PNAS

Supplementary Information for
Global variation in diversification rate and species richness are unlinked in plants

Melanie Tietje*, Alexandre Antonelli, William J. Baker, Rafaël Govaerts, Stephen A. Smith, Wolf L. Eiserhardt*

Melanie Tietje
Email: melanie.tietje@bio.au.dk
Wolf L. Eiserhardt
Email: wolf.eiserhardt@bio.au.dk

This PDF file includes:

Supplementary Information Text

Taxonomy matching. We use the World Checklist of Vascular Plants (WCVP, (1) as a reference species list to ensure consistent taxonomic classification across different data sets.

Matching phylogeny tip labels with WCVP taxa was done using exact string matching of the taxonomic information in subsequent steps. In contrast to the WCVP, the phylogeny tip labels do not provide all required taxonomic information, hence this needs to be obtained first from a third source. Tip label names were based on either Open Tree of Life (OTL) or NCBI taxonomy (2). The source for each tip label name (OTL or NCBI) was identified by matching tip labels with the OTL taxonomy ott3.0 (available at https://tree.opentreeoflife.org/about/taxonomy-version/ott3.0) and extracted the taxon label source, filtering for words "ncbi" and "gbif" IDs. Using these IDs we obtained downloads from NCBI Genbank and GBIF (using rgbif::name_usage()) that included the required taxonomic information to be matched with the WCVP. The required fields for matching are: taxon name, author name, genus hybrid indicator, species hybrid indicator, genus name, species name, taxon rank, infraspecific name.

Taxonomic information was transferred into a common format for both NCBI and GBIF sets and fed into the final taxonomy merging procedure (3) and returned an accepted WCVP plant species name ID. Names that could not be matched or resulted in multiple matching results were kept as blank and removed from the phylogeny. Prior to running the matching, family names in all datasets were corrected for the latest accepted version based on the APG IV system (4).

The three major steps in matching process are:
1. Matching WCVP with common input format (left join) using all columns except author name. Author names were excluded in this first step since they are notoriously prone to faulty spelling or punctuation.
    1.1. Tip labels with exactly one matching WCVP entry were considered resolved and removed from further steps.
    1.2. Tip labels with >1 matching WCVP entries were matched again, this time using all columns including author names. Tip labels with one match were stored as resolved, tip labels with multiple matches were conservatively considered unmatchable since no additional information is available for further comparison.
2. Tip labels with no matches in step 1 were matched again, this time excluding taxonomic rank and infraspecific name. Single matches were marked as solved as before, multiple matches were further resolved if possible (e.g. if pointing to the same accepted plant name ID), else treated as unmatchable.
3. Remaining tip labels were matched using all columns except family. The most common source for family mismacht was missing entries, hence we added this option. As before, single matches were marked as solved, multiple matches per tip label treated as unsolvable.

Phylogeny tip labels were replaced with accepted WCVP plant name IDs. If tip labels have been assigned the same WCVP plant name ID, we kept just one of them, following a three-step hierarchy: 1) keep the tip with molecular data 2) if no molecular data is available, keep the tip which has the same genus name as the species it links to in the WCVP 3) if none of the former applies, select one tip randomly.

By this procedure, 76% of the tip labels could be matched. Tips unmatched were removed from the phylogeny. Species that are considered accepted by the WCVP, but not present in the phylogeny (57124), were added taxonomically.

Structural Equation Model fitting. The optimal set of predictor variables for the structural equation model (SEM) was found using a randomized approach combined with subsequent manual model modification. The SEM structure was built based on variable correlation and theoretical consideration and assumed that 1) species richness is influenced by diversification and environmental variables, 2) diversification is influenced by environmental variables, 3) soil diversity and other potentially scale-dependent variables are influenced by area (see conditional paths), 4) tropical rainforest coverage is influenced by climate. This structure was kept fixed and certain variables of interest or high correlation were always included, whereas remaining variables were selected by randomly including them and evaluating the resulting model fit. Fixing selected variables had the purpose to reduce the overall number of possible combinations, and to ensure variables of special interest were always included in the SEM. Conditional paths were only included if the respective response variable was part of the model. All possible model paths are presented in Fig. S9.

Fixed model paths:
>   species richness ~ soil + (sub)trop mbf + area + diversification
>   diversification ~ annual precipitation + (sub)trop mbf + precipitation seasonality
>   soil ~ area
>   (sub)trop mbf ~ annual precipitation + annual temperature range + mean annual temperature

Soil type diversity (soil) and the biome (sub)tropical moist broadleaf forest ((sub)trop mbf) coverage were fixed since they were of high importance as assessed by the GBMs and good model performance is highly unlikely without them; area was fixed to account for spatial non-uniformity of botanical countries, and diversification was fixed as essential part of our hypotheses.
Fixed variables for the diversification regression were annual precipitation, precipitation seasonality and (sub)trop mbf coverage due to their high importance according to the GBMs and our hypotheses.

Conditional model paths:
>   precipitation SD ~ area
>   mean annual temperature SD ~ area
>   precipitation seasonality SD ~ area
>   annual temperature range SD ~ area

Conditional paths were only included if the respective dependent variable was part of the model. They represent the influence of area on spatial variability in variables as we expected variability to be higher in larger botanical countries. If e.g., precipitation standard deviation (SD) gets included in the model, the likely effect of area on precipitation standard deviation (SD) gets included as well.

We used variable influence estimated by the Generalized Boosted Models (GBM) as a starting point to choose which variables to include in this random selection process (Fig. S8). The maximum number of additional variables to include was set to ten, which provides a reasonable trade-off between computational feasibility and including all variables with considerable importance.

Ten variables for species richness:
tropical dry broadleaf forest coverage, annual precipitation SD, montane grass- and shrublands coverage, mean average temperature SD, precipitation seasonality SD, annual temperature range, annual precipitation, terrain ruggedness, elevational range, LGM temperature anomaly.

Ten variables for diversification:
Miocene temperature anomaly, mean average temperature, elevational range, annual temperature range, terrain ruggedness, soil, precipitation seasonality SD, annual temperature range SD, annual precipitation SD, temperate broadleaf/mixed forest coverage.

Within these model path and variable constraints, we tested every possible combination. We ran SEMs using the lavaan R-package (5) for all possible variable combinations (1,046,529 combinations, ranging from a minimum of 8 variables to a maximum of 26 variables included) and collected model specification, goodness of fit indices (comparative fit index (CFI), root mean square error of approximation (RMSEA) and Akaike information criterion (AIC)) and R-squared for species richness and diversification regressions. From the resulting model set, we selected models with an acceptable initial model fit (CFI > 0.9, RMSEA < 0.11). From this set of promising models, we chose the model with the largest number of variables included and manually applied meaningful modifications suggested by the lavaan::modificationsindices() function (5), further improving model fit, resulting in our final model.

Univariate and SEM path coefficients. Three variables showed different effect directions in the SEM than their univariate regressions/correlations with species richness or diversification.

Seasonality
Precipitation seasonality had a weak negative effect on species richness, while annual temperature range showed a weak positive effect on species richness, contrasting the initially observed correlations (Fig. S3, Fig. S10). Closer inspection of these relationships point to a Simpson's paradox for these variables, a case of reversed trends for all data and subgroups within the data (6). Both annual temperature range and precipitation seasonality showed dependence with mean annual temperature (Fig. S4, Fig. S5). While in warm regions seasonality in temperature and precipitation had a negative influence on species richness, this effect changed in colder regions. The negative effect of seasonality on species richness matches our expectations about narrower climatic niches that allow more species to coexist, the beneficent effect of seasonality in cold regions on species richness is more puzzling. A possible explanation could be that within rather cold environments, periodic oscillations in precipitation and temperature create temporarily limited acceptable conditions for more species than stable conditions (7).

Area
In contrast to the scale dependency of seasonality, the effect of area on diversification represents a special case of mediation (competitive partial mediation; contrasting direct and indirect effects, (8)). While the positive influence of area on diversification is mediated through soil (Fig. 2), the model detects a direct negative influence of area on diversification. As discussed in the main text, this direct effect likely reflects a non-mechanistic correlation caused by colinearity of area with other factors, which together result in an apparent negative effect of area on diversification.

Sensitivity analysis. We performed several sensitivity analyses to test if our model was biased by spatial autocorrelation, scale dependencies, or the biogeographic particularities of remote oceanic islands. We also tested an alternative measure for diversification for our model, the average DR statistic ($\overline{DR}$, (9)).

Spatial autocorrelation
To avoid underestimating standard errors in the SEM due to spatial autocorrelation (10), standard error estimates were corrected using effective sample size based on the observed global Moran's I of model residuals (11). We used the function spdep::moran.mc()(12) to calculate Moran's I of the SEM residuals as global estimate and for different distance bands. Observed spatial

autocorrelation in model residuals for species richness and diversification was strongest for neighboring botanical countries and weaker on a global scale (Fig. S). Correcting standard errors based on the global Moran's I depending effective sample size only resulted in minor decreases in path coefficient error estimates (Table S2).

Scale dependencies
We fitted our best model with additional area interaction effects for each environmental variable using the piecewiseSEM::psem function (13, 14). Fitting our model with this different function was necessary to allow for interaction effects. We found significant interaction effects for terrain ruggedness, mean annual temperature, precipitation and precipitation seasonality , however these interactions did not affect the influence direction of each main variable. but merely altered the effect strength (Table S4).

Islands
Islands often exhibit different dynamics due to their remoteness and / or limited area.We chose to test for differences between mainland botanical countries and oceanic islands of volcanic origin and ran a multigroup SEM to identify significantly different path estimates between the groups. We used island categories used in (15) and picked only volcanic islands to avoid including islands that are close to the shore and therefore potentially influenced by the mainland. This model showed slightly different dynamics on islands compared to mainland, while the mainland model closely resembled our complete final model, suggesting insularity did not influence our results (Table S5).

Diversification Rate
Third, we performed the complete variable and model selection procedure using the average DR statistic ($\overline{DR}$, (9)) as an alternative diversification measure. This measure for diversification is more influenced by younger diversification events, and hence might show different dynamics with environmental variables.
$\overline{DR}$ relies on branch lengths to calculate the diversification rates, hence we used TACT (16) to substitute missing species in our phylogeny for a more sensible placement regarding branch lengths of the missing species. Because TACT introduces a stochastic component into the taxonomic addition process, we ran TACT 50 times for a reasonable trade-off between computational resources and capturing stochasticity (TACT ran for approx. 90 hours and required ca. 150GB memory per run).
Variable selection showed minor differences compared to our main model (Fig. S8). Our best model using $\overline{DR}$ as diversification metric showed no effect of $\overline{DR}$ on species richness (Fig. S6, Table S3). The most important differences in this model include a lack of precipitation influence on $\overline{DR}$, which appears to be substituted by mean annual temperature and the Miocene precipitation anomaly, and the already present, yet not significant, negative precipitation seasonality effect becoming stronger and significant.
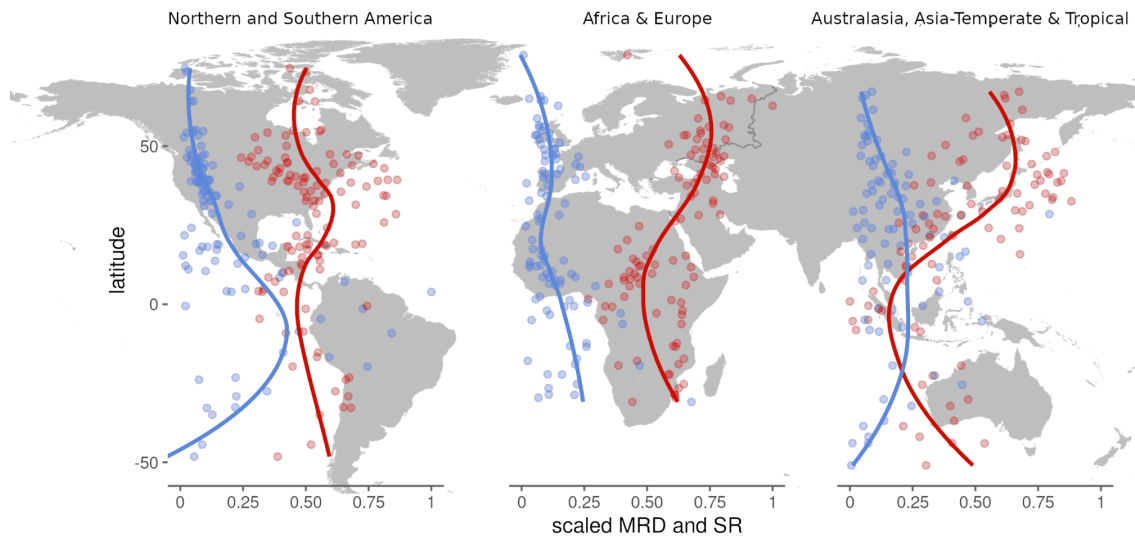
Fig. S1. Scaled species richness (SR, blue) and diversification (mean root distance; MRD, red) plotted for centroid latitudes of botanical countries per longitudinally grouped regions. Region names are as defined in (82). SR and MRD were scaled to values between 0 and 1 for comparability between regions and metrics. We observed a comparable increase towards low latitudes for the southern and northern hemisphere (Pearson correlation southern hemisphere ρ = 0.35, P < 0.01; northern hemisphere ρ = -0.36, P < 0.001). In contrast, diversification rates, (mean phylogenetic root distance; MRD) decreased towards equatorial regions in the northern hemisphere (Fig. 1b, ρ=0.44 (P < 0.001). Correlations were robust to sample size differences between hemispheres (Fig. S2). The latitudinal SR gradient was most pronounced in South America, while latitudinal differences in MRD were strongest in tropical Asia.
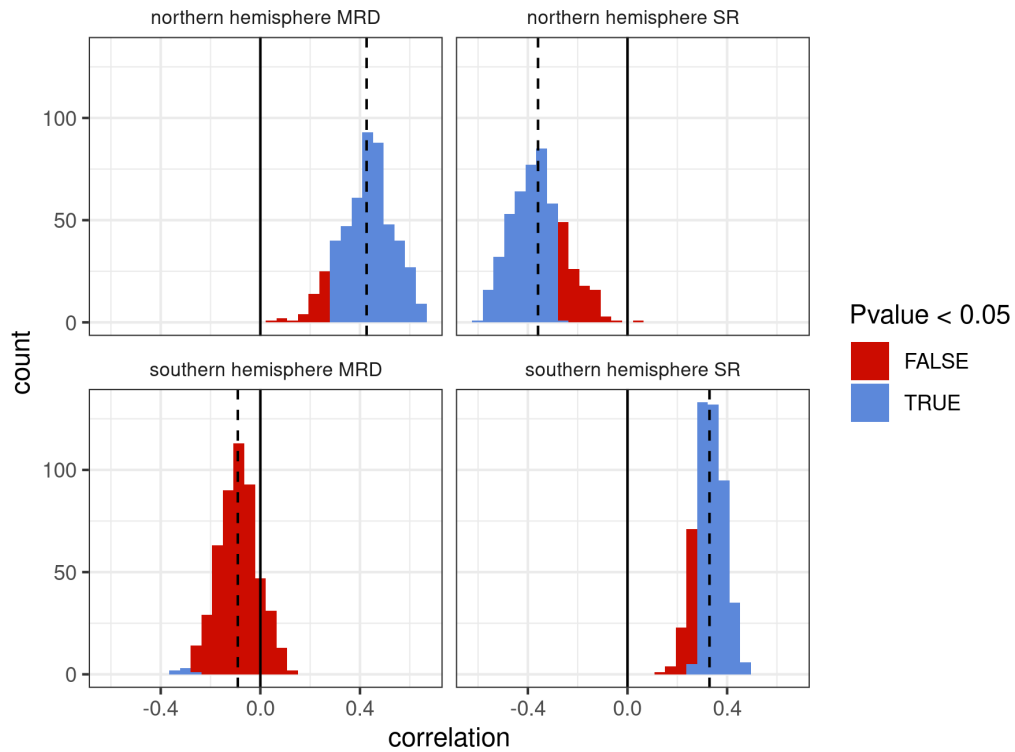
Fig. S2. Bootstrapped Pearson correlation of latitude with diversification (mean root distance; MRD) and species richness (SR) for northern and southern hemisphere. For 500 repetitions, each sample draws 50 botanical countries and correlates their MRD and SR with latitude. Solid line marks the zero intercept, dashed line indicates the mean value from all 500 repetitions. Bars are color coded for significance as indicated in figure legend.
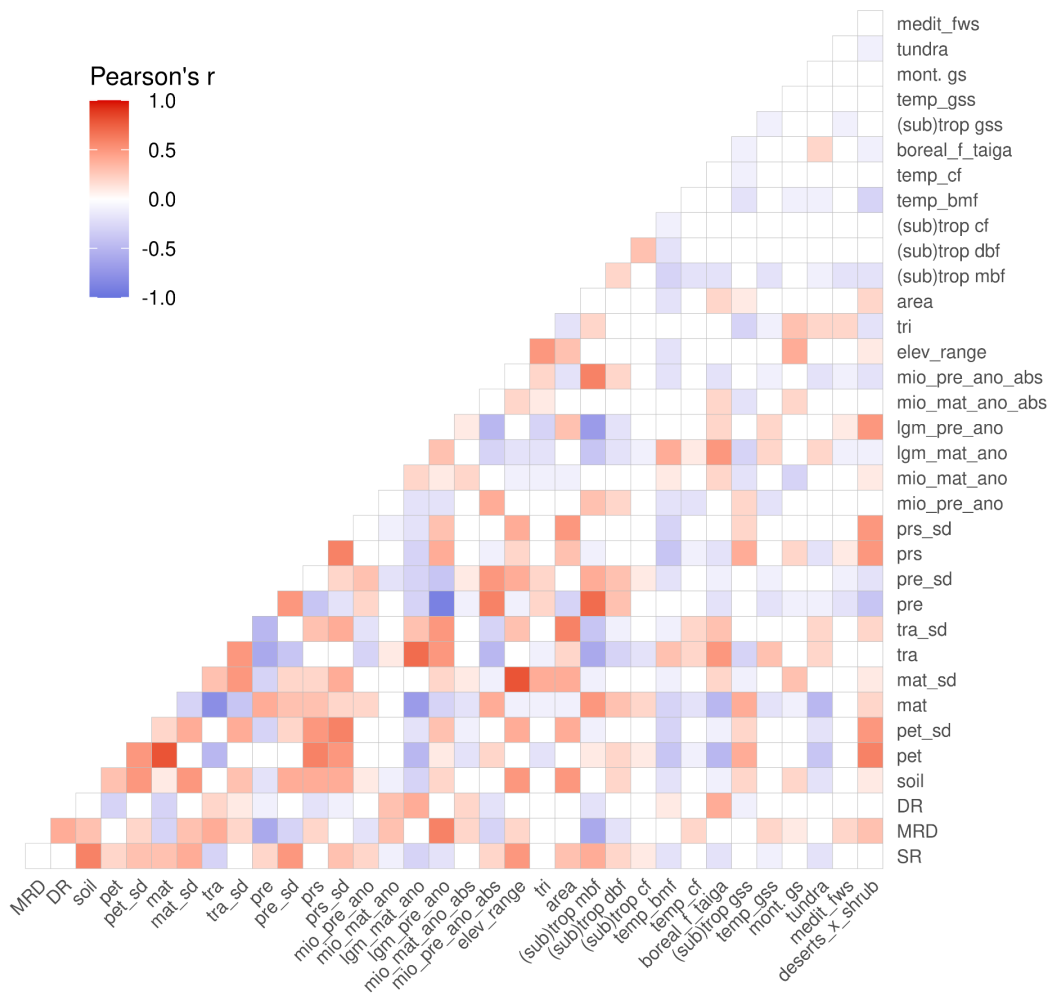
Fig. S3. Correlation matrix (Pearson correlation) of species richness (SR), mean root distance (MRD), diversification rate metric (DR) and environmental variables. Empty cells indicate nonsignificant correlations. Correlations >= 0.7 (absolute r) are highlighted in bold. Climatic variables marked with _sd are standard deviation. Abbreviations are: pet = potential evapotranspiration, mat = mean annual temperature, tra = annual temperature range, pre = total annual precipitation, prs = precipitation seasonality, tri = terrain ruggedness index, elev_range = elevational range. Variables starting with mio_ or lgm_ refer to the respective late Miocene or Last Glacial Maximum climate variable anomalies. Abbreviations for biome types are: (sub)trop mbf = (sub)tropical moist broadleaf forest, (sub)trop dbf = (sub)tropical dry broadleaf forest, (sub)trop cf = (sub)tropical coniferous forest, temp_bmf = temperate broadleaf/mixed forest, temp_cf = temperate coniferous forest, boreal_f_taiga = boreal forests/taiga, (sub)trop_gss = (sub)tropical grasslands, savannas, shrublands, temp_gss = temperate grasslands, savannas, shrublands, mont. gs = montane grasslands and shrublands, medit_fws = mediterranean forests, woodlands, scrub, desert_x_shrub = deserts and xeric shrublands.

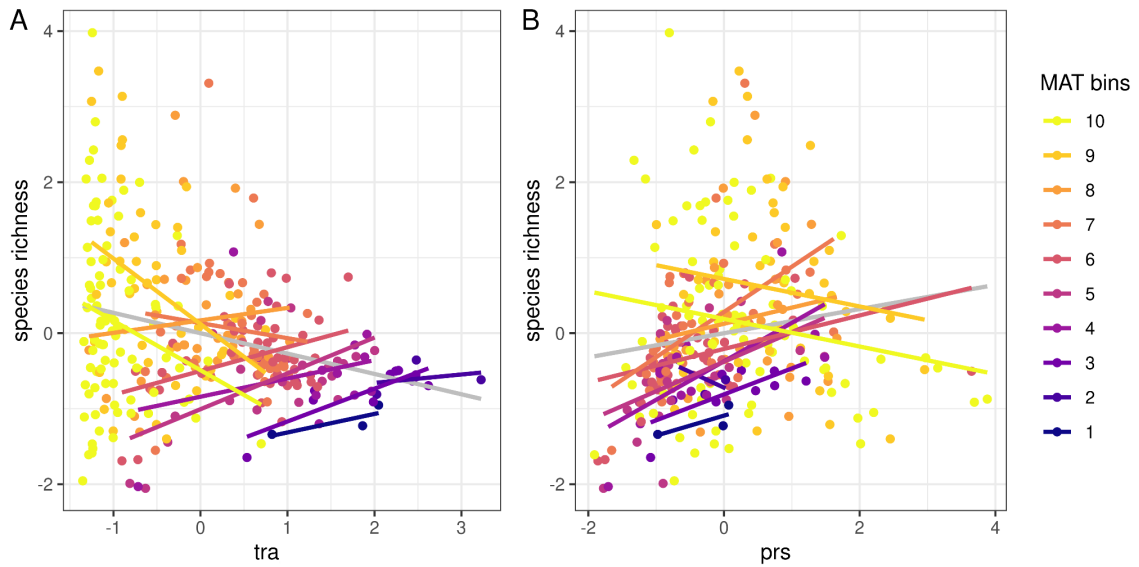**Fig. S4.** Scatterplots of species richness with A) annual temperature range (tra) and B) precipitation seasonality (prs), colored for their corresponding mean annual temperature values. The gray line indicates the univariate linear regression for all data, the colored lines show regressions for a total of 10 equal-sized mean annual temperature bins. Increasing bin numbers indicate higher mean annual temperatures.
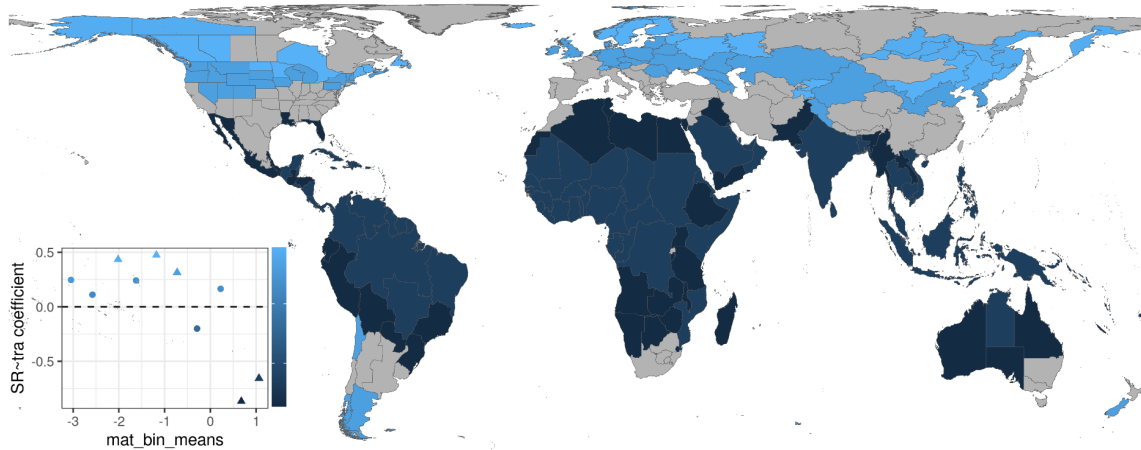
Fig. S5. Geographic representation of the influence direction of annual temperature range on species richness (compare Fig. S4). Gray areas have no significant connection between annual temperature range and species richness. The small figure shows slopes of each regression in Fig. S4A, circles coding for nonsignificant and triangles coding for significant regression coefficients.
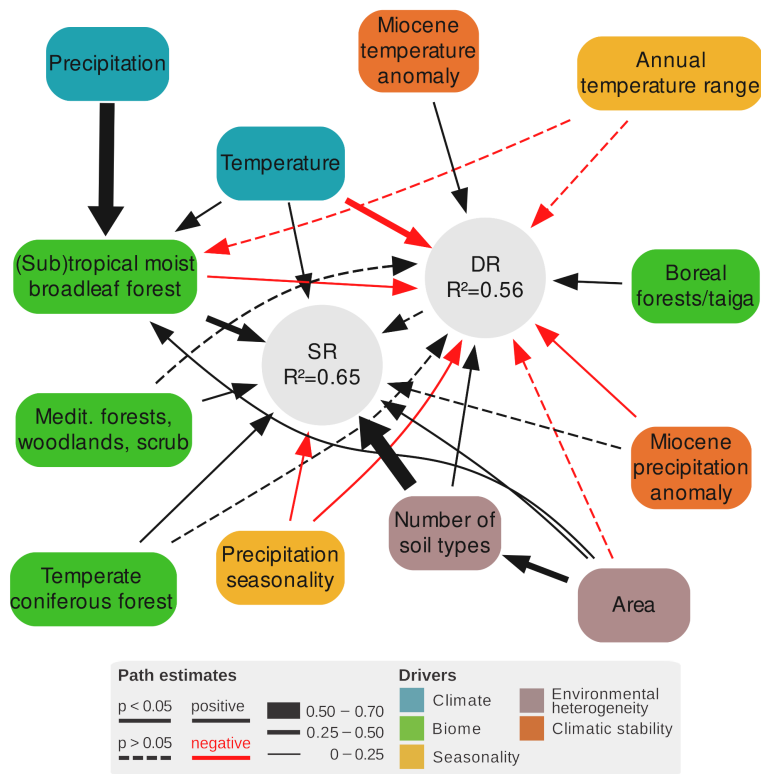
Fig. S6. Structural Equation Model depicts direct and indirect drivers of species richness (SR) and diversification rate (=DR). The width of arrows is proportional to relative effect size (see Table S3). Black arrows represent positive effects, and red arrows represent negative effects; nonsignificant effects are shown as dashed lines. Drivers are color coded for the hypothesis they address.

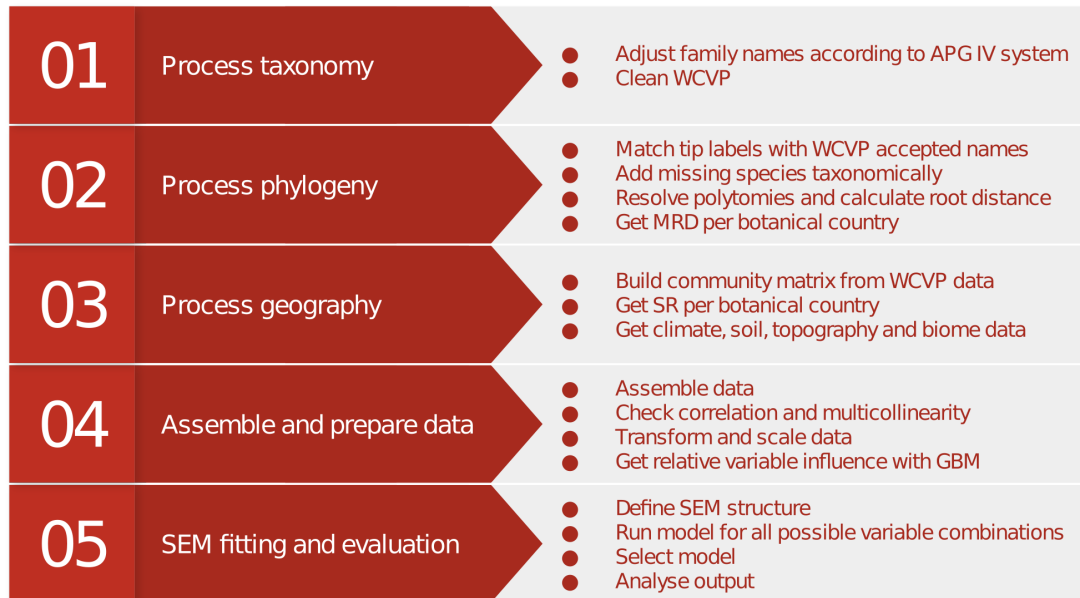| 01	Process taxonomy	● Adjust family names according to APG IV system ● Clean WCVP
02	Process phylogeny	● Match tip labels with WCVP accepted names ● Add missing species taxonomically ● Resolve polytomies and calculate root distance ● Get MRD per botanical country
03	Process geography	● Build community matrix from WCVP data ● Get SR per botanical country ● Get climate, soil, topography and biome data
04	Assemble and prepare data	● Assemble data ● Check correlation and multicollinearity ● Transform and scale data ● Get relative variable influence with GBM
05	SEM fitting and evaluation	● Define SEM structure ● Run model for all possible variable combinations ● Select model ● Analyse output

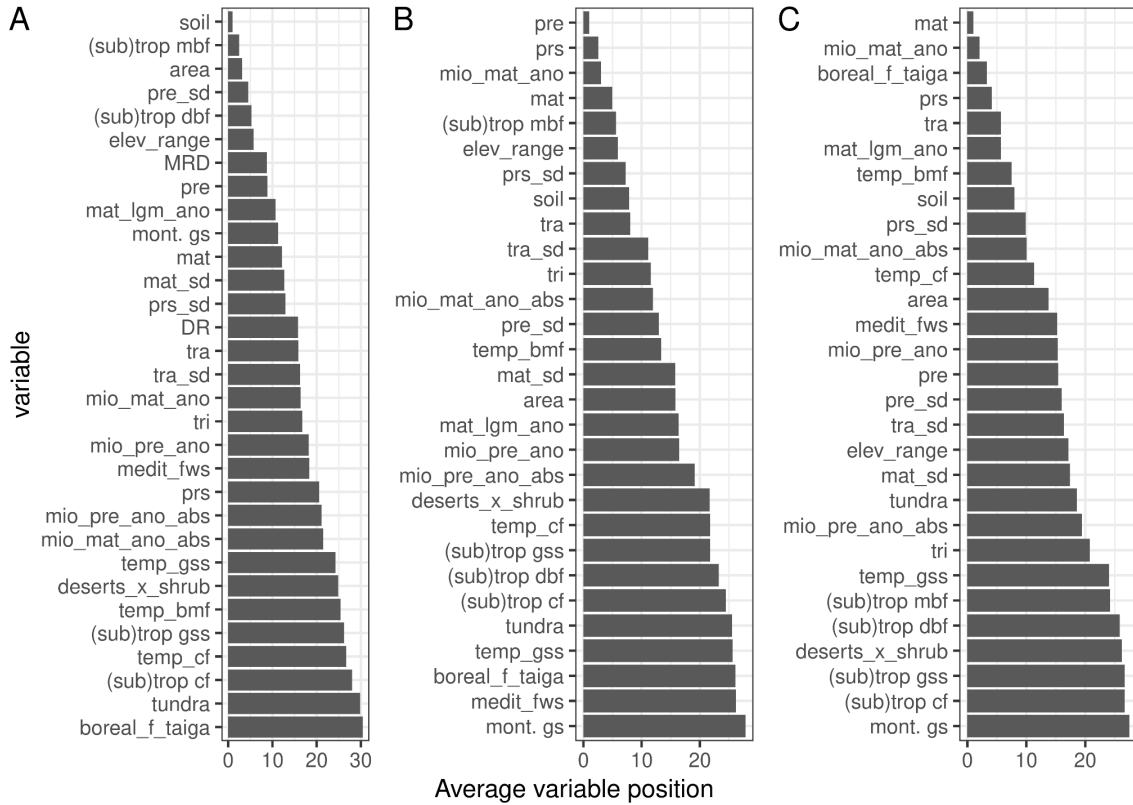**Fig. S7.** Flowchart data processing and analysis steps.

Fig. S8. Relative variable influence estimated with generalized boosted models (GBM). Influence is calculated as the sum of each variable position (i.e. most important, second important, etc) that a variable achieved in each model run, divided by the number of repetitions. Based on 100 GBMs to account for stochasticity. A) Relative variable influence on species richness, B) Relative variable influence on mean root distance, C) Relative variable influence on DR metric. Variable abbreviations are as described in Fig. S3.
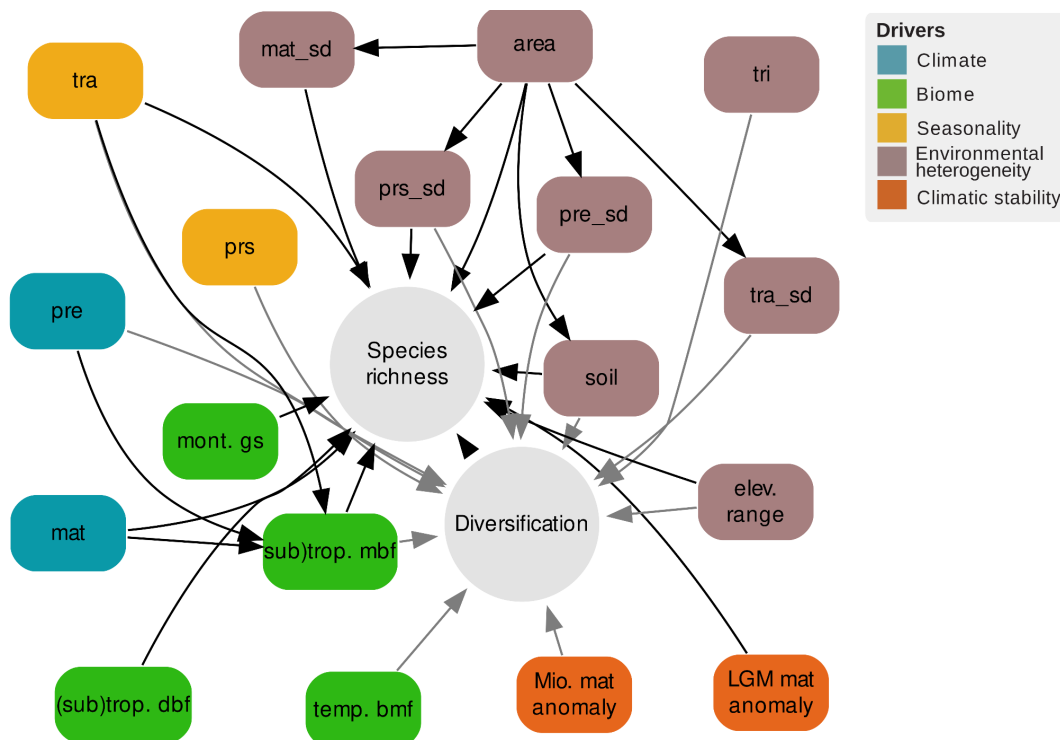
Fig. S9. Conceptual model depicting all hypothesized direct and indirect effects of the variables selected for SEM model selection. Drivers are color coded for the hypothesis they address. Pre = mean annual precipitation; mat = mean annual temperature; prs = precipitation seasonality; tra= annual temperature range; (sub)trop. mbf = coverage with (sub)tropical moist broadleaf forest; (sub)trop. dbf = coverage with (sub)tropical dry broadleaf forest; mont. gs = coverage with montane grasslands and shrublands; temp. bmf = temperate broadleaf/mixed forest; soil = number of soil types; tri = terrain ruggedness index, elev. range = elevational range, LGM mat anomaly = Last Glacial Maximum mean annual temperature anomaly, Mio. mat anomaly = late Miocene mean annual temperature anomaly. _sd refers to the variables' standard deviation per botanical country.
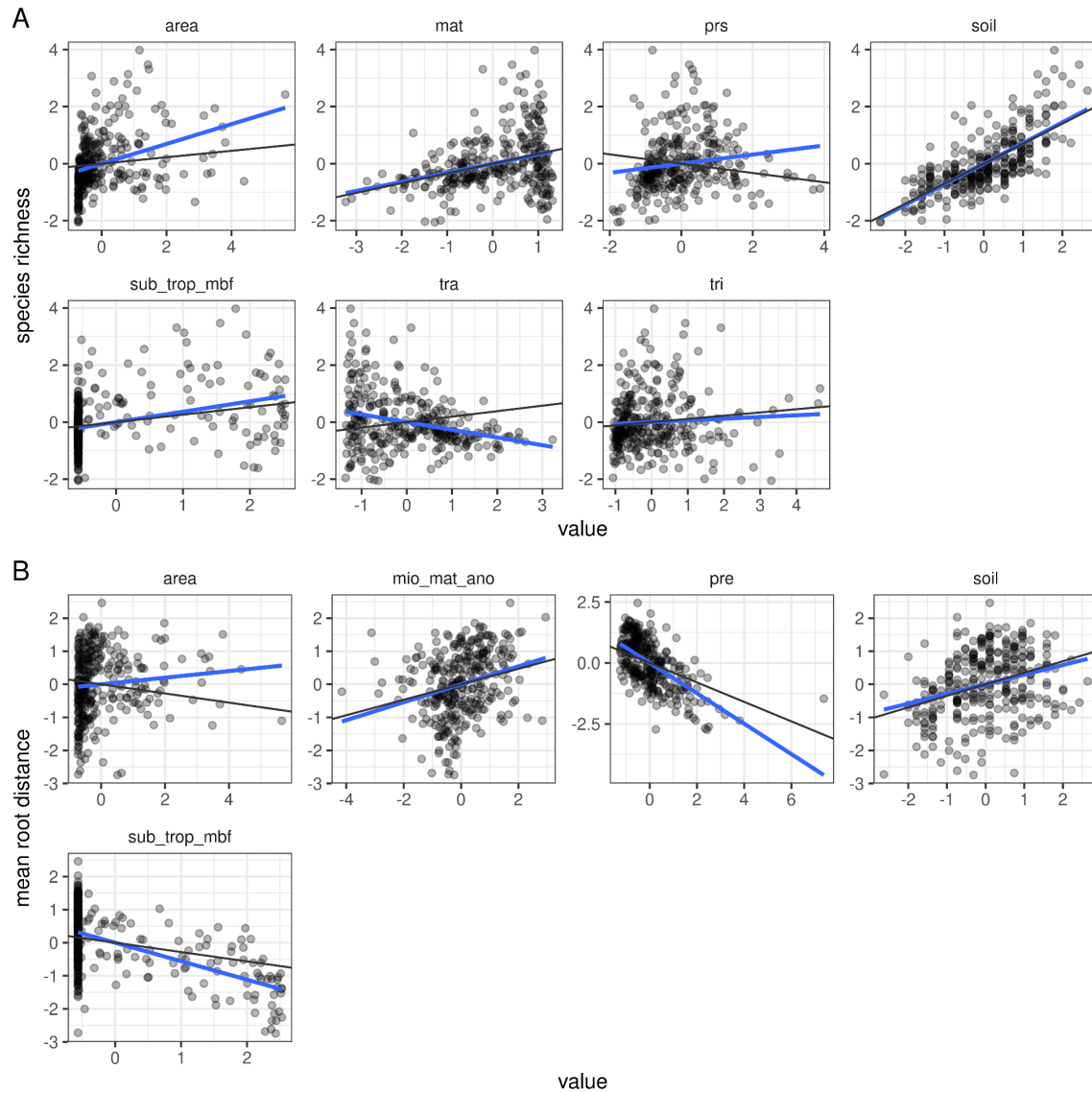
Fig. S10. Univariate and structural equation model regression scatterplots of A) species richness and B) diversification (mean root distance) with all variables that show significant direct influence in the structural equation model. Blue line indicates the univariate linear regression for each variable with species richness or mean root distance, gray line shows the regression coefficient as estimated in the structural equation model. Note the sign change between the univariate and multivariate regression lines for seasonality variables in A) and in area B).
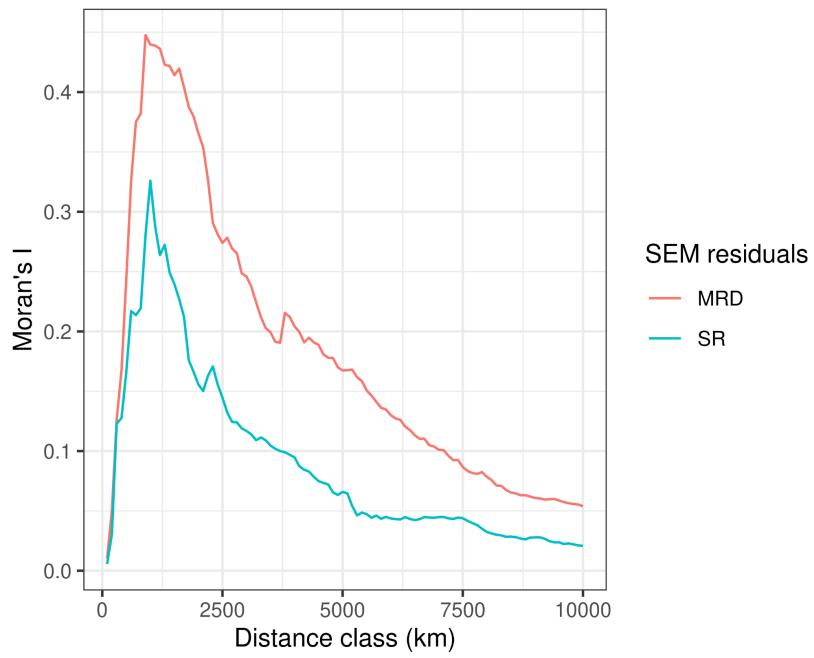
Fig. S11. Correlogram (Moran's I) for structural equation model (SEM) residuals for species richness (red) and diversification (mean root distance, MRD, blue) per botanical country, for different distance classes using spdep::moran.mc().

Table S1. Environmental variables. Climatic and seasonality variables are based on CRU TS v4 data (Climatic Research Unit gridded Time Series; (17), obtained 25. June 2020), covering years 1979 to 2019. For each climate and seasonality variable we calculated mean and standard deviation for each botanical country. Climate stability variables are derived from Last Glacial Maximum (18, 19) and late Miocene (20) climate models. Soil data was derived from SoilGrids global dataset (21). Terrain ruggedness and elevational range were derived from WorldClim elevation data (22), obtained 26 June 2020. Biome coverage was based on Olson's biomes (23) and was calculated for each botanical country as the botanical countries' proportion covered with each biome type. Mean and standard deviation refer to the botanical country level.

Hypothesis	Variable name	Description
Climate	mat	Mean temperature over one year, mean
	pet	Total potential evapotranspiration over one year, mean
	pre	Total precipitation in one year, mean
Climatic stability	LGM temperature anomaly	Difference between present mean annual temperature and Pleistocene: Last Glacial Maximum (ca. 21 ka), mean annual temperature.
	LGM precipitation anomaly	Difference between present annual precipitation and Pleistocene: Last Glacial Maximum (ca. 21 ka), annual precipitation.
	Miocene temperature anomaly	Difference between present mean annual temperature and late Miocene (11.61–5.33 Ma) mean annual temperature. We used the high CO_2 concentration model.
	Miocene precipitation anomaly	Difference between present annual precipitation and late Miocene (11.61–5.33 Ma) annual precipitation. We used the high CO_2 concentration model.
Seasonality	tra	Temperature ranges over one year (maximum temperature of warmest month minus minimum temperature of coldest month), mean
	prs	Coefficient of variation of monthly precipitation, mean
Environmental heterogeneity	pre_sd	Total precipitation in one year, standard deviation
	mat_sd	Mean temperature over one year, standard deviation
	tra_sd	Temperature ranges over one year, standard deviation
	pet_sd	Total potential evapotranspiration over one year, standard deviation
	prs_sd	Coefficient of variation of monthly precipitation, standard deviation
	soil	Number of soil types in a botanical country. Data was obtained from https://files.isric.org/soilgrids/latest/data/wrb/ on 22. June 2020, comprising a global map of the most probable soil type at a resolution of 1kmx1km
	tri	Terrain ruggedness index, following (24)
	Elevational range	Difference between min and max elevation in a botanical country
	area	Area in m^2 of each botanical country using the geosphere::areaPolygon() function (25) on the botanical countries spatial polygons object with WGS84 coordinate reference system.
Biome	Biome coverage	Proportion of botanical country covered with a biome, excluding mangroves and flooded grasslands. Biomes included are: (sub)tropical moist broadleaf forest, (sub)tropical dry broadleaf forest, (sub)tropical coniferous forest, temperate broadleaf and mixed forest, temperate coniferous forest, (sub)tropical grasslands, savannas and shrublands, temperate grasslands, savannas and shrublands, montane grasslands and shrublands, tundra, mediterranean forest, woodlands and scrub, deserts and xeric shrublands

Table S2. Standardized effect sizes and test statistics (SE=standard error) of the structural equation model. Effects are ranked by magnitude and highlighted bold when p < 0.05. Spatial autocorrelation for estimate errors (corr. SE) has been performed based on global Moran's I of 0.07 (effective sample size:276) and 0.05 (effective sample size: 240) for species richness (SR) and diversification (MRD), respectively. Abbreviations are: mat = mean annual temperature, tra = annual temperature range, pre = total annual precipitation, prs = precipitation seasonality, tri = terrain ruggedness index, (sub)trop mbf = (sub)tropical moist broadleaf forest, mont. gs = montane grasslands and shrublands.

Variable	Estimate	SE	p-value	corr. SE	corr. p-value
Effect on SR					
soil	**0.70**	**0.04**	**<0.001**	**0.05**	**<0.001**
mat	**0.34**	**0.07**	**<0.001**	**0.07**	**<0.001**
(sub)trop mbf	**0.26**	**0.06**	**<0.001**	**0.06**	**<0.001**
tra	**0.19**	**0.07**	**<0.001**	**0.07**	**0.005**
prs	**-0.16**	**0.04**	**<0.001**	**0.04**	**<0.001**
tri	**0.11**	**0.04**	**<0.001**	**0.04**	**0.004**
area	**0.11**	**0.04**	**<0.001**	**0.04**	**0.005**
mont. gs	0.03	0.04	0.48	0.04	0.502
MRD	0	0.05	0.96	0.05	0.962
area via (sub)trop MBF	**0.04**	**0.01**	**<0.001**	-	-
area via soil	**0.33**	**0.04**	**<0.001**	-	-
pre via (sub)trop mbf	**0.16**	**0.05**	**0.001**	-	-
mat via (sub)trop mbf	**0.05**	**0.02**	**0.004**	-	-
mont. gs via soil	**0.14**	**0.03**	**<0.001**	-	-
tri via (sub)trop mbf	**0.03**	**0.01**	**0.044**	-	-
Effect on MRD					
pre	**-0.40**	**0.09**	**<0.001**	**0.12**	**0.001**
soil	**0.35**	**0.05**	**<0.001**	**0.05**	**<0.001**
(sub)trop mbf	**-0.28**	**0.08**	**<0.001**	**0.10**	**0.003**
Miocene mat anomaly	**0.23**	**0.04**	**<0.001**	**0.05**	**<0.001**
area	**-0.14**	**0.05**	**<0.001**	**0.05**	**0.010**
tri	0.06	0.05	0.240	0.06	0.302
prs	-0.05	0.05	0.340	0.05	0.400
area via soil	**0.17**	**0.03**	**<0.001**	-	-
pre via (sub)trop mbf	**-0.17**	**0.03**	**<0.001**	-	-
Effect on soil					
area	**0.48**	**0.04**	**<0.001**	**0.06**	**<0.001**
mont. gs	**0.20**	**0.03**	**<0.001**	**0.04**	**<0.001**
Effect on (sub)trop mbf					
pre	**0.60**	**0.14**	**<0.001**	**0.17**	**0.001**
mat	**0.19**	**0.05**	**<0.001**	**0.06**	**0.001**
area	**0.16**	**0.03**	**<0.001**	**0.03**	**<0.001**
tra	-0.11	0.10	0.250	0.10	0.256
tri	**0.10**	**0.04**	**0.030**	**0.05**	**0.032**

Table S3. Standardized effect sizes and test statistics (SE=standard error) of the structural equation model using the average DR statistic ($\overline{DR}$) as a diversification rate metric. Effects are ranked by magnitude and highlighted bold when p < 0.05. Abbreviations are: mat = mean annual temperature, tra = annual temperature range, pre = total annual precipitation, prs = precipitation seasonality, tri = terrain ruggedness index, (sub)trop mbf = (sub)tropical moist broadleaf forest, mont. gs = montane grasslands and shrublands, temp cf = temperate coniferous forest, boreal f/t = boreal forests/taiga, medit. fws = mediterranean forests, woodlands, scrub. CFI of this model is 0.955, RMSEA 0.083.

Variable	Estimate	SE	p-value
Effect on SR			
soil	**0.69**	**0.03**	**<0.000**
(sub)trop mbf	**0.29**	**0.04**	**<0.000**
mat	**0.15**	**0.05**	**0.001**
area	**0.12**	**0.04**	**0.002**
prs	**-0.11**	**0.04**	**0.005**
medit. fws	**0.08**	**0.02**	**<0.000**
temp cf	**0.08**	**0.03**	**0.004**
Miocene mat anomaly	0.05	0.05	0.325
$\overline{DR}$	0.03	0.03	0.315
Effect on $\overline{DR}$			
mat	**-0.41**	**0.10**	**<0.000**
Miocene mat anomaly	**0.25**	**0.05**	**<0.000**
boreal f/t	**0.23**	**0.09**	**0.010**
soil	**0.23**	**0.05**	**<0.000**
prs	**-0.20**	**0.04**	**<0.000**
(sub)trop mbf	**-0.20**	**0.03**	**<0.000**
tra	-0.16	0.12	0.156
temp cf	0.07	0.04	0.090
Miocene pre anomaly	**-0.07**	**0.03**	**0.038**
area	-0.06	0.04	0.145
medit. fws	0.06	0.04	0.102
Effect on soil			
area	**0.48**	**0.04**	**<0.000**
Effect on (sub)trop mbf			
pre	**0.60**	**0.14**	**<0.000**
tra	-0.16	0.09	0.089
area	**0.14**	**0.03**	**<0.000**
mat	**0.12**	**0.05**	**0.008**
Miocene pre anomaly	0.08	0.07	0.246

Table S4. Standardized effect sizes and test statistics (SE=standard error) of the structural equation model with area interaction effects for each environmental variable. Effects are ranked by magnitude and highlighted bold when p < 0.05. Abbreviations are as introduced in Table S2.

Variable	Estimate	SE	p-value
Effect on SR			
soil	**0.53**	**0.05**	**<0.001**
area	**0.44**	**0.08**	**<0.001**
area:tri	**0.31**	**0.06**	**<0.001**
mat	**0.31**	**0.08**	**<0.001**
tri	**0.27**	**0.05**	**<0.001**
(sub)trop mbf	**0.25**	**0.05**	**<0.001**
area:mat	**0.20**	**0.07**	**0.034**
tra	**0.15**	**0.07**	**0.028**
prs	**-0.14**	**0.04**	**0.001**
area:prs	**-0.14**	**0.04**	**0.005**
MRD	0.03	0.04	0.448
mont. gs	-0.01	0.03	0.716
Effect on MRD			
pre	**-0.57**	**0.10**	**<0.001**
soil	**0.37**	**0.05**	**<0.001**
pre:area	**-0.31**	**0.13**	**0.011**
Miocene mat anomaly	**0.24**	**0.04**	**<0.001**
area	**-0.24**	**0.09**	**0.008**
(sub)trop mbf	**-0.16**	**0.08**	**0.041**
area:prs	**-0.16**	**0.04**	**0.001**
prs	-0.06	0.05	0.219
tri	0.02	0.05	0.620
Effect on soil			
area	**0.48**	**0.05**	**<0.001**
mont. gs	**0.20**	**0.05**	**<0.001**
Effect on (sub)trop mbf			
pre	**0.64**	**0.05**	**<0.001**
mat	**0.28**	**0.07**	**<0.001**
area	**0.22**	**0.05**	**<0.001**
pre:area	**0.19**	**0.06**	**<0.001**
tri	**0.13**	**0.05**	**0.004**
tra	-0.07	0.08	0.37

Table S5. Standardized effect sizes and test statistics (SE=standard error) of the structural equation model with island as grouping variable. Effects are ranked by magnitude and highlighted bold when p < 0.05. Abbreviations are as introduced in Table S2. We follow the island categories used in (15). Sample size per group is 283 and 27 for mainland and islands, respectively. Adding the grouping factor changes fit parameters to a CFI of 0.93 and a RMSEA of 0.134, indicating a worse fit than the global model.

Mainland (n=283)			Islands (n=27)		
Variable	**Estimate**	**p-value**	**Variable**	**Estimate**	**p-value**
Effect on SR					
soil	**0.65**	**<0.000**	**area**	**0.45**	**<0.000**
mat	**0.4**	**<0.000**	**(sub)trop mbf**	**0.44**	**<0.000**
(sub)trop mbf	**0.25**	**<0.000**	**soil**	**0.39**	**<0.000**
prs	**-0.21**	**<0.000**	**prs**	**0.20**	**0.015**
tri	**0.15**	**<0.000**	mat	-0.12	0.413
tra	**0.15**	**0.035**	tra	0.12	0.184
area	**0.14**	**<0.000**	MRD	-0.09	0.354
MRD	0.02	0.744	tri	0.09	0.249
mont. gs	0.02	0.586	mont. gs	0	0.995
Effect on MRD					
pre	**-0.47**	**<0.000**	**Miocene mat anomaly**	**0.40**	**0.011**
soil	**0.38**	**<0.000**	**pre**	**-0.28**	**0.002**
(sub)trop mbf	**-0.22**	**0.002**	**area**	**-0.25**	**0.006**
Miocene mat anomaly	**0.22**	**<0.000**	prs	0.21	0.120
area	**-0.17**	**0.001**	tri	0.16	0.114
prs	**-0.11**	**0.016**	(sub)trop mbf	-0.13	0.237
tri	0.04	0.499	soil	0.12	0.389
Effect on soil					
area	**0.46**	**<0.000**	**area**	**0.51**	**<0.000**
mont. gs	**0.20**	**<0.000**	mont. gs	-0.02	0.851
Effect on (sub)trop mbf					
pre	**0.73**	**<0.000**	**mat**	**0.50**	**0.001**
area	**0.17**	**<0.000**	**area**	**0.28**	**0.003**
mat	**0.16**	**<0.000**	tri	0.24	0.133
tra	-0.07	0.219	tra	-0.16	0.283
tri	0.07	0.123	pre	0.10	0.594

SI References

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

2. S. A. Smith, J. W. Brown, Constructing a broadly inclusive seed plant phylogeny. *Am. J. Bot.* **105**, 302–314 (2018).

3. M. Sun, W. Eiserhardt, M. Tietje, *taxonomy_matcher: tax match 1.4* (2021) https:/doi.org/10.5281/zenodo.5656763.

4. M. W. Chase, *et al.*, An update of the Angiosperm phylogeny group classification for the orders and families of flowering plants: APG IV. *Bot. J. Linn. Soc.* **181**, 1–20 (2016).

5. Y. Rosseel, Lavaan: An R package for structural equation modeling and more. Version 0.5--12 (BETA). *J. Stat. Softw.* **48**, 1–36 (2012).

6. C. R. Blyth, On Simpson's Paradox and the Sure-Thing Principle. *J. Am. Stat. Assoc.* **67**, 364–366 (1972).

7. J. D. Tonkin, M. T. Bogan, N. Bonada, B. Rios-Touma, D. A. Lytle, Seasonality and predictability shape temporal species diversity. *Ecology* **98**, 1201–1216 (2017).

8. G. C. Carrión, C. Nitzl, J. L. Roldán, "Mediation Analyses in Partial Least Squares Structural Equation Modeling: Guidelines and Empirical Examples" in *Partial Least Squares Path Modeling: Basic Concepts, Methodological Issues and Applications*, H. Latan, R. Noonan, Eds. (Springer International Publishing, 2017), pp. 173–195.

9. W. Jetz, G. H. Thomas, J. B. Joy, K. Hartmann, A. O. Mooers, The global diversity of birds in space and time. *Nature* **491**, 444–448 (2012).

10. J. C. Davis, R. J. Sampson, *Statistics and data analysis in geology* (John Wiley and Sons (WIE), 1986).

11. J. Byrnes, Correcting Standard Errors in SEMs fit with Covariance Matrices and ML using Moran's I, GitHub repository (2021).

12. R. S. Bivand, D. W. S. Wong, Comparing implementations of global and local indicators of spatial association. *Test* **27**, 716–748 (2018).

13. J. S. Lefcheck, piecewiseSEM : Piecewise structural equation modelling in r for ecology, evolution, and systematics. *Methods Ecol. Evol.* **7**, 573–579 (2016).

14. 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).

15. W. D. Kissling, *et al.*, Quaternary and pre-Quaternary historical legacies in the global distribution of a major tropical plant lineage. *Glob. Ecol. Biogeogr.* **21**, 909–921 (2012).

16. J. Chang, D. L. Rabosky, M. E. Alfaro, Estimating Diversification Rates on Incompletely Sampled Phylogenies: Theoretical Concerns and Practical Solutions. *Syst. Biol.* **69**, 602–611 (2020).

17. I. Harris, T. J. Osborn, P. Jones, D. Lister, Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset. *Sci Data* **7**, 109 (2020).

18. D. N. Karger, M. P. Nobis, S. Normand, C. H. Graham, N. E. Zimmermann, CHELSA-TraCE21k v1.0. Downscaled transient temperature and precipitation data since the last glacial maximum. *Clim. Past*, 1–27 (2021).

19. J. L. Brown, D. J. Hill, A. M. Dolan, A. C. Carnaval, A. M. Haywood, Paleoclim, high spatial resolution paleoclimate surfaces for global land areas. *Scientific Data* **5**, 1–9 (2018).

20. C. D. Bradshaw, *et al.*, The relative roles of CO_2 and palaeogeography in determining late Miocene climate: results from a terrestrial model–data comparison. *Clim. Past* **8**, 1257–1285 (2012).

21. T. Hengl, *et al.*, SoilGrids1km--global soil information based on automated mapping. *PLoS One* **9**, e105992 (2014).

22. 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).

23. D. M. Olson, *et al.*, Terrestrial ecoregions of the world: a new map of life on earth. *Bioscience* **51**, 933 (2001).

24. S. J. Riley, S. D. DeGloria, R. Elliot, Index that quantifies topographic heterogeneity. *Intermt. J. Sci.* **5**, 23–27 (1999).

25. R. J. Hijmans, geosphere: Spherical Trigonometry (2019).