Dispersal and niche evolution jointly shape the geographic turnover of phylogenetic clades across continents

Wolf L. Eiserhardt, Jens-Christian Svenning, William J. Baker, Thomas L.P. Couvreur & Henrik Balslev

Supplementary Table S1: Slopes of the relationships between phylogenetic turnover and spatial distance, barrier distance, and environmental dissimilarities — single predictor models. The P-value is calculated as $n_0/999$, where n_0 is the number of randomized phylogenetic turnover matrices that produced a more extreme slope than the non-randomized matrix. Variables marked ∗ have a significant slope compared to the null model at the α=0.05 level.

	Slope	P	
Spatial distance	0.521	0	∗
Barrier distance	0.586	0	∗
Environmental dissimilarity (all variables PCA)	0.438	0	∗
Temperature			
All variables (PCA)	0.445	0	∗
Temperature seasonality	0.627	0	∗
Temperature annual range	0.579	0	∗
Isothermality	0.514	0	∗
Mean temperature of coldest quarter	0.410	0.001	∗
Min temperature of coldest month	0.403	0	∗
Mean temperature of driest quarter	0.249	0.022	∗
Mean diurnal range	0.248	0	∗
Annual mean temperature	0.216	0.014	∗
Mean temperature of wettest quarter	0.149	0.042	∗
Max temperature of warmest month	0.093	0.133	
Mean temperature of warmest quarter	0.081	0.173	
Soil			
All variables (PCA)	0.297	0	∗
pH	0.343	0	∗
Base saturation	0.328	0	∗
Total exchangeable bases	0.261	0	∗
$CaCO_3$	0.217	0.002	∗
Cation exchange capacity	0.142	0	∗
Sand content	-0.013	0.180	
Vegetation structure			
All variables (PCA)	0.164	0.009	∗
Mean annual QSCAT	0.231	0	∗
Mean annual NDVI	0.142	0.065	

NDVI intraannual variability	0	0.384	
QSCAT intraannual variability	-0.016	0.294	
Topography			
All variables (PCA)	0.147	0.031	*
Slope	0.148	0.036	*
Altitudinal range	0.129	0.040	*
Precipitation			
All variables (PCA)	0.129	0.046	*
Precipitation of wettest quarter	0.249	0	*
Precipitation of wettest month	0.235	0	*
Annual precipitation	0.195	0	*
Precipitation seasonality	0.055	0.333	
Precipitation of coldest quarter	0.020	0.672	
Precipitation of driest quarter	0.009	0.775	
Precipitation of driest month	0.003	0.808	
Precipitation of warmest quarter	-0.013	0.196	

Supplementary Table S2. Variation in phylogenetic turnover among palm assemblages explained by environmental dissimilarity and A) spatial distance or B) barrier distance. R_t = total amount of variation explained by the model. R_{pd} = amount of variation independently explained by spatial distance (A) or barrier distance (B). R_{mx} = amount of variation jointly explained by environmental dissimilarity and spatial distance (A)/barrier distance (B). P(x) is the proportion of null model runs producing $x_{null}>x_{observed}$, and can be interpreted as a P value. C): ΔR_{pd} is R_{pd}(barrier distance) - R_{pd}(spatial distance). ΔR_{pe} is R_{pe}(barrier distance) - R_{pe}(spatial distance).

	A) spatial distance						**B) barrier distance**						**C) comparison**	
Environmental dissimilarity	**R_t**	**R_{pd}**	**$P(R_{pd})$**	**R_{mx}**	**R_{pe}**	**$P(R_{pe})$**	**R_t**	**R_{pd}**	**$P(R_{pd})$**	**R_{mx}**	**R_{pe}**	**$P(R_{pe})$**	**ΔR_{pd}**	**ΔR_{pe}**
Overall														
All variables (PCA)	0.353	0.161	0	0.111	0.081	0	0.382	0.190	0	0.154	0.038	0.007	0.029	-0.043
Temperature														
All variables (PCA)	0.366	0.168	0	0.104	0.094	0.003	0.399	0.201	0	0.143	0.055	0.012	0.033	-0.039
Temperature seasonality	0.470	0.077	0.001	0.195	0.198	0.006	0.488	0.095	0.001	0.249	0.144	0.012	0.018	-0.054
Temperature annual range	0.438	0.103	0	0.169	0.166	0.002	0.465	0.130	0	0.214	0.121	0.005	0.027	-0.045
Isothermality	0.413	0.149	0	0.122	0.142	0.002	0.447	0.183	0	0.160	0.104	0.007	0.034	-0.038
Mean temperature of coldest quarter	0.352	0.184	0	0.088	0.080	0.015	0.395	0.227	0	0.117	0.051	0.034	0.043	-0.030
Min temperature of coldest month	0.352	0.189	0	0.083	0.080	0.005	0.395	0.233	0	0.111	0.052	0.016	0.044	-0.028
Mean temperature of driest quarter	0.307	0.245	0	0.027	0.035	0.059	0.366	0.304	0	0.040	0.022	0.108	0.059	-0.013
Mean diurnal range	0.291	0.229	0	0.042	0.019	0.029	0.355	0.293	0	0.051	0.011	0.089	0.064	-0.008
Annual mean temperature	0.294	0.248	0	0.024	0.023	0.051	0.354	0.307	0	0.037	0.010	0.144	0.060	-0.013
Mean temperature of wettest quarter	0.276	0.254	0	0.018	0.005	0.317	0.344	0.322	0	0.022	0.000	0.898	0.067	-0.005
Max temperature of warmest month	0.276	0.268	0	0.004	0.004	0.277	0.344	0.336	0	0.008	0.000	0.722	0.068	-0.004
Mean temperature of warmest quarter	0.275	0.268	0	0.004	0.003	0.394	0.344	0.337	0	0.007	0.000	0.91	0.069	-0.003
Soil														
All variables (PCA)	0.292	0.203	0	0.069	0.020	0.005	0.351	0.263	0	0.081	0.007	0.114	0.059	-0.013
pH	0.322	0.204	0	0.067	0.051	0.001	0.380	0.262	0	0.082	0.036	0.003	0.057	-0.015
Base saturation	0.312	0.204	0	0.068	0.040	0	0.377	0.270	0	0.074	0.034	0	0.066	-0.006
Total exchangeable bases	0.284	0.216	0	0.055	0.012	0.024	0.349	0.281	0	0.063	0.005	0.164	0.065	-0.007
$CaCO_3$	0.278	0.231	0	0.041	0.006	0.229	0.344	0.297	0	0.047	0.000	0.807	0.066	-0.006
Cation exchange capacity	0.273	0.253	0	0.019	0.001	0.316	0.344	0.324	0	0.020	0.000	0.929	0.071	-0.001
Sand content	0.272	0.272	0	0.000	0.000	0.556	0.344	0.344	0	0.000	0.000	0.442	0.072	0.000

Vegetation structure														
All variables (PCA)	0.279	0.252	0	0.020	0.007	0.174	0.345	0.318	0	0.026	0.001	0.554	0.067	-0.005
Mean annual QSCAT	0.300	0.247	0	0.025	0.028	0.001	0.358	0.304	0	0.040	0.014	0.042	0.058	-0.014
Mean annual NDVI	0.275	0.254	0	0.017	0.003	0.486	0.344	0.324	0	0.020	0.000	0.991	0.069	-0.003
NDVI intraannual variability	0.273	0.273	0	-0.001	0.001	0.592	0.344	0.344	0	0.000	0.000	0.729	0.072	0.000
QSCAT intraannual variability	0.272	0.272	0	0.000	0.000	0.878	0.344	0.344	0	0.000	0.000	0.929	0.072	0.000
Topography														
All variables (PCA)	0.280	0.258	0	0.013	0.008	0.17	0.348	0.327	0	0.017	0.004	0.314	0.068	-0.004
Slope	0.278	0.256	0	0.015	0.006	0.221	0.347	0.325	0	0.018	0.003	0.378	0.069	-0.003
Altitudinal range	0.280	0.263	0	0.008	0.008	0.145	0.348	0.332	0	0.012	0.004	0.273	0.068	-0.004
Precipitation														
All variables (PCA)	0.282	0.266	0	0.006	0.010	0.126	0.348	0.331	0	0.013	0.004	0.383	0.066	-0.006
Precipitation of wettest quarter	0.307	0.245	0	0.027	0.035	0	0.366	0.303	0	0.040	0.022	0.004	0.058	-0.014
Precipitation of wettest month	0.301	0.246	0	0.026	0.030	0.002	0.362	0.306	0	0.037	0.018	0.011	0.060	-0.012
Annual precipitation	0.299	0.261	0	0.011	0.027	0.009	0.360	0.322	0	0.022	0.016	0.052	0.060	-0.012
Precipitation seasonality	0.272	0.269	0	0.003	0.000	0.704	0.346	0.343	0	0.001	0.002	0.393	0.074	0.002
Precipitation of coldest quarter	0.275	0.275	0	-0.003	0.003	0.401	0.347	0.346	0	-0.002	0.003	0.441	0.071	-0.001
Precipitation of driest quarter	0.273	0.273	0	-0.001	0.001	0.61	0.344	0.344	0	0.000	0.000	0.913	0.071	-0.001
Precipitation of driest month	0.273	0.273	0	-0.001	0.001	0.621	0.344	0.344	0	0.000	0.000	0.944	0.071	-0.001
Precipitation of warmest quarter	0.273	0.272	0	-0.001	0.001	0.553	0.346	0.345	0	-0.002	0.002	0.381	0.073	0.001

Supplementary Table S3: Correlations between environmental variables for all 1701 1°×1° grid cells with palm occurrences in the Americas (Spearman's ρ). Only the fifteen variables with the largest pure environmental fractions (Fig. 5 of the main article) are shown. Correlations $|\rho| \geq 0.7$ are highlighted.

	1. Temperature Seasonality	2. Temperature Annual Range (P5-P6)	3. Isothermality	4. Mean Temperature of Coldest Quarter	5. Min Temperature of Coldest Period	6. soil pH	7. base saturation	8. Precipitation of Wettest Quarter	9. Mean Temperature of Driest Quarter	10. Precipitation of Wettest Period	11. QSCAT_mean	12. Annual Precipitation	13. Annual Mean Temperature	14. Mean Diurnal Range	15. total exchangeable bases
1		0.81	-0.92	-0.80	-0.76	0.65	0.58	-0.72	-0.70	-0.71	-0.72	-0.69	-0.65	0.49	0.42
2			-0.81	-0.77	-0.90	0.45	0.38	-0.59	-0.72	-0.60	-0.57	-0.62	-0.68	0.86	0.23
3				0.73	0.73	-0.55	-0.47	0.62	0.65	0.62	0.67	0.59	0.59	-0.44	-0.33
4					0.95	-0.53	-0.47	0.67	0.94	0.67	0.55	0.59	0.96	-0.57	-0.38
5						-0.47	-0.39	0.64	0.92	0.65	0.53	0.61	0.92	-0.76	-0.28
6							0.89	-0.63	-0.49	-0.60	-0.60	-0.65	-0.41	0.30	0.75
7								-0.59	-0.41	-0.56	-0.50	-0.57	-0.35	0.22	0.87
8									0.63	0.99	0.71	0.90	0.57	-0.41	-0.47
9										0.63	0.49	0.56	0.93	-0.57	-0.34
10											0.69	0.88	0.57	-0.42	-0.44
11												0.74	0.43	-0.33	-0.36
12													0.49	-0.51	-0.38
13														-0.54	-0.29
14															0.09
15															

Supplementary Table S4: Environmental GIS layers

source	layers	source URL	reference
Worldclim	19 bioclimatic variables	http://www.worldclim.org/current	Hijmans et al. 2005
Harmonized World Soil Database	six soil variables	http://www.iiasa.ac.at/Research/LUC/External-World-soil-database/HTML/	Fischer et al. 2008
HYDRO1k Elevation Derivative Database	altitude, slope	http://eros.usgs.gov/#/Find_Data/Products_and_Data_Available/gtopo30/hydro	—
NASA Scatterometer Climate Record Pathfinder	QSCAT microvave backscatter*‡	http://www.scp.byu.edu/	Long et al. 2001
EDIT Geoplatform	Normalised Difference Vegetation Index (NDVI)†‡	http://edit.csic.es/GISdownloads.html	—

*QSCAT is related to forest structure and biomass (Buermann et al. 2008)
†NDVI reflects vegetation density ('greenness')
‡For QSCAT and NDVI, the mean and coefficient of variance were included as separate variables

References:

Buermann, W., et al. 2008. Predicting species distributions across the Amazonian and Andean regions using remote sensing data. - *J. Biogeogr.* 35: 1160-1176.
Fischer, G., et al. 2008. Global Agro-ecological Zones Assessment for Agriculture (GAEZ 2008). - IIASA, Laxenburg, Austria and FAO, Rome, Italy.
Hijmans, R. J., et al. 2005. Very high resolution interpolated climate surfaces for global land areas. - *International Journal of Climatology* 25: 1965-1978.
Long, D. G., et al. 2001. Global Ice and Land Climate Studies Using Scatterometer Image Data. - *EOS, Transaction of American Geophysical Union* 82: 503-509.

Supplementary Methods

Calculation of phylogenetic turnover. We derived an R function for computing the comdist measure (Webb *et al*., 2008), modified from the *comdist* function of the *picante* package (Kembel *et al*., 2010) to improve computational performance:

```
comdist <- function(comm, dis)
{
    dat <- match.comm.dist(comm, dis)
    pcmatrix  <- as.matrix(dat$comm)
    dstmatrix <- as.matrix(dat$dist)
    tax <- colnames(pcmatrix)

    N <- dim(pcmatrix)[1]
    samples <- row.names(pcmatrix)

    comdist <- matrix(nrow=N,ncol=N,dimnames=list(samples,samples))

    for(i in 1:(N-1)){
      for(j in (i+1):N){
          a <- tax[pcmatrix[i, ] > 0]
          b <- tax[pcmatrix[j, ] > 0]
          comdist[i,j] <- sum(dstmatrix[a,b]) / (length(a) * length(b))
      }
    }

    return(comdist)
}
```

Barrier distance. The likelihood of a species to disperse from A to B depends not only on the geographic distance between A and B, but also on the environment of the interjacent areas. A species is unlikely to cross areas with an environment that is very different from the environment in which it occurs (dispersal barriers *sensu* Wiens & Graham, 2005) because it is unlikely to establish reproducing populations in those areas. We thus computed a "barrier distance" that takes both geographic distance and the (un-)suitability of the interjacent areas into account.

First, we extracted the mean values of all 31 environmental variables (see Materials & Methods) for all 1×1 degree grid cells overlapping with land surface between 40º N and 40º S, and 30º W and 124º W. This rectangular space fully encompasses the distribution of palms in the Americas, plus an additional five degree buffer to the north and the south to allow for the circumvention of potential dispersal barriers. As in the main analysis, range was used for elevation instead of the mean.

Second, we carried out a Principal Components Analysis (PCA) on the centered and standardised environmental variables, and computed euclidean distance on the PCA scores. The distance matrix was subsequently rescaled to a range of 0–1 by dividing it by the maximum distance.

For each cell, the unsuitabilty of all other cells was quantified on a scale from 0–1 by mapping the environmental distance between the focal cell and all other cells to the coordinates of the other cells (Figure S1). Mapping of distance values and all further spatial processing was carried out in ArcGis 10.

A small positive value (1×10^{-4}) was added to all cells, because the subsequent analyses cannot handle zero values. All cells that do not overlap with land surface were arbitrarily set to a value of one (= maximum unsuitability). Two examples of such unsuitability maps are shown in Fig. S1 A, C.

We then used the “Cost Distance” function in ArcGis 10 to generate maps of cumulative unsuitability on the least cost path between the focal cell and each other cell in the dataset. I.e., the algorithm finds the path on which the least cumulative amount of unsuitability is crossed (the “least cost path”, which is often indirect, see Supplementary Fig. S2) and sums up the unsuitabilty of all passed cells.

The least cost path, and cumulative unsuitability, between two cells A and B depends on whether the unsuitability surface is computed based on the environment of cell A or of cell B (Supplementary Fig. S2 B,C). To obtain a non-directional measure of barrier, we calculated the average of the minimum cumulative suitability a species from A must cross to reach B, and vice versa.

The resulting barrier distance was quite strongly correlated to spatial distance (Supplementary Fig. S3) but the transformation had a clear effect on relationships among grid cells (Supplementary Fig. S4). Note for example that relationships among gridcells located in the relatively homogeneous Amazon remain relatively unchanged, whereas gridcells from different sides of the Andes mountains become widely separated (Supplementary Fig. S4).

References:

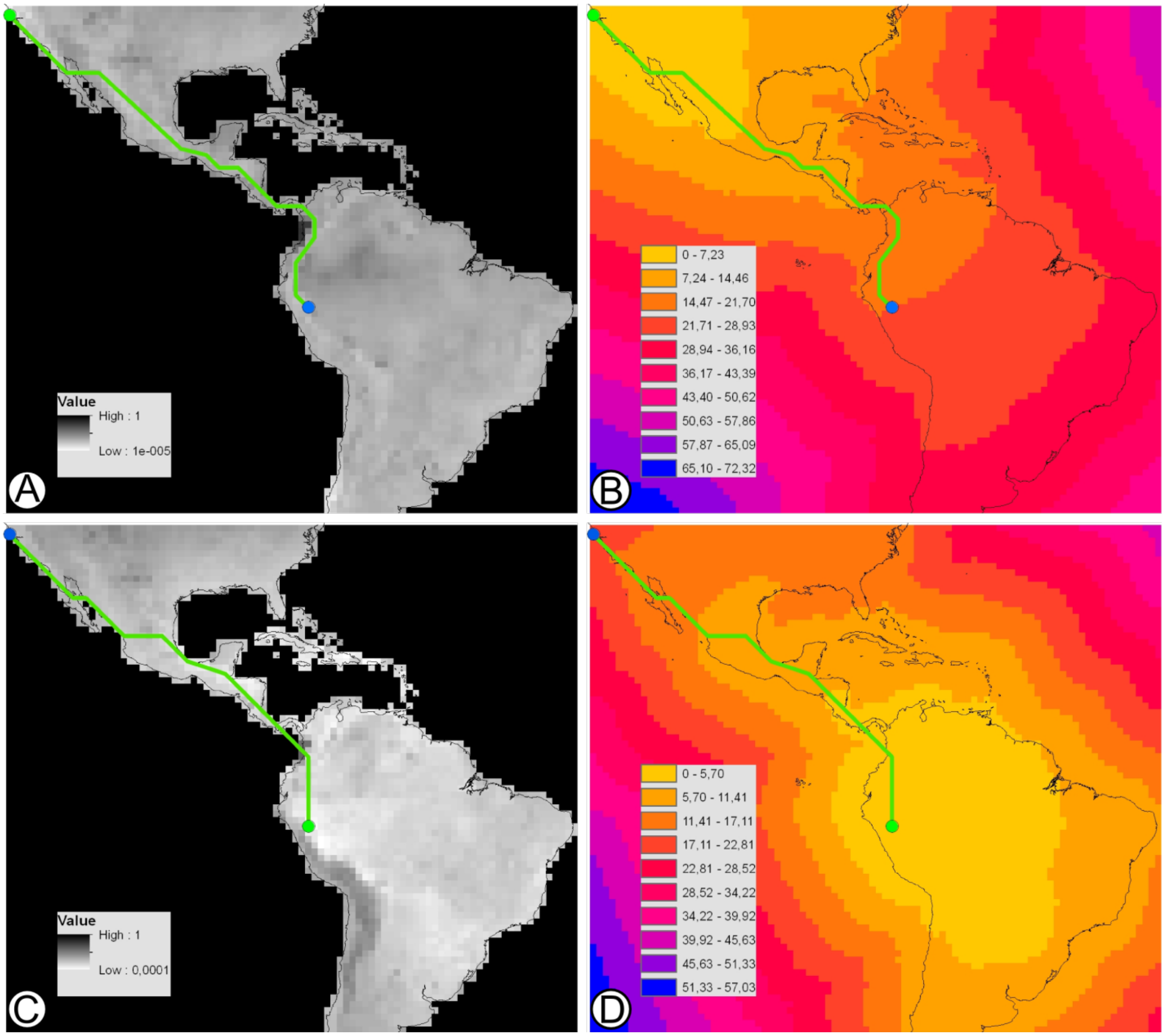
Kembel, S.W., Cowan, P.D., Helmus, M.R., Cornwell, W.K., Morlon, H., Ackerly, D.D., Blomberg, S.P. & Webb, C.O. (2010) Picante: R tools for integrating phylogenies and ecology. *Bioinformatics*, **26**, 1463-1464.

Webb, C.O., Cannon, C.H. & Davies, R.G. (2008) Ecological organization, biogeography, and the phylogenetic structure of tropical forest tree communities. *Tropical forest community ecology* (ed. by W.P. Carson and S.A. Schnitzer). Wiley & Sons, Chichester.

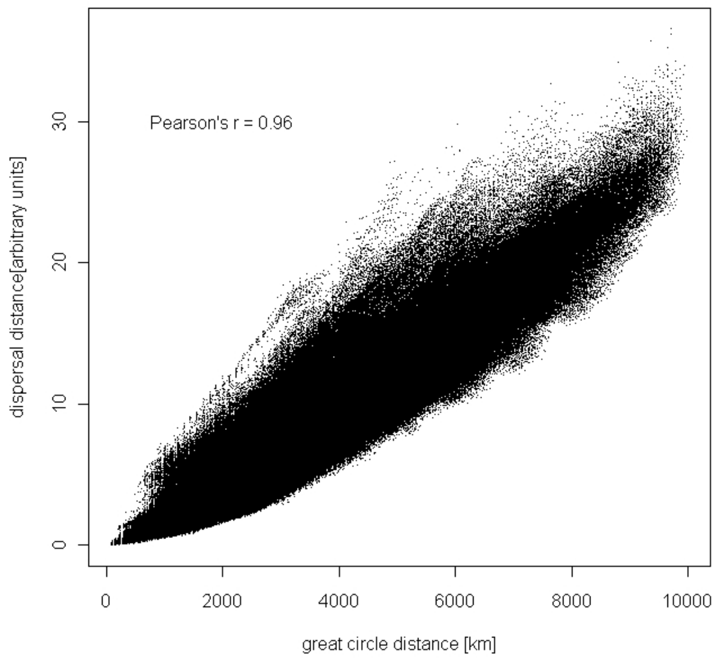
Wiens, J.J. & Graham, C.H. (2005) Niche conservatism: Integrating evolution, ecology, and conservation biology. *Annual Review of Ecology Evolution and Systematics*, **36**, 519-539.

cell #	1	2	3	4	5	…	2342	2343
1	0	0.068574	0.120092	0.12447	0.167302	…	0.339158	0.321754
2	0.068574	0	0.159378	0.117325	0.137886	…	0.372642	0.351516
3	0.120092	0.159378	0	0.120715	0.176813	…	0.331961	0.323342
4	0.12447	0.117325	0.120715	0	0.07281	…	0.362818	0.341223
5	0.167302	0.137886	0.176813	0.07281	0	…	0.393114	0.368781
…	…	…	…	…	…	0	…	…
2342	0.339158	0.372642	0.331961	0.362818	0.393114	…	0	0.052394
2343	0.321754	0.351516	0.323342	0.341223	0.368781	…	0.052394	0
Longitude	-123.5	-123.5	-122.5	-122.5	-122.5		-34.5	-34.5
Latitude	38.5	39.5	37.5	38.5	39.5		-8.5	-7.5

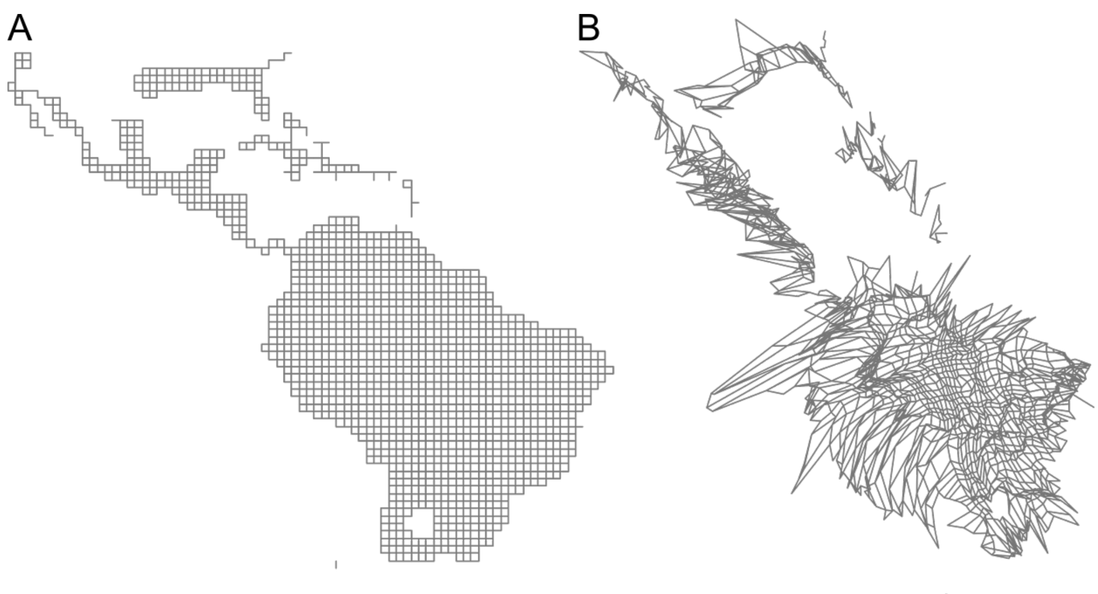
Supplementary Figure S1: Part of the environmental distance matrix used to quantify unsuitability for dispersal. The green shading indicates the values that were mapped using the corresponding Lat-Lon coordinates to represent unsuitability for dispersal of species found in cell no. 2.



Supplementary Figure S2 A,C: Cost surfaces. Cost equals environmental dissimilarity to the green cell on a scale from 0–1. B,D: Cost distance between each grid cell and the green cell, based on the cost surface on the left. Green line = least cost path from the green to the blue cell based on the cost surface on the left.



Supplementary Figure S3. Correlation between geographic (great circle) distance and a measure of dispersal distance that takes environmental conditions of interjacent areas into account.



Supplementary Figure S4: Effect of taking into account the environment of interjacent areas on relationships between grid cells. A: grid cells plotted in geographic space. B: grid cells plotted in the space spanned by the first two axes of a NMDS analysis on "barrier distance" — rotated. The lines connect each grid cell to its immediate neighbours (±1°) in geographic space (i.e., based on A).