multiple mountain systems with high-resolution data. However, there is a general trade-off between the extent and resolution of biodiversity data (König *et al.*, 2019). Species range maps primarily result from synthesising approaches using global biodiversity databases (e.g. World Checklist of Vascular Plants (WCVP), Govaerts *et al.*, 2021, Botanical Information and Ecology Network (BIEN, https://bien.nceas.ucsb.edu/bien/), Global Biodiversity Information

Facility (GBIF, https://www.gbif.org/)), and they are often available only at coarse resolution (e.g. generalised polygons (Fig. 1a) or large pixels (i.e. 1°)). Such data do not accurately represent biodiversity distributions within mountain regions (Antonelli *et al.*, 2018). The WCVP database is currently the most comprehensive inventory of vascular plants at the global scale. The plant distributions in this database are polygon based at the province-to-country level (hereafter botanical country), the product of a large effort on the part of taxonomists, and local experts, yet the resolution remains coarse (Govaerts *et al.*, 2021). By contrast, high-resolution species distribution data are only available where sampling has been extended using statistical techniques to extrapolate and interpolate based on environmental factors (i.e. using species distribution models, Guisan & Zimmermann, 2000). Such better-resolved data on species' distributions are only available for specific regions (e.g. European Alps, Chauvier *et al.*, 2021) or globally for specific clades (e.g. Fagales and Pinales; Lyu *et al.*, 2022). It is computationally demanding to generate high-resolution diversity patterns over large regions for multiple species, due to sampling bias and inconsistency in collection efforts (Hortal *et al.*, 2015). The integration of large-scale yet locally less precise biogeographic patterns (Tietje *et al.*, 2022) and high-resolution spatial data from small areas (Chauvier *et al.*, 2021) could be instrumental to connect our knowledge across scales. Thus, combining global checklists with high-resolution data in a focal area could allow us to better understand the mechanisms that may have shaped biodiversity distribution at both continental and regional scales.

Here, we took advantage of the comprehensive WCVP database and combined it with high-resolution mapping using species distribution models to study biodiversity drivers across spatial scales. Specifically, we aimed to assess the novelty of the diversity of the HDM nested within the Eurasian region and to determine to what degree and which families within the HDM display
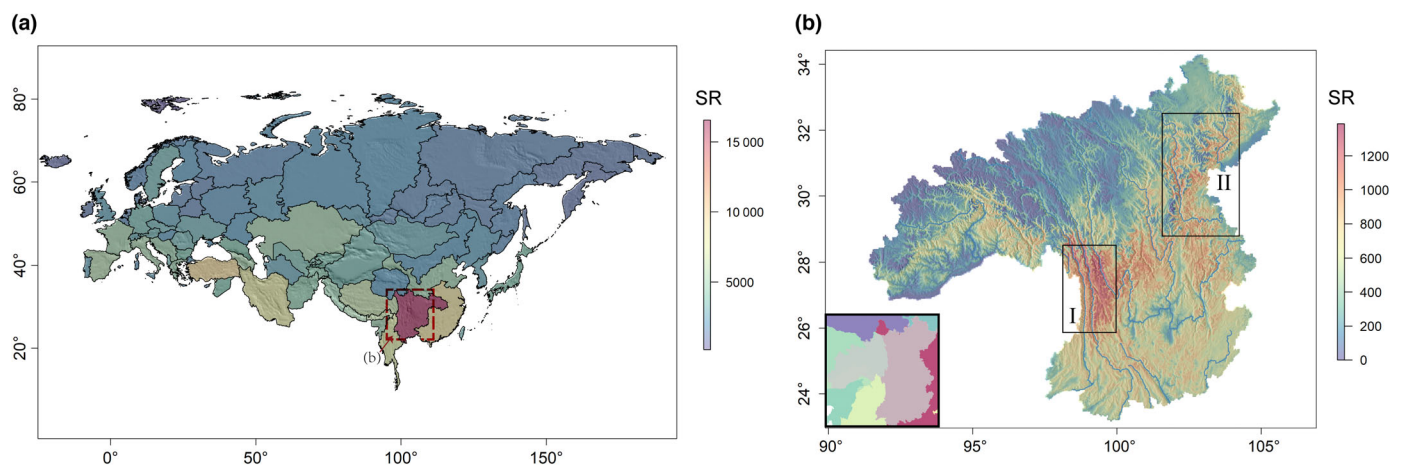

Fig. 1 Vascular plant species richness at the Eurasian scale (a) and the Hengduan Mountains (HDM) scale (b). At the Eurasian scale, we extracted the plant species richness (SR) of 97 selected families from the World Checklist of Vascular Plants (WCVP; Govaerts *et al.*, 2021) and the botanical countries shown with grey borders. We applied species distribution modelling at the HDM scale to map species ranges at a 1-km spatial resolution, based on data derived from Lyu *et al.* (2021) and the Plant Science Data Center (2021). The HDM region mainly covers western China-South-Central (CHC) and Tibet (CHT) botanical countries (polygons), highlighted with a dashed box in (a) and a bolded square box in (b). We specified the HDM as the CHC polygon alone in the WCVP database to perform analyses at the Eurasian scale (a) but kept the original HDM definition to perform analyses at the HDM scale (b). At the HDM scale (b), we identified two hotspots, which we refer to as the Three Rivers Region (I), and the Longmenshan region (II). See Supporting Information Fig. S1 for a detailed comparison of the spatial extents.

outstanding levels of diversity. We then explained the species richness patterns of the exceptionally rich families in the HDM across spatial scales using habitat heterogeneity as a predictor. More specifically, we addressed the following questions:

(**1**) Is habitat heterogeneity associated with species richness at different scales, and does a compound index integrating growing degree days, site water balance, and bedrock type perform better than habitat heterogeneity based on single variables in explaining species richness patterns?

(**2**) Compared with the other mountain regions in Eurasia, is the biodiversity of the HDM region exceptionally high, or does it simply follow the Eurasian relationship between species richness and habitat heterogeneity (i.e. 95th quantile)? If exceptional, which families contribute most to this hotspot? Are there any phylogenetic signals or ecological similarities among outlier families?

(**3**) To what extent do different habitat heterogeneity predictors explain the observed species richness distribution, and are there specific geographic patterns in the residuals at both Eurasia and the HDM scale?

We first focused on an analysis of Eurasia-wide data using the WCVP database, evaluating the heterogeneity–richness relationships for each family. After identifying exceptionally rich families, we used a 1 km resolution richness map at the scale of the HDM to investigate the extent to which habitat heterogeneity explains richness patterns at different local scales for these families within the HDM.

Materials and Methods

Study areas at the Eurasian and Hengduan Mountains scales

To study the mechanisms shaping the HDM biodiversity hotspot, we used the polygon-based WCVP distribution data at the Eurasian scale (Fig. 1a), as well as a high-resolution compilation of distribution maps of seed plant species within the HDM (Fig. 1b). At the Eurasian scale, we confined our study to regions of subtropical to polar climates, based on the Köppen–Geiger classification (Beck *et al.*, 2018). We started with the WCVP polygon map at the scale of the whole Eurasian continent, in which each polygon corresponds to a botanical country (Brummitt *et al.*, 2001). We excluded polygons that include > 50% tropical rainforest (Af), tropical monsoon (Am), tropical savanna (Aw), and arid, hot desert (Bwh) based on the 1 km resolution Köppen–Geiger climate map (Beck *et al.*, 2018). The final study region included 87 polygons within Eurasia. At the Eurasian scale, the HDM region is represented as the China-South-Central (CHC) polygon in the botanical country map (red colour polygon in Fig. 1a). For the higher-resolution analysis, we defined the HDM as the counties in southeastern Tibet, western Sichuan, and northern and central Yunnan provinces of China (Fig. 1b). This definition follows several studies about HDM biodiversity (Ding *et al.*, 2020; Li *et al.*, 2021), but is expanded slightly to the south to capture sufficient niche space in the species distribution modelling procedure. The spatial extent of the Eurasia and HDM scale comparison is shown in Supporting Information Fig. S1.

World Checklist of Vascular Plants database

To investigate the species richness pattern of seed plants at the Eurasian scale, we used the updated WCVP database (Govaerts *et al.*, 2021) from Royal Botanic Gardens, Kew, which represents a compilation of the geographical distribution of each plant species. After removing hybrid species and merging forms, varieties, and subspecies to the species level, we selected cosmopolitan seed plant families based on the following criteria: (1) the family occupies > 15 polygons in the WCVP map (Fig. 1a); (2) the distribution of these families includes the CHC polygon; (3) the family is composed of > 50 species. Based on these criteria, we kept 100 families for further analyses (see Table S1 for the family-level richness summaries at both Eurasia and HDM scales).

Hengduan Mountains biodiversity data

At the scale of the HDM, we developed a species distribution modelling pipeline to map species diversity from county-level polygon distribution data and information on the elevational distribution per species at a 1 km spatial resolution. We compiled county-level species distribution data from Lyu *et al.* (2021) and province-level distribution data from the Plant Science Data Center (List of plant species in China, 2021 Edition; https://www.plantplus.cn/doi/10.12282/plantdata.0021). For each species, we used the intercept of the two distributions as the final county-level species distribution input to construct a more conservative county-level distribution (Fig. S2). We standardised the Latin names of species from different data sources following the Catalogue of Life (https://www.catalogueoflife.org/). We cleaned synonyms and removed cultivated species, ferns, and invasive and aquatic species. We then compiled elevation information from the Flora of China (http://www.efloras.org/flora_page.aspx?flora_id=2) and local floras (Wu, 1986; Wu, 1987; Zhou, 1994; Wang, 1994). We used species elevation range information to mask areas where individual species are unlikely to occur within the county-level distribution data (Li *et al.*, 2021). Next, we used species distribution models (SDMs) to downscale distribution data from the county scale to a 1 km resolution to account for detection probability in the county-level distribution in the HDM. Not all counties were inventoried with the same intensity, and thus, the counties vary in the completeness of their inventories. We set up the species distribution modelling pipeline as follows: (1) we rasterised the elevation-masked county-level distribution data at 1 km resolution; (2) we sampled the presence pixels spatially at random and used the distribution maps to randomly sample pseudo-absence pixels from the non-presence background area (see Fig. S3 for the detailed sampling procedure and Table S2 for the presence/pseudo-absence summary); (3) we selected six variables representing climate and soil conditions to build SDMs; (4) we built SDMs for each species using generalised linear models (GLMs; Nelder & Wedderburn, 1972), generalised additive models (GAMs; Hastie & Tibshirani,

1986), and gradient boosting machines (GBMs; Friedman, 2001); and (5) we mapped species distributions by model ensembles using the committee averaging method and by clipping the modelled distribution with a buffer around presence pixels and an ecoregion polygon (https://geospatial.tnc.org/datasets/TNC::terrestrial-ecoregions/about) to avoid excessively wide extrapolation (Fig. S2). This is a conservative approach to mapping species distributions at finer spatial scales without extrapolating far beyond the counties in which the species was observed.

We derived the climate variables for species distribution modelling from the Climatologies at High resolution for the Earth's Land Surface Areas (CHELSA; Karger et al., 2017): mean temperature from 1981 to 2010 (bio1), mean precipitation from 1981 to 2010 (bio12), and precipitation seasonality from 1981 to 2010 (bio15). We extracted the soil conditions from SoilGrids (Poggio et al., 2021): soil pH at 5 cm, soil coarse fragment content at 5 cm, and soil silt content at 5 cm (the detailed predictor selection procedure is described in Notes S1; Figs S4, S5). As our presence data came from nonstandardised sampling, we built SDMs with relatively low degrees of complexity to describe the general relationship between occurrence and environmental data (Merow et al., 2014; Brun et al., 2020). We used second-order polynomials for GLMs, fitted GAMs with three degrees of freedom, and limited the number of trees for training to 1000 trees for GBMs. After fitting the models, we evaluated model quality using the true skill statistic (Fig. S6; TSS; Allouche et al., 2006) in a fivefold cross-validation procedure (see detailed SDM downscaling pipeline in Notes S2). We removed models with a TSS < 0.35 before assembling species distributions, and ensemble presence–absence species range maps using TSS as a maximisation threshold criterion (Allouche et al., 2006. We layered all species range maps to generate the species richness distribution of the HDM at a 1 km resolution (Fig. S7).

As the spatial delineation of the HDM mapped in the WCVP database differs from the regional definition of the HDM, we quantified spatial and taxonomic overlaps between these two scales. Although the two scales overlap spatially by only 44.1% (Fig. S1), the family-level taxonomic overlap is 97%. To avoid bias when analysing the two scales, we restricted all further analyses to 97 families that occurred in both datasets. These families comprise 74 919 species at the Eurasia scale and 12 356 species at the HDM scale. This represents 91.8% and 92.8% of all species at the Eurasia and HDM scale, respectively (see Table S1 for the family-level data summary and Fig. S8 for the unselected plant richness pattern).

Habitat heterogeneity data

To compute 'habitat heterogeneity' at the Eurasian and HDM scales, we created an index at 1 km resolution based on climate and lithology maps, as these variables have been identified as important components of the habitat heterogeneity in mountain systems (Rahbek et al., 2019a,b). For climate variables, we selected annual mean growing degree days above a 0°C threshold (gdd) and site water balance (swb) from 1981 to 2010 from the CHELSA climate layers (Karger et al., 2017; Brun et al., 2022).

We selected gdd as a proxy for the required thermal energy of plants (Zimmermann & Kienast, 1999) and swb as a proxy for the water availability to plants (Woodward & Williams, 1987). Because most of the flora in the HDM is composed of mountain species, we set the required gdd threshold to 0°C. We then reclassified gdd and swb into nine classes spread evenly across Eurasia. As soils originating from different bedrock types can, through geochemical processes, act as diverse filters for plant specialisation and diversity Rahbek et al., 2019a, we used bedrock type as one heterogeneity dimension in our study. We rasterised a 1 : 1 000 000 lithology map (Hartmann & Moosdorf, 2012) at a 1 km spatial resolution and set cells with no information, glaciers, or water bodies to NA, which left 13 types of bedrock. Based on these three classified maps, we computed habitat heterogeneity as the number of unique combinations of the three layers per WCVP polygon. To do so, we first defined a 'compound index' integrating gdd, swb, and bedrock type maps to obtain a unique combination of each dimension (Guisan et al., 2017). Thus, each value represented a unique combination of these environmental predictors. To compute habitat heterogeneity at both Eurasian and HDM scales, we calculated the Shannon diversity index (Shannon, 2001) for each WCVP polygon based on the compound index, gdd, swb, and bedrock type habitat maps. For a detailed flowchart describing how we generated the heterogeneity map, see Fig. S9. At the HDM scale, we computed the Shannon heterogeneity for the compound index defined above, as well as for gdd, swb, and bedrock types individually within different neighbourhood sizes using the 'focal' function in the RASTER package (Hijmans & van Etten, 2016) in R v.4.2.0 (R Core Team, 2022). We chose window sizes with a range of 5–285 km, with intervals of 20 km as neighbourhoods (See Video S1 for compound index changes).

Relationship between heterogeneity and species richness across scales

We investigated the relationship between species richness and heterogeneity for each family at the Eurasian scale. To eliminate the confounding effect of the area on heterogeneity, we first built a model relating heterogeneity to the area and then computed the model residuals as a measure of heterogeneity with the effect of area removed. This approach has been shown to give unbiased coefficients (Freckleton, 2002). As heterogeneity increases with the increasing area, reaching a plateau (Fig. S10), we used the 'SSarrhenius' function in the R package VEGAN (Oksanen et al., 2013) with the Arrhenius relationship (i.e. $S = k \times A^z$), where S represents the habitat heterogeneity per WCVP polygon, A represents the WCVP polygon area, k represents the expected number of species in the polygon area, and z is the slope of the heterogeneity-area curve (Arrhenius, 1921). After using the model residuals as area-corrected habitat heterogeneity at the Eurasian scale, we built the heterogeneity–richness relationship using quantile GAMs, as there is no uniform and a priori expected shape of the heterogeneity–richness relationship. Previous mountain studies (e.g. Rahbek et al., 2019b and a heterogeneity–richness synthesis (Stein et al., 2014) have

suggested a positive relationship. We used the 'qgam' function with the 5th, 50th, and 95th quantiles in the R package QGAM (Fasiolo *et al.*, 2021) to fit the habitat heterogeneity–richness relationships for each family. The QGAM package is an extension of the MGCV (Wood & Wood, 2015) package developed to construct GAMs for different quantiles (Fasiolo *et al.*, 2020) and has been applied in ecological research (He *et al.*, 2022; Nunes *et al.*, 2022).

To evaluate the performance of the models, we calculated the adjusted variance (R^2) at the 50th quantile, as this metric corrects the number of predictors which can be compared across different models. We further performed a one-way Analysis of Variance (ANOVA) to assess whether the adjusted R^2 of the heterogeneity–richness models differed significantly among compound, *gdd*, *swb*, bedrock, and multivariate indices and to assess which model offered better species richness explanations. To further investigate the heterogeneity–richness relationship at the HDM scale, we aggregated the 1 km species richness maps per outlier family at 20 km intervals for 5–285 km window sizes within the HDM. To minimise pseudo-replication generated in the focal analyses and to conserve the same number of analysis windows, we randomly allocated a regularly spaced lattice with a point distance of 125 km across the study area. We then extracted the habitat heterogeneity predictors and species richness data randomly for all window sizes at the lattice point locations. Next, we performed the same quantile regression and evaluated the 50th quantile for heterogeneity–richness models for the compound, individual (*gdd, swb*, and bedrock types), and multivariate (SR $\sim$ *gdd* + *swb* + bedrock) heterogeneity indices. We extracted the adjusted R^2 for each model and computed the mean and standard error. With this procedure, only window sizes larger than 125 km overlapped, which allowed us to analyse at least 50 points per random replicate in the heterogeneity–richness regression analyses. We additionally performed a robustness test with a minimum sampling distance of 185 km and a window size gradient from 5 to 185 km to fully avoid pseudo-replication (25 lattice points per replicate, Fig. S11). We analysed windows ranging from 5 to 185 km and again assessed the adjusted R^2 of heterogeneity–richness models across the window sizes. The results were similar between the 125 km lattice with 50 sampling points and the 185 km lattice with 25 sampling points, yet the latter showed slightly more variation in adjusted R^2 values among replicates, due to the small sample size. We, therefore, accepted the slight overlap with the 125 km lattice (Fig. S12). We also performed sensitivity analyses for nonoutlier families and the species richness map solely based on elevation to evaluate the robustness of our findings (Figs S13, S14).

Spatial residual analyses

At both the Eurasian and the HDM scales, we investigated the spatial distribution of residuals. At the Eurasian scale, we identified not only the regions with the highest residuals but also the families that contributed to the residuals. To do so, we highlighted the regions with the highest residuals of the 95th quantile in the heterogeneity–richness relationship. We did so to assess whether

the HDM belongs to the regions with the highest species richness compared with the general expectation of species richness predicted by habitat heterogeneity. We then identified the families which exceeded the 95th species richness prediction in the CHC region and denoted them as HDM-outlier families (Fig. 2c). At the HDM scale, we generated predicted richness maps from all habitat heterogeneity models for the entire HDM at a resolution of 205 km. We then obtained residual richness maps by subtracting the predicted SR maps from actual SR maps.

We highlighted the outlier families in the plant tree of life developed by Smith & Brown (2018) to test the strength of the phylogenetic signal of the outlier families identified at the Eurasian scale (Fig. 2c). To this end, we calculated the Fritz and Purvis *D* statistic on outlier/nonoutlier families (Fritz & Purvis, 2010). In addition, we collected information about the main biomes of the families at the Eurasian scale, as well as the elevation ranges of those families at the HDM scale. We further investigated the environmental preferences of those families, mainly characterised by elevation distribution.

Results

Mapping plant species richness in the HDM

At the HDM scale, all GLMs, GAMs, and GBMs for all 13 221 species perform well, yielding a mean TSS of 0.85 ± 0.11 in GLMs, 0.86 ± 0.10 in GAMs, and 0.89 ± 0.08 in GBMs (Fig. S6). The comparison of the TSS between outlier and nonoutlier families reveals that both categories yield excellent performance. The proposed SDM approach increases the detection probability and thus increases the species range continuity and the number of species per polygon, as indicated by species accumulation curves (Fig. S15; Notes S3). For instance, in the Sichuan and Tibet regions, the modelled species richness increases from 9740 to 10 791 and from 7634 to 8879, respectively, when SDM maps are used instead of the original county-level maps.

Relationship between habitat heterogeneity and species richness across scales

At the Eurasian scale, compound heterogeneity and habitat heterogeneity computed solely from *gdd* have significantly higher explanatory power than other habitat heterogeneity models (compound adjusted R^2: $12.67\% \pm 14.20\%$; *gdd* adjusted R^2: $12.53\% \pm 12.83\%$; *swb* adjusted R^2: $7.37\% \pm 12.48\%$; bedrock adjusted R^2: $1.27\% \pm 3.71\%$; multivariate model adjusted R^2: $7.33\% \pm 12.52\%$; Fig. 2b; Table S3, see Tables S4–S8 for family-level model results). At the HDM scale, all heterogeneity indices, except *swb*, perform better at larger window sizes. The compound Shannon habitat heterogeneity index has higher explanatory power than any of the individual indices at large scales (Fig. 3). The explanatory power of the compound index starts to exceed other habitat heterogeneity indices at 145 km, with an adjusted R^2 of $38.2\% \pm 0.23\%$, and reaches an asymptote at a window size of *c.* 205 km ($45.8\% \pm 0.34\%$ explained, see Table S9 for detailed scale comparisons). This result holds
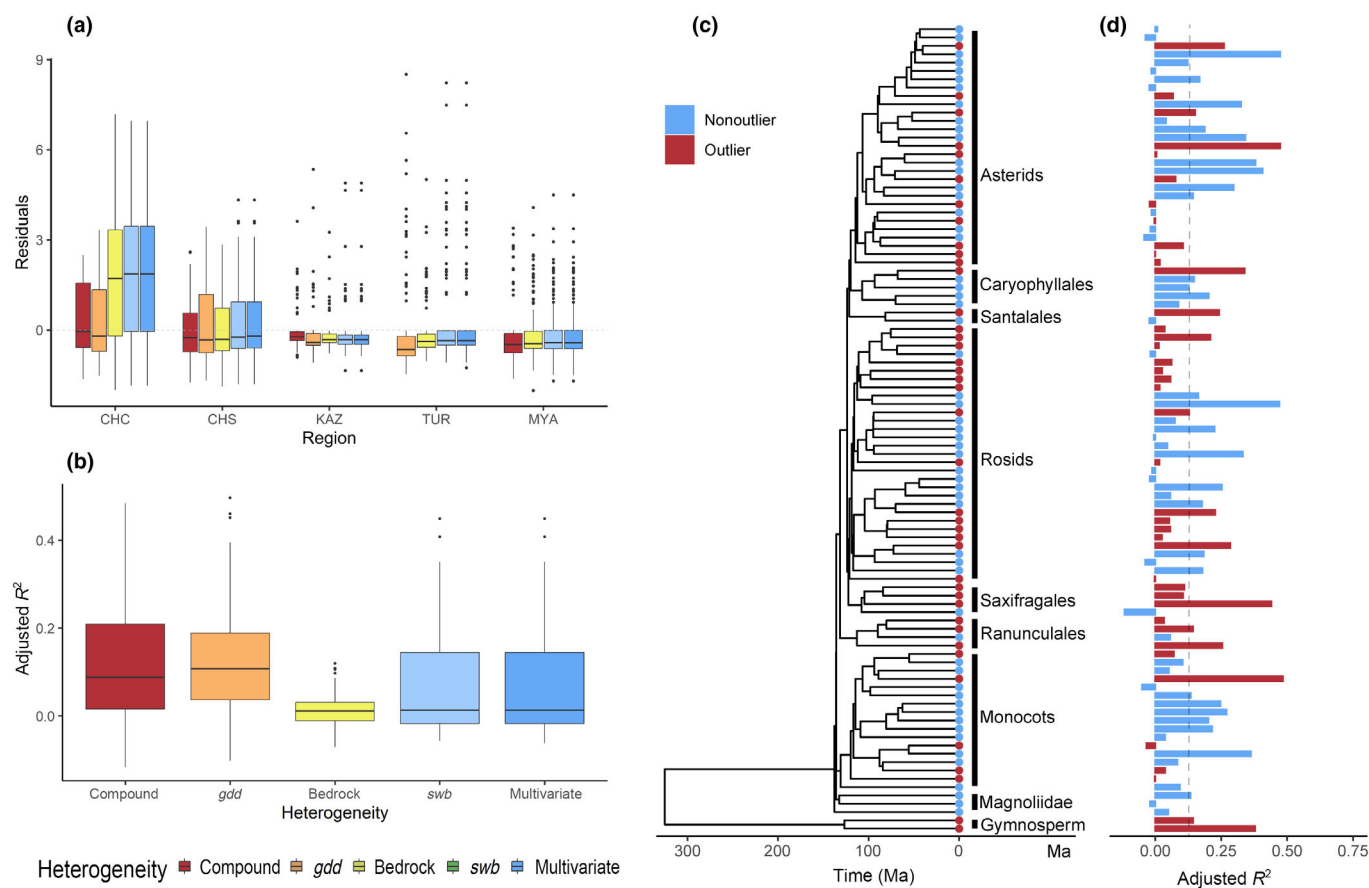
Fig. 2 Habitat heterogeneity model residuals across different geographic regions (a, see Supporting Information Table S10 for Tukey comparisons of residuals in different regions); explanatory power of different habitat heterogeneity models at the Eurasian scale (b, see Table S3 for Tukey comparisons of different habitat heterogeneity models); families included in this study and their position in a plant phylogeny (c); and adjusted R^2 of the heterogeneity–richness relationship (d). The top 10 residual-ranked polygons selected in at least four habitat heterogeneity models are plotted in (a). For the geographic regions, CHC represents China-South-Central, CHS represents China-Southeast, KAZ represents Kazakhstan,TUR represents Turkey, and MYA represents Myanmar. We evaluated the performance of a model considering a compound heterogeneity index; models considering growing degree day (*gdd*), bedrock type, and site water balance (*swb*) heterogeneity separately; and finally, a multivariate model considering *gdd*, *swb*, and bedrock type together (b). Bolded black horizontal lines within boxplots in (a) and (b) represent median values for residuals and adjusted R^2 respectively. The upper and lower boundaries of the box show the 25th and 75th percentile of data and the whiskers show the minimum (25th quantile minus 1.5 times interquartile range) and maximum (75th quantile add 1.5 times interquartile range) respectively. Finally, the dots above the whiskers represent outliers. In (c) and (d) panels, we defined species richness above the 95th percentile prediction interval for all habitat heterogeneity models for the Hengduan Mountains (HDM) in the heterogeneity–richness relationship as HDM-outlier families (highlighted in red). Nonoutlier families are defined as those with family-level species richness below the 95th percentile prediction interval (highlighted in blue and See Tables S5–S9 for model details).

when even larger sampling distances, with no overlap in the analysed windows, are used (Fig. S12). By contrast, the *gdd* and multivariate heterogeneity models, although having a high explanatory power at a small window size, reach the plateau at 45 km, with adjusted R^2 values of 28.7% ± 0.25% and 29.3% ± 0.26%, respectively. The bedrock and *swb* models have low explanatory power across scales. Spatially, the highest values of the compound heterogeneity index are found in the Three Rivers and Longmenshan regions (Figs 1bI–II, 4). By contrast, the highest values of individual heterogeneity of bedrock type are situated along the Mekong and Yangtze rivers (Fig. S16a). The individual *swb* index shows the highest values in the northern Three Rivers Region and a cold spot at the Yarlung Tsangpo River (Fig. S16b). By contrast, the individual *gdd* heterogeneity distribution has the highest values at the Yarlung Tsangpo River, followed by the western Three Rivers and Longmenshan regions

(Fig. S16c). When smaller window sizes are used, the spatial distribution of all habitat heterogeneity indices peaks along river valley bottoms (Fig. S17). By contrast, with increasing window size, areas of high heterogeneity values emerge from high relief and highly complex topography and bedrock variation, which is mainly concentrated in the Three Rivers Region (i.e. western HDM; Fig. 4aI,bI).

Spatial residual analyses

At the scale of Eurasia, five regions display systemically high residuals among the family-level heterogeneity–richness models. These regions are CHC (i.e. HDM), east China (CHS), Myanmar (MYA), Kazakhstan (KAZ), and Turkey (TUR; Fig. S18). These regions have significantly different residuals among the family-level heterogeneity–richness models, except the *gdd* model
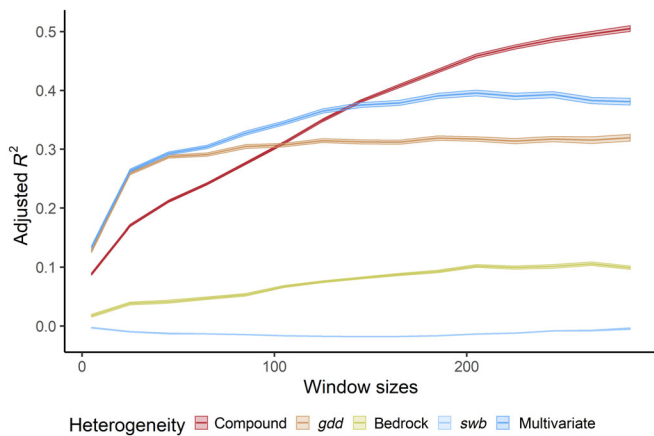
Fig. 3 Cross-scale determinants of Hengduan Mountains (HDM) outlier-family species richness are assessed using the adjusted R^2 value by habitat heterogeneity indices across different window sizes. The figure illustrates how the explained deviance of compound vs individual (growing degree day (*gdd*), bedrock type, site water balance (*swb*)) indices alone and all three individual indices (*gdd*, bedrock, and *swb*) in a multivariate model explain outlier species richness across different moving window sizes. The shaded areas represent the standard errors of the adjusted R^2 values across 200 resampling lattices.

(one-way ANOVA: compound $P = 0.00$, bedrock $P = 0.00$, *swb* $P = 0.00$, multivariate $P = 0.00$), and the CHC region has significant high residuals in these models (Fig. 2a; see Table S10 for the Tukey HSD comparisons). Similarly, the residuals of the heterogeneity–richness models show geographic coherence patterns in the HDM. In particular, the Three Rivers Region displays positive residuals across all habitat heterogeneity models (Figs 1bI, 4cI, S18). The Longmenshan region (Fig. 4cII) has zero to negative residuals in the compound heterogeneity model, whereas the western Longmenshan shows large positive residuals (Fig. 4cII). The southeastern HDM has positive residuals in the models including the compound (Fig. 4c), bedrock (Fig. S19b), or *gdd* (Fig. S19d) heterogeneity, and in the multivariate model (Fig. S19a).

Characterisation of outlier families

We identified outlier families in the CHC region by family-level species richness over the 95th prediction interval in the heterogeneity–richness models (Fig. 2c). The mean adjusted R^2

of outlier families does not differ significantly from nonoutlier families in any habitat heterogeneity model ($P > 0.05$) except the bedrock model (outlier 0.03 ± 0.04; nonoutlier 0.00 ± 0.03). The phylogenetic signal tests, with a Fritz and Purvis D-value of 0.79, indicate that the distribution of outlier families is not clustered, but rather falls between clumped allocation and random in the phylogeny. At the HDM scale, we show that the outlier families occur mainly at the middle elevation of 2466.7 m, but with a large elevational variation of 1111.7 m, while the nonoutlier families occur slightly lower than the outlier families (2136.7 m $\pm$ 1127.2 m). Families not identified as outliers of the HDM (falling within or below the blue shading in Fig. S20) may be outliers in different hotspots (e.g. Iran; Fig. S18) or may simply not be species rich in any of the regions.

Discussion

The distribution of species is determined by ecological preferences along multiple environmental dimensions (Grinnell, 1917; Hutchinson, 1957). Given that most species show specialised ecological niches (Ricklefs *et al.*, 2014), areas of high habitat heterogeneity generally support a greater species richness compared with more homogeneous habitats (MacArthur & MacArthur, 1961). We found that habitat heterogeneity can explain much of the variability in plant diversity among the WCVP polygons after accounting for the area effect. However, the species richness of 41 out of 97 seed plant families analysed in the HDM exceeds that predicted by the Eurasian heterogeneity–richness relationship. High-resolution mapping within the HDM shows that the multidimensional (compound) habitat heterogeneity index explains species richness variation, especially at a larger scale, pointing to the role of regional processes in creating and/or maintaining diversity. The residuals of the heterogeneity–richness relationship retain spatial coherence (Figs 4c, S19), however, suggesting that additional ecological or biogeographical drivers may help shape the exceptional species richness of the HDM.

Habitat heterogeneity generally determines species richness from the regional (Dufour *et al.*, 2006; Báldi, 2008) to the global scale (Udy *et al.*, 2021) and across multiple taxonomic groups (MacArthur & MacArthur, 1961; Johnson & Simberloff, 1974; Suissa *et al.*, 2021). Habitat heterogeneity increases available
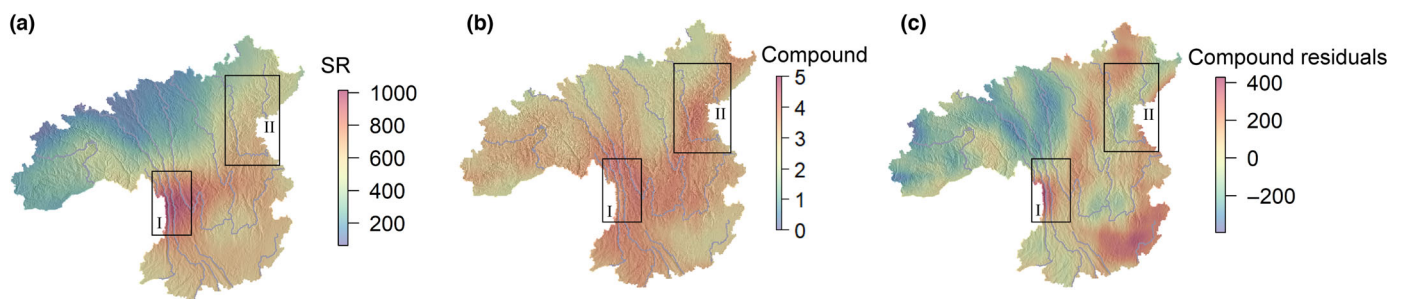

Fig. 4 Spatial pattern of species richness (SR), compound heterogeneity index, and residual richness of Hengduan Mountain (HDM)-outlier families at 205 km resolution. Species richness of outlier families (a); spatial distribution of Shannon heterogeneity for the compound habitat heterogeneity index (b); and residual richness of outlier families in the compound heterogeneity index model (c). Note that I represents the Three Rivers Region and II represents the Longmenshan region.

environmental niches in a region, allowing more species to diversify and coexist (Stein & Kreft, 2015). In mountain regions, different microhabitats occur within short geographical distances, providing numerous niches for species to colonise (Körner, 2004). Niche diversity combined with reduced dispersal and gene flow between ecologically similar habitats can drive the diversification of species in mountain regions (Pyron *et al.*, 2015). Polygons across Eurasia display a range of topographic complexities, with mountainous regions such as Anatolia and the HDM. The observation that these heterogeneous mountain ranges harbour higher species richness than expected is consistent with previous documentation of the unique diversity of these regions (Noroozi *et al.*, 2018). Our results regarding the HDM imply that there are distinct ecological or evolutionary processes active in this region that contribute to the high species richness. By contrast, other mountain ranges, such as the European Alps, exhibit relatively low residuals in the Eurasian scale analyses, indicating that habitat heterogeneity is sufficient to explain the species richness (Wohlgemuth, 1998; Gentili *et al.*, 2010; Tordoni *et al.*, 2020).

Habitat heterogeneity reflects the number of niches available to plant species (Hutchinson, 1957), which should be quantified considering multiple ecological dimensions. Previous ecological studies associating habitat heterogeneity with species richness focused on single dimensions of habitat heterogeneity, such as topography (Muellner-Riehl *et al.*, 2019; Li *et al.*, 2021), vegetation structure (Qian & Kissling, 2010), and climate heterogeneity (Durães & Loiselle, 2004), or they used separate multivariate analyses that captured different dimensions of habitat heterogeneity (Nichols *et al.*, 2008). Our study demonstrates that integrating climate conditions and bedrock type into a compound heterogeneity index can explain species richness in mountain areas better than a single predictor-based heterogeneity index. Bedrock type offers an important habitat heterogeneity axis by capturing different edaphic microhabitats. A wide range of geological substrates allows species to diversify under variable soil structural and chemical conditions (Hulshof & Spasojevic, 2020). While a Shannon heterogeneity index for each predictor represents the heterogeneity of these predictors individually, the species niche is determined by the intersection across all dimensions. Hence, our results are consistent with Hutchinson's niche concept, which defines a species' niche as an n-dimensional hypervolume along multiple axes representing the biological requirements of that species (Hutchinson, 1957). The variance of habitat heterogeneity is predicted best at larger spatial scales (i.e. 205 km, Fig. 3), where the compound habitat heterogeneity index provides a good representation of the regional geographic, geological, and climatic variability (Qi *et al.*, 1994; Cheng *et al.*, 2018), enabling better predictions of species richness.

The residual pattern from the heterogeneity–richness models can not only infer the hotspots that have large positive residuals but also identify the families that contribute to these hotspots. In the Eurasian scale habitat heterogeneity–richness models, outlier families (41/97) in the CHC are generally temperate families. These families show a prevalence of species with habitat tending towards higher elevations (Figs 1b, S21, S22). Our findings confirm that the HDM region is a temperate–climate hotspot (Ding

et al., 2020). Moreover, the phylogenetic signal shows that these outlier families are distributed with a pattern between clumped allocation and random in the phylogeny (Fig. 2c). Therefore, the diversification of these families may be partly associated with ecological radiations (Weigelt *et al.*, 2015; Whittaker *et al.*, 2017). For example, one distinct cluster in the phylogeny comprises the *Saxifragales* and *Ranunculales*, characterised as mid- to high-elevation herbaceous species, which originated *in situ* and diversified during the uplift of the HDM (Ebersbach *et al.*, 2017; Sun *et al.*, 2017). However, the generally random pattern of anomalous families in the phylogeny may indicate an extrinsic effect of landscape dynamics on their diversification over the last 15 Ma (Ding *et al.*, 2020).

The HDM is associated with distinct tectonic–geomorphic domains, consistent with the idea that tectonic–geomorphic processes create complex and fragmented habitats (Favre *et al.*, 2015). The spatial pattern of residuals in the heterogeneity–richness model shows a distinct region of high richness and residuals extending from the narrow neck of the Three Rivers Region (Figs 1b–I, 4c) across the mid-elevation domain of the HDM, suggesting that additional mechanisms may be important in explaining the unique diversity of this region. The Three Rivers Region near the largest diversity hotspot is unique on Earth for the proximity of large, parallel rivers with deeply incised valleys and surrounding high mountains, creating a variety of isolated habitats. The north–south orientation of the major river valleys provides multiple effective barriers and environmental gradients, inhibiting east–west species dispersion but enhancing north–south genetic exchange and providing dispersal corridors along the river valleys during periods of climate oscillation (Rana *et al.*, 2021). We expect that the tectonic activity leading to high rates of river incision and to major river reorganisation is also important for the geomorphic processes shaping the Three Rivers Region (Yang *et al.*, 2015) and may have altered habitat gradients and connectivity (Albert *et al.*, 2021). Moreover, the proximity of the HDM to other high mountains in Tibet, as well as the Himalayas, may have enhanced species colonisation (Ding *et al.*, 2020) according to island biogeography theory (MacArthur & Wilson, 1967).

Our study is limited by the accuracy of the species range maps at both the Eurasian and the HDM scales. WCVP distribution data are the most complete geographical dataset for seed plants currently available, but the relatively low spatial resolution at the province-to-country level limits our understanding of the finer-scale distribution of species. In addition, there are important domain differences for the analyses at different scales analyses. The CHC polygon in the WCVP definition includes the Yunnan, Sichuan, Chongqing, Guizhou, and Hubei provinces, while the local-scale analysis of the HDM includes the Yunnan, Sichuan, and Tibet provinces. The partial overlap of the defined regions reduces species-level taxonomic overlap at both scales (i.e. the CHC polygon has 16 705 species, the HDM have 12 356 species, and the overlap is 9162 species). Nevertheless, the species-level overlap is relatively high, that is taking account of 74.2% of species at the HDM polygon and 54.8% of species at the CHC polygon, which ensures a sufficient comparison in these scales. Moreover, at the HDM scale, inaccuracy in the original

county-level original distribution data may have led to overprediction of the SDM-based and downscaled species richness pattern and may have additionally lowered the explained deviance of habitat heterogeneity at a smaller scale extent. Although our maps have been designed to increase possible detection probability, some regions – such as the Yarlung Tsangpo River in the western HDM (i.e. Tibet) – may still suffer from an underestimation of species richness due to the undersampling in this ecologically distinct region (Fig. S15). Nevertheless, our downscaling approach to mapping all plant species in the HDM produces a smoother pattern than downscaling solely based on elevation (Li *et al.*, 2021) and therefore offers an improved database for ecological and evolutionary analyses of the plant biodiversity in the HDM.

The integration of the plant distribution data at the Eurasian scale with high-resolution mapping at the HDM scale reveals an important role of habitat heterogeneity in explaining species richness. However, the large residuals in the HDM region point to its anomalous richness and thus other mechanisms of diversification. The complex topography and tectonic settings of the HDM have produced a highly perturbed, transient landscape with an enhanced opportunity for habitat creation and fragmentation, which would impact richness in ways not captured by habitat heterogeneity. However, the transient components are difficult to quantify at the multiple scales needed to establish a linkage to biodiversity. For instance, the erosion rate has been proposed to be an important process in generating topographical relief and is associated with biodiversity at the global scale (Antonelli *et al.*, 2018). Erosion may also shape the dissected landscape and biodiversity of the HDM (Ding *et al.*, 2020), but quantification of erosion across the region is still needed (Yang *et al.*, 2016). To better understand the geological processes driving biodiversity patterns in the HDM, we recognise a need for additional collaborations between geologists, biologists, and climatologists. Such collaborations could help us to develop biologically meaningful characterisations of landscape processes, such as uplift, erosion histories, and landscape/climate stability, to enhance our ability to explain HDM biodiversity at smaller spatial scales.

Acknowledgements

Competing interests

None declared.

Author contributions

YC, LP, NEZ and SDW had the original idea for the study. YC, LP and NEZ planned the research. KG provided the bedrock map. AL and ZW provided biodiversity data at the HDM scale. XS provided statistical suggestions about modelling at different scales. YC led the analyses, with support from LP, NEZ and CA. YC led the writing, together with NEZ and LP, with contributions from all co-authors. NEZ and LP contributed equally to this work.

ORCID

Camille Albouy https://orcid.org/0000-0003-1629-2389
Yaquan Chang https://orcid.org/0000-0003-2753-5665
Katrina Gelwick https://orcid.org/0000-0001-8266-7412
Ao Luo https://orcid.org/0000-0003-1270-6353
Loïc Pellissier https://orcid.org/0000-0002-2289-8259
Xinwei Shen https://orcid.org/0000-0002-7801-4803
Zhiheng Wang https://orcid.org/0000-0003-0808-7780
Sean D. Willett https://orcid.org/0000-0002-8408-0567
Niklaus E. Zimmermann https://orcid.org/0000-0003-3099-9604

Data availability

Full access to the World Checklist of Vascular Plants data was granted in November 2022. The heterogeneity and plant distribution map of the Hengduan Mountains scale can be obtained from doi: 10.16904/envidat.424, code for related models is available at https://github.com/YaquanChang/heterogeneity-model.git.

References

Albert JS, Bernt MJ, Fronk AH, Fontenelle JP, Kuznar SL, Lovejoy NR. 2021. Late Neogene megariver captures and the Great Amazonian Biotic Interchange. *Global and Planetary Change* 205: 103554.

Allouche O, Tsoar A, Kadmon R. 2006. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *The Journal of Applied Ecology* 43: 1223–1232.

Antonelli A, Kissling WD, Flantua SGA, Bermúdez MA, Mulch A, Muellner-Riehl AN, Kreft H, Linder HP, Badgley C, Fjeldså J *et al.* 2018. Geological and climatic influences on mountain biodiversity. *Nature Geoscience* 11: 718–725.

Arrhenius O. 1921. Species and area. *Journal of Ecology* 9: 95–99.

Báldi A. 2008. Habitat heterogeneity overrides the species–area relationship. *Journal of Biogeography* 35: 675–681.

Beck HE, Zimmermann NE, McVicar TR, Vergopolan N, Berg A, Wood EF. 2018. Present and future Köppen-Geiger climate classification maps at 1-km resolution. *Scientific Data* 5: 180214.

Brummitt RK, Pando F, Hollis S, Brummitt NA. 2001. World geographical scheme for recording plant distributions. *International working group on taxonomic databases for plant science (TDWG)*: 951–952.

Brun P, Thuiller W, Chauvier Y, Pellissier L, Wüest RO, Wang Z, Zimmermann NE. 2020. Model complexity affects species distribution projections under climate change. *Journal of Biogeography* 47: 130–142.

Brun P, Zimmermann NE, Hari C, Pellissier L, Karger DN. 2022. Global climate-related predictors at kilometre resolution for the past and future. *Earth System Science Data Discussions* 14: 5573–5603.

Chauvier Y, Thuiller W, Brun P, Lavergne S, Descombes P, Karger DN, Renaud J, Zimmermann NE. 2021. Influence of climate, soil, and land cover

on plant species distribution in the European Alps. *Ecological Monographs* 91: e01433.

Cheng Z, Weng C, Guo J, Dai L, Zhou Z. 2018. Vegetation responses to late Quaternary climate change in a biodiversity hotspot, the Three Parallel Rivers region in southwestern China. *Palaeogeography, Palaeoclimatology, Palaeoecology* 491: 10–20.

Ding W-N, Ree RH, Spicer RA, Xing Y-W. 2020. Ancient orogenic and monsoon-driven assembly of the world's richest temperate alpine flora. *Science* 369: 578–581.

Dufour A, Gadallah F, Wagner HH, Guisan A, Buttler A. 2006. Plant species richness and environmental heterogeneity in a mountain landscape: effects of variability and spatial configuration. *Ecography* 29: 573–584.

Durães R, Loiselle BA. 2004. Inter-scale relationship between species richness and environmental heterogeneity: a study case with antbirds in the Brazilian Atlantic Forest. *Ornitologia Neotropical* 15: 127–135.

Ebersbach J, Muellner-Riehl AN, Michalak I, Tkach N, Hoffmann MH, Röser M, Sun H, Favre A. 2017. In and out of the Qinghai–Tibet Plateau: divergence time estimation and historical biogeography of the large arctic-alpine genus *Saxifraga* L. *Journal of Biogeography* 44: 900–910.

Fasiolo M, Wood SN, Zaffran M, Nedellec R, Goude Y. 2020. QGAM: Bayesian non-parametric quantile regression modelling in R. *arXiv*: 2007.03303.

Fasiolo M, Wood SN, Zaffran M, Nedellec R, Goude Y. 2021. Fast calibrated additive quantile regression. *Journal of the American Statistical Association* 116: 1402–1412.

Favre A, Päckert M, Pauls SU, Jähnig SC, Uhl D, Michalak I, Muellner-Riehl AN. 2015. The role of the uplift of the Qinghai–Tibetan Plateau for the evolution of Tibetan biotas. *Biological Reviews of the Cambridge Philosophical Society* 90: 236–253.

Freckleton RP. 2002. On the misuse of residuals in ecology: regression of residuals vs. multiple regression. *Journal of Animal Ecology* 71: 542–545.

Friedman JH. 2001. Greedy function approximation: a gradient boosting machine. *Annals of Statistics* 29: 1189–1232.

Fritz SA, Purvis A. 2010. Selectivity in mammalian extinction risk and threat types: a new measure of phylogenetic signal strength in binary traits. *Conservation Biology* 24: 1042–1051.

Gentili R, Armiraglio S, Rossi G, Sgorbati S, Baroni C. 2010. Floristic patterns, ecological gradients and biodiversity in the composite channels (Central Alps, Italy). *Flora – Morphology, Distribution, Functional Ecology of Plants* 205: 388–398.

Gourbet L, Yang R, Fellin MG, Paquette J-L, Willett SD, Gong J, Maden C. 2020. Evolution of the Yangtze River network, southeastern Tibet: insights from thermochronology and sedimentology. *Lithosphere* 12: 3–18.

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

Grinnell J. 1917. Field tests of theories concerning distributional control. *The American Naturalist* 51: 115–128.

Guisan A, Thuiller W, Zimmermann NE. 2017. Data acquisition, sampling design, and spatial scales. In: *Habitat suitability and distribution models: with applications in R*. Cambridge, UK: Cambridge University Press, 59–150.

Guisan A, Zimmermann NE. 2000. Predictive habitat distribution models in ecology. *Ecological Modelling* 135: 147–186.

Hartmann J, Moosdorf N. 2012. The new global lithological map database GLiM: A representation of rock properties at the Earth surface. *Geochemistry, Geophysics, Geosystems* 13: 12.

Hastie T, Tibshirani R. 1986. Generalized additive models. *Statistical Models in S* 1: 249–307.

He P, Jiang L, Li F. 2022. Evaluation of parametric and non-parametric stem taper modeling approaches: a case study for Betula platyphylla in Northeast China. *Forest Ecology and Management* 525: 120535.

Hijmans RJ, van Etten J. 2016. RASTER: geographic data analysis and modeling. R package v.2: 8.

Hortal J, Bello F, Diniz-Filho JAF. 2015. Seven shortfalls that beset large-scale knowledge of biodiversity. *Annual Review of Ecology, Evolution, and Systematics* 46: 523–549.

Hulshof CM, Spasojevic MJ. 2020. The edaphic control of plant diversity. *Global Ecology and Biogeography* 29: 1634–1650.

Hutchinson GE. 1957. Concluding remarks. population studies: animal ecology and demography. *Cold Spring Harbor Symposia on Quantitative Biology* 22: 415–427.

Jiménez-Alfaro B, Abdulhak S, Attorre F, Bergamini A, Carranza ML, Chiarucci A, Ćušterevska R, Dullinger S, Gavilán RG, Giusso del Galdo G *et al.* 2021. Post-glacial determinants of regional species pools in alpine grasslands. *Global Ecology and Biogeography* 30: 1101–1115.

Johnson MP, Simberloff DS. 1974. Environmental determinants of Island species numbers in the British Isles. *Journal of Biogeography* 1: 149–154.

Karger DN, Conrad O, Böhner J, Kawohl T, Kreft H, Soria-Auza RW, Zimmermann NE, Linder HP, Kessler M. 2017. Climatologies at high resolution for the earth's land surface areas. *Scientific Data* 4: 170122.

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

Körner C. 2000. Why are there global gradients in species richness? Mountains might hold the answer. *Trends in Ecology & Evolution* 15: 513–514.

Körner C. 2004. Mountain biodiversity, its causes and function. *Ambio* Spec No 13: 11–17.

Li Q, Sun H, Boufford DE, Bartholomew B, Fritsch PW, Chen J, Deng T, Ree RH. 2021. Grade of membership models reveal geographical and environmental correlates of floristic structure in a temperate biodiversity hotspot. *New Phytologist* 232: 1424–1435.

Lyu L, Leugger F, Hagen O, Fopp F, Boschman LM, Strijk JS, Albouy C, Karger DN, Brun P, Wang Z *et al.* 2022. An integrated high-resolution mapping shows congruent biodiversity patterns of Fagales and Pinales. *New Phytologist* 235: 759–772.

Lyu T, Wang Y, Luo A, Li Y, Peng S, Cai H, Zeng H, Wang Z. 2021. Effects of climate, plant height, and evolutionary age on geographical patterns of fruit type. *Frontiers in Plant Science* 12: 604272.

MacArthur RH, MacArthur JW. 1961. On bird species diversity. *Ecology* 42: 594–598.

MacArthur RH, Wilson EO. 1967. *The theory of Island biogeography.* Princeton, NJ, USA: Princeton University Press.

Merow C, Smith MJ, Edwards TC Jr, Guisan A, McMahon SM, Normand S, Thuiller W, Wüest RO, Zimmermann NE, Elith J. 2014. What do we gain from simplicity versus complexity in species distribution models? *Ecography* 37: 1267–1281.

Muellner-Riehl AN, Schnitzler J, Kissling WD, Mosbrugger V, Rijsdijk KF, Seijmonsbergen AC, Versteegh H, Favre A. 2019. Origins of global mountain plant biodiversity: testing the 'mountain-geobiodiversity hypothesis'. *Journal of Biogeography* 46: 2826–2838.

Mulch A, Chamberlain CP. 2006. The rise and growth of Tibet. *Nature* 439: 670–671.

Nelder JA, Wedderburn RWM. 1972. Generalized Linear Models. *Journal of the Royal Statistical Society: Series A* 135: 370.

Nichols WF, Killingbeck KT, August PV. 2008. The influence of geomorphological heterogeneity on biodiversity II. A landscape perspective. *Conservation Biology* 12: 371–379.

Noroozi J, Talebi A, Doostmohammadi M, Rumpf SB, Linder HP, Schneeweiss GM. 2018. Hotspots within a global biodiversity hotspot - areas of endemism are associated with high mountain ranges. *Scientific Reports* 8: 10345.

Nunes CA, Berenguer E, França F, Ferreira J, Lees AC, Louzada J, Sayer EJ, Solar R, Smith CC, Aragão LEOC *et al.* 2022. Linking land-use and land-cover transitions to their ecological impact in the Amazon. *Proceedings of the National Academy of Sciences, USA* 119: e2202310119.

Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'hara RB, Simpson GL, Solymos P, Stevens MHH, Wagner H *et al.* 2013. *Package "VEGAN".* Community ecology package. 2: 1–295.

Poggio L, de Sousa LM, Batjes NH, Heuvelink GBM, Kempen B, Ribeiro E, Rossiter D. 2021. SOILGRIDS 2.0: producing soil information for the globe with quantified spatial uncertainty. *The Soil* 7: 217–240.

Pyron RA, Costa GC, Patten MA, Burbrink FT. 2015. Phylogenetic niche conservatism and the evolutionary basis of ecological speciation. *Biological Reviews of the Cambridge Philosophical Society* 90: 1248–1262.

Qi Z, Dejin Z, Dasheng Z, Zhongxiang H, Song H, Xiuqin J, Jinquan D. 1994. Ophiolites of the Hengduan Mountains, China: characteristics and tectonic settings. *Journal of Southeast Asian Earth Sciences* 9: 335–344.

Qian H, Kissling WD. 2010. Spatial scale and cross-taxon congruence of terrestrial vertebrate and vascular plant species richness in China. *Ecology* 91: 1172–1183.

R Core Team. 2022. *R: a language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing. [WWW document] URL https://www.R-project.org/ [accessed 22 April 2022].

Rahbek C, Borregaard MK, Antonelli A, Colwell RK, Holt BG, Nogues-Bravo D, Rasmussen CMØ, Richardson K, Rosing MT, Whittaker RJ *et al.* 2019a. Building mountain biodiversity: geological and evolutionary processes. *Science* 365: 1114–1119.

Rahbek C, Borregaard MK, Colwell RK, Dalsgaard B, Holt BG, Morueta-Holme N, Nogues-Bravo D, Whittaker RJ, Fjeldså J. 2019b. Humboldt's enigma: what causes global patterns of mountain biodiversity? *Science* 365: 1108–1113.

Rana SK, Luo D, Rana HK, O'Neill AR, Sun H. 2021. Geoclimatic factors influence the population genetic connectivity of *Incarvillea arguta* (Bignoniaceae) in the Himalaya–Hengduan Mountains biodiversity hotspot. *Journal of Systematics and Evolution* 59: 151–168.

Ricklefs RE, Relyea R, Richter C. 2014. *Ecology: the economy of nature.* New York, NY, USA: WH Freeman.

Ricklefs RE. 2010. Evolutionary diversification, coevolution between populations and their antagonists, and the filling of niche space. *Proceedings of the National Academy of Sciences, USA* 107: 1265–1272.

Shannon CE. 2001. A mathematical theory of communication. *ACM SIGMOBILE Mobile Computing and Communications Review* 5: 3–55.

Silvertown J. 2004. Plant coexistence and the niche. *Trends in Ecology & Evolution* 19: 605–611.

Smith SA, Brown JW. 2018. Constructing a broadly inclusive seed plant phylogeny. *American Journal of Botany* 105: 302–314.

Spehn EM, Körner C. 2005. A global assessment of mountain biodiversity and its function. In: Huber UM, Bugmann HKM, Reasoner MA, eds. *Global change and mountain regions: an overview of current knowledge.* Dordrecht, the Netherlands: Springer Netherlands, 393–400.

Stein A, Gerstner K, Kreft H. 2014. Environmental heterogeneity as a universal driver of species richness across taxa, biomes and spatial scales. *Ecology Letters* 17: 866–880.

Stein A, Kreft H. 2015. Terminology and quantification of environmental heterogeneity in species-richness research. *Biological Reviews of the Cambridge Philosophical Society* 90: 815–836.

Suissa JS, Sundue MA, Testo WL. 2021. Mountains, climate and niche heterogeneity explain global patterns of fern diversity. *Journal of Biogeography* 48: 1296–1308.

Sun H, Zhang J, Deng T, Boufford DE. 2017. Origins and evolution of plant diversity in the Hengduan Mountains, China. *Plant Diversity* 39: 161–166.

Tietje M, Antonelli A, Baker WJ, Govaerts R, Smith SA, Eiserhardt WL. 2022. Global variation in diversification rate and species richness are unlinked in plants. *Proceedings of the National Academy of Sciences, USA* 119: e2120662119.

Tordoni E, Casolo V, Bacaro G, Martini F, Rossi A, Boscutti F. 2020. Climate and landscape heterogeneity drive spatial pattern of endemic plant diversity within local hotspots in South-Eastern Alps. *Perspectives in Plant Ecology, Evolution and Systematics* 43: 125512.

Udy K, Fritsch M, Meyer KM, Grass I, Hanß S, Hartig F, Kneib T, Kreft H, Kukunda CB, Pe'er G *et al.* 2021. Environmental heterogeneity predicts global species richness patterns better than area. *Global Ecology and Biogeography* 30: 842–851.

Wang W. 1994. *Vascular plants of the Hengduan Mountains. The series of the scientific expedition to Hengduan Mountains.* Qinghai-Xizang Plateau directed by Institute of Botany.

Weigelt P, Kissling WD, Kisel Y, Fritz SA, Karger DN, Kessler M, Lehtonen S, Svenning J-C, Kreft H. 2015. Global patterns and drivers of phylogenetic structure in Island floras. *Scientific Reports* 5: 12213.

Whittaker RJ, Fernández-Palacios JM, Matthews TJ, Borregaard MK, Triantis KA. 2017. Island biogeography: taking the long view of nature's laboratories. *Science* 357: eaam8326.

Wohlgemuth T. 1998. Modelling floristic species richness on a regional scale: a case study in Switzerland. *Biodiversity and Conservation* 7: 159–177.

Wood S, Wood MS. 2015. *Package 'MGCV'.* R package v.1: 729.

Woodward FI, Williams BG. 1987. Climate and plant distribution at global and local scales. *Vegetatio* 69: 189–197.

Wu ZY. 1986. *Flora Yunnanica.* Beijing, China: Science Press.

Wu ZY. 1987. *Flora xizangica.* Beijing, China: Science Press.

Wu ZY. 1988. Hengduan mountain flora and her significance. *The Journal of Japanese Botany* 63: 297–311.

Xing Y, Ree RH. 2017. Uplift-driven diversification in the Hengduan Mountains, a temperate biodiversity hotspot. *Proceedings of the National Academy of Sciences, USA* 114: E3444–E3451.

Yang J, El-Kassaby YA, Guan W. 2020. The effect of slope aspect on vegetation attributes in a mountainous dry valley, Southwest China. *Scientific Reports* 10: 16465.

Yang R, Fellin MG, Herman F, Willett SD, Wang W, Maden C. 2016. Spatial and temporal pattern of erosion in the Three Rivers Region, southeastern Tibet. *Earth and Planetary Science Letters* 433: 10–20.

Yang R, Willett SD, Goren L. 2015. *In situ* low-relief landscape formation as a result of river network disruption. *Nature* 520: 526–529.

Zhou BK. 1994. *Flora Sichuanica.* Chengdu, China: Sichuan Publishing House of Science and Technology.

Zimmermann NE, Kienast F. 1999. Predictive mapping of alpine grasslands in Switzerland: species versus community approach. *Journal of Vegetation Science* 10: 469–482.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information section at the end of the article.

Fig. S1 Overlap between the China-South-Central polygon in the World Checklist of Vascular Plants database and the HDM polygon in the local-scale map.

Fig. S2 Example of species distribution model construction.

Fig. S3 Decision tree of the sampling strategies used in the species distribution modelling.

Fig. S4 Pearson correlations between environmental predictors.

Fig. S5 Spatial distribution of different predictors in the sdm and the soil predictors changes along the elevation bands.

Fig. S6 Boxplots for the true skill statistic in outlier and nonoutlier families.

Fig. S7 Species richness map based on the species distribution model.

Fig. S8 Distribution of the species richness of families in the Hengduan Mountains that were not selected in this study.

Fig. S9 Flowchart used to generate the Shannon habitat heterogeneity map.

Fig. S10 Relationship between habitat heterogeneity and area at the Eurasian scale.

Fig. S11 Flowchart of the model processes applied at the Hengduan Mountains scale.

Fig. S12 Cross-scale determinants of outlier-family richness at Hengduan Mountain scale.

Fig. S13 Cross-scale determinants of nonoutlier-family richness at Hengduan Mountain scale.

Fig. S14 Cross-scale determinants of elevation downscaled outlier-family species richness at Hengduan Mountain scale.

Fig. S15 Species accumulation curves for the species distribution model.

Fig. S16 Spatial distribution of Shannon heterogeneity of bedrock types.

Fig. S17 Congruence analysis between the species richness of outlier families and habitat heterogeneity using a 25 km focal window size.

Fig. S18 Spatial distribution of the top 10 residual-ranked polygons selected in at least four heterogeneity models.

Fig. S19 Residual species richness of Hengduan Mountains outlier families based on different heterogeneity models at 205 km resolution.

Fig. S20 Examples of outlier and nonoutlier families.

Fig. S21 Elevation density plot for the species richness of outlier and nonoutlier families and for the families not included in this study.

Fig. S22 Elevation distribution in the Hengduan Mountains.

Notes S1 Preselection of the species distribution model predictors.

Notes S2 Species distribution model downscaling pipeline.

Notes S3 Completeness of species distribution model.

Table S1 Full family list with mapping summaries at the Eurasian and Hengduan Mountains scales.

Table S2 Summary of the number of presence and pseudo-absence pixels.

Table S3 Probabilities for multiple pairwise Tukey comparisons of adjusted R^2 values in the different heterogeneity models shown in Fig. 2(b).

Table S4 Model performance summary for the species richness models considering the compound habitat heterogeneity index.

Table S5 Model performance summary for the species richness models considering the growing degree days heterogeneity index.

Table S6 Model performance summary for the species richness models considering the bedrock type heterogeneity index.

Table S7 Model performance summary for the species richness models considering the site water balance heterogeneity index.

Table S8 Model performance summary for the multivariate species richness model.

Table S9 Performance in the generalised additive model of different heterogeneity predictors at the Hengduan Mountains scale.

Table S10 Probabilities for multiple pairwise Tukey comparisons of the residuals in the different regions.

Video S1 Compound heterogeneity pattern from 5 to 285 km window sizes at the Hengduan Mountains scale.

Please note: Wiley is not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.