Science Advances
AAAS

Supplementary Materials for

Occurrence-based diversity estimation reveals macroecological and conservation knowledge gaps for global woody plants

Buntarou Kusumoto *et al.*

Corresponding author: Buntarou Kusumoto, kusumoto.buntarou@gmail.com

Sci. Adv. **9**, eadh9719 (2023)
DOI: 10.1126/sciadv.adh9719

The PDF file includes:

Supplementary Text
Figs. S1 to S27
Tables S1, S4, and S5
Legends for tables S2 and S3
References

Other Supplementary Material for this manuscript includes the following:

Tables S2 and S3

Supplementary Text

1. Sample coverage levels of standardisation of species diversity

To fairly compare diversity values among locations, we need to consider the spatial inequality of sample completeness (*6*). In the case of the global diversity of woody plants, the observed number of species showed strong correlation with the number of species incidences (fig. S1), indicating that the spatial distribution of sample completeness contaminated the observed species diversity patterns. Because the relationships between observed species diversity and number of species incidences were often unsaturated in our dataset, we could not use the asymptotic diversity value to 'fairly' compare the diversities among grid cells (*15*). Instead, we applied a non-asymptotic approach whereby species diversity is standardised using rarefaction (interpolation) or extrapolation based on sample completeness (*14*).

Sample completeness can be evaluated using sample coverage, i.e. the proportion of the number of individuals (or frequency of incidences) detected in the focal assemblage. The sample coverage is more accurate than the conventional richness ratio (i.e. observed species richness/Chao-2 estimator), which is positively biased when sample size is inadequate (*14*). As a guideline, Chao and colleagues recommended using the minimum sample coverage value of doubled reference sample size for reliable extrapolation (*15*). However, in macro-scale studies, the magnitude of difference in true diversity could be huge (10^1 –10^4), and species-poor sites (e.g. temperate regions) tend to be explored better (higher sample coverage) than species-rich sites (e.g. tropical regions). Indeed, in our data set, sample coverage showed a large spatial variation, especially at the finest spatial resolution (fig. S2). In such a case, if we applied a too small sample coverage, we would miss diversity gradients in regions showing less species richness. To deal with this, we relaxed the restriction to extrapolation by using percentiles of sample coverage values of doubled reference sample size as the level of standardization (*17*). When we used the p-th percentile as the level of standardization, extrapolation to more than double its reference sample size will be applied to p% of the grid cells, while rarefaction or extrapolation to less than double its reference sample size will be applied to (100 - p) % of the grid cells. There is no prescribed percentile value. Therefore, we set several percentiles (1st, 5th, 10th, 20th, 30th, 40th and 50th percentiles; see table S4 for the specific values of sample coverage at each percentile) and assessed the impact of arbitrary choice of the level of standardization on descriptions, interpretations, and modelling of geographical diversity patterns. We also included the asymptotic diversity (i.e., assuming ∞ sampling effort or sample coverage = 1) for comparison.

Impacts of sample coverage-based standardisation on geographical diversity patterns

Standardisation based on sample coverage influenced the shape of latitudinal diversity gradients (LatDGs) differently for the three longitudinal zones: Americas, Africa-Europe, and Asia-Oceania (fig S3). The shape of the LatDG was relatively robust to the level of sample coverage-based standardisation in the Americas zone (fig. S3). A relatively minor difference was observed in the steepness of the LatDG slope in the transition from tropical to temperate regions; the observed species richness tended to underestimate the slope owing to the undervalue in the tropical regions. In the Africa-Europe zone, although the LatDG was unclear for the observed species richness owing to large intra-regional variation in Africa and super-high representation in Europe (fig. S3), sample coverage-based standardisation clarified consistently high species richness within the tropical zone and a decreasing trend from the equator to the northern pole. In the Asia-Oceania zone, the observed LatDG peaked in the Southern Hemisphere (around the Tropic of Capricorn) and decreased towards the northern pole (fig. S3); the sample coverage-based standardisation toned down the southern peak, and depicted alternative peak around the northern middle latitudes which represents the transition from tropical to temperate regions. Although the level of standardization affected the absolute values of diversity (i.e., the vertical position of LatDGs), they did not have a large influence on the shape of LatDGs.

Longitudinal diversity gradients (LonDGs) were also influenced by sample coverage-based standardisation, although they were less sensitive than LatDGs (fig. S4). In the north extratropics, the standardised species richness became too low (almost zero) at the smallest level of standardization (1st percentile of sample coverage), obscuring the LonDG and generating a dip in the eastern margin (>100 °). In the tropics and the south extratropic, the standardised species richness was highest in the western parts (South America and/or South Africa) and decreased towards the east, whereas the observed species richness was over-represented in the eastern regions (fig. S4).

Impacts of sample coverage-based standardisation on environmental driver analysis

We analysed relationships between the diversities and environmental variables using three regression models: ordinary least squares (OLS), generalised additive model (GAM), and random forest (RF). We used log-scaled species richness as the response variable, and 11 environmental variables as the explanatory variables associated with energy, water, their seasonality, topography, habitat heterogeneity, and historical climatic stability: mean annual temperature (Bio1), temperature seasonality (Bio4), annual precipitation (Bio12), precipitation seasonality (Bio15), actual evapotranspiration (AET), potential evapotranspiration (PET), aridity index (AI), average elevation (Elv), standard deviation of elevation (Elv.sd), and differences in temperature (D_{temp}) and precipitation (D_{prec}) between the Last Glacial Maximum and the present day.

The influence of sample coverage-based standardisation depends on the environmental variables and the modelling frameworks. In the OLS model, the linear relationship with the log-scaled species richness (residuals after controlling for the effect of the other explanatory factors) was maintained well for Bio 12 and AET (fig. S5), whereas the reverse relationship was observed for Bio 4, Bio 15, PET, AI and D_{temp}. The effect size was magnified in the standardised species richness for Bio1 and Elv, but weakened for Elv.sd and D_{prec}.

In the GAM model, the response curves (with the effect of the other explanatory factors fixed as their respective means) to individual explanatory factors differed between the standardised and observed species richness, but were qualitatively similar among the sample coverage levels (fig. S6). The curve for AET was the most robust to presence/absence of sample coverage-based standardisation; a consistent positive monotonic linear relationship was observed. Meanwhile, sample coverage-based standardisation was most influential on the relationships with Bio1, PET and Elv.sd. The observed (also asymptotic) species richness was high at extremely low Bio 1 and high PET (>2500 mm). This suggests that the observed diversity suffered from a multicollinearity associated with the biased geographical representation (e.g., higher latitudes). Such unrealistic patterns were not observed for the sample coverage-based standardised species richness.

In the RF model, the partial dependency curves were generally stable to presence/absence of sample coverage-based standardisation and its levels (fig. S7). An exception was Bio 1 for which the partial dependency tended to be higher between 5 and 15 °C for the observed (and asymptotic) species richness, compared with that for the standardised diversities (fig. S7). The species diversities were higher in regions with higher Bio 1, Bio 12, AET, AI, Elv and Elv.sd, lower Bio 4, Bio-15, D_{temp} and D_{prec}, and intermediate PET.

We evaluated the potential impacts of sample coverage-based standardisation on relative importance of the explanatory variables using the coefficients of partial determination (r^2) and the mean squared error in out-of-bag data, for the OLS and RF models, respectively. The most important variable was consistently AET, regardless of sample coverage-based standardisation and the modelling approach (figs. S8 and S9). In OLS, the model with the observed species richness tended to overvalue the relative importance of Bio 15, but undervalued Bio 1. In RF, the overall importance ranking was relatively stable against sample coverage-based standardisation, whereas the rank of D_{temp} (historical temperature change) changed depending on the level of standardization: the importance increased with reduction in the level, suggesting that D_{temp} would have better explanatory power when ignoring the diversity variation in species-poor regions, and may act as a dichotomous factor at the global scale (i.e., historically stable vs unstable sites).

The RF model showed the highest explanatory and prediction performance among the three modelling frameworks (figs. S10 and S11). In all approaches, the explanatory power (R^2) was higher and the predictive error (root mean squared error based on a 10-fold cross-validation test) was smaller for the sample coverage-based standardised diversities than for the observed diversity. The explanatory and predictive performance was generally comparable among the level of standardisation, but slightly better in the intermediate levels (20th–40th percentiles).

Finally, we projected the regression models on the geographical space to visually check how the influence of sample coverage-based standardisation transmitted to spatial predictions. The spatial predictions based on the observed and asymptotic species richness were highly influenced by the geographical pattern of sample coverage (figs. 1 and S12). The prediction based on low level standardisation (1st and 5th percentiles) failed to capture the diversity gradients in species-poor, higher-latitudinal regions. The spatial congruence among the predicted values derived from the three modelling frameworks exhibited a slight improvement when using standardized species richness for the spatial projection, compared with using observed species richness (fig. S12). For the predictions based on the intermediate levels of sample coverage-based standardisations, the OLS model reflected a large-scale diversity trend from species-rich lower latitudes to species-poor higher latitudes, but obscured intraregional diversity variation. The nonlinear frameworks (GAM and RF) were successful in visualizing local diversity peaks (South America, central Africa, and south China) as well as the large-scale diversity trends from lower to higher latitudes (fig. S12).

2. Impacts of spatial resolution on species diversity analyses

The sample coverage is dependent on spatial resolution (grid-cell size): the coarser the spatial resolution, the higher the sample coverage on average, and the smaller its variance (fig. S2). To test potential impacts of spatial resolution on the species richness estimations and biogeographical patterns, we repeated the above-mentioned suite of analyses at coarser spatial resolutions (~200 km × 200 km, 400 km × 400 km, and 800 km × 800 km).

Even at the coarser spatial resolutions (≥ 200 km × 200 km), we observed LatDGs (fig. S13). Species richness was highest in the tropics (the Americas and the Africa-Europe zones) or the tropics–extratropics boundary (the Asia-Oceania zone), and decreased toward higher latitudes. The LonDGs were relatively sensitive to the spatial resolution, especially for the north extratropics (fig. S14). At the finest resolution (100 km × 100 km), species richness was higher in the eastern region (East Asia) than in the western regions (Europe and North America), whereas at the coarser resolutions, a bimodal diversity pattern with comparable peaks in the eastern North America and East Asia were observed. In the tropics, at any spatial resolution, the species richness was highest in South America. In the southern extratropics, the species richness was highest between 45°W and 45°E (eastern South America and the tip of South Africa) at any spatial resolution. In LatDGs and LonDGs, the disparity between the observed (and asymptotic) and the sample coverage-based standardised species richness was greatest at the finest resolution (100 km × 100 km) but was reduced at the coarser resolutions.

Overall, the spatial resolution showed only marginal effects on the correlative relationships between species richness and the environmental variables (figs. S15–S17), except for the environmental variables whose range was truncated at the coarser resolutions (Bio 12, Bio 15, AI, Elv.sd, and D_{prec}). In OLS models, the effect size and estimation error of regression coefficients tended to be larger at coarser resolutions in general (fig. S15). In GAM models, the response curves tended to deviate among spatial resolutions, especially at the edges of variable ranges (fig. S16). The shape of partial dependency plots in the RF models was relatively stable to the difference in the spatial resolution (fig. S17).

AET was consistently the most important explanatory variable regardless of the spatial resolution, whereas the ranking of the other variables were depending on the spatial resolution (figs. S8 and S9). The relative importance ranking (except for AET) was more sensitive to the spatial resolution in OLS (fig. S8) than in RF models. In the RF models, the relative importance ranking, particularly for the four most important variables (AET, PET, Bio 1 and Bio 4) was stable (fig. S9). Interestingly, at coarse resolutions (≥400 km × 400 km), the relative importance of historical temperature change (D_{temp}) was higher in OLS and RF models (figs. S8 and S9).

In general, the explanatory power (R^2) was slightly better, but the prediction error was greater, at coarser resolutions (figs. S10 and S11). The RF models consistently showed the best explanatory and predictive performances across all spatial resolutions (figs. S10 and S11). Exceptionally, the GAM models showed the highest R^2 at the 800 km × 800 km, which is likely to be an artefact of overfitting owing to a small sample size (*80*).

3. Comparison of geographical patterns between diversities at different orders

To test the potential influence of species incidence frequencies on biogeographical patterns, we checked the behaviour of species diversities at different orders. In a Hill number-based approach, species diversity (*D*) is represented by the following equation (*14*):

when $q \neq 1$, ${}^{q}D = \left(\sum_{i=1}^{S} p_i^{q}\right)^{1/(1-q)}$;

when $q = 1$, ${}^{1}D = exp\left(-\sum_{i=1}^{S} p_i \log p_i\right)$,

where *S* is the total number of species in an assemblage, and p_i is the relative abundance of the *i*-th species. The parameter *q* controls the weighting for relative abundance: the larger *q* is, the larger the weight for abundant (dominant) species (*15*). At $q = 0$, all species are treated equally, then ${}^{0}D$ represents number of species (or species richness); at $q = 1$, species are weighted by their relative abundance, then ${}^{1}D$ corresponds to the exponential of Shannon entropy; at $q = 2$, species are weighted by their squared relative abundance (i.e., relatively dominant species will receive a higher weight), then ${}^{2}D$ becomes the inverse of Simpson's diversity index. This formulation can be straightforwardly extended to species incidence data by replacing relative abundance with relative frequency of incidence (*15*).

We compared the geographical patterns (LatDGs and LonDGs) and the results of environmental driver analysis between species richness ($q = 0$), Shannon diversity ($q = 1$) and Simpson diversity ($q = 2$) at the scale of 100 km × 100 km grid cells. To reduce the volume of supplementary materials, we only show the results of RF models for the observed and standardised (at the 40th percentile of sample coverage = 0.82) diversities.

The orders of diversity ($q = 0$, 1 and 2) did not change the biogeographical patterns of global woody angiosperm diversity, as shown in the global maps (fig. S18). All diversities showed similar latitudinal and longitudinal gradients (fig. S19 and S20). While the absolute values were different (${}^{0}D > {}^{1}D > {}^{2}D$), the patterns of partial dependency curves along the environmental gradients in RF models were similar between the orders (fig. S21). The relative importance ranking of environmental variables in RF models was also similar among the orders (fig. S22). Because the rarefaction/extrapolation estimators for high orders ($q > 0$) are nearly unbiased and valid for a wide range of prediction (*15*), the consistency of geographical patterns among different orders of diversities ($q = 0$: species richness, $q = 1$: Shannon diversity, $q = 2$: Simpson diversity) suggests a robustness of our findings.

4. Regional difference in environmental drivers of species diversity

To test the generality or region-specificity of environmental drivers of species richness, we conducted a regional-scale environmental driver analysis (with the RF approach) using the standardised species richness (at the 40th percentile of sample coverage = 0.82) at the scale of 100 km × 100 km grid cells. We used the three longitudinal (Americas, Africa-Europe, and Asia-Oceania) and three latitudinal (north extratropics, tropics, and south extratropics) zones as the unit of regions.

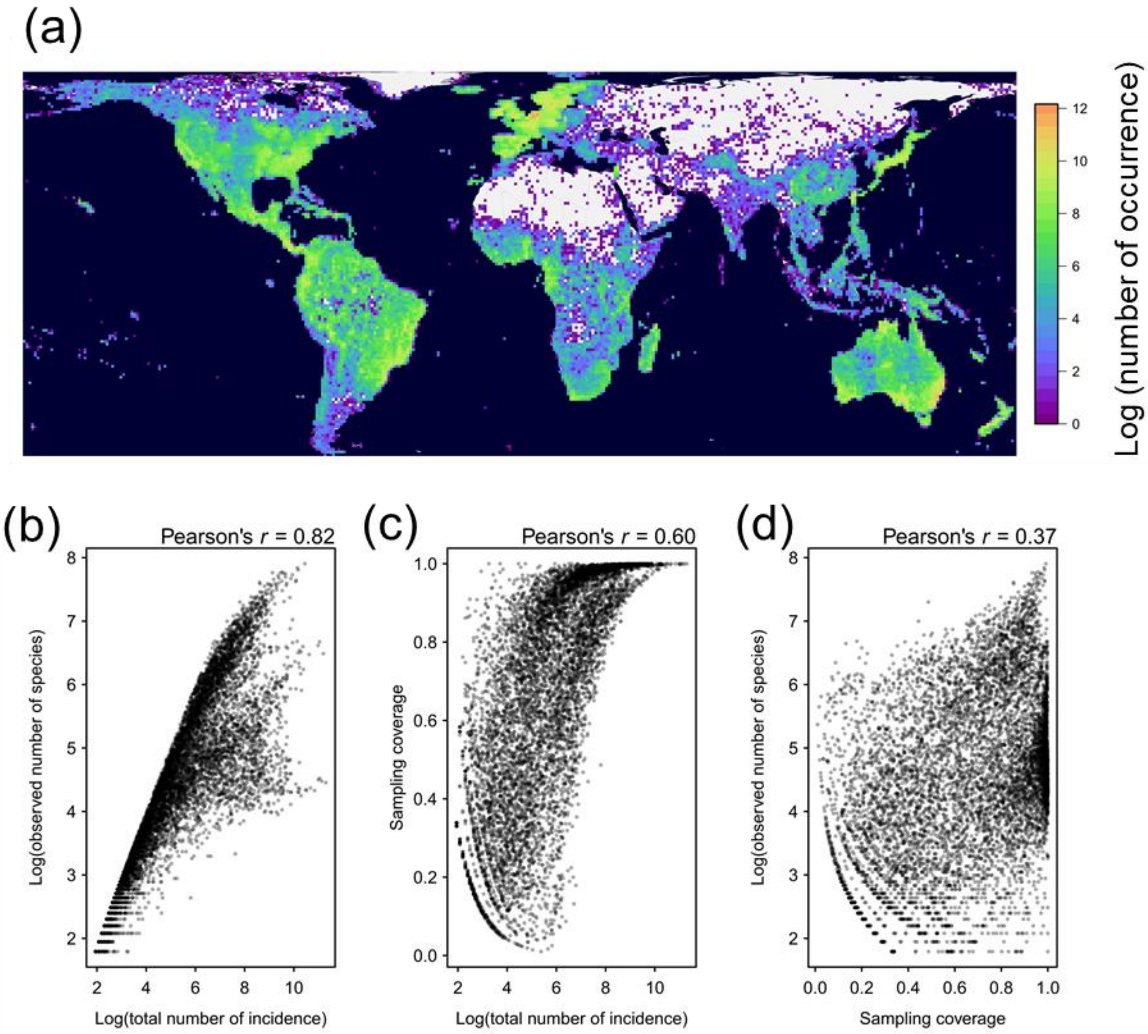
Consistent relationships among the regions were observed for AET (positive monotonic), and Bio 12, AI, and Elv.sd (positive saturation) (fig. S23). For the other variables, the relationships differed among the latitudinal and longitudinal zones (fig. S23). Along the temperature gradient (Bio 1), the standardised species richness was saturated in the south and north extratropics, Americas, and Asia-Oceania zones, whereas it steeply dipped at higher temperature in the tropics (>25 °C) and the Africa-Europe zone (>20 °C). Negative relationships with temperature seasonality (Bio 4) were evident in the Americas and Africa-Europe zones, but not in the Asia-Oceania, and longitudinal zones. Precipitation seasonality (Bio 15) showed high regional variation: unimodal patterns in the tropics, south extratropics, and Asia-Oceania zone, a saturating pattern in the north extratropics, and a negative relationship in the Africa-Europe zone. PET showed a strong negative relationship in the tropics, a unimodal pattern in the Americas zone, and a saturating pattern in the south extratropics; notably, the peak and saturating points were common among the regions (about 1500 mm). With regard to elevation (Elv), the standardised species richness increased rapidly from lowlands (Elv = 0 m) to several hundred metres in all regions except the north extratropics where a rapid increase was observed only above 1,000 m. Historical temperature change (D_{temp}) showed positive relationships within a small degree (<5 °C), but changed to negative with greater change in temperature; this trend was most prominent in the Americas zone. Historical precipitation change (D_{prec}) showed an abrupt decline of diversity at ~1,000 mm in the tropics, and a weak negative relationship in the Asia-Oceania zone, whereas no clear trends were observed in the other regions.

The relative importance ranking of the environmental variables differed among the zones (fig. S24). For the LatDGs (Americas, Africa-Europe, and Asia-Oceania zones), energy (AET), the climatic seasonality (Bio 4 and Bio 15) and historical temperature change (D_{temp}) showed high relative importance. For the LonDGs (north extratropics, tropics, and south extratropics), the factors relevant to availability of energy and water (AET, PET, AI, Bio 1 and/or Bio 12) were the most important drivers of species richness.

5. Software

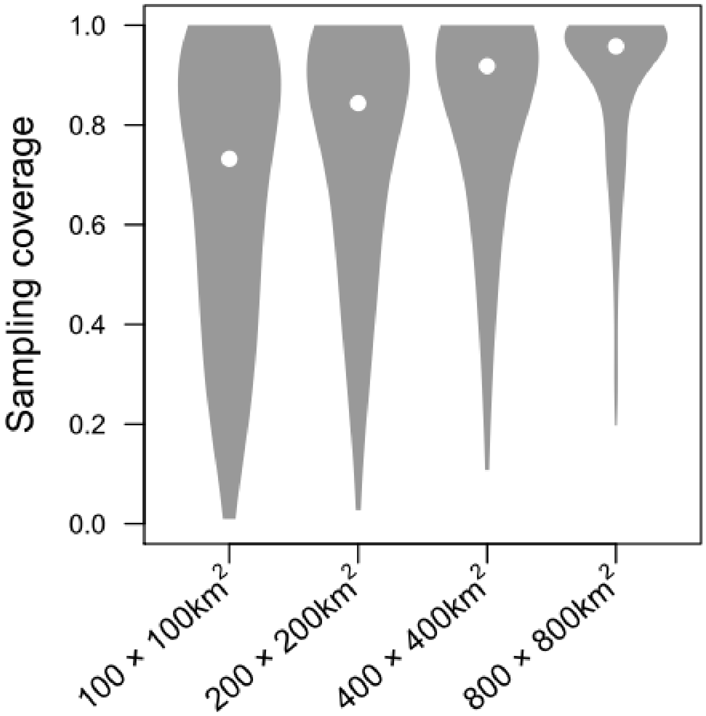
All analyses were performed in the R statistical environment (ver. 4.1.1) (*81*) with the following packages: 'data.table' (*82*), 'dplyr' (*83*), 'tidyr' (*84*) for handling data; 'doParallel' (*85*), 'foreach' (*86*) for processing parallel computation; 'geodata' (*87*), 'maps' (*88*), 'maptools' (*89*), 'raster' (*90*), 'rasterVis' (*91*), 'rgdal' (*92*), 'rgeos' (*93*), 'sf' (*94*), 'stars' (*95*), 'terra' (*96*), for editing spatial data; 'ggplot2' (*97*), 'colorRamps' (*98*), 'pals' (*99*), 'RcolorBrewer' (*100*), 'sm' (*101*), 'TeachingDemos' (*102*) for graphic working; 'rgbif' (*103*) for downloading species occurrence records from GBIF; 'iNEXT' (*104*) for estimating and standardizing species diversity; 'mgcv' (*75*) for conducting GAM analysis; 'ranger' (*105*) for conducting Random Forest analysis; 'pdp' (*106*) for calculating partial dependence plots of Random Forest model; 'spm' (*107*) for cross-validation of Random Forest model; 'plotbiomes' (*108*) for drawing Whittaker biome; 'car' (*109*) for checking multicollinearity, 'pgirmess' (*110*) for Moran's I test 'SpatialPack' (*111*) for checking spatial correlations; 'htmlwidgets' (*112*), 'networkD3' (*113*), 'webshot' (*114*) for making sanky diagrams.

Fig. S1.



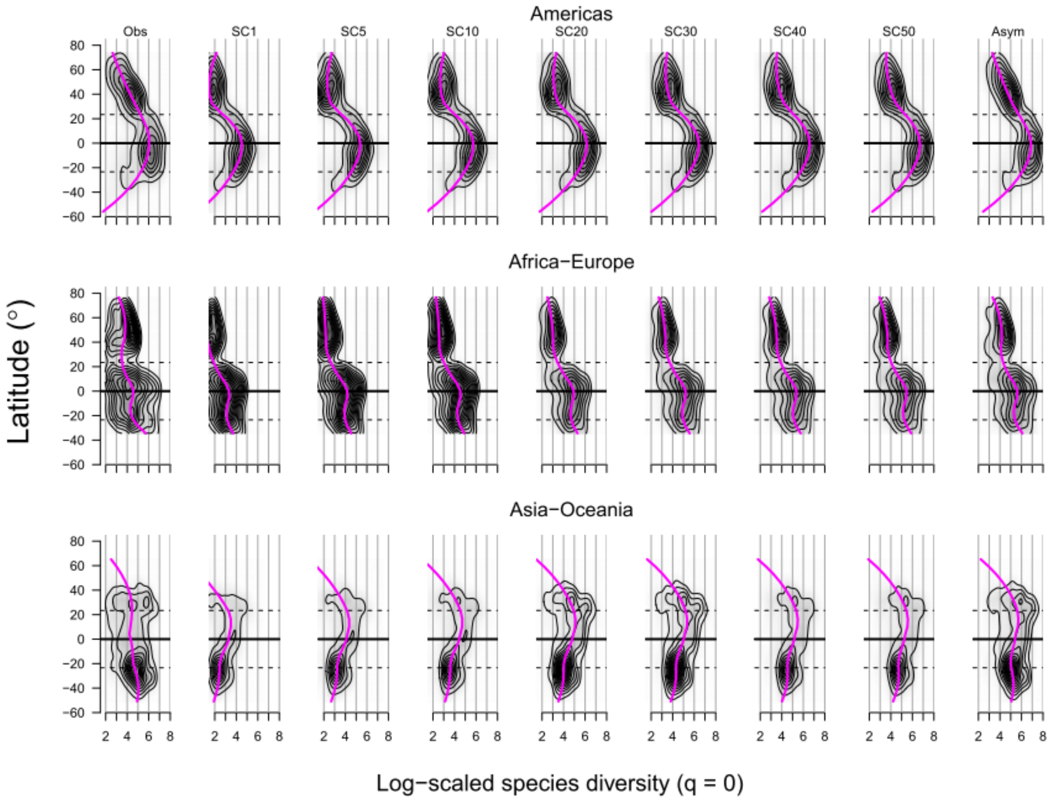
Global maps of log-scaled total number of occurrence records (a) and the relationships between observed number of species, total number of incidence, and sample coverage (b-d) at the level of 100 km × 100 km grid cells.

Fig. S2.



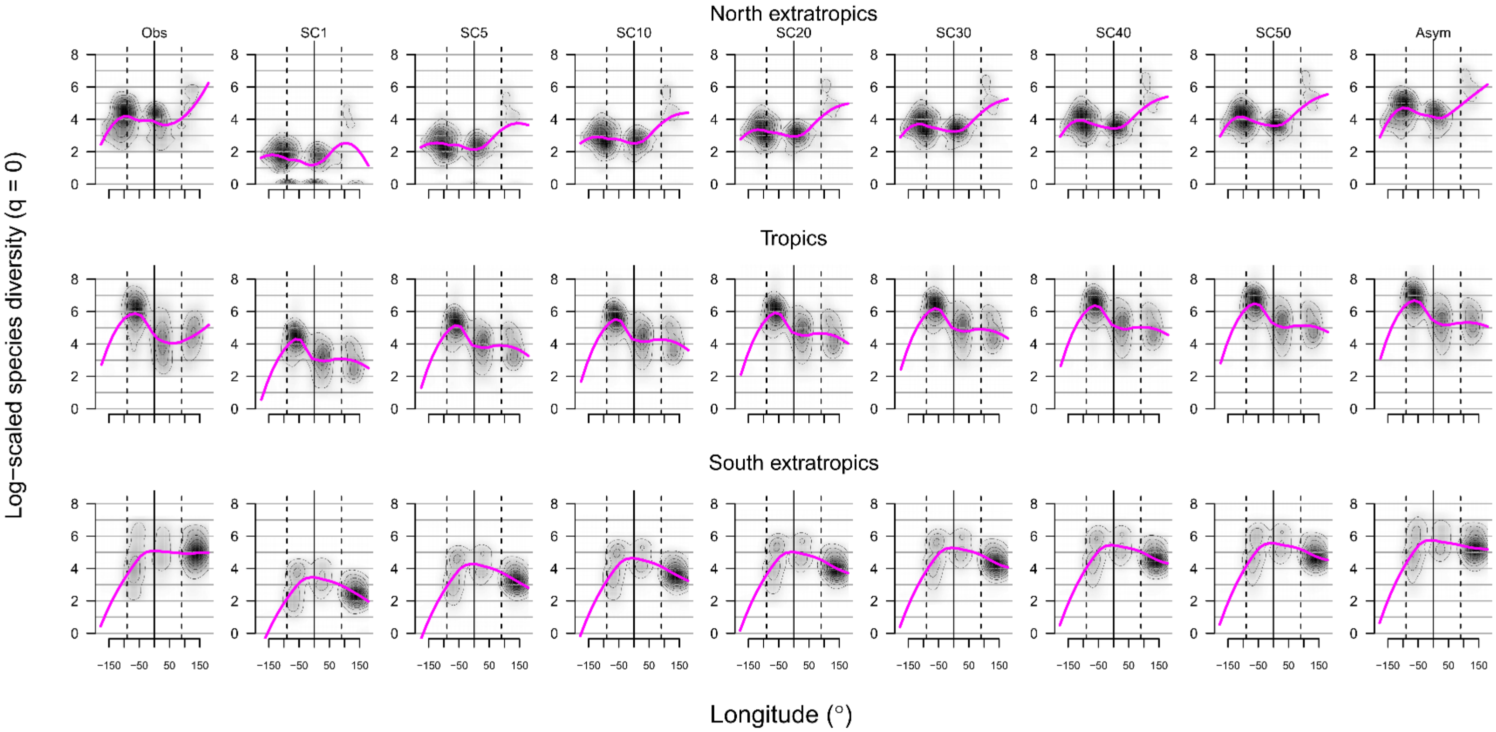
Violin plots of sample coverage values at four different spatial resolutions. White point shows mean value.

Fig. S3.



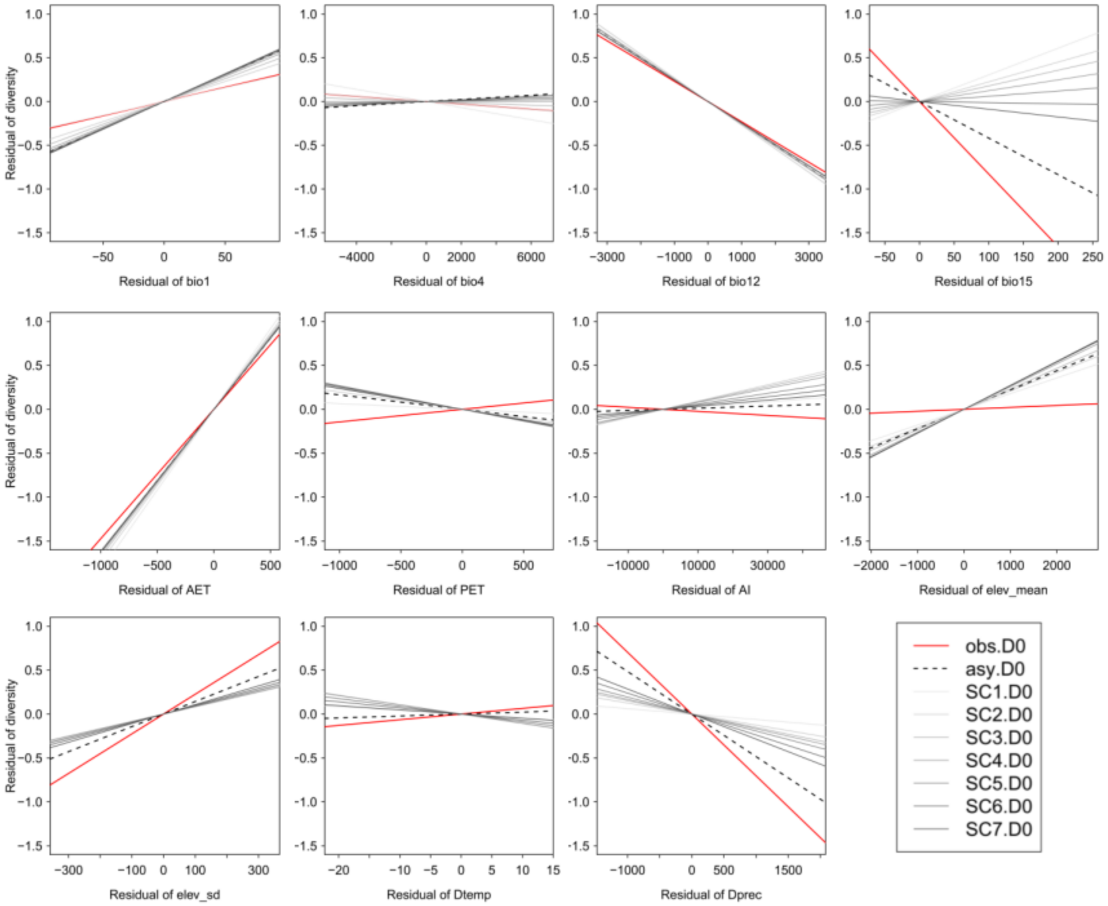
Latitudinal gradients of species richness at three longitudinal zones for the 100 km × 100 km grid cells with different levels of standardization by sample coverage (SC): observed sample coverage (Obs), standardised at 1st, 5th, 10th, 20th, 30th, 40th and 50th percentiles of SC (SC1–50; see table S4 for corresponding SC values), and asymptotic diversity (Asym; corresponding to SC = 100%). The diversity values are log-scaled. Loess (locally estimated scatterplot smoothing) curve (scaling parameter alpha = 0.6) is shown (pink line). Thick vertical line represents the equator. Dashed lines represent the tropics of Capricorn and Cancer.

Fig. S4.



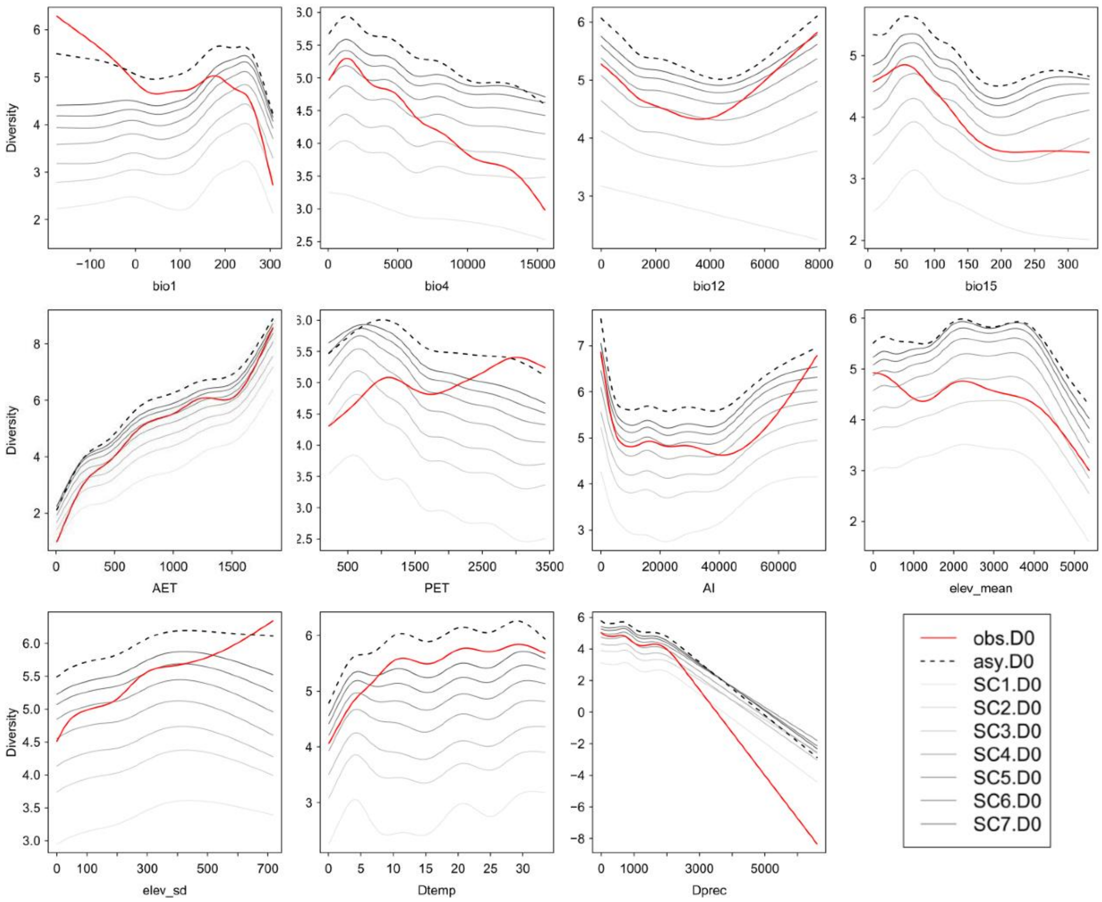
Longitudinal gradients of species richness at three latitudinal zones (tropics and north and south extratropics) for the 100 km × 100 km grid cells with different levels of standardization by sample coverage (SC): observed sample coverage (Obs), standardised at 1st, 5th, 10th, 20th, 30th, 40th and 50th percentiles of SC (SC1–50; see table S4 for corresponding SC values), and asymptotic diversity (Asym; corresponding to SC = 100%). The diversity values are log-scaled. Loess (locally estimated scatterplot smoothing) curve (scaling parameter alpha = 0.6) is shown (pink line). The dashed vertical lines represent ±90 degrees in longitude.

Fig. S5.



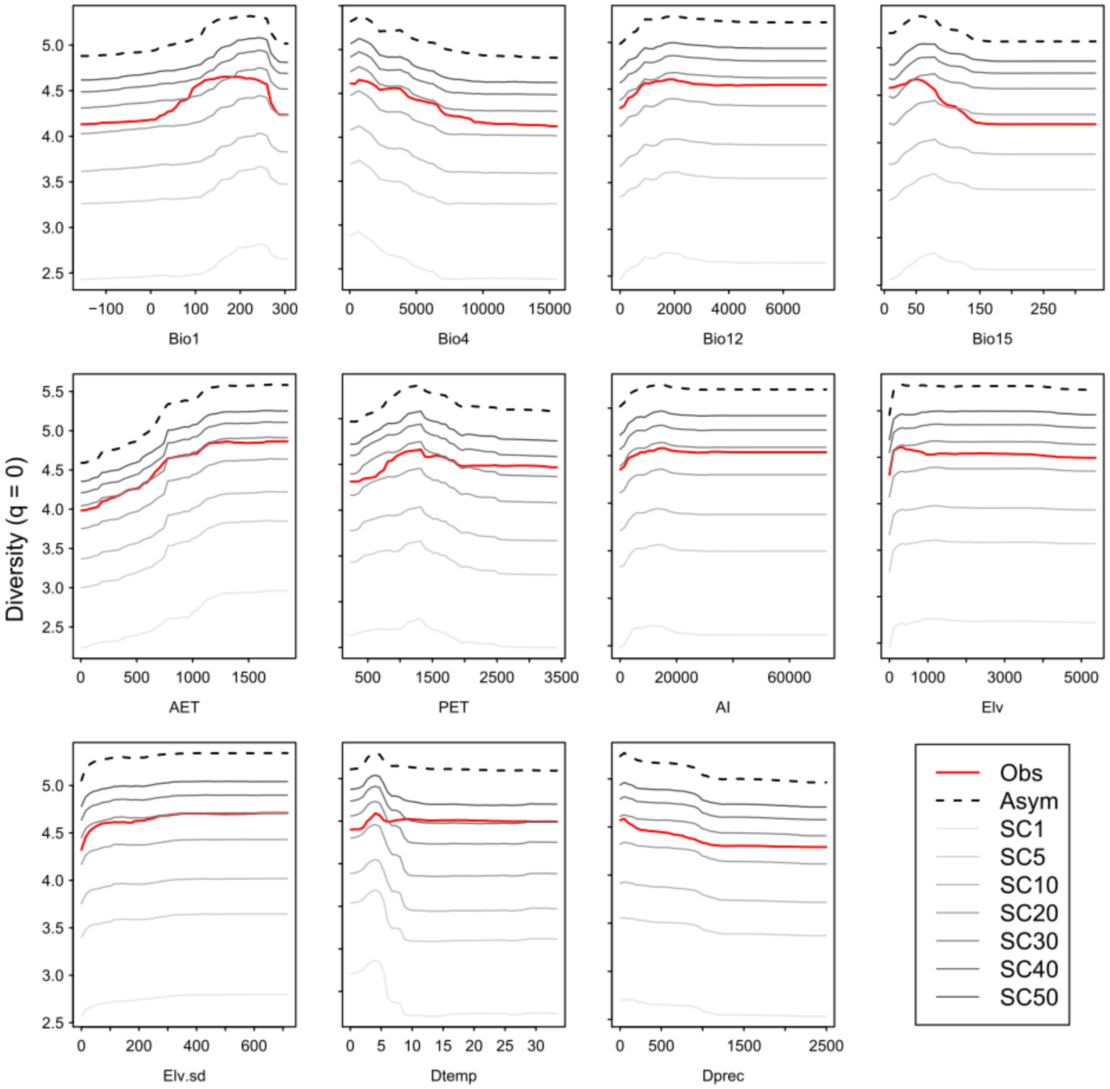
Linear relationship between species richness and environmental variables analysed by ordinary least squares model for the 100 km × 100 km grid cells. The partial effect of each variable was evaluated by the residual regression. Species richness is standardised with different levels of sample coverage (SC): observed sample coverage (Obs), standardised at 1st, 5th, 10th, 20th, 30th, 40th and 50th percentiles of SC (SC1–50; see table S4 for corresponding SC values), and asymptotic diversity (Asym; corresponding to SC = 100%). The environmental variables are mean annual temperature (Bio 1), temperature seasonality (Bio4), annual precipitation (Bio 12), precipitation seasonality (Bio 15), actual evapotranspiration (AET), potential evapotranspiration (PET), aridity index (AI), average elevation (Elv), standard deviation of elevation (Elv.sd), and differences in temperature (D_{temp}) and precipitation (D_{prec}) between the Last Glacial Maximum and the present.

Fig. S6.



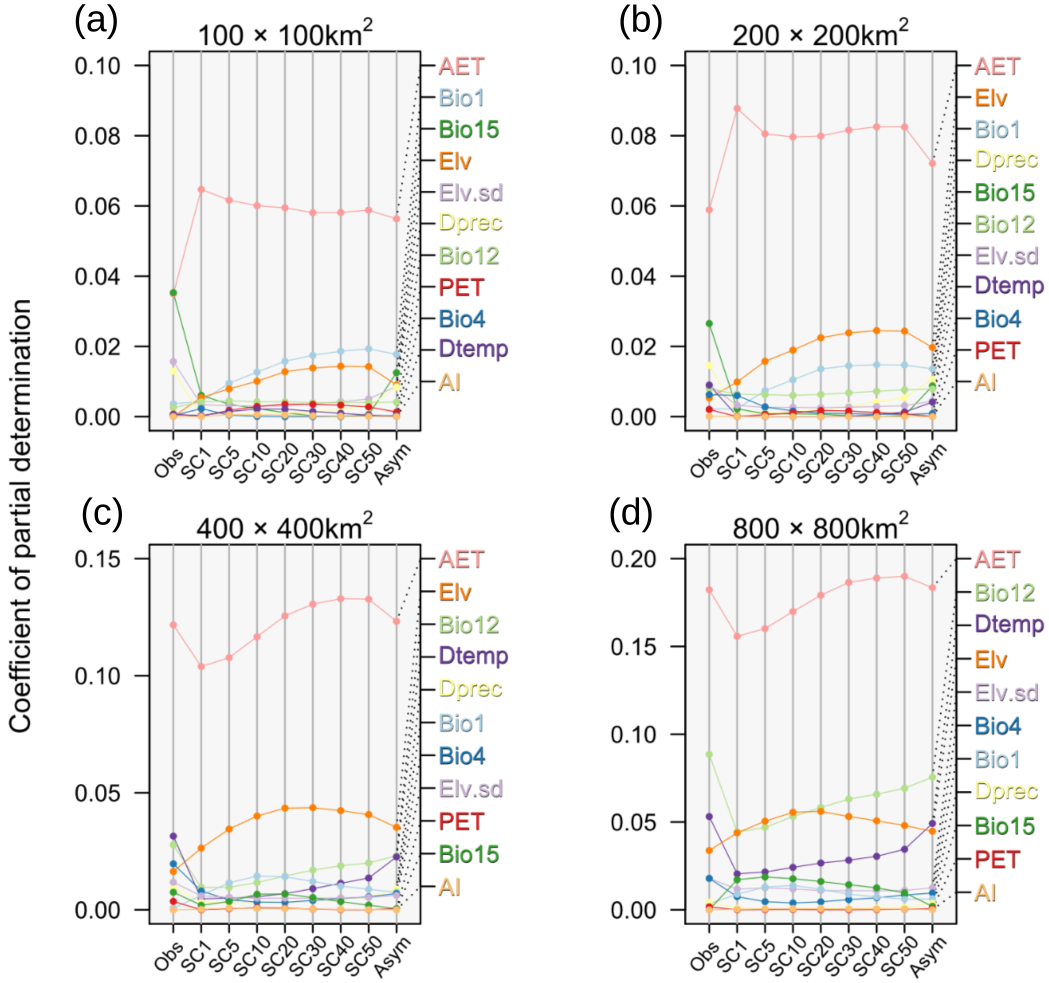
Relationship between species richness and environmental factors analysed by generalized additive model for the 100 km × 100 km grid cells. Species richness is standardised with different levels of sample coverage (SC): observed sample coverage (Obs), standardised at 1st, 5th, 10th, 20th, 30th, 40th and 50th percentiles of SC (SC1–50; see table S4 for corresponding SC values), and asymptotic diversity (Asym; corresponding to SC = 100%). The environmental variables are mean annual temperature (bio1), temperature seasonality (Bio 4), annual precipitation (Bio 12), precipitation seasonality (Bio 15), actual evapotranspiration (AET), potential evapotranspiration (PET), aridity index (AI), average elevation (Elv), standard deviation of elevation (Elv.sd), and differences in temperature (D_{temp}) and precipitation (D_{prec}) between the Last Glacial Maximum and the present.

Fig. S7.



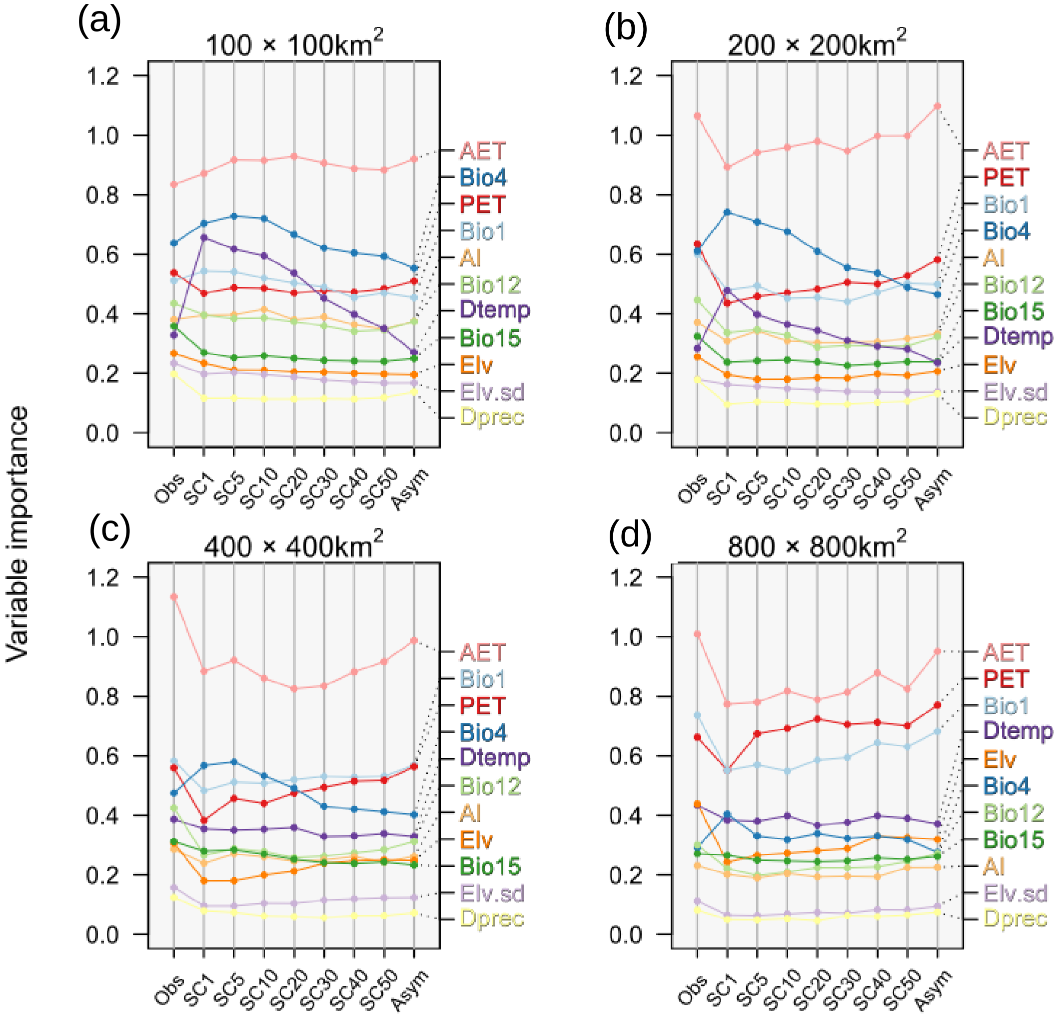
Partial dependency plot between species richness and environmental factors analysed by random forest model for the 100 km × 100 km grid cells. Species richness is standardised with different levels of sample coverage (SC): observed sample coverage (Obs), standardised at 1st, 5th, 10th, 20th, 30th, 40th and 50th percentiles of SC (SC1–50; see table S4 for corresponding SC values), and asymptotic diversity (Asym; corresponding to SC = 100%). The environmental variables are mean annual temperature (Bio 1), temperature seasonality (Bio 4), annual precipitation (Bio 12), precipitation seasonality (Bio 15), actual evapotranspiration (AET), potential evapotranspiration (PET), aridity index (AI), average elevation (Elv), standard deviation of elevation (Elv.sd), and differences in temperature (D_{temp}) and precipitation (D_{prec}) between the Last Glacial Maximum and the present.

Fig. S8.



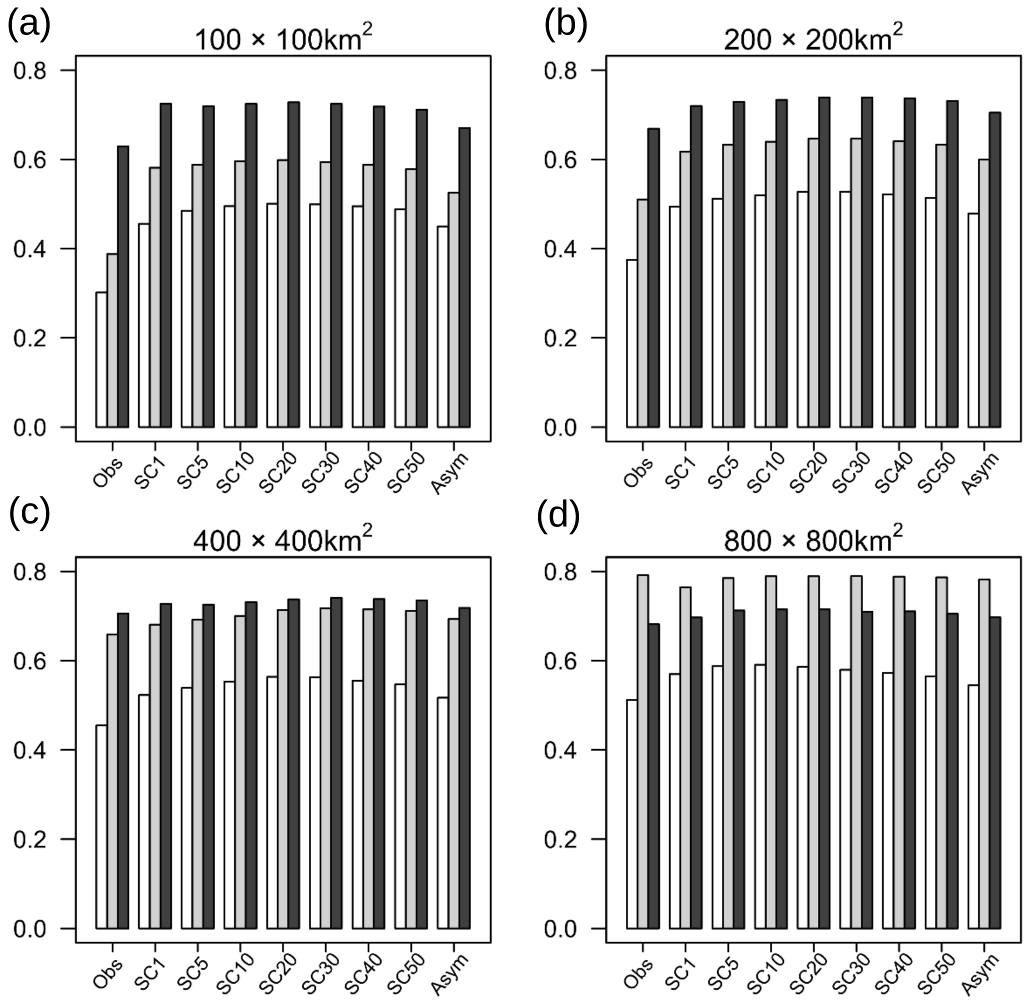
Relative importance (coefficients of partial determination) of environmental variables in the ordinary least squares model explaining species richness for four spatial resolutions ((a) 100 km × 100 km, (b) 200 km × 200 km, (c) 400 km × 400 km, and (d) 800 km × 800 km) with different levels of standardization by sample coverage (SC): observed sample coverage (Obs), standardised at 1st, 5th, 10th, 20th, 30th, 40th and 50th percentiles of SC (SC1–50; see table S4 for corresponding SC values), and asymptotic diversity (Asym; corresponding to SC = 100%) at different spatial resolutions. The environmental variables are mean annual temperature (Bio 1), temperature seasonality (Bio 4), annual precipitation (Bio 12), precipitation seasonality (Bio 15), actual evapotranspiration (AET), potential evapotranspiration (PET), aridity index (AI), average elevation (Elv), standard deviation of elevation (Elv.sd), and differences in temperature (D_{temp}) and precipitation (D_{prec}) between the Last Glacial Maximum and the present.

Fig. S9.



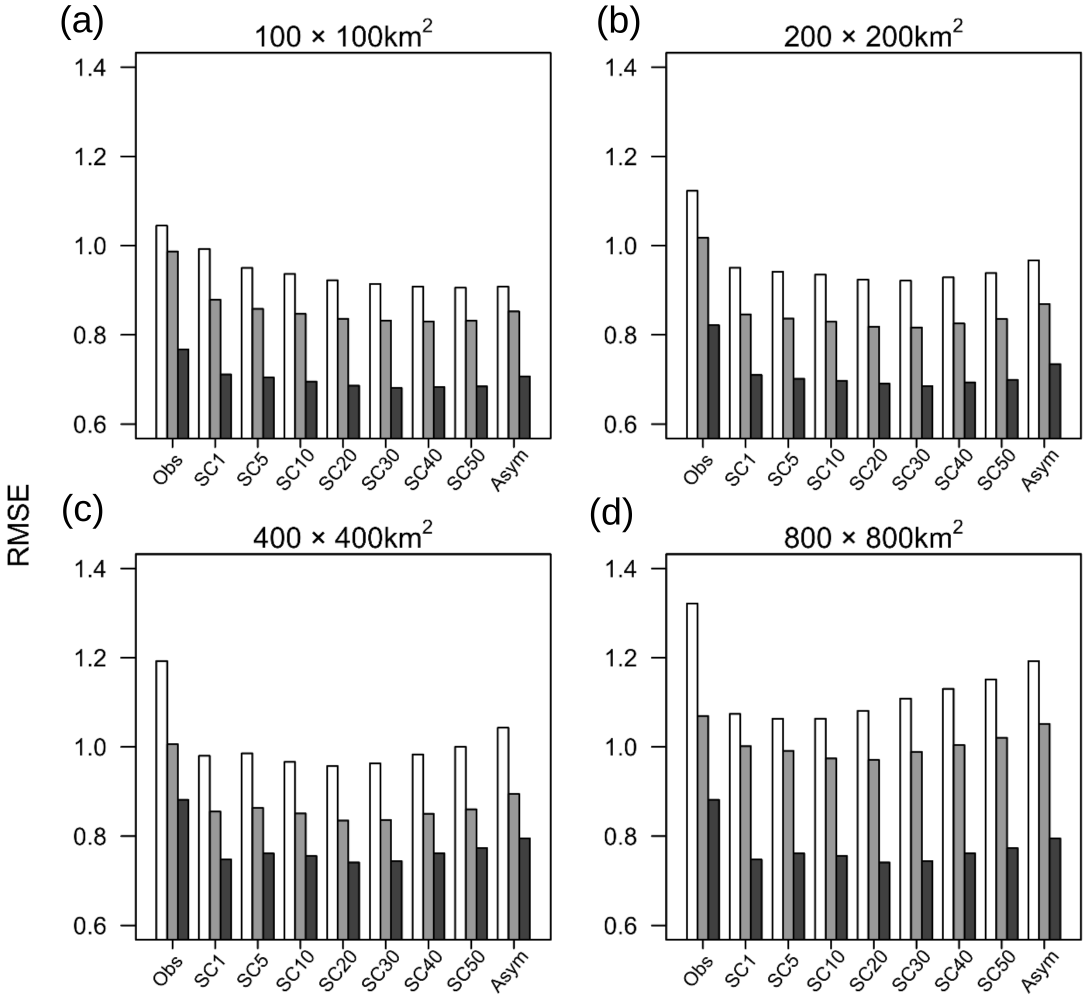
Relative importance (permutation importance) of environmental variables in the random forest models explaining species richness for four spatial resolutions ((a) 100 km × 100 km, (b) 200 km × 200 km, (c) 400 km × 400 km, and (d) 800 km × 800 km) with different levels of standardization by sample coverage (SC): observed sample coverage (Obs), standardised at 1st, 5th, 10th, 20th, 30th, 40th and 50th percentiles of SC (SC1–50; see table S4 for corresponding SC values), and asymptotic diversity (Asym; corresponding to SC = 100%) at different spatial resolutions. The environmental variables are mean annual temperature (Bio 1), temperature seasonality (Bio 4), annual precipitation (Bio 12), precipitation seasonality (Bio 15), actual evapotranspiration (AET), potential evapotranspiration (PET), aridity index (AI), average elevation (Elv), standard deviation of elevation (Elv.sd), and differences in temperature (D_{temp}) and precipitation (D_{prec}) between the Last Glacial Maximum and the present.

Fig. S10.



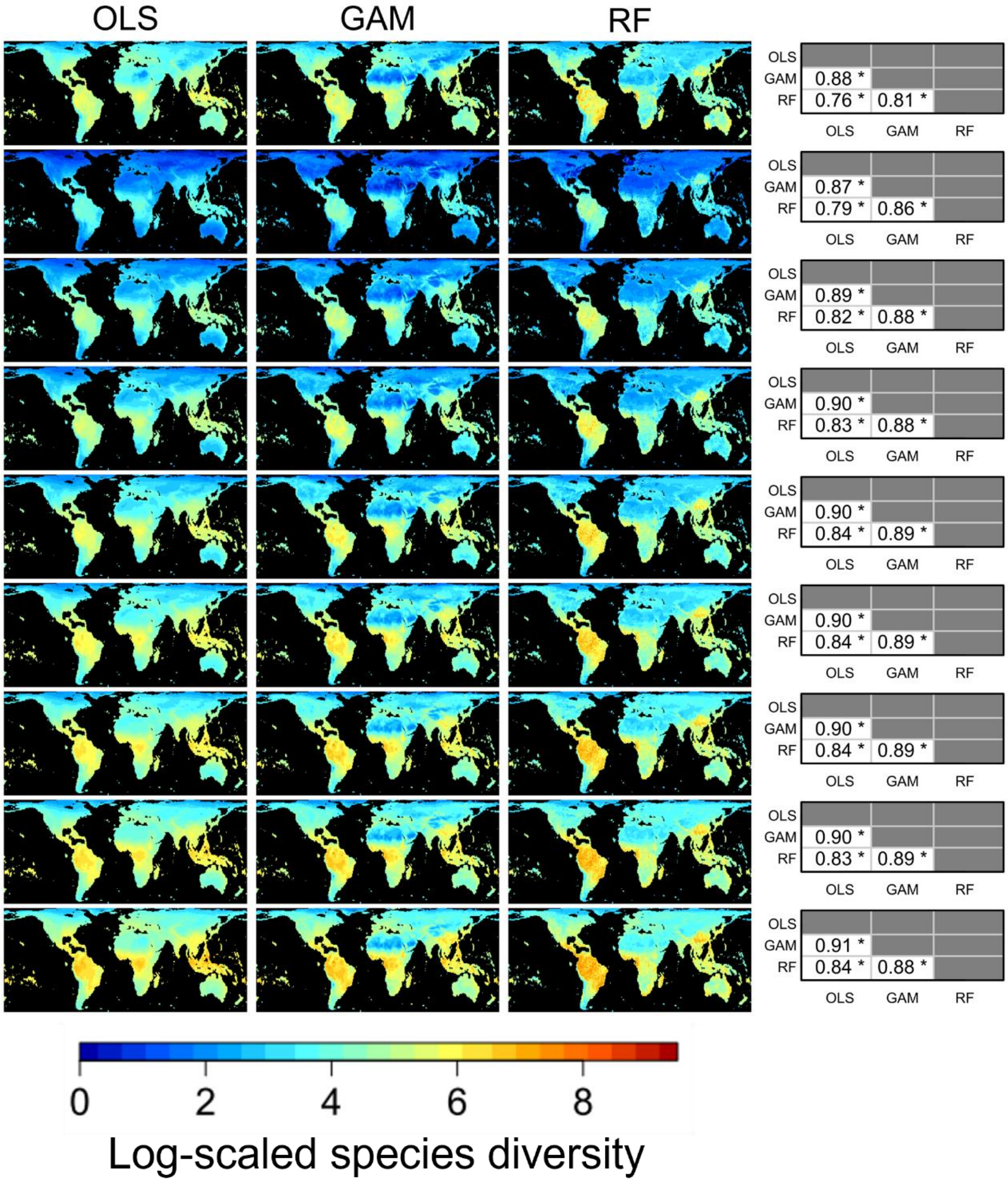
Explanatory power (R^2) of regression models (ordinary least squares = white, generalized additive model = light grey, random forest = dark grey) for four spatial resolutions ((a) 100 km × 100 km, (b) 200 km × 200 km, (c) 400 km × 400 km, and (d) 800 km × 800 km) explaining the relationship between environmental variables and species richness with different levels of standardization by sample coverage (SC): observed sample coverage (Obs), standardised at 1st, 5th, 10th, 20th, 30th, 40th and 50th percentiles of SC (SC1–50; see table S4 for corresponding SC values), and asymptotic diversity (Asym; corresponding to SC = 100%).

Fig. S11.



Predictive performance (root mean squared error from 10-fold cross-validation) of regression models (ordinary least squares = white, generalized additive model = light grey, random forest = dark grey) for four spatial resolutions ((a) 100 km × 100 km, (b) 200 km × 200 km, (c) 400 km × 400 km, and (d) 800 km × 800 km), explaining the relationship between environmental variables and species richness with different levels of standardization by sample coverage (SC): observed sample coverage (Obs), standardised at 1st, 5th, 10th, 20th, 30th, 40th and 50th percentiles of SC (SC1–50; see table S4 for corresponding SC values), and asymptotic diversity (Asym; corresponding to SC = 100%).

Fig. S12.



	OLS	GAM	RF
OLS			
GAM	0.88 *		
RF	0.76 *	0.81 *	

	OLS	GAM	RF
OLS			
GAM	0.87 *		
RF	0.79 *	0.86 *	

	OLS	GAM	RF
OLS			
GAM	0.89 *		
RF	0.82 *	0.88 *	

	OLS	GAM	RF
OLS			
GAM	0.90 *		
RF	0.83 *	0.88 *	

	OLS	GAM	RF
OLS			
GAM	0.90 *		
RF	0.84 *	0.89 *	

	OLS	GAM	RF
OLS			
GAM	0.90 *		
RF	0.84 *	0.89 *	

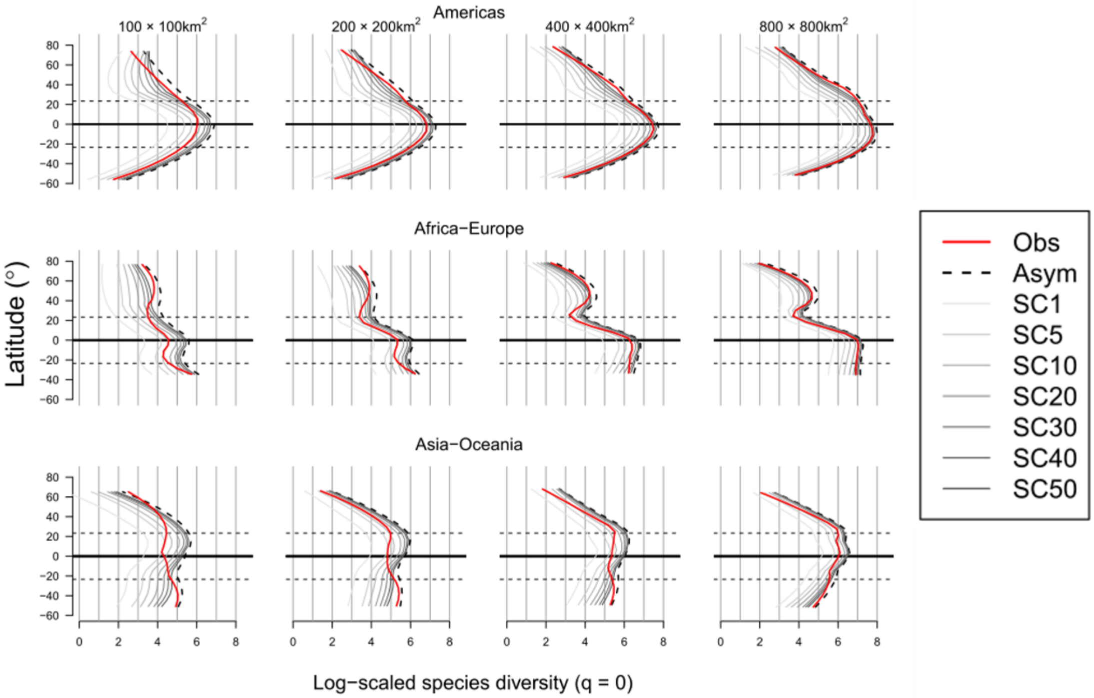
	OLS	GAM	RF
OLS			
GAM	0.90 *		
RF	0.84 *	0.89 *	

	OLS	GAM	RF
OLS			
GAM	0.90 *		
RF	0.83 *	0.89 *	

	OLS	GAM	RF
OLS			
GAM	0.91 *		
RF	0.84 *	0.88 *	

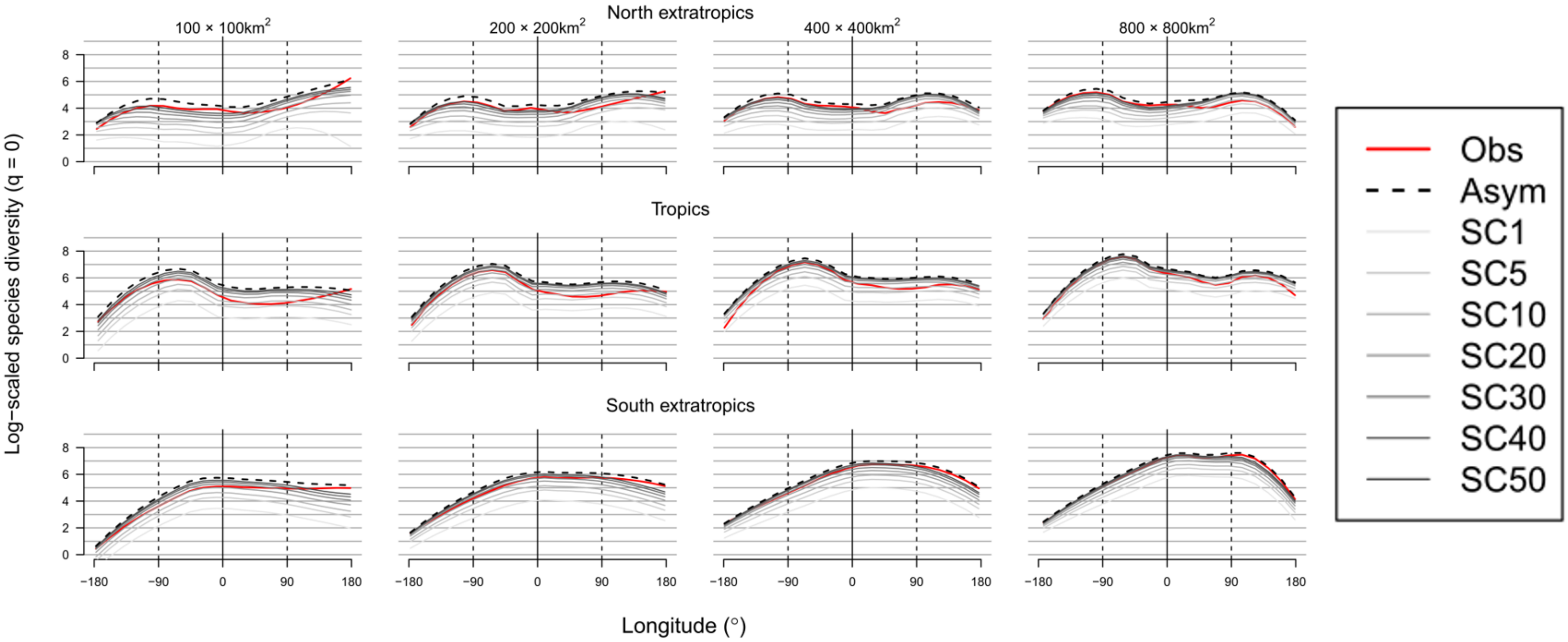
Projection of species richness with different levels of standardization by sample coverage (SC; see table S4 for corresponding SC values) predicted using ordinary least squares (OLS), generalized additive model (GAM) and Random Forest (RF) with environmental variables at 100 km × 100 km grid cell level. The right-hand column presents the table of correlation coefficients among the predicted values of three modelling frameworks. Asterisk indicates statistically significant correlation at p-value < 0.05 in the modified-t-test (*97*).

Fig. S13.



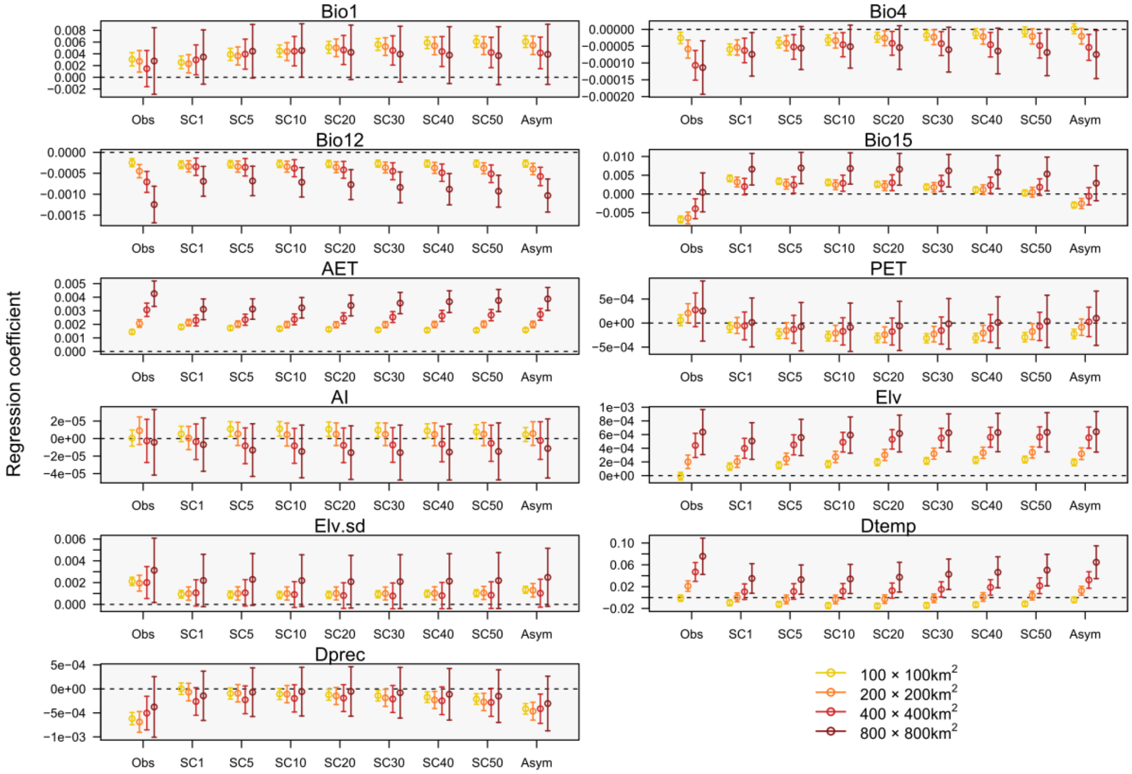
Latitudinal gradient of species richness at different spatial resolutions for three longitudinal bands: the Americas, the Africa-Europe, and the Asia-Oceania zones. Loess (locally estimated scatterplot smoothing) curves (scaling parameter alpha = 0.6) are shown per species richness estimate with different levels of standardization by sample coverage (SC): observed sample coverage (Obs), standardised at 1st, 5th, 10th, 20th, 30th, 40th and 50th percentiles of SC (SC1–50; see table S4 for corresponding SC values), and asymptotic diversity (Asym; corresponding to SC = 100%). The diversity values are log-scaled. Thick vertical line represents the equator. Shaded area represents the tropical zone between the tropics of Capricorn and Cancer.

Fig. S14.



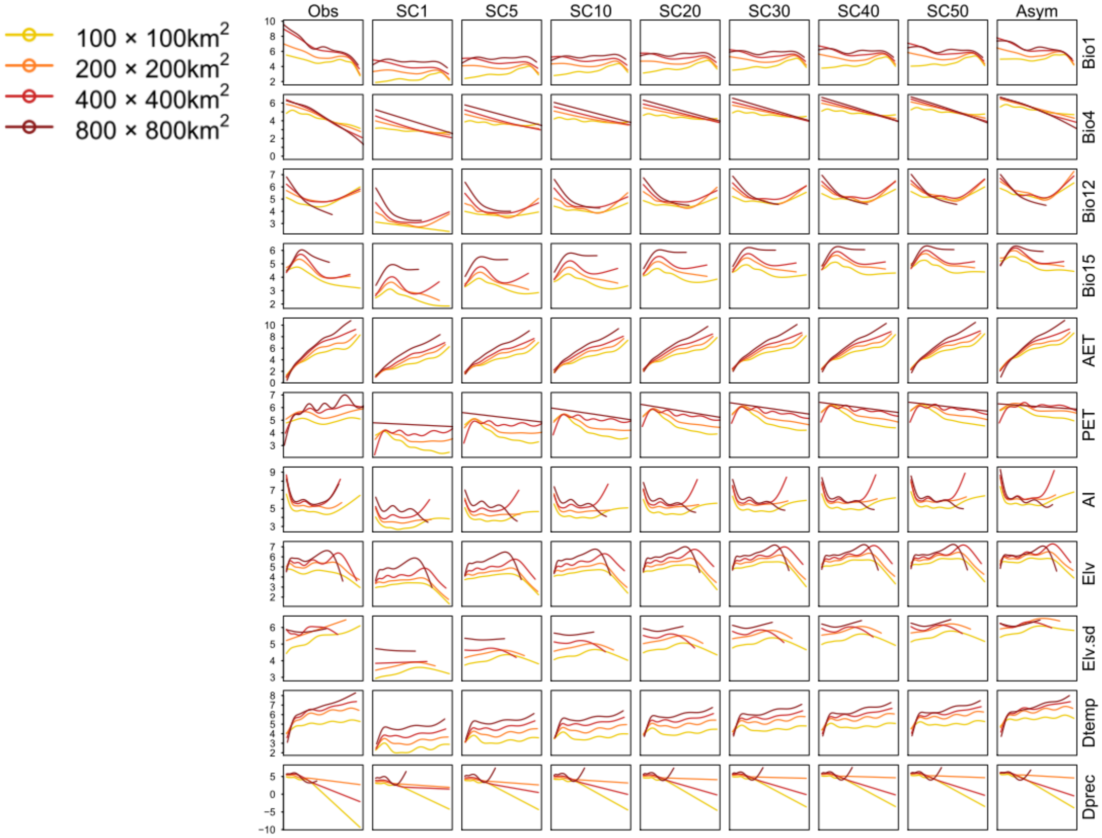
Longitudinal gradients of species richness at different spatial resolutions for three latitudinal zones (tropics and north and south extratropics) with different levels of standardization by sample coverage (SC): observed sample coverage (Obs), standardised at 1st, 5th, 10th, 20th, 30th, 40th and 50th percentiles of SC (SC1–50; see table S4 for corresponding SC values), and asymptotic diversity (Asym; corresponding to SC = 100%). The diversity values are log-scaled. Loess (locally estimated scatterplot smoothing) curves (scaling parameter alpha = 0.6) are shown per species richness estimate. Thick vertical line represents the Prime meridian.

Fig. S15.



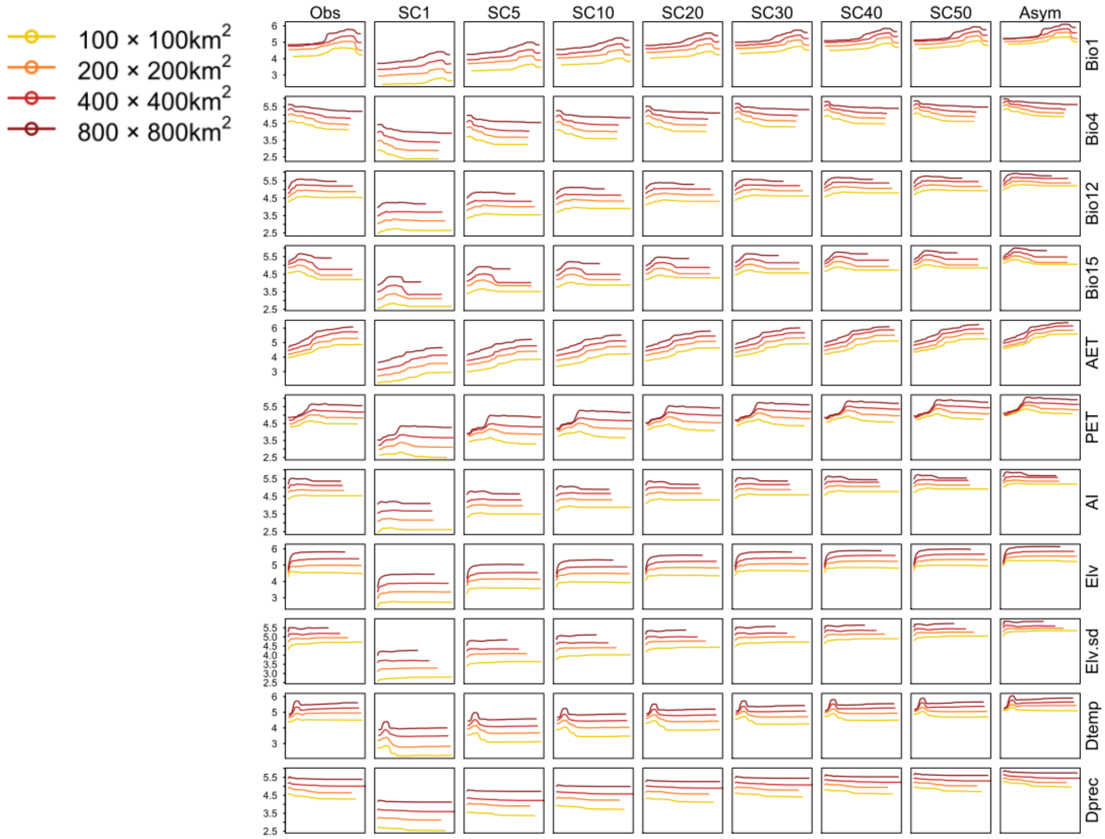
Impact of spatial-resolutions on the standardised regression coefficient of the environmental variables in ordinary least square regression model explaining species richness with different levels of standardization by sample coverage (SC): observed sample coverage (Obs), standardised at 1st, 5th, 10th, 20th, 30th, 40th and 50th percentiles of SC (SC1–50; see table S4 for corresponding SC values), and asymptotic diversity (Asym; corresponding to SC = 100%) analysed using generalized additive model. The environmental variables are mean annual temperature (Bio 1), temperature seasonality (Bio 4), annual precipitation (Bio 12), precipitation seasonality (Bio 15), actual evapotranspiration (AET), potential evapotranspiration (PET), aridity index (AI), average elevation (Elv), standard deviation of elevation (Elv.sd), and differences in temperature (D_{temp}) and precipitation (D_{prec}) between the Last Glacial Maximum and the present.

Fig. S16.



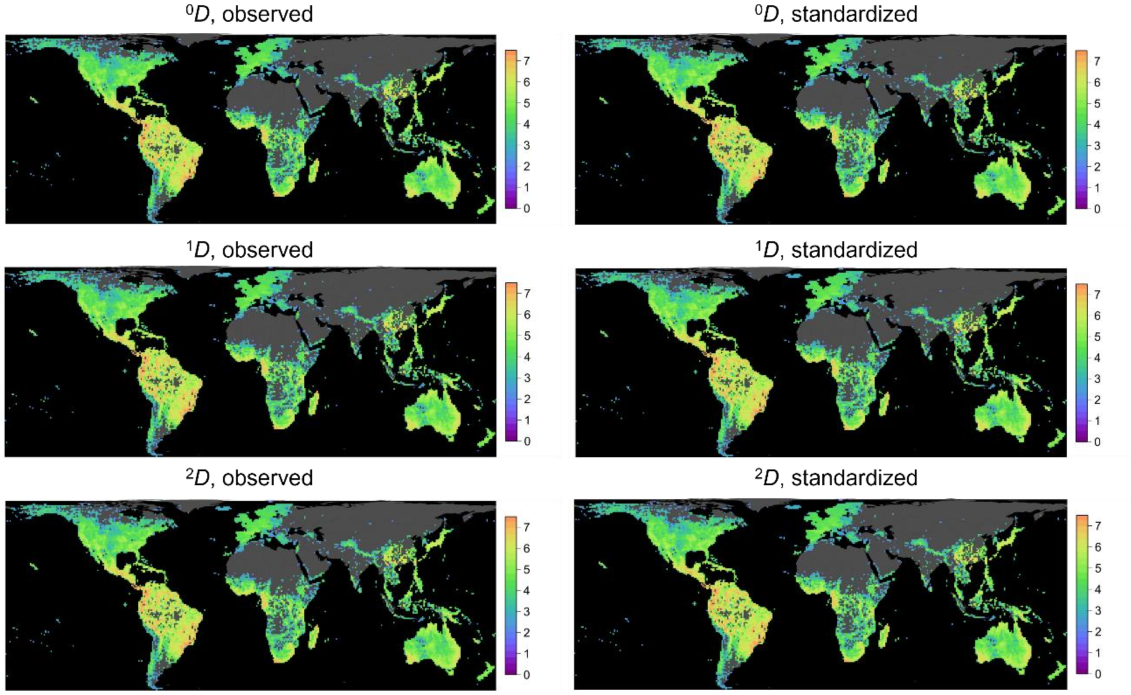
Impact of spatial-resolutions on the relationships between the environmental variables and species richness with different levels of standardization by sample coverage (SC): observed sample coverage (Obs), standardised at 1st, 5th, 10th, 20th, 30th, 40th and 50th percentiles of SC (SC1–50; see table S4 for corresponding SC values), and asymptotic diversity (Asym; corresponding to SC = 100%) analysed using generalized additive model. The environmental variables are mean annual temperature (Bio 1), temperature seasonality (Bio 4), annual precipitation (Bio 12), precipitation seasonality (Bio 15), actual evapotranspiration (AET), potential evapotranspiration (PET), aridity index (AI), average elevation (Elv), standard deviation of elevation (Elv.sd), and differences in temperature (D_{temp}) and precipitation (D_{prec}) between the Last Glacial Maximum and the present.

Fig. S17.



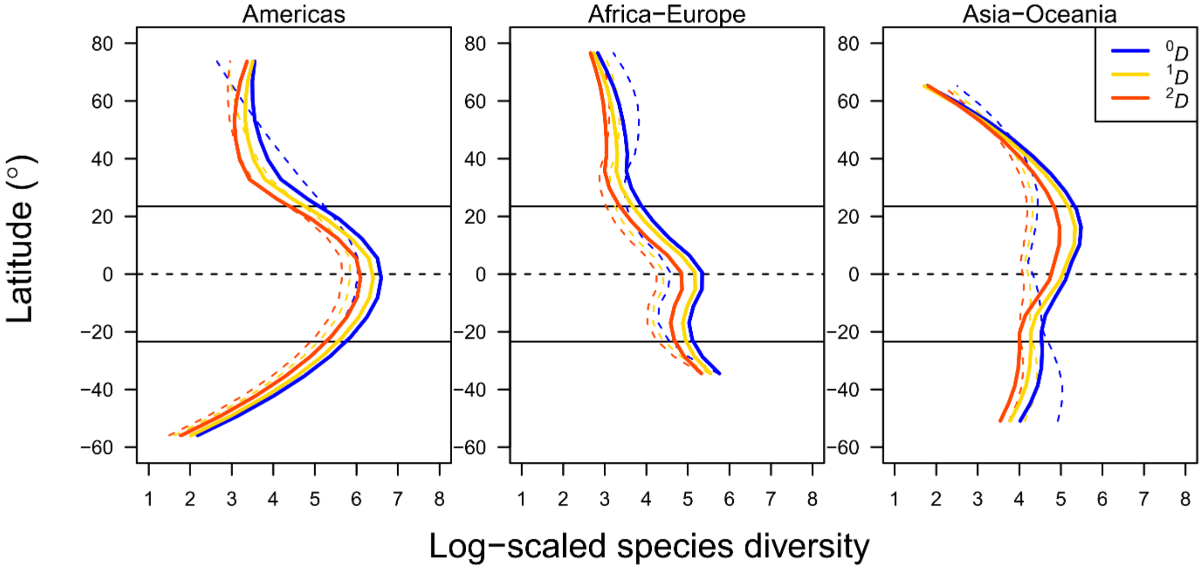
Impact of spatial-resolutions on the relationships between the environmental variables and species richness with different levels of standardization by sample coverage (SC): observed sample coverage (Obs), standardised at 1st, 5th, 10th, 20th, 30th, 40th and 50th percentiles of SC (SC1-50; see table S4 for corresponding SC values), and asymptotic diversity (Asym; corresponding to SC = 100%) analysed using random forest model. The environmental variables are mean annual temperature (Bio 1), temperature seasonality (Bio 4), annual precipitation (Bio 12), precipitation seasonality (Bio 15), actual evapotranspiration (AET), potential evapotranspiration (PET), aridity index (AI), average elevation (Elv), standard deviation of elevation (Elv.sd), and differences in temperature (D_{temp}) and precipitation (D_{prec}) between the Last Glacial Maximum and the present.

Fig. S18.



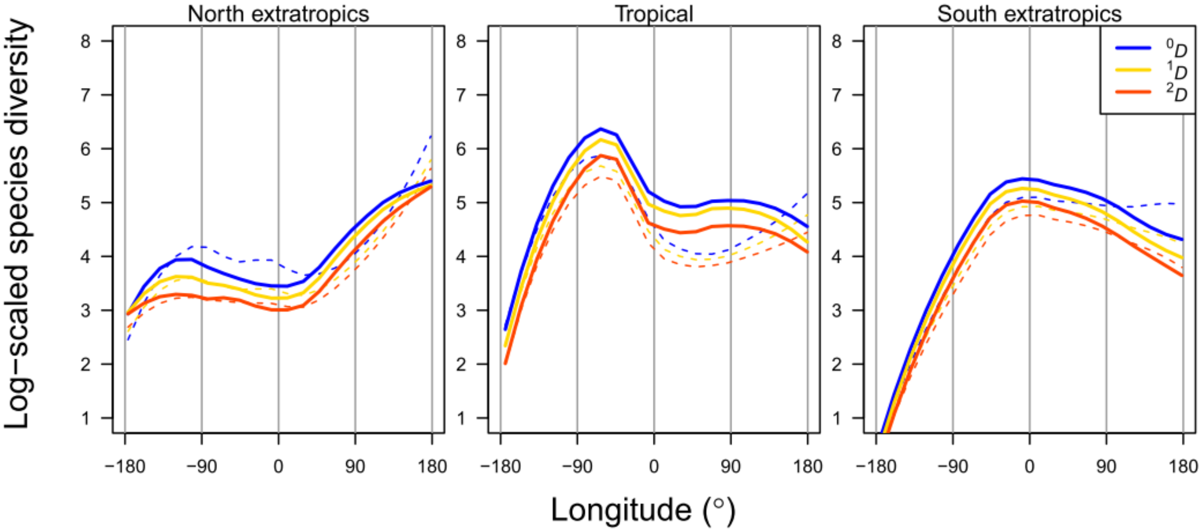
Geographical maps for the observed and standardised (at 40th percentile of sample coverage = 0.82) species diversity based on Hill numbers: species richness (^0D), Shannon diversity (^1D), and Simpson diversity (^2D), at approximately 100 km × 100 km grid cell level.

Fig. S19.



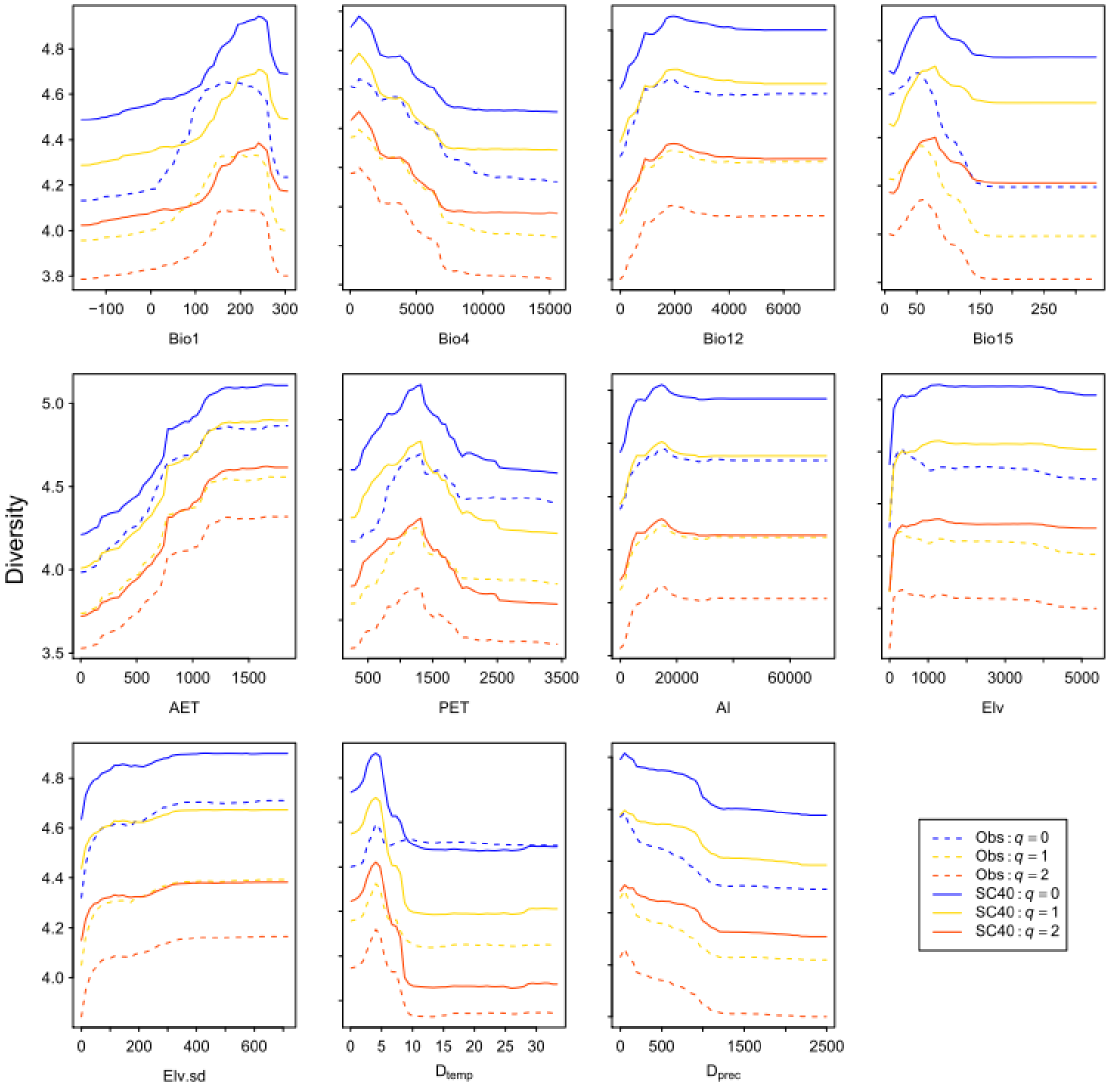
Latitudinal gradient of species diversity based on Hill numbers (^qD; $q = 0, 1, 2$) at approximately 100 km × 100 km grid cell level. Dashed and solid lines represent observed and standardised (at 40th percentile of sample coverage = 0.82) values, respectively. Loess (locally estimated scatterplot smoothing) curves (scaling parameter alpha = 0.6) are shown.

Fig. S20.



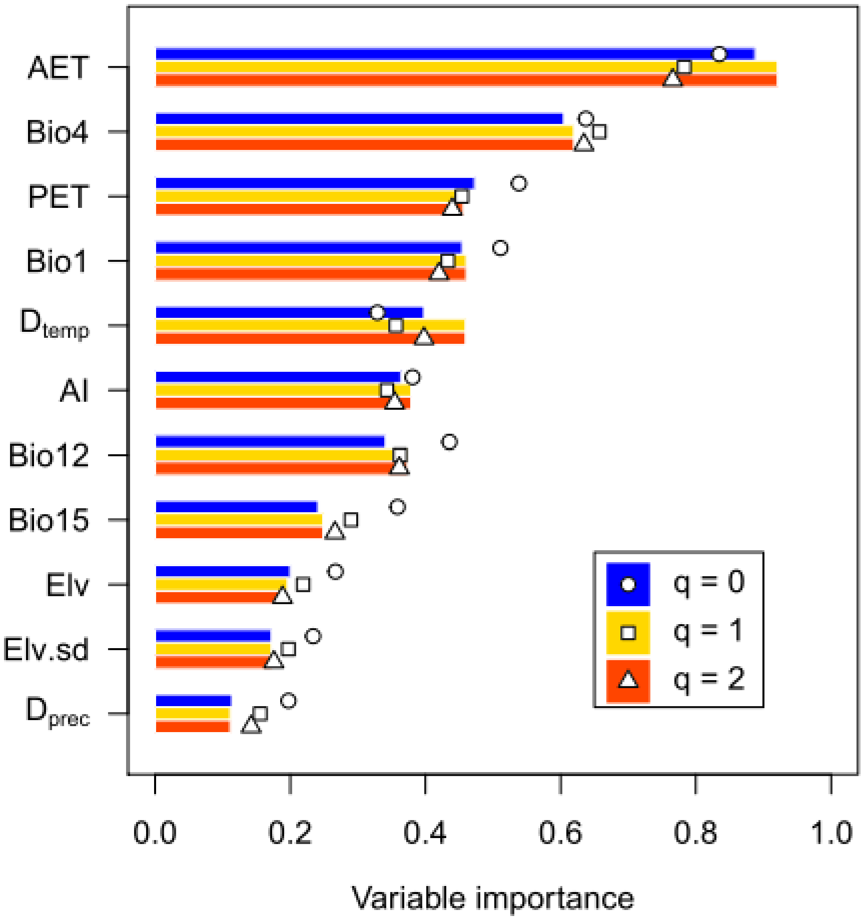
Longitudinal gradient of species diversity based on Hill numbers (^qD; $q = 0, 1, 2$) at approximately 100 km × 100 km grid cell level. Dashed and solid lines represent observed and standardised (at 40th percentile of sample coverage = 0.82) values, respectively. Loess (locally estimated scatterplot smoothing) curves (scaling parameter alpha = 0.6) are shown

Fig. S21.



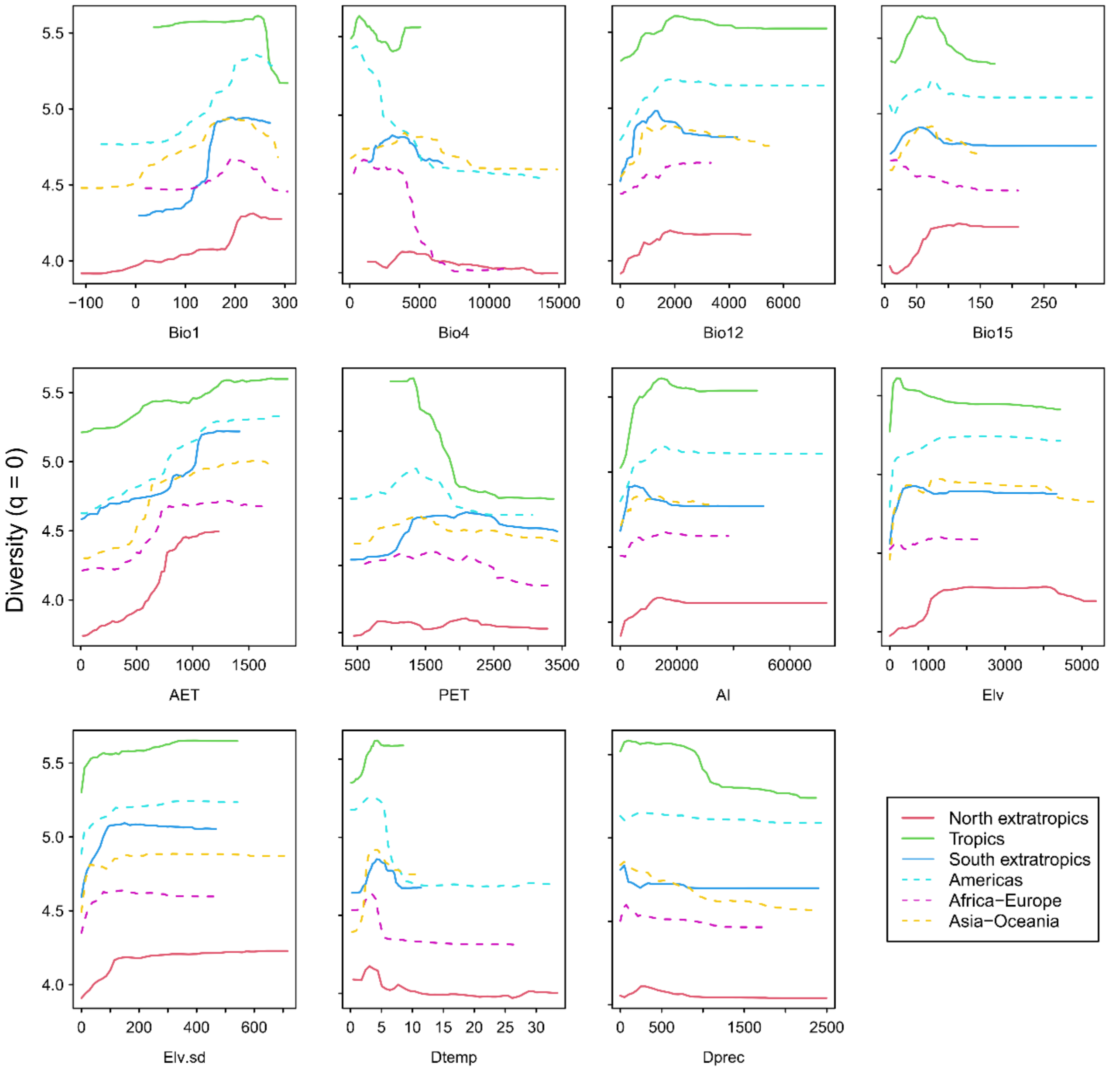
Comparison of partial dependency of explanatory variables in the Random Forest models explaining species diversity based on Hill numbers (${}^{q}D$; $q = 0, 1, 2$) at approximately 100 km × 100 km grid cell level. Observed diversity (Obs) and standardized (at 40th percentile of sample coverage = 0.82) species diversity (SC40) are shown. The environmental variables are mean annual temperature (Bio 1), temperature seasonality (Bio 4), annual precipitation (Bio 12), precipitation seasonality (Bio 15), actual evapotranspiration (AET), potential evapotranspiration (PET), aridity index (AI), average elevation (Elv), standard deviation of elevation (Elv.sd), and differences in temperature (D_{temp}) and precipitation (D_{prec}) between the Last Glacial Maximum and the present.

Fig. S22.



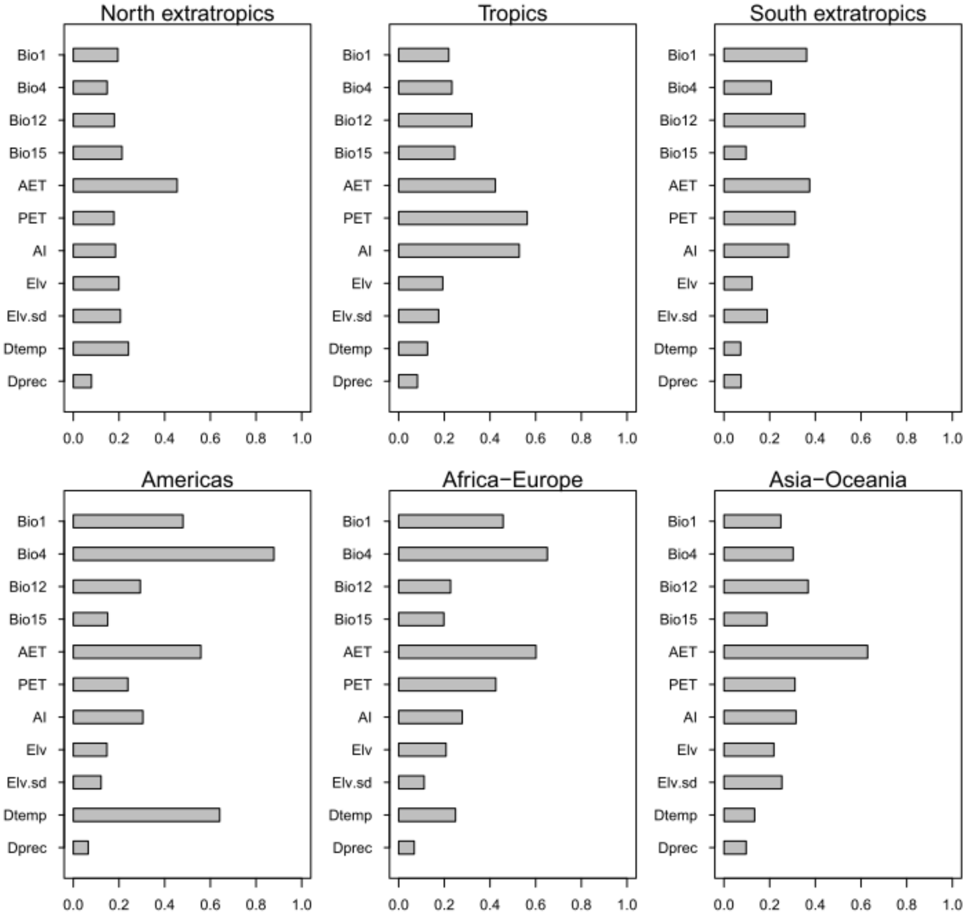
Comparison of relative importance (permutation importance) of explanatory variables in the random forest models explaining species diversity based on Hill numbers (^{q}D; $q = 0, 1, 2$) at approximately 100 km × 100 km grid cell level. Bars and points represent observed and standardised (at 40th percentile of sample coverage = 0.82) values, respectively. The environmental variables are mean annual temperature (Bio 1), temperature seasonality (Bio 4), annual precipitation (Bio 12), precipitation seasonality (Bio 15), actual evapotranspiration (AET), potential evapotranspiration (PET), aridity index (AI), average elevation (Elv), standard deviation of elevation (Elv.sd), and differences in temperature (D_{temp}) and precipitation (D_{prec}) between the Last Glacial Maximum and the present.

Fig. S23.



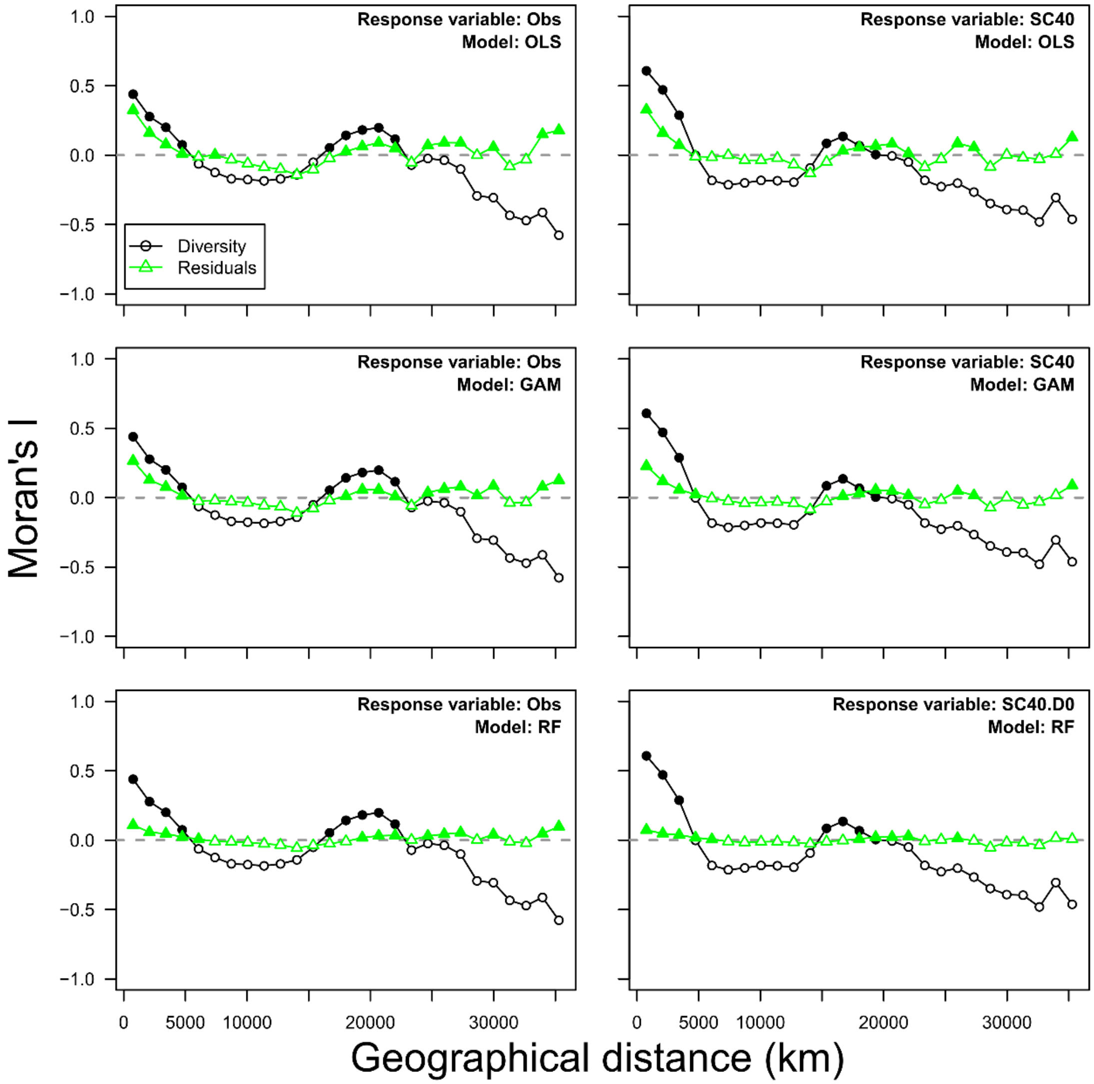
Comparison of partial dependency of explanatory variables in the random forest models explaining species richness among latitudinal (tropics and north and south extratropics) or latitudinal (Americas, Africa-Europe, and Asia-Oceania) zones. The species richness at approximately 100 km × 100 km grid cell was standardised at the 40th percentile of sample coverage (0.82). The environmental variables are mean annual temperature (Bio1), temperature seasonality (Bio4), annual precipitation (Bio12), precipitation seasonality (Bio15), actual evapotranspiration (AET), potential evapotranspiration (PET), aridity index (AI), average elevation (Elv), standard deviation of elevation (Elv_sd), and differences in temperature (D_{temp}) and precipitation (D_{prec}) between the Last Glacial Maximum and the present.

Fig. S24.



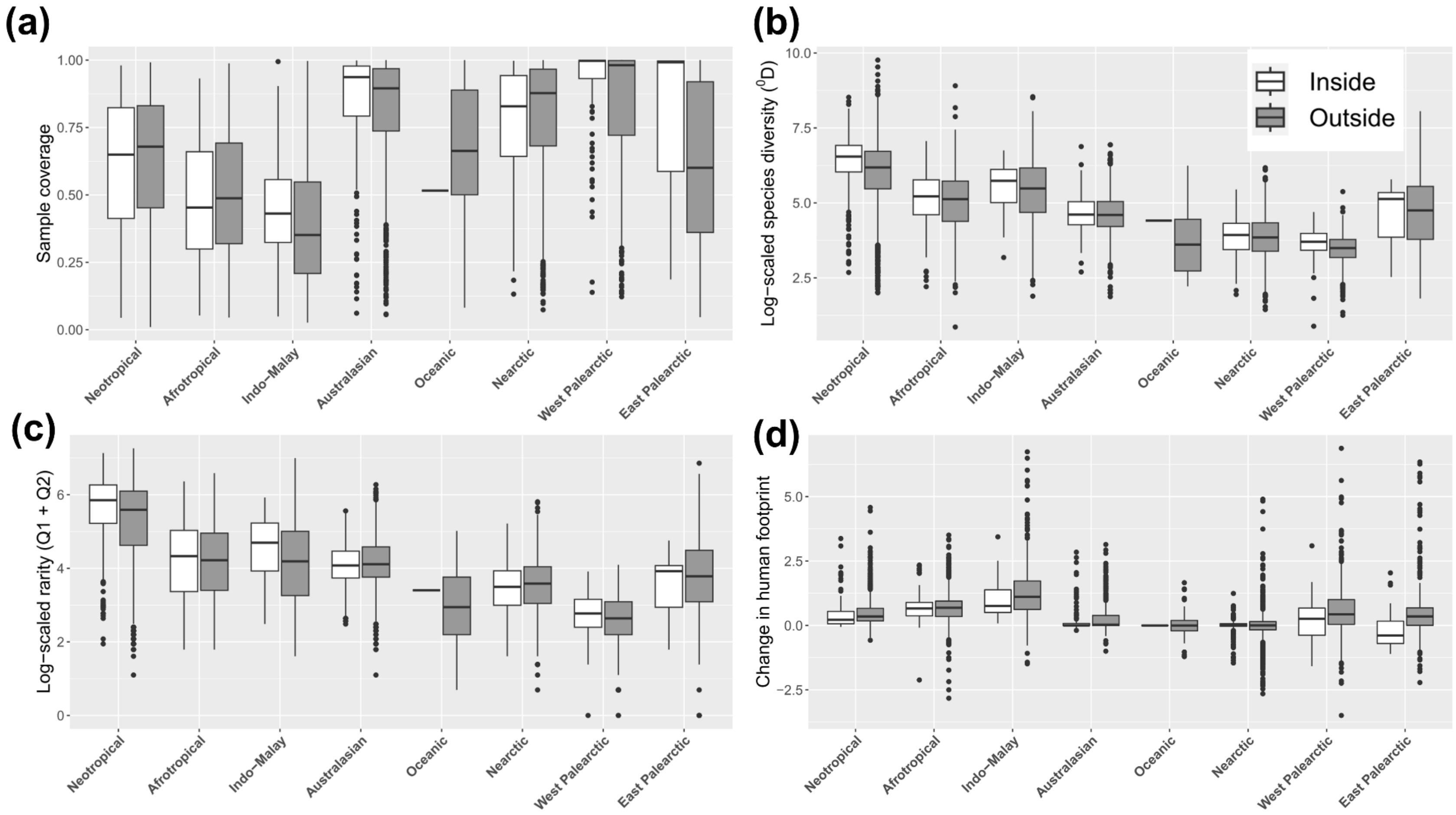
Comparison of relative importance of explanatory variables in the random forest models explaining species richness among latitudinal (tropical and north and south extratropics) or latitudinal (Americas, Africa-Europe, and Asia-Oceania) zones. The species richness at approximately 100 km × 100 km grid cell was standardised at the 40th percentile of sample coverage (0.82). The environmental variables are mean annual temperature (Bio1), temperature seasonality (Bio4), annual precipitation (Bio12), precipitation seasonality (Bio15), actual evapotranspiration (AET), potential evapotranspiration (PET), aridity index (AI), average elevation (Elv), standard deviation of elevation (Elv.sd), and differences in temperature (D_{temp}) and precipitation (D_{prec}) between the Last Glacial Maximum and the present.

Fig. S25



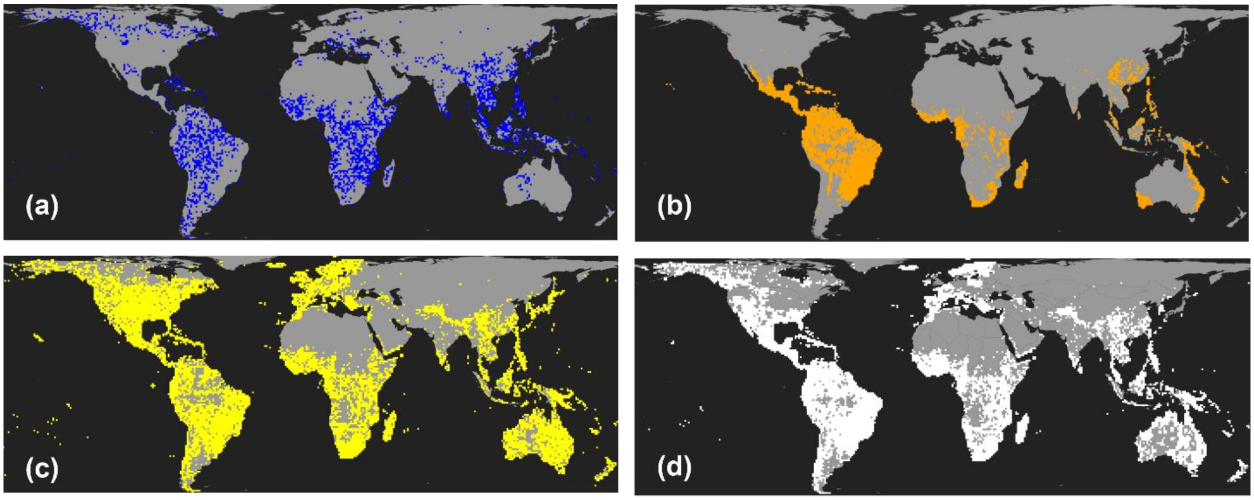
Spatial correlogram of species richness and of the residuals of regression models analyzed at 100 km × 100 km grid cell level: ordinary least squares (OLS), generalized additive model (GAM), Random Forest model (RF). The observed (Obs) and the sample coverage-based standardized (SC40) species richness (sample coverage = 0.82) are shown.

Fig. S26



Comparison of grid-cell properties associated with woody angiosperm diversity between inside and outside the existing protected areas (PA): (a) sample coverage, (b) log-scaled standardized species richness, (c) rarity defined as the number of unique and duplicated species, and (d) change in the Human Footprint between 2000 and 2018.

Fig. S27



Global map of binarized variables used to determine spatial priority areas for improving sampling completeness of species occurrence records of woody angiosperms: a) low sampling coverage areas (lower than the 30th percentile; blue); b) high species rarity areas (upper than the 70th percentile; orange); c) unprotected areas (yellow); d) areas experiencing escalating human pressure between 2000-2018 (white).

Table S1.

Overview of literature sources used to compile the woody angiosperm data

Type	**Sources**
Country flora list	Refs. 115-211
Botanical literature	Refs. 119, 124, 130, 131, 148, 171, 176, 202, 213-237
Species occurrence records	Refs. 145, 238-242; GBIF*

* See table S3 for doi for the occurrence data downloads

Table S2 (separate file)

Global woody angiosperm species list

Table S3 (separate file)

Digital object identifiers (doi) for the download of species occurrence data from GBIF

Table S4

Sample coverage (SC) values at 1st, 5th, 10th, 20th, 30th, 40th and 50th percentiles at different spatial resolution

Spatial resolution	SC1	SC5	SC10	SC20	SC30	SC40	SC50
ca 100 km × 100 km	0.142	0.313	0.431	0.596	0.729	0.820	0.889
ca 200 km × 200 km	0.227	0.449	0.579	0.734	0.836	0.905	0.944
ca 400 km × 400 km	0.336	0.559	0.697	0.829	0.902	0.95	0.972
ca 800 km × 800 km	0.470	0.69	0.799	0.896	0.951	0.972	0.986

Table S5

Correlations among environmental variables and their variance inflation factors (VIF) used in environmental driver analysis

	Bio1	Bio4	Bio12	Bio15	AET	PET	AI	Elv	Elv.sd	D_{temp}	D_{prec}
Bio1											
Bio4	-0.80										
Bio12	0.37	-0.54									
Bio15	0.33	-0.23	-0.23								
AET	0.51	-0.63	0.86	-0.18							
PET	0.68	-0.32	-0.23	0.47	-0.17						
AI	-0.02	-0.27	0.84	-0.41	0.64	-0.55					
Elv	-0.33	0.12	-0.23	0.23	-0.24	-0.03	-0.22				
Elv.sd	-0.32	0.05	0.03	0.07	-0.09	-0.22	0.09	0.69			
D_{temp}	-0.68	0.73	-0.29	-0.33	-0.36	-0.48	0.01	-0.06	-0.12		
D_{prec}	-0.02	-0.13	0.54	-0.17	0.32	-0.25	0.53	-0.06	0.16	0.04	
VIF	20.88	6.78	13.10	1.64	9.15	9.15	7.56	3.32	2.59	3.08	1.73

Bio1: mean annual temperature; Bio 4: temperature seasonality, Bio12: annual precipitation; Bio15: precipitation seasonality; AET: actual evapotranspiration; PET: potential evapotranspiration; AI: aridity index (AI); Elv: average elevation; Elv.sd: standard deviation of elevation, D_{temp}: differences in temperature between the Last Glacial Maximum and the present; D_{prec} differences in precipitation between the Last Glacial Maximum and the present.

REFERENCES AND NOTES

1. J. Soberón, T. Peterson, Biodiversity informatics: Managing and applying primary biodiversity data. *Philos. Trans. R. Soc. London. Ser. B Biol. Sci.* **359**, 689–698 (2004).

2. J. M. Serra-Diaz, B. J. Enquist, B. Maitner, C. Merow, J.-C. Svenning, Big data of tree species distributions: How big and how good? *For. Ecosyst.* **4**, 30 (2017).

3. A. Guisan, W. Thuiller, N. E. Zimmermann, *Habitat Suitability and Distribution Models: With Applications in R* (Cambridge Univ. Press, 2017).

4. R. Cazzolla Gatti, P. B. Reich, J. G. P. Gamarra, T. Crowther, C. Hui, A. Morera, J.-F. Bastin, S. De-Miguel, G.-J. Nabuurs, J.-C. Svenning, J. M. Serra-Diaz, C. Merow, B. Enquist, M. Kamenetsky, J. Lee, J. Zhu, J. Fang, D. F. Jacobs, B. Pijanowski, A. Banerjee, R. A. Giaquinto, G. Alberti, A. M. Almeyda Zambrano, E. Alvarez-Davila, A. Araujo-Murakami, V. Avitabile, G. A. Aymard, R. Balazy, C. Baraloto, J. G. Barroso, M. L. Bastian, P. Birnbaum, R. Bitariho, J. Bogaert, F. Bongers, O. Bouriaud, P. H. S. Brancalion, F. Q. Brearley, E. N. Broadbent, F. Bussotti, W. Castro da Silva, R. G. César, G. Češljar, *V.* Chama Moscoso, H. Y. H. Chen, E. Cienciala, C. J. Clark, D. A. Coomes, S. Dayanandan, M. Decuyper, L. E. Dee, J. Del Aguila Pasquel, G. Derroire, M. N. K. Djuikouo, T. Van Do, J. Dolezal, I. Đ. Đorđević, J. Engel, T. M. Fayle, T. R. Feldpausch, J. K. Fridman, D. J. Harris, A. Hemp, G. Hengeveld, B. Herault, M. Herold, T. Ibanez, A. M. Jagodzinski, B. Jaroszewicz, K. J. Jeffery, V. K. Johannsen, T. Jucker, A. Kangur, V. N. Karminov, K. Kartawinata, D. K. Kennard, S. Kepfer-Rojas, G. Keppel, M. L. Khan, P. K. Khare, T. J. Kileen, H. S. Kim, H. Korjus, A. Kumar, A. Kumar, D. Laarmann, N. Labrière, M. Lang, S. L. Lewis, N. Lukina, B. S. Maitner, Y. Malhi, A. R. Marshall, O. V. Martynenko, A. L. Monteagudo Mendoza, P. V. Ontikov, E. Ortiz-Malavasi, N. C. Pallqui Camacho, A. Paquette, M. Park, N. Parthasarathy, P. L. Peri, P. Petronelli, S. Pfautsch, O. L. Phillips, N. Picard, D. Piotto, L. Poorter, J. R. Poulsen, H. Pretzsch, H. Ramírez-Angulo, Z. Restrepo Correa, M. Rodeghiero, R. D. P. Rojas Gonzáles, S. G. Rolim, F. Rovero, E. Rutishauser, P. Saikia, C. Salas-Eljatib, D. Schepaschenko, M. Scherer-Lorenzen, V. Šebeň, M. Silveira, F. Slik, B. Sonké, A. F. Souza, K. J. Stereńczak, M. Svoboda, H. Taedoumg, N. Tchebakova, J. Terborgh, E. Tikhonova, A. Torres-Lezama, F. van der Plas, R. Vásquez, H. Viana, A. C. Vibrans, E. Vilanova, V. A. Vos, H.-F. Wang, B. Westerlund, L. J. T. White, S. K. Wiser, T. Zawiła-Niedźwiecki, L. Zemagho, Z.-X. Zhu, I. C. Zo-Bi, J. Liang, The number of tree species on Earth. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2115329119 (2022).

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

6. A. Chao, L. Jost, Coverage-based rarefaction and extrapolation: Standardizing samples by completeness rather than size. *Ecology* **93**, 2533–2547 (2012).

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

8. W. Jetz, M. A. McGeoch, R. Guralnick, S. Ferrier, J. Beck, M. J. Costello, M. Fernandez, G. N. Geller, P. Keil, C. Merow, C. Meyer, F. E. Muller-Karger, H. M. Pereira, E. C. Regan, D. S. Schmeller, E. Turak, Essential biodiversity variables for mapping and monitoring species populations. *Nat. Ecol. Evol.* **3**, 539–551 (2019).

9. P. Keil, A. M. Wilson, W. Jetz, Uncertainty, priors, autocorrelation and disparate data in downscaling of species distributions. *Divers. Distrib.* **20**, 797–812 (2014).

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

11. E. H. Boakes, P. J. K. McGowan, R. A. Fuller, D. Chang-qing, N. E. Clark, K. O’Connor, G. M. Mace, Distorted views of biodiversity: Spatial and temporal bias in species occurrence data. *PLOS Biol.* **8**, e1000385 (2010).

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

13. J. Grand, M. P. Cummings, T. G. Rebelo, T. H. Ricketts, M. C. Neel, Biased data reduce efficiency and effectiveness of conservation reserve networks. *Ecol. Lett.* **10**, 364–374 (2007).

14. A. Chao, Y. Kubota, D. Zelený, C.-H. Chiu, C.-F. Li, B. Kusumoto, M. Yasuhara, S. Thorn, C.-L. Wei, M. J. Costello, R. K. Colwell, Quantifying sample completeness and comparing diversities among assemblages. *Ecol. Res.* **35**, 292–314 (2020).

15. A. Chao, N. J. Gotelli, T. C. Hsieh, E. L. Sander, K. H. Ma, R. K. Colwell, A. M. Ellison, Rarefaction and extrapolation with Hill numbers: A framework for sampling and estimation in species diversity studies. *Ecol Monogr.* **84**, 45–67 (2014).

16. C. Chaudhary, H. Saeedi, M. J. Costello, Bimodality of latitudinal gradients in marine species richness. *Trends Ecol. Evol.* **31**, 670–676 (2016).

17. B. Kusumoto, M. J. Costello, Y. Kubota, T. Shiono, C. Wei, M. Yasuhara, A. Chao, Global distribution of coral diversity: Biodiversity knowledge gradients related to spatial resolution. *Ecol. Res.* **35**, 315–326 (2020).

18. A. Menegotto, T. F. Rangel, Mapping knowledge gaps in marine diversity reveals a latitudinal gradient of missing species richness. *Nat. Commun.* **9**, 4713 (2018).

19. CBD Secretariat, Final text of Kunming-Montreal Global Biodiversity Framework; published 22 December 2022; https://prod.drupal.www.infra.cbd.int/sites/default/files/2022-12/221222-CBD-PressRelease-COP15-Final.pdf.

20. T. Shiono, Y. Kubota, B. Kusumoto, Area-based conservation planning in Japan: The importance of OECMs in the post-2020 Global Biodiversity Framework. *Global Ecol. Conserv.* **30**, e01783 (2021).

21. R. T. Corlett, Plant diversity in a changing world: Status, trends, and conservation needs. *Plant Divers.* **38**, 10–16 (2016).

22. W. Barthlott, A. Hostert, G. Kier, W. Kueper, H. Kreft, J. Mutke, M. D. Rafiqpoor, J. H. Sommer, Geographic patterns of vascular plant diversity at continental to global scales. *Erdkunde* **61**, 305–315 (2007).

23. H. Kreft, W. Jetz, Global patterns and determinants of vascular plant diversity. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 5925–5930 (2007).

24. P. Keil, J. M. Chase, Global patterns and drivers of tree diversity integrated across a continuum of spatial grains. *Nat Ecol Evol.* **3**, 390–399 (2019).

25. M. O. Hill, Diversity and evenness: A unifying notation and its consequences. *Ecology* **54**, 427–432 (1973).

26. W. Yang, K. Ma, H. Kreft, Geographical sampling bias in a large distributional database and its effects on species richness-environment models. *J. Biogeogr.* **40**, 1415–1426 (2013).

27. A. Zizka, O. Rydén, D. Edler, J. Klein, A. Perrigo, D. Silvestro, S. C. Jagers, S. I. Lindberg, A. Antonelli, Bio-Dem, a tool to explore the relationship between biodiversity data availability and socio-political conditions in time and space. *J. Biogeogr.* **48**, 2715–2726 (2021).

28. D. M. C. C. Alves, A. A. Eduardo, E. V. da Silva Oliveira, F. Villalobos, R. Dobrovolski, T. C. Pereira, A. de Souza Ribeiro, J. Stropp, J. F. M. Rodrigues, J. A. F. Diniz-Filho, S. F. Gouveia, Unveiling geographical gradients of species richness from scant occurrence data. *Glob. Ecol. Biogeogr.* **29**, 748–759 (2020).

29. R. K. Colwell, J. A. Coddington, Estimating terrestrial biodiversity through extrapolation. *Philos. Trans. R Soc. Lond. B Biol. Sci.* **345**, 101–118 (1994).

30. M. S. Sousa-Baena, L. C. Garcia, A. T. Peterson, Completeness of digital accessible knowledge of the plants of Brazil and priorities for survey and inventory. *Divers. Distrib.* **20**, 369–381 (2014).

31. R. Field, E. M. O'Brien, R. J. Whittaker, Global models for predicting woody plant richness from climate: Development and evaluation. *Ecology* **86**, 2263–2277 (2005).

32. E. M. O'Brien, R. Field, R. J. Whittaker, Climatic gradients in woody plant (tree and shrub) diversity: Water-energy dynamics, residual variation, and topography. *Oikos* **89**, 588–600 (2000).

33. A. P. Francis, D. J. Currie, A globally consistent richness-climate relationship for angiosperms. *Am. Nat.* **161**, 523–536 (2003).

34. N. L. Kinlock, L. Prowant, E. M. Herstoff, C. M. Foley, M. Akin-Fajiye, N. Bender, M. Umarani, H. Y. Ryu, B. Şen, J. Gurevitch, J. Pither, Explaining global variation in the latitudinal diversity gradient: Meta-analysis confirms known patterns and uncovers new ones. *Glob. Ecol. Biogeogr.* **27**, 125–141 (2018).

35. J. F. Colville, C. M. Beale, F. Forest, R. Altwegg, B. Huntley, R. M. Cowling, Plant richness, turnover, and evolutionary diversity track gradients of stability and ecological opportunity in a megadiversity center. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 20027–20037 (2020).

36. J. Mutke, W. Barthlott, Patterns of vascular plant diversity at continental to global scales. *Biol. Skr.* **55**, 305–315 (2005).

37. R. E. Ricklefs, H. Qian, P. S. White, The region effect on mesoscale plant species richness between eastern Asia and eastern North America. *Ecography (Cop.).* **27**, 129–136 (2004).

38. T. L. P. Couvreur, Odd man out: Why are there fewer plant species in African rain forests? *Plant Syst. Evol.* **301**, 1299–1313 (2015).

39. H. Qian, R. E. Ricklefs, Large-scale processes and the Asian bias in species diversity of temperate plants. *Nature* **407**, 180–182 (2000).

40. A. H. Gentry, Changes in plant community diversity and floristic composition on environmental and geographical gradients. *Ann. Missouri Bot. Gard.* **75**, 1–34 (1988).

41. Y. Kubota, B. Kusumoto, T. Shiono, W. Ulrich, Environmental filters shaping angiosperm tree assembly along climatic and geographic gradients. *J. Veg. Sci.* **29**, 607–618 (2018).

42. G. Feng, Z. Ma, B. Sandel, L. Mao, S. Normand, A. Ordonez, J. Svenning, Species and phylogenetic endemism in angiosperm trees across the Northern Hemisphere are jointly shaped by modern climate and glacial–interglacial climate change. *Glob. Ecol. Biogeogr.* **28**, 1393–1402 (2019).

43. T. Shiono, B. Kusumoto, M. Yasuhara, Y. Kubota, Roles of climate niche conservatism and range dynamics in woody plant diversity patterns through the Cenozoic. *Glob. Ecol. Biogeogr.* **27**, 865–874 (2018).

44. F.M. Sabatini, B. Jiménez-Alfaro, U. Jandt, M. Chytrý, R. Field, M. Kessler, J. Lenoir, F. Schrodt, S. K. Wiser, M. A. S. Arfin Khan, F. Attorre, L. Cayuela, M. De Sanctis, J. Dengler, S. Haider, M. Z. Hatim, A. Indreica, F. Jansen, A. Pauchard, R. K. Peet, P. Petřík, V. D. Pillar, B. Sandel, M. Schmidt, Z. Tang, P. van Bodegom, K. Vassilev, C. Violle, E. Alvarez-Davila, P. Davidar, J. Dolezal, B. Hérault, A.

Galán-de-Mera, J. Jiménez, S. Kambach, S. Kepfer-Rojas, H. Kreft, F. Lezama, R. Linares-Palomino, A. M. Mendoza, J. K. N'Dja, O. L. Phillips, G. Rivas-Torres, P. Sklenář, K. Speziale, B. J. Strohbach, R. V. Martínez, H.-F. Wang, K. Wesche, H. Bruelheide, Global patterns of vascular plant alpha diversity. *Nature Comm.* **13**, 4683 (2022).

45. N. V Page, K. Shanker, Climatic stability drives latitudinal trends in range size and richness of woody plants in the Western Ghats, India, *PLOS ONE* **15**, e0235733 (2020).

46. P. V. A. Fine, Ecological and evolutionary drivers of geographic variation in species diversity, *Ann. Rev. Ecol. Evol. Syst.* **46**, 369–392 (2015).

47. B. J. Seliger, B. J. McGill, J. Svenning, J. L. Gill, Widespread underfilling of the potential ranges of North American trees. *J. Biogeogr.* **48**,359–371 (2020).

48. D. Montoya, M. A. Rodríguez, M. A. Zavala, B. A. Hawkins, Contemporary richness of holarctic trees and the historical pattern of glacial retreat. *Ecography (Cop.).* **30**, 173–182 (2007).

49. B. Sandel, L. Arge, B. Dalsgaard, R. Davies, K. Gaston, W. Sutherland, J.-C. Svenning, The influence of Late Quaternary climate-change velocity on species endemism. *Science* **334**, 660–664 (2011).

50. R. I. Milne, R. J. Abbott, The origin and evolution of tertiary relict floras. *Advance. Bot. Res.* **38**, 281–314 (2002).

51. M. Berdugo, M. Delgado-Baquerizo, S. Soliveres, R. Hernández-Clemente, Y. Zhao, J. J. Gaitán, N. Gross, H. Saiz, V. Maire, A. Lehmann, M. C. Rillig, R. V Solé, F. T. Maestre, Global ecosystem thresholds driven by aridity. *Science* **367**, 787–790 (2020).

52. S. A. Ford, M. R. Jepsen, N. Kingston, E. Lewis, T. M. Brooks, B. MacSharry, O. Mertz, Deforestation leakage undermines conservation value of tropical and subtropical forest protected areas. *Glob. Ecol. Biogeogr.* **29**, 2014–2024 (2020).

53. W. Li, W.-Y. Guo, M. Pasgaard, Z. Niu, L. Wang, F. Chen, Y. Qin, J.-C. Svenning, Human fingerprint on structural density of forests globally. *Nat Sust.*, **6**, 368–379(2023).

54. M. M. A. Boehm, Q. C. B. Cronk, Dark extinction: The problem of unknown historical extinctions. *Biol. Lett.* **17**, 20210007 (2021).

55. G. T. Prance, Floristic Inventory of the tropics: Where do we stand? *Ann. Missouri Bot. Gard.* **64**, 659 (1977).

56. C. Meyer, H. Kreft, R. Guralnick, W. Jetz, Global priorities for an effective information basis of biodiversity distributions. *Nat. Commun.* **6**, 8221 (2015).

57. R. A. Mittermeier, W. R. Turner, F. W. Larsen, T. M. Brooks, C. Gascon, *in Biodiversity Hotspots* (Springer Berlin, 2011), pp. 3–22.

58. A. H. Ariño, V. Chavan, N. King, The biodiversity informatics potential index. *BMC Bioinformatics* **12**, S4 (2011).

59. I. Jiménez, R. E. Ricklefs, Diversity anomalies and spatial climate heterogeneity. *Glob. Ecol. Biogeogr.* **23**, 988–999 (2014).

60. M. Yasuhara, C.-L. Wei, M. Kucera, M. J. Costello, D. P. Tittensor, W. Kiessling, T. C. Bonebrake, C. R. Tabor, R. Feng, A. Baselga, K. Kretschmer, B. Kusumoto, Y. Kubota, Past and future decline of tropical pelagic biodiversity. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 12891, 12896(2020).

61. A. Zacaï, C. Monnet, A. Pohl, G. Beaugrand, G. Mullins, D. M. Kroeck, T. Servais, Truncated bimodal latitudinal diversity gradient in early Paleozoic phytoplankton. *Sci. Adv.* **7**, eabd6709 (2021).

62. M. J. Donoghue, A phylogenetic perspective on the distribution of plant diversity. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 11549–11555 (2008).

63. F. Grattarola, J. A. Martínez-Lanfranco, G. Botto, D. E. Naya, R. Maneyro, P. Mai, D. Hernández, G. Laufer, L. Ziegler, E. M. González, I. da Rosa, N. Gobel, A. González, J. González, A. L. Rodales, D. Pincheira-Donoso, Multiple forms of hotspots of tetrapod biodiversity and the challenges of open-access data scarcity. *Sci. Rep.* **10**, 22045 (2020).

64. H. Gaisberger, T. Fremout, C. J. Kettle, B. Vinceti, D. Kemalasari, T. Kanchanarak, E. Thomas, J. M. Serra-Diaz, J.-C. Svenning, F. Slik, W. Eiadthong, K. Palanisamy, G. Ravikanth, V. Bodos, J. Sang, R. R. Warrier, A. K. S. Wee, C. Elloran, L. T. Ramos, M. Henry, M. A. Hossain, I. Theilade, S. Laegaard, K. M. A. Bandara, D. P. Weerasinghe, S. Changtragoon, V. Yuskianti, P. Wilkie, N. H. Nghia, S. Elliott, G. Pakkad, P. Tiansawat, C. Maycock, C. Bounithiphonh, R. Mohamed, M. Nazre, B. N. Siddiqui, S.-L. Lee, C.-T. Lee, N. F. Zakaria, I. Hartvig, L. Lehmann, D. B. D. David, J.-P. B. Lillesø, C. Phourin, Z. Yongqi, H. Ping, H. A. Volkaert, L. Graudal, A. Hamidi, S. Thea, S. Sreng, D. Boshier, E. Tolentino Jr., W. Ratnam, M. M. Aung, M. Galante, S. F. M. Isa, N. Q. Dung, T. T. Hoa, T. C. Le, M. D. Miah, A. L. M. Zuhry, D. Alawathugoda, A. Azman, G. Pushpakumara, N. Sumedi, I. Z. Siregar, H. K. Nak, J. Linsky, M. Barstow, L. P. Koh, R. Jalonen, Tropical and subtropical Asia’s valued tree species under threat. *Conserv. Biol.* **36**, e13873 (2022).

65. A. Chao, S.-M. Lee, Estimating the number of classes via sample coverage. *J. Am. Stat. Assoc.* **87**, 210–217 (1992).

66. D. J. Currie, G. G. Mittelbach, H. V. Cornell, R. Field, J.-F. Guégan, B. A. Hawkins, D. M. Kaufman, J. T. Kerr, T. Oberdorff, E. O'Brien, J. R. G. Turner, Predictions and tests of climate-based hypotheses of broad-scale variation in taxonomic richness. *Ecol. Lett.* **7**, 1121–1134 (2004).

67. F. I. Woodward, M. R. Lomas, C. K. Kelly, Global climate and the distribution of plant biomes. *Phil. Trans. R. Soc. B.* **359**, 1465–1476 (2004).

68. M. B. Araújo, D. Nogués-Bravo, J. A. F. Diniz-Filho, A. M. Haywood, P. J. Valdes, C. Rahbek, Quaternary climate changes explain diversity among reptiles and amphibians. *Ecography* **31**, 8–15 (2008).

69. A. Stein, K. Gerstner, H. Kreft, Environmental heterogeneity as a universal driver of species richness across taxa, biomes and spatial scales. *Ecol. Lett.* **17**, 866–880 (2014).

70. S. E. Fick, R. J. Hijmans, WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.* **37**, 4302–4315 (2017).

71. A. Trabucco, R. J. Zomer, Global high-resolution soil-water balance (2019); doi: 10.6084/m9.figshare.7707605.v3.

72. R. J. Zomer, A. Trabucco, D. A. Bossio, L. V Verchot, Climate change mitigation: A spatial analysis of global land suitability for clean development mechanism afforestation and reforestation. *Agric. Ecosyst. Environ.* **126**, 67–80 (2008).

73. A. Chao, R. K. Colwell, C.-H. Chiu, D. Townsend, Seen once or more than once: Applying Good–Turing theory to estimate species richness using only unique observations and a species list. *Method. Ecol. Evol.* **8**, 1221–1232 (2017).

74. R. K. Colwell, A. Chao, N. J. Gotelli, S.-Y. Lin, C. X. Mao, R. L. Chazdon, J. T. Longino, Models and estimators linking individual-based and sample-based rarefaction, extrapolation and comparison of assemblages. *J. Plant Ecol.* **5**, 3–21 (2012).

75. S. N. Wood, Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *J. R. Stat. Soc. B.* **73**, 3–36 (2011).

76. L. Breiman,Random forest. *Machine Learn.* **45**, 5–32 (2001).

77. C. F. Dormann, J. Elith, S. Bacher, C. Buchmann, G. Carl, G. Carré, J. R. G. Marquéz, B. Gruber, B. Lafourcade, P. J. Leitão, T. Münkemüller, C. McClean, P. E. Osborne, B. Reineking, B. Schröder, A. K. Skidmore, D. Zurell, S. Lautenbach, Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography* **36**, 27–46 (2013).

78. P. A. P. Moran, Notes on continuous stochastic phenomena. *Biometrika* **37**, 17–23 (1950).

79. H. Mu, X. Li, Y. Wen, J. Huang, P. Du, W. Su, S. Miao, M. Geng, A global record of annual terrestrial Human Footprint dataset from 2000 to 2018. *Sci. Data* **9** 176 (2022).

80. M. A. Babyak, What you see may not be what you get: A brief, nontechnical introduction to overfitting in regression-type models. *Psychosom. Med.* **66**, 411–421 (2004)

81. R Core Team, R: A language and environment for statistical computing (2021); www.r-project.org/.

82. M. Dowle, A. Srinivasan, Data.table: Extension of 'data.frame'. R package version 1.14.8 (2023); https://CRAN.R-project.org/package=data.table.

83. H. Wickham, R. Francois, L. Henry, K. Muller, D. Vaughan, dplyr: A Grammar of Data Manipulation. R package version 1.1.2 (2023); https://CRAN.R-project.org/package=dplyr.

84. H. Wickham, D. Vaughan, M. Girlich, tidyr: Tidy Messy Data. R package version 1.3.0 (2023); https://CRAN.R-project.org/package=tidyr.

85. Microsoft Corporation, S. Weston, doParallel: Foreach Parallel Adaptor for the 'parallel' Package. R package version 1.0.17 (2022); https://CRAN.R-project.org/package=doParallel.

86. Microsoft Corporation, S. Weston, foreach: Provides Foreach Looping Construct. R package version 1.5.2 (2022); https://CRAN.R-project.org/package=foreach.

87. R. J. Hijmans, M. Barbosa, A. Ghosh, A. Mandel, geodata: Download Geographic Data. R package version 0.5-8 (2023); https://CRAN.R-project.org/package=geodata.

88. Original S code by R. A. Becker, A. R. Wilks. R version by R. Brownrigg. Enhancements by T. P. Minka, A. Deckmyn, maps: Draw Geographical Maps. R package version 3.4.1 (2022); https://CRAN.R-project.org/package=maps.

89. R. Bivand, N. Lewin-Koh, maptools: Tools for handling spatial objects. R package version 1.1-6 (2022); https://CRAN.R-project.org/package=maptools.

90. R. J. Hijmans, raster: Geographic data analysis and modeling. R package version 3.6-20 (2023); https://CRAN.R-project.org/package=raster.

91. O. P. Lamigueiro, R. Hijmans, rasterVis. R package version 0.51.5 (2023); https://oscarperpinan.github.io/rastervis/.

92. R. Bivand, T. Keitt, B. Rowlingson, rgdal: Bindings for the 'Geospatial' data abstraction library. R package version 1.6-6 (2023); https://CRAN.R-project.org/package=rgdal.

93. R. Bivand, C. Rundel, rgeos: Interface to Geometry Engine - Open Source ('GEOS'). R package version 0.6-2 (2023); https://CRAN.R-project.org/package=rgeos.

94. E. Pebesma, Simple features for R: standardized support for spatial vector data. *R. J.* **10**, 439–446, (2018.

95. E. Pebesma, R. Bivand, *Spatial Data Science: With Applications in R* (Chapman and Hall/CRC, 2023).

96. R. J. Hijmans, terra: Spatial Data Analysis. R package version 1.7-23 (2023); https://CRAN.R-project.org/package=terra.

97. H. Wickham, *ggplot2: Elegant Graphics for Data Analysis* (Springer-Verlag, 2016).

98. T. Keitt, colorRamps: Builds Color Tables. R package version 2.3.1 (2022); https://CRAN.R-project.org/package=colorRamps.

99. K. Wright, pals: Color Palettes, Colormaps, and Tools to Evaluate Them. R package version 1.8 (2023); https://CRAN.R-project.org/package=pals.

100. E. Neuwirth, RColorBrewer: ColorBrewer Palettes. R package version 1.1-3 (2022); https://CRAN.R-project.org/package=RColorBrewer.

101. A. W. Bowman, A. Azzalini, R package 'sm': Nonparametric smoothing methods. R package version 2.2-5.7 (2021); www.stats.gla.ac.uk/~adrian/sm.

102. G. Snow, TeachingDemos: Demonstrations for Teaching and Learning. R package version 2.12 (2020); https://CRAN.R-project.org/package=TeachingDemos.

103. S. Chamberlain, K. Ram, V. Barve, Package ‘rgbif’ Interface to the Global “Biodiversity” Information Facility “API” (2016).

104. T. C. Hsieh, K. H. Ma, A. Chao, iNEXT: An R package for rarefaction and extrapolation of species diversity (Hill numbers). *Methods Ecol. Evol.* **7**, 1451–1456 (2016).

105. M. N. Wright, A. Ziegler, ranger: A fast implementation of random forests for high dimensional data in C++ and R. *J. Stat. Softw.* **77**, 1–17 (2017).

106. B. M. Greenwell, pdp: An R package for constructing partial dependence plots. *R. J.* **9**, 421–436 (2017).

107. J. Li, spm: Spatial Predictive Modeling. R package version 1.2.0 (2019); https://cran.r-project.org/package=spm.

108. S. V Levin, _plotbiomes: Plot Whittaker biomes with ggplot2_. R package version 0.0.0.9001 (2023); https://valentinitnelav.github.io/plotbiomes/.

109. J. Fox, S. Weisberg, *An R Companion to Applied Regression* (Sage Publications, ed. 3, 2019); https://socialsciences.mcmaster.ca/jfox/Books/Companion/.

110. P. Giraudoux, pgirmess: Spatial Analysis and Data Mining for Field Ecologists. R package version 2.0.2 (2023); https://CRAN.R-project.org/package=pgirmess.

111. R. Vallejos, F. Osorio, M. Bevilacqua, *Spatial Relationships Between Two Georeferenced Variables: With Applications in R* (Springer, 2020).

112. R. Vaidyanathan, Y. Xie, J. J. Allaire, J. Cheng, C. Sievert, K. Russell htmlwidgets: HTML Widgets for R. R package version 1.6.2 (2023); https://CRAN.R-project.org/package=htmlwidgets.

113. J. J. Allaire, C. Gandrud, K. Russell, C. J. Yetman, networkD3: D3 JavaScript Network Graphs from R. R package version 0.4 (2017); https://cran.r-project.org/web/packages/networkD3/index.html.

114. W. Chang, webshot: Take Screenshots of Web Pages. R package version 0.5.5 (2023); https://CRAN.R-project.org/package=webshot.

115. A.-W. A. Al-Khulaidi, *Flora of Yemen*. The Sustainable Natural Resource, Management Project (SNRMP II), EPA and UNDP, Republic of Yemen (2013).

116. Z. Al-Kuwari, Bahrain First National Report to the Convention on Biological Diversity (Public Commission for the Protection of Marine Resources, Environment and Wildlife, General Directorate for Environment and Wildlife Protection, 2006).

117. M. Arechavaleta, N. Zurita, M. C. Marrero, J. L. Martín, Lista preliminar de especies silvestres de Cabo Verde (hongos, plantas y animales terrestres) (Consejería de Medio Ambiente y Ordenación Territorial, Gobierno de Canarias, Santa Cruz de Tenerife, 2005).

118. J. Audru, J. Cesar, J. P. Lebrun, Les plantes vasculaires de la république de Djibouti: flore illustrée (CIRAD, Département d'Élevage et de Médicine Vétérinaire, Djibouti, 1994).

119. Australian Biological Resources Study, Flora of Australia Online (2011); www.awe.gov.au/science-research/abrs/online-resources/flora-of-australia-online.

120. BGCI, GlobalTreeSearch online database (2017); https://tools.bgci.org/global_tree_search.php.

121. M. L. Baker, M. F. Duretto, *A Census of the Vascular Plants of Tasmania* (Tasmanian Herbarium, Tasmanian Museum and Art Gallery, 2011).

122. M. J. Balick, M. H. Nee, D. E. Atha, *Checklist of the Vascular Plants of Belize, with Common Names and Uses* (The New York Botanical Garden Press, 2000).

123. P. Bamps, *Flore d'Afrique Centrale (Zaïre-Rwanda-Burundi): Répertoire des lieux de récolte* (Jardin Botanique National de Belgique, 1982).

124. H. J. Beentje, C. M. Whitehouse, *Flora of Tropical East Africa* (Routledge, 2017); https://doi.org/10.1201/9780203755761.

125. M. Belov, Chileflora (2005–2012); www.chileflora.com/.

126. W. G. Berendsohn, A. K. Gruber, J. A. Monterrosa Salomón, N. S. Cuscatlanica. Árboles nativos e introducidos de El Salvador. Parte 1: Angiospermae—Familias A a L. *Englera* **29**, 1–438 (2009).

127. W. G. Berendsohn, A. K. Gruber, J. A. Monterrosa Salomón, N. S. Cuscatlanica. Árboles nativos e introducidos de El Salvador. Parte 2: Angiospermae - Familias M a P y Pteridophyta. *Englera* **29**, 1–300 (2012).

128. R. Bernal, S. R. Gradstein, M. Celis, Catálogo de plantas y líquenes de Colombia (2016); http://catalogoplantasdecolombia.unal.edu.co/en/.

129. Bhutan Biodiversity Portal, Bhutan Biodiversity Portal (2013); https://biodiversity.bt/.

130. M. G. Bingham, A. Willemen, B. T. Wursten, P. Ballings, M. A. Hyde, Flora of Zambia (2018); www.zambiaflora.com.

131. J. K. Boggan, V. A. Funk, C. L. Kelloff, M. Hoff, G. Cremers, C. Feuillet, *Checklist of the Plants of the Guianas (Guyana, Surinam, French Guiana)* (National Museum of Natural History Smithsonian Institution, 1997).

132. M. Bou Dagher-Kharrat, Lebanon FLORA (2013); www.lebanon-flora.org/default.php.

133. L. Boulos, M. Al-Dosari, Checklist of the flora of Kuwait. *J. Univ. Kuwait* **21**, 203–217 (1994).

134. Brazil Flora Group, *Brazilian Flora 2020 project - Projeto Flora do Brasil 2020. Version 393* (Instituto de Pesquisas Jardim Botanico do Rio de Janeiro, 2021).

135. I. Breitwieser, P. J. Brownsey, W. A. Nelson, A. D. Wilton, Eds., *Flora of New Zealand Online* (2010); www.nzflora.info/.

136. G. Brundu, I. Camarda, The Flora of Chad: A checklist and brief analysis. *PhytoKeys* **23**, 1–18 (2013).

137. L. Catarino, E. S. Martins, M. F. Basto, M. A. Diniz, An annotated checklist of the vascular flora of Guinea-Bissau (West Africa). *Blumea-Biodiversity, Evol. Biogeogr. Plant.* **53**, 1–222 (2008).

138. C.-S. Chang, H. Kim, K. Chang, *Provisional Checklist of the Vascular Plants for the Korea Peninsular Flora (KPF)* (Designpost, 2014).

139. K. Chayamarit, R. Pooma, N. Pattharahirantricin, *A Checklist of Plants in Thailand*, volume 1 (Office of Natural Resources and Environmental Policy and Planning, 2014).

140. K. Y. Chong, H. T. Tan, R. T. Corlett, *A Checklist of the Total Vascular Plant Flora of Singapore: Native, Naturalised and Cultivated Species* (Raffles Museum of Biodiversity Research, 2009).

141. B. J. Conn, R. Banka, L. L. Lee, *Plants of Papua New Guinea* (2006); www.pngplants.org.

142. Conservatoire et Jardin botaniques, South African National Biodiversity Institute, African Plant Database (2012); www.ville-ge.ch/musinfo/bd/cjb/africa/recherche.php.

143. P. Craven, H. Kolberg, Plants of Namibia (BRAHMS online, 2018); https://herbaria.plants.ox.ac.uk/bol/namibia.

144. I. Darbyshire, M. Kordofani, I. Farag, R. Candiga, H. A. Pickering, Eds., *The Plants of Sudan and South Sudan: An Annotated Checklist* (Royal Botanic Gardens, 2015).

145. G. Dauby, R. Zaiss, A. Blach-Overgaard, L. Catarino, T. Damen, V. Deblauwe, S. Dessein, J. Dransfield, V. Droissart, M. C. Duarte, H. Engledow, G. Fadeur, R. Figueira, R. E. Gereau, O. J. Hardy, D. J. Harris, J. de Heij, S. Janssens, Y. Klomberg, A. C. Ley, B. A. MacKinder, P. Meerts, J. L. van de Poel, B. Sonké, M. S. M. Sosef, T. Stévart, P. Stoffelen, J.-C. Svenning, P. Sepulchre, X. van der Burgt, J. J. Wieringa, T. L. P. Couvreur, RAINBIO: A mega-database of tropical African vascular plants distributions. *PhytoKeys* **74**, 1–18 (2016).

146. A. Deomurari, M. Jani, K. Matieda, J. KaPatel, IBIS-Flora Beta Version—Angiosperm Flora of India (2014); https://flora.indianbiodiversity.org/.

147. Editorial Committee of the Flora of Taiwan, Ed., *Flora of Taiwan* (Department of Botany, National Taiwan University, Taipei, 1975).

148. Endemia, Faune et Flore de Nouvelle-Calédonie (2016); https://endemia.nc/.

149. Eswatini National Trust Commission, Eswatini’s Flora Database (2020); http://eswatininaturereserves.com/flora/index.asp.

150. Euro+Med, Euro+Med PlantBase: The information resource for Euro-Mediterranean plant diversity (2006); www.emplantbase.org/home.html.

151. E. Figueiredo, G. F. Smith, *Plants of Angola/Plantas de Angola* (South African National Biodiversity Institute, SANBI Publishing, 2008).

152. E. Figueiredo, J. Paiva, T. Stevart, F. Oliveira, G. F. Smith, Annotated catalogue of the flowering plants of São Tomé and Príncipe. *Bothalia* **41**, 41–82 (2011).

153. Flora do Brasil, Brazilian Flora 2020 (2016); http://floradobrasil.jbrj.gov.br/.

154. Flora of China Editorial Committee, *Flora of China* (Science Press, 1994).

155. A. Ghahreman, F. Attar, Biodiversity of Plant Species in Iran: The vegetation of Iran, plant species, red data of Iran, endemic species, rare species, species threatened by extinction (vol. 1) (Central Herbarium of Tehran University, Faculty of Science, Tehran, 1999).

156. J. F. Gillet, J. L. Doucet, A commented checklist of woody plants in the Northern Republic of Congo. *Plant. Ecol. Evol.* **145**, 258–271 (2012).

157. M. Hassler, “World plants: Synonymic checklists of the vascular plants of the world (version Nov 2018)” in *Species 2000 & ITIS Catalogue of Life*, Y. Roskov, L. Abucay, T. Orrell, D. Nicolson, N. Bailly, P. Kirk, T. Bourgoin, R. E. DeWalt, W. Decock, E. van Nieukerken, J. Zarucchi, L. Penev, Eds. (Species 2000: Naturalis, 2018), ISSN 2405-884X; www.catalogueoflife.org/annual-checklist/2019.

158. J. Hutchinson, J. M. Dalziel, *Flora of West Tropical Africa* (Crown Agents For Oversea Governments and Administrations, ed. 2, 1954).

159. M. A. Hyde, B. T. Wursten, P. Ballings, M. Coates Palgrave, Flora of Malawi (2016); www.malawiflora.com/index.php.

160. M. A. Hyde, B. T. Wursten, P. Ballings, M. Coates Palgrave, Flora of Mozambique (2016); www.mozambiqueflora.com/.

161. M. A. Hyde, B. T. Wursten, P. Ballings, M. Coates Palgrave, Flora of Zimbabwe (2018); www.zimbabweflora.co.zw/index.php.

162. J. Kellermann, Ed., *Flora of South Australia* (State Herbarium of South Australia, 2011).

163. A. H. Kitalong, R. DeMeo, T. Holm, Native Trees of Palau, A *Field Guide* (The Environment, Koror, 2008).

164. K. Kobisi, Preliminary checklist of the plants of Lesotho. *SABONET* **34**, 1–84 (2005).

165. G. A. Lazkov, B. A. Sultanova, *Checklist of Vascular Plants of Kyrgyzstan. Norrlinia*, vol. 24. (Botanical Museum, Finnish Museum of Natural History, 2011).

166. Missouri Botanical Garden, Catalogue of the vascular plants of Madagascar (2013); http://legacy.tropicos.org/Project/Madagascar.

167. Missouri Botanical Garden, Pakistan Plant Database (1981); http://legacy.tropicos.org/project/Pakistan.

168. Missouri Botanical Garden, Bolivia Catalogue (2017); http://legacy.tropicos.org/Project/BC.

169. Missouri Botanical Garden, Catalogue of the vascular plants of Ecuador (2017); http://legacy.tropicos.org/Project/CE.

170. Missouri Botanical Garden, Flora de Nicaragua (2017); http://legacy.tropicos.org/Project/FC.

171. Missouri Botanical Garden, Manual de Plantas de Costa Rica (2017); http://legacy.tropicos.org/Project/Costa%20Rica.

172. Missouri Botanical Garden, Panama Checklist (2022); http://legacy.tropicos.org/Project/PAC.

173. Missouri Botanical Garden, Peru Checklist (2022); http://legacy.tropicos.org/Project/PEC.

174. M. Modzelevich, Flowers in Israel (2005–2021); www.flowersinisrael.com/.

175. National Museum of Natural History (U.S.), *Botanical Exploration in Myanmar* (Smithsonian, National Museum of Natural History, 2003).

176. National Museum of Natural Science, Flora of Solomon Islands (2022); http://siflora.nmns.edu.tw/.

177. New Zealand Plant Conservation Network, New Zealand's Flora (2011); www.nzpcn.org.nz/flora/.

178. M. Newman, S. Ketphanh, B. Svengsuksa, P. Thomas, K. Sengdala, V. Lamxay, K. Armstrong, *A Checklist of the Vascular Plants of Lao PDR* (Royal Botanic Garden Edinburgh, 2007).

179. J. Norton, S. A. Majid, D. Allan, M. Al Safran, B. Böer, R. Richer, *An Illustrated Checklist of the Flora of Qatar* (Browndown Publications Gosport, 2009).

180. PVNH, The herbarium of Vanuatu (PVNH) (2015); http://publish.plantnet-project.org/project/vanuaflora_en.

181. P. B. Pelser, J. F. Barcelona, D. L. Nickrent, Co's digital flora of the Philippines (2011); www.philippineplants.org/.

182. PlantNET, New South Wales Flora Online (2017); http://plantnet.rbgsyd.nsw.gov.au/.

183. D. Podlech, *Checklist of the Flowering Plants of Afghanistan* (Ludwig-Maximilians Universität Publ, 2012).

184. J. R. Press, K. K. Shrestha, D. A. Sutton, *Annotated Checklist of the Flowering Plants of Nepal* (Natural History Museum Publications, 2000).

185. Queensland Government, Census of the Queensland flora 2014 (2014); https://data.qld.gov.au/dataset/census-of-the-queensland-flora-2014.

186. RBG Kew, World checklist of selected plant families (2016); http://wcsp.science.kew.org/home.do.

187. RMI Office of Environmental Planning and Policy Coordination (OEPPC), Native plants of the Marshalls (2008); https://web.archive.org/web/20150619131813/http://biormi.org/index.shtml?en%2Fnative_plants.shtml.

188. S. A. Robertson, *Flowering Plants of Seychelles* (Royal Botanic Gardens, 1989).

189. R. Rolland, V. Boullet, J.-P. Quod, Eds., *Mayotte: Biodiversité et Evaluation Patrimoniale* (DAF Mayotte et CBN Mascarin (coord.), 2005), pp. 115–187.

190. V. E. Selvam, *Trees and Shrubs of the Maldives* (RAP Publication, 2007).

191. L. K. Senaratne, *A Check List of the Flowering Plants of Sri Lanka* (National Science Foundation, 2001).

192. M. Setshogo, *Preliminary Checklist of the Plants of Botswana* (Southern African Botanical Diversity Network Report, 2005).

193. P. S. Short, D. E. Albrecht, I. D. Cowie, D. L. Lewis, B. M. Stuckey, *Checklist of the Vascular Plants of the Northern Territory* (Department of Natural Resources, Environment, The Arts and Sport, 2011).

194. K. U. Siddiqui, M. A. Islam, Z. U. Ahmed, S. M. H. Kabir, Eds., *Encyclopedia of flora and fauna of Bangladesh* (Asiatic Society of Bangladesh, 2007).

195. Smithsonian Institution, Flora of the West Indies (2007); https://naturalhistory2.si.edu/botany/WestIndies/.

196. South African National Biodiversity Institute, Botanical Database of Southern Africa (BODATSA) (2016); https://posa.sanbi.org/.

197. W. R. Sykes, *Contributions to the Flora of Niue* (New Zealand Department of Scientific and Industrial Research Bulletin 200, 1970), pp. 1–321.

198. R. R. Thaman, The Flora of Tuvalu: Lakau Mo Mouku o Tuvalu. *Atoll. Research Bulletin* **611**, 1–129 (2016).

199. R. R. Thaman, F. R. Fosberg, H. I. Manner, D. C. Hassall, The flora of Nauru. *Atoll Research Bulletin* **392**, 1–223 (1994).

200. J. Thomas, Plant diversity of Saudi Arabia: Flora checklist (2011); www.plantdiversityofsaudiarabia.info/Biodiversity-Saudi-Arabia/Flora/Checklist/Cheklist.htm.

201. M. Thulin, *Flora of Somalia* (Royal Botanic Gardens, 2006).

202. USDA, NRCS, The PLANTS database (2017); https://plants.usda.gov/home.

203. University of Greifswald, Virtual Guide to the Flora of Mongolia (2010); https://floragreif.uni-greifswald.de/floragreif/.

204. Vietnam Plant Data Center, Plant Database (2007–2018); www.botanyvn.com/?lg=en.

205. J. L. Villaseñor, Checklist of the native vascular plants of Mexico. *Revista Mexicana de Biodiversidad* **87**, 559–902 (2016).

206. W. L. Wagner, D. H. Lorence, Flora of the Marquesas Islands (2002–2018); https://naturalhistory2.si.edu/botany/marquesasflora/index.htm.

207. W. L. Wagner, D. R. Herbst, M. W. Tornabene, A. Weitzman, D. H. Lorence, Flora of Micronesia website (2012); http://botany.si.edu/pacificislandbiodiversity/micronesia/index.htm.

208. Western Australian Herbarium, FloraBase – the Western Australian Flora (2017); https://florabase.dpaw.wa.gov.au/.

209. K. Yonekura, T. Kajita, BG plants wamei-gakumei index (YList) (2003); http://ylist.info/.

210. M. Zohary, N. Feinbrun-Dothan, *Flora Palaestina* (Israel Academy of Sciences and Humanities, 1966).

211. F. O. Zuloaga, O. Morrone, M. Belgrano, Catálogo de las Plantas Vasculares del Cono Sur (2014); www.floraargentina.edu.ar/.

212. BioScripts, Flora vascular (2014); www.floravascular.com/

213. Botanical Research Institute of Texas & San Diego Zoo Global, Atrium Biodiversity Information System (2005–2017); https://atrium.andesamazon.org/index.php.

214. California Native Plant Society, Rare Plant Program, Inventory of Rare and Endangered Plants of California (online edition, v9-01 1.0) (2021); www.rareplants.cnps.org.

215. E. Castro, La selva florula digital website (2013); https://sura.ots.ac.cr/florula4/.

216. W. D. Clayton, M. S. Vorontsova, K. T. Harman, H. Williamson, GrassBase-the online world grass flora (2016); www.kew.org/data/grasses-db.html.

217. Dave’s Garden, Dave’s Garden (2017); https://davesgarden.com/..

218. Department of the Environment, Australia, Species profile and threats database (2016); www.environment.gov.au/cgi-bin/sprat/public/sprat.pl.

219. K. Fern, A. Fern, R. Morris, Useful tropical plants database (2017); http://tropical.theferns.info/.

220. Flora Malesiana, Flora Malesiana (2017); http://portal.cybertaxonomy.org/flora-malesiana/.

221. Flora of North America Editorial Committee, Eds., *Flora of North America North of Mexico*, 20+ vols (Oxford Univ. Press, 1993).

222. T. Girija, Flowers of India (2016); www.flowersofindia.net.

223. W. H. Harvey, O. W. Sonder, *Flora Capensis* (Hodges Smith & Co., 1865).

224. Hortipedia, Hortipedia (2015); https://en.hortipedia.com/Main_page.

225. IUCN, The IUCN Red List of Threatened Species, version 2017 (2017); www.iucnredlist.org.

226. K. Iwatsuki, T. Yamazaki, D. E. Boufford, H. Ohba, *Flora of Japan*, vols. 1–4 (Kodansha, 1993–2020).

227. Missouri Botanical Garden, Flora of Pakistan (2017); http://legacy.tropicos.org/Project/Pakistan.

228. D. Oliver, *Flora of Tropical Africa*, vols. 1–4) (Lovell Reeve and Co. Publishers, 1904).

229. Plant Lust, Plant Lust (2007); https://plantlust.com/.

230. RBG Kew, Seed information database (SID) (2016); http://data.kew.org/sid/. The webpage was retired in May 2022. The information will be migrated to several of RBG Kew's other databases; www.kew.org/about-us/reports-and-policies/website-accessibility-science.

231. J.P. Roux, *Flora of Southern Africa* (South African National Biodiversity Institute, 2003).

232. Smithsonian Institution's National Museum of Natural History, Encyclopedia of Life (2018); https://eol.org/.

233. South Australian Seed Conservation Centre, Seeds of South Australia, (2017); www.saseedbank.com.au/.

234. State Herbarium of South Australia, eFloraSA: Electronic flora of South Australia (2007); www.flora.sa.gov.au/census.shtml.

235. T. Vattakaven, R. George, D. Balasubramanian, M. Réjou-Méchain, G. Muthusankar, B. Ramesh, R. Prabhakar, India biodiversity portal: An integrated, interactive and participatory biodiversity informatics platform. *Biodivers. Data J.*, **4**, e10279 (2016).

236. Xycol database, Xycol (2021); https://xycol.net/.

237. CRIA (Centro de Referência em Informação Ambiental), SpeciesLink (2019); https://specieslink.net/.

238. Y. Kubota, B. Kusumoto, T. Shiono, W. Ulrich, F. Jabot, Non-neutrality in forest communities: Evolutionary and ecological determinants of tree species abundance distributions. *Oikos* **125**, 237–244 (2016).

239. Y. Kubota, T. Shiono, B. Kusumoto, Role of climate and geohistorical factors in driving plant richness patterns and endemicity on the east Asian continental islands. *Ecography* **38**, 639–648 (2015).

240. B. S. Maitner, B. Boyle, N. Casler, R. Condit, J. Donoghue, S. M. Durán, D. Guaderrama, C. E. Hinchliff, P. M. Jørgensen, N. J. B. Kraft, B. McGill, C. Merow, N. Morueta-Holme, R. K. Peet, B. Sandel, M. Schildhauer, S. A. Smith, J. C. Svenning, B. Thiers, C. Violle, S. Wiser, B. J. Enquist, Thebien rpackage: A tool to access the Botanical Information and Ecology Network (BIEN) database. *Method. Ecol. Evol.* **9**, 373–379 (2018).

241. D. L. Nickrent, M. Costea, J. F. Barcelona, P. B. Pelser, K. Nixon, PhytoImages (2006); www.phytoimages.siu.edu.