A biome-dependent distribution gradient of tree species range edges is strongly dictated by climate spatial heterogeneity

David Lerner[1], Marcos Fernández Martínez[2], Stav Livne-Luzon[1], Jonathan Belmaker[3,4], Josep Peñuelas[2,5], Tamir Klein[1]

[1]Department of Plant & Environmental Sciences, Weizmann Institute of Science. Rehovot 76100, Israel.

[2]CREAF, Cerdanyola de Valles, Catalonia E08193, Spain.

[3]School of Zoology, Faculty of Life Sciences, Tel Aviv University, Tel Aviv, Israel.

[4]Steinhardt Museum of Natural History, Tel Aviv University, Tel Aviv, Israel.

[5]CSIC, Global Ecology Unit CREAF-CSIC-UAB, Bellaterra, Catalonia E08193, Spain

Abstract

Understanding the causes of the arrest of species distributions has been a fundamental question in ecology and evolution. These questions are of particular interest for trees due to their long lifespan and sessile nature. A surge in data-availability evokes for a macro-ecological analysis to determine the underlying forces limiting distributions. Here we analyse the spatial distribution of >3600 major tree species to determine geographical areas of range-edge hotspots and find drivers for their arrest. We confirmed biome edges to be strong delineators of distributions. Importantly, we identified a stronger contribution of temperate than tropical biomes to range edges, adding strength to the notion that tropical areas are centers of radiation. We subsequently identified a strong association of range-edge hotspots with steep spatial climatic gradients. We linked spatial and temporal homogeneity and high potential evapotranspiration in the tropics as the strongest predictors for this phenomenon. We propose that the poleward migration of species in light of climate change might be hindered due to steep climatic gradients.

Main Text

Introduction

The geographical distributions of species are marked by their range limits. Understanding the causes of distribution arrest has been a fundamental question in ecology and evolution[1–4]. Given the strong interplay between biotic, abiotic, demographic, physical and historical forces in predicting range-edges, it has been challenging to find underpinnings for their

formation. Two main environmental forces seem to play a major role in the formation of range edges; spatial environmental heterogeneity and habitat quality[4,54,5]. Most models on the formation of species-range edges rely on the interplay between either one of these two forces with non-climatic pressures to explain their formation. For example, steep climate gradients combined with high dispersal and gene flow reduces species' fitness and genetically constrains their evolution into novel environments[6,7]. Likewise, low habitat quality reduces population size[2,8], increasing drift and migration load[9–11]. Nevertheless, the significance of climate in the interplay between these two environmental components in defining range edges for an array of species or on a wide biogeographical scale remains elusive[12,13].

The field of biogeography has long sought understanding of species ranges despite having limited tools (e.g. [3,14]). Given the surge of large-scale datasets, it is now possible to better identify the underpinnings of species distributions by studying the macro eco-evolutionary processes involved in their formation[11,12,15–17]. Although methods and results are disparate between studies, there is almost a consensus that the presence of large-scale biogeographical units confine species with climate as their primary predictor. For example, Bontager et al.[11] suggested distinct characteristics for range edge populations dependent on their latitude. Likewise, niche conservatism and strong beta-diversity patterns seem to withhold at large macro-ecological scales[16–18]. This is generally true for plant species, with biomes being the most consistent classifier based on structural and functional similarity[19,20]. Recent efforts have been made to understand how accurate and substantial biome entities are at defining species distributions [20,21]. Nevertheless, it remains an open question if, and to what extent, the intersection between biomes is a source of species range edge hotspots. Deciphering such patterns will enable the proper understanding of how communities of species redistribute and are structured geographically, and if similar biomes in distinct geographic areas have similar effects on the distribution of species.

However, although there has been an increased interest in defining the biogeographical underpinnings of species distributions, most techniques have used species relatedness and diversity metrics to test for the existence of shared niche space between species and communities. The direct analysis of the ecological and climatic limitations to geographic space, although trivial, remain elusive. Here, we look at the universal set of climatic factors and geographical patterns of species distributions by focusing directly on species' range edge distributions.

We present the first-to-date global study of tree species range edges, applying a novel, simple, yet effective method of delineating range edges (REs) in order to: (1) identify deterministic patterns of REs, as seen by RE-dense areas (RE hotspots: REH); (2) determine whether the classification of biomes as distinct community-level patterns of biodiversity properly delineate the niche of species; (3) identify global-scale REH patterns, and; (4) discern the underlying niche factors responsible for RE formation. In particular, we investigate whether spatial heterogeneity of abiotic factors or a universal predictor for habitat quality are determinants of RE formation. We focus specifically on tree species, as they are an exemplary group of species in the study of the ecological changes predicted to occur at the peripheries of distributions, given their long-lived characteristics and fundamental role in many ecosystems[25,26], specifically biomes [18,27].

Although we do not study the interplay between climate and other ecological and evolutionary limiting factors to species distributions (e.g., seed dispersal, plasticity, and adaptation), discerning these patterns and the climatic components leading to such distributions will enable a better understanding of interplays between biotic and abiotic factors in future studies. A better understanding of the climatic factors affecting dispersal enable for better predictions of the success of species to track changing climates, and in turn, if they will be subject of migration lags[22–24].

Results

Global data set and REH distribution

We present the first report of the global distribution of tree range edge hotspots (REH) (Fig. 1), marked by the hexagons with significant clustering of REs (Supplementary Fig. 1). We did not identify any significant coldspots, given the baseline presence of REs around the globe. The visual patterns emerging from these distributions indicated that distributions stopped disproportionally more at the edges of biomes than within them. For example, northern REH occurred mostly at the intersection between a montane (Himalayas) and a desert biome (Gobi Desert) or at the edge of the tundra in North America. Southern REH in Africa and southern Eurasia tended to stop at the edges of desert biomes (e.g. Sahara Desert) but stopped mainly at the edges of temperate and montane grasslands in the Southern Hemisphere (pampas and Andes, respectively). Eastern and western REH were notably

mostly in similar geographical locations, e.g. at the intersections between the Himalayas and central Asian deserts or at the edge of the Atacama Desert in South America.

In order to identify the underlying niche factors that define REs, we focused only on inland REs, since REs at the edge of a water source is probably due to an obvious geographical barrier rather than an ecological effect. The fraction of inland REs was globally similar between the continents (between 65 and 75%), except for Australia (~45%, Supplementary Fig. 2a). The large fraction of inland REs (Supplementary Fig. 2b) was mainly associated with the edges of biomes (e.g. northern REs in Africa, Asia, and North America were associated with the Sahara Desert, montane grasslands of the Himalayas, and the tundra biome, respectively; Fig. 1). A significant fraction of eastern inland REs were in Europe, but almost no significant REH were identified, implying the sparse distribution of REs throughout the continent, or as an effect from a smaller area for distribution compared to other continents.

Contribution of biome-biome intersections to REHs

We quantitatively identified the global patterns of arrest by analysing the fraction of REH that stopped at biome-biome intersections (14 central global ecological regions best distinguished by their climate, fauna, and flora obtained from World Wild Life (WWF, http://www.worldwildlife.org/; Fig. 1) and by delineating a buffer zone at points of intersection between two or more biomes. Although not significant, the number of REH was strongly associated with the intersection between biomes (Fig. 2a, first panel), indicating that climatic conditions were a probable cause for the REs of tree species at biome edges. Our results, however, identified an unequal contribution of the different biome edges to the fraction of REH (Fig 2a). We then analysed the individual biome-biome intersections normalised by a *global* permutation (i.e., from permutation of the global distribution of REH, see Materials and Methods). Here we identified (1) a strong contribution of REH at the intersections between temperate and desert biomes (Fig. 2b, left panel) in comparison to (2) a weaker contribution at the intersections of tropical and subtropical biomes (between themselves and with temperate biomes). The significance of the contribution of REH (p<0.05) (see Materials and Methods) was attributed almost exclusively to the intersections within temperate biomes (asterisks in Fig. 2b). This unequal contribution between the temperate vs tropical and subtropical biomes was globally consistent. Nevertheless, under a

143 *per-biome* permutation (i.e., normalising each intersect by a selective permutation from the
144 respective biome combinations separately; Fig 2b right panel), we observed a much weaker
145 contribution of range edges to the formation of biomes. Only a selected number of biome-
146 biome intersections had a significant contribution to the formation of REH, yet no specific
147 pattern between temperate and tropical biomes was observed. Indeed, the strong positive
148 correlation between the number of REs between biomes with the number of REs within
149 biomes (Fig. 2c; $R = 0.9$, linear regression, $p < 0.001$) reflects such discrepancy between
150 panels in Fig 2b, as the number of REH at the edge of biomes had a strong linear association
151 with the number of REH within that biome. We identified five biomes, however, outside the
152 95% confidence interval (CI) of the regression (Fig. 2c). The two biomes above the
153 regression CI (desert; and boreal) were biomes where the number of REH at the edges was
154 larger than predicted (Fig 2c). In contrast, temperate coniferous forest, mangrove and flooded
155 grasslands fell below the CI of the regression, indicating a much larger number of REH
156 within the biome compared to the edge.

157

158 A parallel analysis using the distribution of REs (rather than REH) was conducted to identify
159 similarities and differences between the distributions. The results indicated a similar pattern
160 of distribution (Supplementary Fig. 4a-b compared to Fig. 2b-c), suggesting global forces
161 associated with REH.

162

163 Climatic predictors of RE formation

164 Intersections between biomes were a significant cause of RE formation, so we also
165 investigated the dependence of RE formation on climate. We tested both the '*absolute*
166 *climate*' (i.e. annual and seasonal average temperature and precipitation) and the '*spatial*
167 *heterogeneity of climate*' (*SH*, i.e. spatial variability of absolute climate) at each of the global
168 hexagons using 19 bioclimatic variables obtained from WorldClim (see Materials and
169 Methods). Elevation (*absolute* and *SH*) and latitude were also accounted for at each hexagon
170 resolution. Generalised linear models of regression between each climatic variable as an
171 independent predictor indicated that all 40 climatic variables were significant predictors of
172 RE formation ($p<0.05$) (Supplementary Fig5a). Interestingly, most *absolute* climatic
173 variables were negatively associated with the global RE distribution (Fig. 3a and b), implying
174 a general prediction of REs occurring in climates with low temperatures and low

precipitation. A positive association was attributed to the four climatic variables that defined the temporal heterogeneity of temperature and precipitation (mean diurnal range, temperature seasonality temperature annual range and precipitation seasonality; BioClim variables 2, 4, 7, and 15, respectively; Fig. 3a, Supplementary Fig. 5) and all of the *SH* climatic variables. A mixed model with biomes and continents as random factors gave a reduced number of variables that significantly predicted the formation of RE (Fig. 3a). A crossed model with both continents and biomes consistently gave a stronger fit (AIC values) than only considering either of these categorical random factors independently (Supplementary Fig 6b). In this more stringent global analysis, latitude, *absolute* temperature and *SH* temperature, precipitation and elevation were significant predictors of REH. The goodness of fit (R^2) of each model, as a measure of predictive strength, indicated that spatial heterogeneity accounted for RE formation better than did their absolute equivalents. A model selection was carried out to identify the most important factors associated with RE formation, followed by a model averaging of the models with a $\Delta AIC < 2$, (Figs. 3c). The *SH* climatic variables again defined RE better than the *absolute* climatic variables. Although *absolute* climatic variables such as isothermality (BioClim3), temperature of the wettest quarter (BioClim8) (both characteristic of tropical and subtropical climates), and annual precipitation (BioClim12) were strongly associated with REs in a generalised linear model, temperature of the warmest quarter (BioClim10) was the only predictor strongly associated with RE under the mixed model (Figs. 3a and S6). This difference in results can be visually seen when comparing between continents (Fig. 3b) e.g. in panel 3; with temperature of wettest quarter (BioClim8) having partial dissociations in Africa and South America, i.e. the two continents with the most tropical biomes. Furthermore, *SH* isothermality (BioClim3), temperature of coldest month (*SH* BioClim 6), precipitation (*SH* BioClim13 and 16) and elevation change (*SH* elevation) were all predictors of REs. Interestingly, the *absolute* climatic predictors (BioClim8 and BioClim10) were strongly negatively correlated with *spatial heterogeneity* at the minimum temperature of the coldest month (*SH* BioClim6) (Supplementary Fig. 7b), indicating its representation of a tropical biome climate. The weaker relative importance of *SH* precipitation (Fig. 3c, left panel) is due to its dissociation with the continents that most strongly represent temperate regions; Europe and North America (Fig 3b, panel 4).

In parallel, we ran models with the ENVIREM dataset, a dataset of environmental variables complementary to WorldClim that are more ecophysiologically meaningful for plant species[28]. Most ENVIREM variables associated with REH are spatially heterogeneous

variables (Fig. 3c, Supplementary Fig. 7). *Absolute* potential evapotranspiration (PET) of the coldest quarter was strongly correlated with both *absolute* temperature of the warmest quarter (BioClim10) and *SH* temperature of the coldest month (BioClim6); the two predictors indicative of a transition between tropical and temperate biomes (Supplementary Fig. 7b). *SH* EmbergerQ and *SH* PET of warmest quarter were strongly associated with *SH* precipitation (BioClim13 and 16) and *SH* temperature of the warmest quarter (BioClim10), respectively. Albeit the strong correlation of all *absolute* ENVIREM variables with *absolute* BioClim variables[28], we identified several *SH* ENVIREM variables to be weakly correlated to BioClim (e.g., *SH* moisture index and PET of driest month). All results in Fig. 3 were robust against spatial autocorrelation. (Supplementary Fig. 5b, see Materials and Methods for further information).

The REH distribution from our generated polygons was compared to a REH distribution from expert polygons (see Materials and Methods), in order to test for the accuracy of our generated dataset. Indeed, results of GLMs converge, indicating the robustness of our generated dataset to describe REs (Supplementary Fig. S8).

Discussion

The results support our hypothesis of a nonrandom distribution of REs. We were able to confirm that many tree species range edges were clustered rather than sparsely (stochastic) distributed, by obtaining a larger number of significant REH. This finding suggests the underlying presence of ecological and evolutionary forces governing REH formation. Similarly, the matching results obtained from the biome analyses when accounting for REs (Supplementary Fig. 4) or REH (Fig. 2) was also indicative of the deterministic clustering of species REs at these specific ecological barriers (biome edges). Nevertheless, we identified case-specific exceptions such as the scattered distribution of eastern REs throughout Europe, indicated by the strong identification of internal REs (Supplementary Fig. 2b) but not REH (Fig. 1c). This could potentially have occurred as a consequence of either the intense anthropogenic activity throughout[29,30] that prevented the distribution of tree species to reach their natural REs, or the effect of smaller land area compared to other continents, altering the effect of how species interact with abiotic factors, thus affecting their distribution and adaptation.

Our results also indicated a strong dependence of REH on the edges of biomes, strongly supporting the many efforts to determine whether the division of the planet into discrete geographical units has been appropriately delineated[21,31–34]. Our findings, however, were unexpectedly biome-specific, identifying the biome borders that most defined tree distributions (Fig. 2). There is no seemingly obvious pattern of differential contribution to REH when analysing each biome independently (*per-biome* bootstrap; Fig 2a). Yet, a clear distinction between a tropical/temperate distribution of REH is strongly observed under a biome-biome pairs analysis (*global* bootstrap; Fig 2b). Specifically, REs were strongly dependent on desert, temperate, and montane edges compared to the weak dependence on tropical and subtropical biome edges. Despite the marginally significant association of range edges to biomes globally (Fig 2a), a tropical-temperate REH distribution is not specifically dependent on biome edges, but rather, it represents a global biome trend (Fig 2b, left panel; Fig. 2c). In addition, the importance of biomes in describing climate's association to REs, as seen from a mixed effects model (Fig. 3a, S6c), as well as the absence of latitude in predicting REs in a global model (Fig 3c) is further indicative of the importance of biomes (rather than a latitudinal effect) in defining REs.

The large differences in the relative number of REs between the tropical and temperate biomes can indicate adaptive mechanisms between species residing in either of these two types of biomes. These nontrivial results may have been related to the Latitude Diversity Gradient[35–37], a well-established pattern in which biodiversity is higher in the tropics than in temperate regions. Firstly, the notion that these differences surged through the effective evolutionary time hypothesis[36,38] suggests that genetic diversity is higher in the tropics due to the possible longer times needed for adaptation and expansion. Similarly, long-term climatic oscillations have been suggested to reduce cladogenesis at higher latitudes, consistent with the observation that tropical areas are centers of evolutionary novelty[36,39,40] and evolve faster than temperate regions[35], leading to the notion of tropics as centers for radial expansion of clades and species. Recent studies have identified mechanisms for this 'out of the tropics expansion' model[41,42] and have reported a higher fraction of bridge species (species violating niche conservatism) from the tropics compared to temperate regions. We conclude that the lower contribution to REs in tropical vs temperate regions is consistent with stronger radial expansions from the tropics than temperate regions.

However, other mechanisms could explain the distinctive REH patterns between the tropic and temperate regions. For example, based on several studies demonstrating the large physiological and evolutionary effects of forest fragmentation[29,43,44], the increased long-term anthropogenic activity, and consequently excessive fragmentation, in forests in temperate regions compared to tropical and subtropical regions may have hindered the adaptation of species to novel climates. The lack of association of REH specifically to temperate grasslands, temperate conifers and Mediterranean biomes specifically (Fig 2a) could be indicative of such an effect, as these biomes have been historically subject to strong anthropogenic activity.

Our analyses confirmed the strong dependence of RE on climate. Lower temperature, higher climatic heterogeneity (both temporal and spatial), and elevation changes were the strongest climatic predictors of REs. The analysis of the best predictive variables indicated a noticeable weak correlation between most of the leading factors determining REs (Supplementary Fig. 7b), suggesting that these factors were site-specific predictors of REs (e.g. desert biome with low annual precipitation, montane grassland with high elevation, and tropical and subtropical biomes with temperature homogeneity (isothermality, temperature of the wettest quarter and spatial heterogeneity of low temperature – *SH6*)). We suggest that the strong negative correlation between these absolute climatic variables and the spatial heterogeneity of low temperature is an indication of different temperature patterns between tropical and temperate regions, i.e. a buffered heterogeneous temperature in the tropics in contrast to the latitudinal effect of decreasing temperatures in the temperate biomes. There was an overall positive trend for the effect of precipitation (BC12-BC19) when controlling for biome since the signal appeared only when biome is included as a random factor (Supplementary Fig. 6). This means there must be other non-accounted confounding factors altering the relationship when analysing all biomes together, suggesting that the effect is not biome dependent. Although this study does not make a note of the complex interplay between biotic and abiotic forces or the ecological traits of tree species (e.g. seed dispersal, phenology) in the formation of RE, we find a consistent global effect of temperature and spatial heterogeneity of temperature and precipitation to predict their formation throughout the different models. This is indicative of their principal role as universal predictors of species distributions.

The strong prediction of REH formation from PET of the coldest quarter and its covariance to the tropical-temperate transition variables (*SH* BioClim6 and *absolute* BioClim10), could

reflect a possible mechanistic evolutionary constrain for the distribution of woody species' at temperate biomes. At higher latitudinal temperate biomes, where evapotranspiration is strongly reduced, especially during the coldest quarter, there is a strong limitation to photosynthesis and growth by the significant time reduction of stomatal conductance[45,46]. Similarly, given the strength of its prediction of REH, *SH* embergerQ (pluviothermic quotient) could indicate a more refined mechanism for RE formation than its covariates of precipitation (*SH* BioClim 13 and 16). This index describes mean annual precipitation in relation to annual changes in temperature. EmbergerQ thus increases the predictability of how precipitation also dictates the formation of REH in the more temperate biomes (Fig 3b, panel 4). The consistency of our results indicates spatial and temporal heterogeneity of climate and topography as overwhelmingly stronger predictors of RE formation than their absolute climatic counterparts (Fig. 3) and, in turn, frail evidence for a universal poor habitat quality. Even in the cases where *mean* climatic variables strongly predict REH, these were strongly associated with this transition between the spatial and temporal climatically homogenous tropical and subtropical biomes to the more heterogeneous temperate biomes. Nevertheless, we note the importance of a lack of evapotranspiration, particularly in cold climates, as a main predictor of RE formation.

These observations have substantial implications for the effects of climate change on tree distributions and its effects on tree migration. Although predictions for future steeper temperature gradient as a result of greenhouse gas emission and climate change has not been trivial[47–49], such an increase in temperature gradients could vastly affect the distribution of species. In particular, our results strengthen the growing understanding that the predicted poleward migration of tree species might not be as successful as previously predicted[23,24]. The increase of stronger spatial gradients (especially in the lower latitudes)[24,48,50] or extreme and spontaneous events might all be causes of migration lags, despite the suitable temperatures at higher latitudes and altitudes. Likewise, the importance of PET from temperate biomes on the formation of REH presented here could also suggest a possible migration lag or loss of adaptation due to the predicted reduction in PET at higher latitudes[51]. Our results thus highlight the importance of accounting for more precise spatial heterogeneity of climate as a critical feature in future models of species distribution and the development of more precise conservation efforts such as assisted migration.

Materials and Methods

343 *Data acquisition and polygon formation*

344 Supplementary Fig. 9 visually summarises the methodologies used to obtain global range
345 edge hotspot (REH) distributions. We downloaded a data set of tree species from the open-
346 source data set using R packages *rgbif* and *taxize*. Global Biodiversity Information Facility
347 (GBIF; 05 July 2021, https://doi.org/10.15468/dl.ajen6k) using the Botanic Gardens
348 Conservation list of 60000 tree species. We downloaded occurrences with entries from 1980
349 onwards, removing any occurrence reported with a *geospatial issue*, species not belonging to
350 the kingdom Plantae (in case of mismatched species names), and any occurrence marked as
351 *unlikely*, *mismatched*, or *invalid*. We removed occurrences that had reported uncertainties of
352 >100 km and records based on fossils and unknown sources. We then used the
353 *CoordinateCleaner* R package[52] to remove any occurrences with zero coordinates, equal x
354 and y coordinates, duplicates, occurrences at sea, coordinates at capitals, and centroids. To
355 finalise, we again removed species with <300 occurrences. The data we used undoubtedly
356 contained sampling bias[53], probably overrepresenting the number of REs in some regions
357 with a reduced or negligible sampling effort. We tried to overcome this issue by basing our
358 filtering steps on several previous studies[52–54]. The strength of the critical filtering steps
359 applied in our analysis resembled those previously presented[53].

360 We converted the georeferenced species occurrences (x and y coordinates) into distributional
361 polygons in parallel using two independent techniques; through concave-hull
362 (Supplementary Methods) and multivariate kernel-density estimation (described below).
363 Given the strong similarity between the two methods (Supplementary Fig. 8), we discuss the
364 methods and results in detail only for the kernel-density estimated polygons. We created
365 polygons using two-dimensional kernel density estimations. We first divided the extent of all
366 the coordinates into 800 grid points in each dimension (longitude and latitude) in order to
367 produce a matrix of 640,000 grid cells for each of the species. Subsequently, we selected for
368 the grid cells with the highest 99% estimation of the species' occurrence and subsequently
369 rasterised these. Polygons were then delineated around the contour of the rasters.

370

371 *Polygon groupings*

372 All polygons belonging to the same species were grouped based on absolute distance from
373 one another. Polygons separated by ≤500 km were grouped together, with the assumption that
374 fragmentation, gene flow, and unreported data could all warrant two nearby populations to be
375 considered as one. We used the hclust function (package stats, agglomeration method:

complete) to hierarchically cluster populations from a sequence of three or more populations by their distances, also using a cutoff of 500 km (cutree function, package stats) for determining the clusters. The final data set comprised >3600 tree species, ranging from one to nine populations (polygons) per species, for a total of 8500 populations. All spatial data was analysed using R packages *sf* and *raster*.

RE determination

RE dense areas were determined by (1) defining distinct global units, (2) identifying the RE of each species and (3) map species' RE to the global units to calculate the density of REs/unit. In detail: (1) We rasterised the world map to spatially bin the density of REs. To overcome the problem of spatial distortion, we used hexagonal bins with the *dggridR* R package, developed using the ISEA Discrete Global Grids system, a repetition of polygons on the surface of an icosahedron, allowing for the projection of equal sized bins onto a 2D plane. We defined the size of each hexagon as ~23 000 km^2 (with an average spacing between center nodes of 165 km). (2) We used coordinates of the cardinal directions (north, south, east, and west) to represent species REs by subdividing each polygon cluster into four quartiles in the four cardinal directions (NE, NW, SE, and SW). The REs for each quartile were determined as the two most-outward coordinates of the corresponding cardinal directions (e.g. north and east cardinal coordinates for the NE quartile). Eight REs were thus determined for each population (two for each cardinal direction). As a filtering step, we accounted for both REs from the same cardinal direction if they were >20 arc-degrees apart, otherwise we only accounted for the farthest point from the centroid. (3) The total number of REs obtained using this method was normalised by the total number of species intersecting its respective hexagon. In parallel, we also defined REs by accounting for the perimeter of the polygon for each species (Supplementary Fig. 10). The 'perimeter' system may be a more realistic and complete system for identifying REs, but the 'cardinal coordinate' system, although more simplistic in nature, (1) provides a clearer visual representation of the distribution of REs, and (2) allows for the directionality of REs to be compared, essential farther along the pipeline by distinguishing between coastline and inland REs and identifying hotspots in the four cardinal directions.

We then classified coastline and inland REs, assuming that the arrest of species distribution at the edge of a water source was probably due to an obvious geographical barrier rather than an ecological effect. Coastline REs were determined by creating a semicircle (buffer of 3 arc-

degrees) around each RE in the direction of its cardinal coordinate and measured the percentage overlap with water. Cardinal coordinates with >50% overlap were considered a "coastal RE".

In order to find the probability distribution of REs at the edge of water source (Supplementary Fig. 2), we permutated the global population of REs (except for the Australian population) and used the mean of this permutation to compare to the number of REs in each of the continents.

Hotspot analysis

Hotspots were identified using the Getis-Ord Gi* hotspot analysis[55,56] to find spatial correlations between hexagon (inland and normalised) RE densities. We initially compiled a list of neighbors between all hexagons using the poly2nb function and then obtained the local G statistic using the weighted density (normalised number of REs) of the global hexagons and their relative distance from each other. The G statistic calculates a Z-score (measure of standard deviation) for each hexagon. *P* values were then determined using the critical Z-scores at 95% confidence levels followed by a Bonferroni correction using the p.adjustSP function (using the number of neighbors between hexagons rather than the total number of hexagons). All analyses were carried out using the R *spdep* package[57].

We compared the analysis from the linear models with expert based polygons from three different sources – IUCN, BIEN and EUFORGEN. A randomised weighted sample of all of this dataset was used to generate a global distribution of REH by running this sample through our pipeline. GLMs were run on the global distribution of expert-based REH in the same way as with our generated polygons. Given the uneven distribution of expert-based polygons globally (Supplementary Fig. 8a), we ran models excluding Asia and Africa, in order to account for this bias. As seen by the strong similarity between the GLMs of expert polygons in a global and filtered model (Supplementary Fig 8b and Fig 8c, panel 3), we observed an overrepresentation of the expert polygons for these continents. Likewise, the GLMs from our generated polygons are much similar to those obtained from the filtered model. In this case, practically all variables showed the same relationship with range edges (either positive or negative β values) as well as similar magnitudes.

Statistical analyses

Contribution of biome edge to RE. We used the 14 biomes defined by the World Wildlife Fund (WWF) for our analyses. The distributions were downloaded from the WWF webpage (http://www.worldwildlife.org/). To identify the intersection between biomes, we reduced the complexity of the polygon edge using the rmapshaper package, which can perform topologically aware polygon simplifications, thus maintaining the intersection between biomes upon reduction of "edginess" of the polygons. The intersections were delineated and subsequently enlarged (with a buffer distance of 0.1 arc-minute, ~185 m at the equator). A:

i. *Global* permutation assay was carried out by randomising (1000 iterations) the global distribution of hexagons with REH (absolute Z-score)
ii. *Per-biome* stratified permutation was carried out by randomising the distribution of hexagons with REH within each biome independently.

The averaged global bootstrap shown in panel 1 of Fig 2a, was calculated using a *per-biome* bootstrap to obtain the probability distribution of REH at biome edges (Supplementary Fig. 3a). Distributions that were not normally distributed as a result of their small size (flooded grassland, mangrove and tropical and subtropical coniferous forest; Supplementary Fig 3b), were removed from the analysis. A general trend for the probability of range edges falling at the intersection of biomes was therefore measured as a unified standardised z-distribution, and compared to the median z-score from the actual percent overlap for each biome.

The density of hotspots at the intersection between biomes was calculated using the sum of Z-scores of the hotspots at that intersection, and the percentage contribution was then calculated using this value over the total Z-score at all biome intersections. The *global* or *per-biome* 1000 permutation means were used as a normalising denominator for the values obtained from our data set (Fig 2b, Supplementary Fig 4a). The denominator could be either larger or smaller than the numerator, so we log-transformed the outcome to obtain a linear-like relationship. Contribution within a biome was calculated the same way as for the contribution at the edge, using the mean from a permutated assay to normalise for the absolute value.

Climatic dependency of RE. We tested the relationship between RE density and climatic features by assigning a set of environmental variables to each hexagon. We used the bioclimatic attributes downloaded from WorldClim Global Climate Data[58] at a resolution of 5

arc-minutes. The 19 BIOCLIM variables and elevations for each hexagon were extracted using the R raster package. We also used the 16 ENVIREM variables described by [28], downloaded from their website, at a resolution of 2.5 arc-minutes. *Absolute* climate for each hexagon was obtained using the mean over all pixels. Climatic *spatial heterogeneity* (*SH*) was calculated using the proportional variability index[59,60] (PV) over all pixels in each hexagon.

We used an array of linear mixed models (Fig 3, Supplementary Fig 5 and 6) to test for the dependence of REs to climate and the robustness of the results. Linear mixed models were carried out to account for biomes and continents. Both of these variables were introduced as random effects in random intercept models. A model selection analysis was used to determine the models that best predicted the formation of RE. Random intercept models using both continent and biome as random variables were run, and models with $\Delta AIC < 2$ were selected for. Given the strong correlation between different predictor variables, we ran models only with variable combinations that had a Pearson's correlation value $r < 0.7$. The relative contribution of the variable included in the model were calculated from the selected models. Analyses were run using R package *MuMIn*[61].

All statistical analyses (individual GLMs, and multiple-predictor GLMs) were tested for their robustness to spatial autocorrelation by creating a spatial autocovariate (autocov_dist function, spdep package), calculated as the distance-weighted average of neighboring dependent variables[62], so hexagons in proximity were averaged and those farther away received a lower weighting average. We set the predetermined distance to 200 km based on the average distance between cells. The spatial autocovariate was then included in the regression model as a dependent variable.

Data Availability

The occurrence points used from GBIF can be found in the GBIF webpage (https://doi.org/10.15468/dl.ajen6k). Polygons generated from occurrence points are provided in the public Zenodo repository 10.5281/zenodo.7613535. Biome polygons were obtained from the WWF webpage (http://www.worldwildlife.org/). Bioclimatic attributes were downloaded from WorldClim Global Climate Data[58]. ENVIREM variables were downloaded from their webpage (https://envirem.github.io/).

Code Availability

Custom codes related to this paper can be found in a GitHub repository at https://github.com/dlernerg/Global-Range-edges

Acknowledgments

TK wishes to thank Edith and Nathan Goldenberg Career Development Chair; Mary and Tom Beck-Canadian Center for Alternative Energy Research; Larson Charitable Foundation New Scientist Fund; Yotam Project; Dana and Yossie Hollander; Estate of Emile Mimran; and the Estate of Helen Nichunsky. DL was supported by the Sustainability and Energy Research Initiative PhD grant. MF-M and JP were supported by the PID2019-110521GB-I00 and TED2021-132627B-I00 grants funded by MCIN, AEI/10.13039/501100011033 and the NextGeneration EU/PRTR. MF-M. was supported by a postdoctoral fellowship from "la Caixa" Foundation (ID 100010434), code: LCF/BQ/PI21/11830010.

Author Contributions: D.L., J.B. and T.K. designed the research; D.L. performed the research and analysed the data; M.F.M, J.P., T.K., J.B., and S.L.L. provided scientific advice; M.F.M. and J.B. advised on statistical aspects; D.L. wrote the paper with special contribution from J.P., J.B., M.F.M. and T.K.

Competing Interest Statement: The authors declare no conflict of interest.

Figure 1. Range-edge hotspots. Hotspots were identified using Getis-Ord Gi* analysis, which returns a Z-score for each hexagon in the world. Only hexagons with $p<0.05$ are considered hotspots and subsequently shown here. The Z-score of each hexagon is represented by the color gradient. Biomes (as defined by the World Wildlife Fund) are marked by colors. 'T&sT' and 'Temp' stand for 'Tropic and Subtropic' and 'Temperate', respectively.

Figure 2. Range-edge hotspots at intersections between biomes. (**a**) Modeled distribution of the percentage of hotspots at the edges of biomes from a permutated (randomised) *per-biome* distribution of hotspots. The first panel represented the median value of all the other biomes in the figure over a standardised z-distribution of biomes. The arrow marks the percentage of RE hotspots at biome intersections in the data set (one-sided *p-values*) (**b**) Heat maps of the percentage of RE hotspots (relative to the total number of hotspots) at the

intersection between two biomes. A biome-pair intersection with a significant number of hotspots (*p-value* < 0.08) is marked with an asterisk. Biome-pairs that have no intersections are gray. Panel 1 is normalised over a *global* bootstrap and panel 2 over a *per-biome* bootstrap. (**c**) Correlations (and regression lines) of the relationships between the number of REs at the edges of biomes and the number of REs within the biomes. The shaded area represents the 95% confidence interval around the regression line. *P-values* are calculated using a two-sided Student's T-test (*degrees of freedom* = 12). See Methods and Materials for further information on the methodology for obtaining significance levels (for (**b**)) and hotspot permutations. 'T&sT' and 'Temp' stand for 'Tropic and Subtropic' and 'Temperate', respectively.

Figure 3. Climatic predictors of range-edge formation. Models of RE formation using the absolute climate (mean) and the spatial heterogeneity (PV index) of the 19 WorldClim variables and elevation. (**a**) Individual (binomial) mixed regression models between each of the predictor variables and number of REs (1-19 are BioClim variables) and accounting for continents and biomes as random effects. The estimated coefficients of the explanatory variables (β) are represented by the color gradient. '*' and '**' represent *p-values* <0.05 and <0.01, respectively (two-sided Student T-test).
BioClim1 – Annual Mean Temperature, BioClim2 - Mean Diurnal Range (Mean of monthly (max temp - min temp), BioClim3 – Isothermality, BioClim4 - Temperature Seasonality, BioClim5 - Max Temperature of Warmest Month, BioClim6 - Min Temperature of Coldest Month, BioClim7 - Temperature Annual Range, BioClim8 - Mean Temperature of Wettest Quarter, BioClim9 - Mean Temperature of Driest Quarter, BioClim10 - Mean Temperature of Warmest Quarter, BioClim11 - Mean Temperature of Coldest Quarter, BioClim12 - Annual Precipitation, BioClim13 – Precipitation of Wettest Month, BioClim14 – Precipitation of Driest Month, BioClim15 – Precipitation Seasonality (PV), BioClim16 – Precipitation of Wettest Quarter, BioClim17 – Precipitation of Driest Quarter, BioClim18 – Precipitation of Driest Quarter, BioClim19 – Precipitation of Coldest Quarter.
(**b**) Violin plots depicting the results from (**a**) for four predictor variables. The distribution of continental climates is shown in gray, contrasted with the climatic distribution specific to the RE hotspots (scaled to the intensity of the hotspot, i.e. the Z- score). (**c**) Forest plot of a model average from the highest predicting LMM with both BioClim and ENVRIEM variables (identified with a model selection). Beta values (log-odds) are shown for each

predictor. Absolute and SH climate was obtained from all of the inland global hexagonal units (*n = 5851*). Error bars represent 95% confidence interval around the average effect.

References

1. Soule, M. The epistasis cycle: A theory of marginal populations. *Annu Rev Ecol Syst* **4**, 165–187 (1973).
2. Brown, J. H. On the Relationship between Abundance and Distribution of Species. *Am Nat* **124**, 255–279 (1984).
3. Gaston, K. J. The Structure and Dynamics of Geographic Ranges. *Oxford University Press* (2003).
4. Sexton, J. P., McIntyre, P. J., Angert, A. L. & Rice, K. J. Evolution and ecology of species range limits. *Annu Rev Ecol Evol Syst* **40**, 415–436 (2009).
5. Gaston, K. J. Geographic range limits: Achieving synthesis. *Proceedings of the Royal Society B: Biological Sciences* **276**, 1395–1406 (2009).
6. Goldberg, E. E. & Lande, R. *Notes and Comments Species' Borders and Dispersal Barriers*. *Am. Nat* vol. 170 (2007).
7. Bachmann, J. C., Rensburg, A. J. van, Cortazar-Chinarro, M., Laurila, A. & Buskirk, J. van. Gene flow limits adaptation along steep environmental gradients. *American Naturalist* **195**, E67–E86 (2020).
8. Hargreaves, A. L., Samis, K. E. & Eckert, C. G. Are species' range limits simply niche limits writ large? A review of transplant experiments beyond the range. *American Naturalist* **183**, 157–173 (2014).
9. Henry, R. C., Bartoń, K. A. & Travis, J. M. J. Mutation accumulation and the formation of range limits. *Biol Lett* **11**, 11DUUMY (2015).
10. Perrier, A., Sánchez-Castro, D. & Willi, Y. Environment dependence of the expression of mutational load and species' range limits. *J Evol Biol* **35**, 731–741 (2022).
11. Bontrager, M. *et al.* Adaptation across geographic ranges is consistent with strong selection in marginal climates and legacies of range expansion. *Evolution (N Y)* 1–18 (2021) doi:10.1111/evo.14231.
12. Santini, L., Pironon, S., Maiorano, L. & Thuiller, W. Addressing common pitfalls does not provide more support to geographical and ecological abundant-centre hypotheses. *Ecography* **42**, 696–705 (2019).
13. Oldfather, M. F., Kling, M. M., Sheth, S. N., Emery, N. C. & Ackerly, D. D. Range edges in heterogeneous landscapes: Integrating geographic scale and climate complexity into range dynamics. *Glob Chang Biol* **26**, 1055–1067 (2020).
14. Janzen, D. H. *Why Mountain Passes are Higher in the Tropics WHY MOUNTAIN PASSES ARE HIGHER IN THE TROPICS*. Source: The American Naturalist* vol. 101.

15. Maxwell, M. F., Leprieur, F., Quimbayo, J. P., Floeter, S. R. & Bender, M. G. Global patterns and drivers of beta diversity facets of reef fish faunas. *J Biogeogr* **49**, 954–967 (2022).

16. Roy, K., Hunt, G., Jablonski, D., Krug, A. Z. & Valentine, J. W. A macroevolutionary perspective on species range limits. *Proceedings of the Royal Society B: Biological Sciences* **276**, 1485–1493 (2009).

17. Loiseau, N. *et al.* Global distribution and conservation status of ecologically rare mammal and bird species. *Nat Commun* **11**, (2020).

18. Kerkhoff, A. J., Moriarty, P. E. & Weiser, M. D. The latitudinal species richness gradient in New World woody angiosperms is consistent with the tropical conservatism hypothesis. *Proc Natl Acad Sci U S A* **111**, 8125–8130 (2014).

19. Donoghue, M. J. & Edwards, E. J. Biome shifts and niche evolution in plants. *Annu Rev Ecol Evol Syst* **45**, 547–572 (2014).

20. Ringelberg, J. J., Zimmermann, N. E., Weeks, A., Lavin, M. & Hughes, C. E. Biomes as evolutionary arenas: Convergence and conservatism in the trans-continental succulent biome. *Global Ecology and Biogeography* **29**, 1100–1113 (2020).

21. Smith, J. R. *et al.* A global test of ecoregions. *Nat Ecol Evol* **2**, 1889–1896 (2018).

22. Zhu, K., Woodall, C. W. & Clark, J. S. Failure to migrate: Lack of tree range expansion in response to climate change. *Glob Chang Biol* **18**, 1042–1052 (2012).

23. Corlett, R. T. & Westcott, D. A. Will plant movements keep up with climate change? *Trends in Ecology and Evolution* vol. 28 482–488 Preprint at https://doi.org/10.1016/j.tree.2013.04.003 (2013).

24. la Sorte, F. A., Butchart, S. H. M., Jetz, W. & Böhning-Gaese, K. Range-Wide Latitudinal and Elevational Temperature Gradients for the World's Terrestrial Birds: Implications under Global Climate Change. *PLoS One* **9**, e98361 (2014).

25. Paquette, A. & Messier, C. The effect of biodiversity on tree productivity: From temperate to boreal forests. *Global Ecology and Biogeography* **20**, 170–180 (2011).

26. Pichancourt, J. B., Firn, J., Chadès, I. & Martin, T. G. Growing biodiverse carbon-rich forests. *Glob Chang Biol* **20**, 382–393 (2014).

27. Pennington, R. T., Lavin, M. & Oliveira-Filho, A. Woody plant diversity, evolution, and ecology in the tropics: Perspectives from seasonally dry tropical forests. *Annu Rev Ecol Evol Syst* **40**, 437–457 (2009).

28. Title, P. O. & Bemmels, J. B. ENVIREM: an expanded set of bioclimatic and topographic variables increases flexibility and improves performance of ecological niche modeling. *Ecography* **41**, 291–307 (2018).

29. Veresoglou, S. D. & Peñuelas, J. Variance in biomass-allocation fractions is explained by distribution in European trees. *New Phytologist* **222**, 1352–1363 (2019).

30. Grantham, H. S. *et al.* Anthropogenic modification of forests means only 40% of remaining forests have high ecosystem integrity. *Nat Commun* **11**, (2020).

31. Holdridge, L. R. Determination of world plant formations from simple climatic data. *Science (1979)* **105**, 367–368 (1947).

32. Whittaker, R. H. Classification of Natural Communities. *Botanical Review* **28**, 1–239 (1962).

33. McDonald, R. *et al.* Species compositional similarity and ecoregions: Do ecoregion boundaries represent zones of high species turnover? *Biol Conserv* **126**, 24–40 (2005).

34. von Humboldt, A. & Bonpland, A. *Essay on the Geography of Plants*. (2013).

35. Cardillo, M. Latitude and rates of diversifcation in birds and butterfies. *Proc. R. Soc. Lond. B* **266**, 1221–1225 (1999).

36. Hillebrand, H. *On the Generality of the Latitudinal Diversity Gradient*. *Am. Nat* vol. 163 (2004).

37. Mittelbach, G. G. *et al.* Evolution and the latitudinal diversity gradient: Speciation, extinction and biogeography. *Ecology Letters* vol. 10 315–331 Preprint at https://doi.org/10.1111/j.1461-0248.2007.01020.x (2007).

38. Hewitt, G. M. Genetic consequences of climatic oscillations in the Quaternary. in *Philosophical Transactions of the Royal Society B: Biological Sciences* vol. 359 183–195 (2004).

39. Crane, P. & Scott, L. Angiosperm Diversification and Paleolatitudinal Gradients in Cretaceous Floristic Diversity. *Science (1979)* **246**, 675–678 (1989).

40. Jablonski, D. The tropics as a source of evolutionary novelty through geological time. *P. J. IEEE Trans. Geosci. Remote Sensing* **361**, 180–191 (1993).

41. Jablonski, D. *et al.* Out of the tropics, but how? Fossils, bridge species, and thermal ranges in the dynamics of the marine latitudinal diversity gradient. *Proc Natl Acad Sci U S A* **110**, 10487–10494 (2013).

42. Antonelli, A. *et al.* An engine for global plant diversity: Highest evolutionary turnover and emigration in the American tropics. *Front Genet* **6**, (2015).

43. Jump, A. S. & Peñuelas, J. Running to stand still: Adaptation and the response of plants to rapid climate change. *Ecol Lett* **8**, 1010–1020 (2005).

44. Morreale, L. L., Thompson, J. R., Tang, X., Reinmann, A. B. & Hutyra, L. R. Elevated growth and biomass along temperate forest edges. *Nat Commun* **12**, 7181 (2021).

45. Wilkinson, S., Clephan, A. L. & Davies, W. J. Rapid Low Temperature-Induced Stomatal Closure Occurs in Cold-Tolerant Commelina communis Leaves But Not in Cold-Sensitive Tobacco Leaves, via a Mechanism That Involves Apoplastic Calcium But Not Abscisic Acid. *Plant Physiol* **126**, 1566–1578 (2001).

46. Brodribb, T. J. & Holbrook, N. M. Stomatal protection against hydraulic failure: a comparison of coexisting ferns and angiosperms. *New Phytologist* **162**, 663–670 (2004).

47. Davis, B. A. S. & Brewer, S. Orbital forcing and role of the latitudinal insolation/temperature gradient. *Clim Dyn* **32**, 143–165 (2009).

48. Seager, R. *et al.* Strengthening tropical Pacific zonal sea surface temperature gradient consistent with rising greenhouse gases. *Nat Clim Chang* **9**, 517–522 (2019).

49. Xu, Y. & Ramanathan, V. Latitudinally asymmetric response of global surface temperature: Implications for regional climate change. *Geophys Res Lett* **39**, n/a-n/a (2012).

50. Colwell, R. K., Brehm, G., Cardelús, C. L., Gilman, A. C. & Longino, J. T. *Global Warming, Elevational Range Shifts, and Lowland Biotic Attrition in the Wet Tropics*. vol. 322 www.sciencemag.org (2008).

51. Basso, B., Martinez-Feria, R. A., Rill, L. & Ritchie, J. T. Contrasting long-term temperature trends reveal minor changes in projected potential evapotranspiration in the US Midwest. *Nat Commun* **12**, 1476 (2021).

52. Zizka, A. *et al.* CoordinateCleaner: Standardized cleaning of occurrence records from biological collection databases. *Methods Ecol Evol* **10**, 744–751 (2019).

53. Serra-diaz, J. M., Enquist, B. J., Maitner, B., Merow, C. & Svenning, J. Big data of tree species distributions : how big and how good ? (2018) doi:10.1186/s40663-017-0120-0.

54. Zizka, A. *et al.* No one-size-fits-all solution to clean GBIF. *PeerJ* **8**, (2020).

55. Getis, A. & Ord, J. K. The Analysis of Spatial Association by Use of Distance Statistics. *Geogr Anal* (1992).

56. Mendez, C. Spatial autocorrelation analysis in R. R Studio/RPubs. *https://rpubs.com/quarcs-lab/spatial-autocorrelation* (2020).

57. Bivand, R. S., Pebesma, E. & Gómez-Rubio, V. *Applied Spatial Data Analysis with R*. (2013).

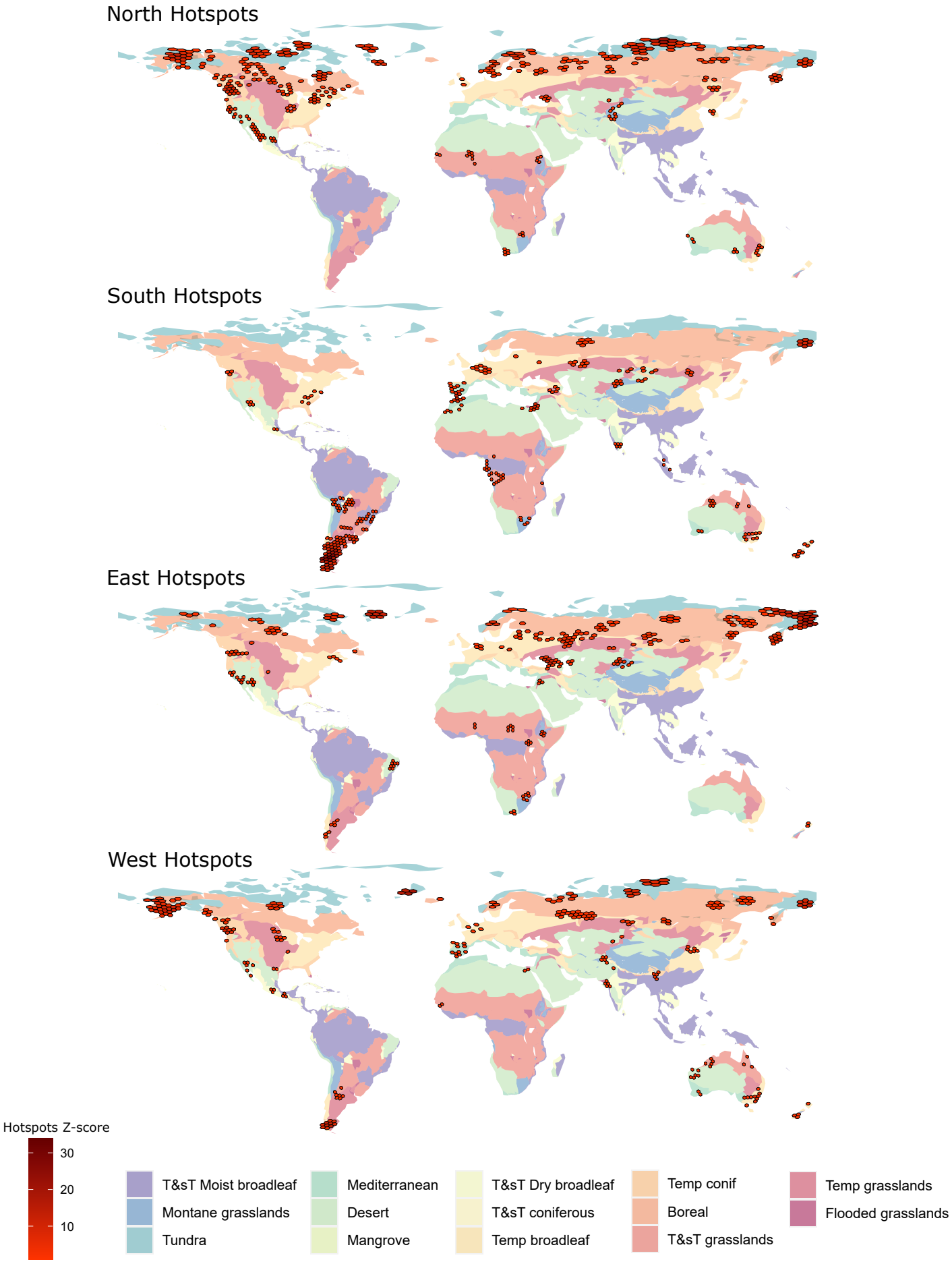
58. Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G. & Jarvis, A. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* **25**, 1965–1978 (2005).

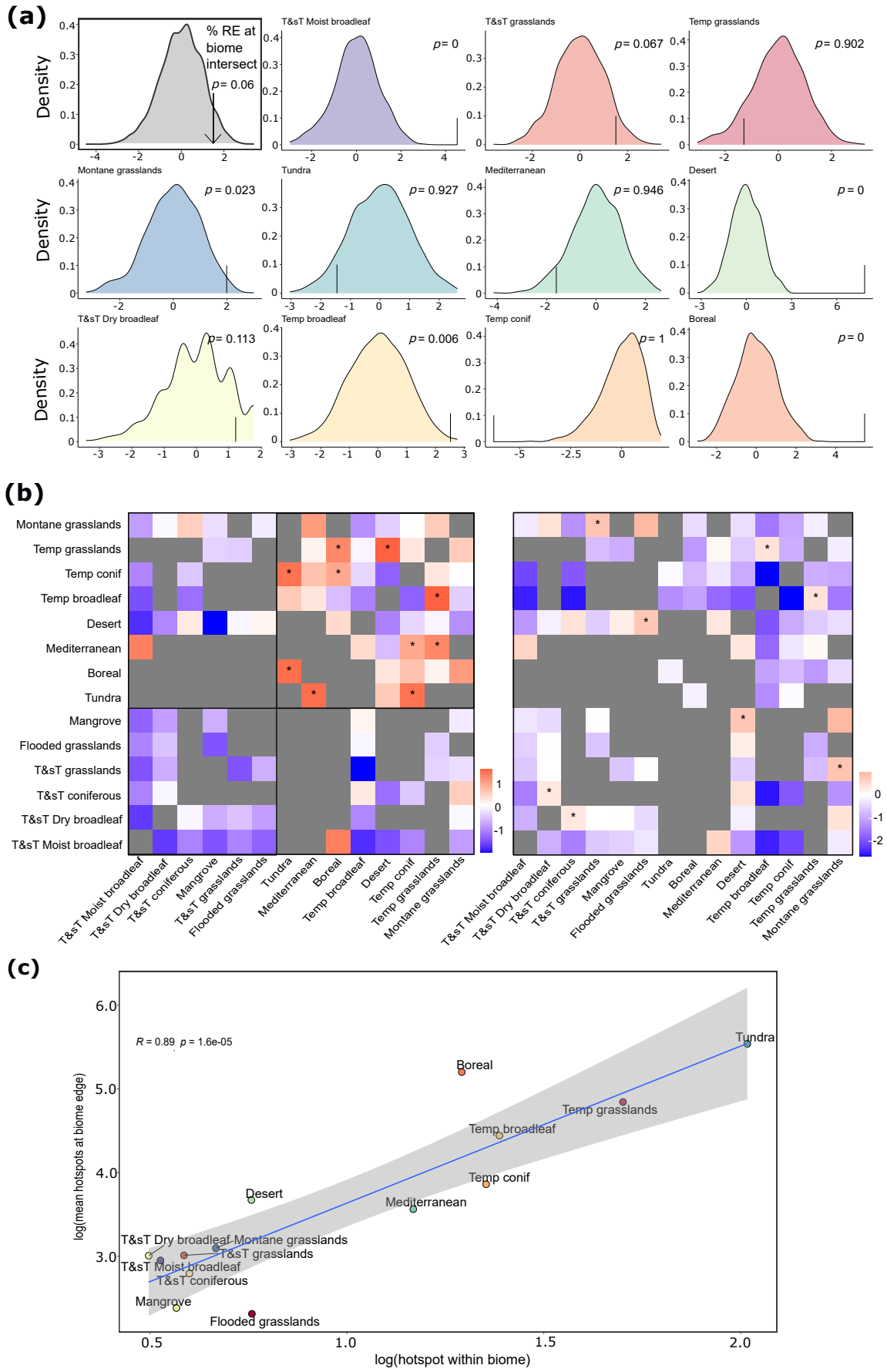
59. Heath, J. P. Quantifying temporal variability in population abundances. *Oikos* vol. 115 573–581 Preprint at https://doi.org/10.1111/j.2006.0030-1299.15067.x (2006).

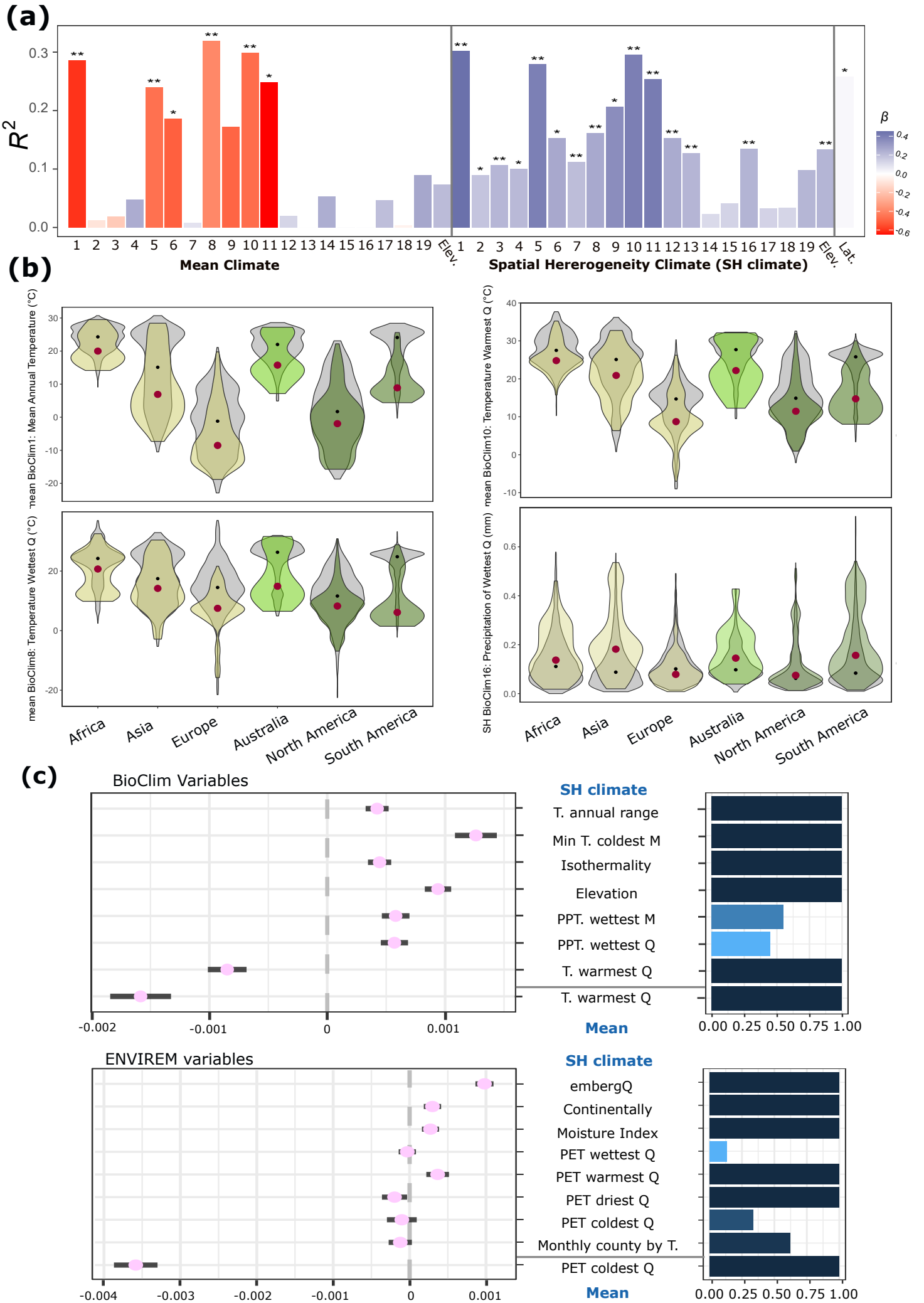
60. Fernández-Martínez, M. *et al.* The consecutive disparity index, D: a measure of temporal variability in ecological studies. *Ecosphere* **9**, (2018).

61. Bartoń, K. MuMIn: Multi-model inference. R package version 1.10.0.. 1. . (2013).

62. F. Dormann, C. *et al.* Methods to account for spatial autocorrelation in the analysis of species distributional data: A review. *Ecography* vol. 30 609–628 Preprint at https://doi.org/10.1111/j.2007.0906-7590.05171.x (2007).


North Hotspots
South Hotspots
East Hotspots
West Hotspots
Hotspots Z-score
30
20
10
T&sT Moist broadleaf
Montane grasslands
Tundra
Mediterranean
Desert
Mangrove
T&sT Dry broadleaf
T&sT coniferous
Temp broadleaf
Temp conif
Boreal
T&sT grasslands
Temp grasslands
Flooded grasslands


(a)
% RE at biome intersect
p = 0.06
T&sT Moist broadleaf
p = 0
T&sT grasslands
p = 0.067
Temp grasslands
p = 0.902
Montane grasslands
p = 0.023
Tundra
p = 0.927
Mediterranean
p = 0.946
Desert
p = 0
T&sT Dry broadleaf
p = 0.113
Temp broadleaf
p = 0.006
Temp conif
p = 1
Boreal
p = 0
Density
(b)
Montane grasslands
Temp grasslands
Temp conif
Temp broadleaf
Desert
Mediterranean
Boreal
Tundra
Mangrove
Flooded grasslands
T&sT grasslands
T&sT coniferous
T&sT Dry broadleaf
T&sT Moist broadleaf
(c)
R = 0.89 p = 1.6e-05
log(mean hotspots at biome edge)
log(hotspot within biome)
Tundra
Boreal
Temp grasslands
Temp broadleaf
Temp conif
Mediterranean
Desert
T&sT Dry broadleaf
Montane grasslands
T&sT grasslands
T&sT Moist broadleaf
T&sT coniferous
Mangrove
Flooded grasslands


(a)
R^2
0.0
0.1
0.2
0.3
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 Elev.
Mean Climate
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 Elev. Lat.
Spatial Hererogeneity Climate (SH climate)
β
0.4
0.2
0.0
-0.2
-0.4
-0.6
(b)
mean BioClim1: Mean Annual Temperature (°C)
mean BioClim8: Temperature Wettest Q (°C)
mean BioClim10: Temperature Warmest Q (°C)
SH BioClim16: Precipitation of Wettest Q (mm)
Africa
Asia
Europe
Australia
North America
South America
(c)
BioClim Variables
SH climate
T. annual range
Min T. coldest M
Isothermality
Elevation
PPT. wettest M
PPT. wettest Q
T. warmest Q
T. warmest Q
-0.002
-0.001
0
0.001
Mean
0.00 0.25 0.50 0.75 1.00
ENVIREM variables
SH climate
embergQ
Continentally
Moisture Index
PET wettest Q
PET warmest Q
PET driest Q
PET coldest Q
Monthly county by T.
PET coldest Q
-0.004
-0.003
-0.002
-0.001
0
0.001
Mean
0.00 0.25 0.50 0.75 1.00

A biome-dependent distribution gradient of tree species range edges is strongly dictated by climate spatial heterogeneity

David Lerner[1], Marcos Fernández Martínez[2], Stav Livne-Luzon[1], Jonathan Belmaker[3,4], Josep Peñuelas[2,5], Tamir Klein[1]

[1]Department of Plant & Environmental Sciences, Weizmann Institute of Science. Rehovot 76100, Israel.

[2]CREAF, Cerdanyola de Vallès, Catalonia E08193, Spain.

[3]School of Zoology, Faculty of Life Sciences, Tel Aviv University, Tel Aviv, Israel.

[4]Steinhardt Museum of Natural History, Tel Aviv University, Tel Aviv, Israel.

[5]CSIC, Global Ecology Unit CREAF-CSIC-UAB, Bellaterra, Catalonia E08193, Spain

Abstract

Understanding the causes of the arrest of species distributions has been a fundamental question in ecology and evolution. These questions are of particular interest for trees due to their long lifespan and sessile nature. A surge in data-availability evokes for a macro-ecological analysis to determine the underlying forces limiting distributions. Here we analyse the spatial distribution of >3600 major tree species to determine geographical areas of range-edge hotspots and find drivers for their arrest. We confirmed biome edges to be strong delineators of distributions. Importantly, we identified a stronger contribution of temperate than tropical biomes to range edges, adding strength to the notion that tropical areas are centers of radiation. We subsequently identified a strong association of range-edge hotspots with steep spatial climatic gradients. We linked spatial and temporal homogeneity and high potential evapotranspiration in the tropics as the strongest predictors for this phenomenon. We propose that the poleward migration of species in light of climate change might be hindered due to steep climatic gradients.

Main Text

Introduction

The geographical distributions of species are marked by their range limits. Understanding the causes of distribution arrest has been a fundamental question in ecology and evolution[1–4]. Given the strong interplay between biotic, abiotic, demographic, physical and historical forces in predicting range-edges, it has been challenging to find underpinnings for their formation. Two main environmental forces seem to play a major role in the formation of

range edges; spatial environmental heterogeneity and habitat quality[4,54,5]. Most models on the formation of species-range edges rely on the interplay between either one of these two forces with non-climatic pressures to explain their formation. For example, steep climate gradients combined with high dispersal and gene flow reduces species' fitness and genetically constrains their evolution into novel environments[6,7]. Likewise, low habitat quality reduces population size[2,8], increasing drift and migration load[9–11]. Nevertheless, the significance of climate in the interplay between these two environmental components in defining range edges for an array of species or on a wide biogeographical scale remains elusive[12,13].

The field of biogeography has long sought understanding of species ranges despite having limited tools (e.g. [3,14]). Given the surge of large-scale datasets, it is now possible to better identify the underpinnings of species distributions by studying the macro eco-evolutionary processes involved in their formation[11,12,15–17]. Although methods and results are disparate between studies, there is almost a consensus that the presence of large-scale biogeographical units confine species with climate as their primary predictor. For example, Bontager et al.[11] suggested distinct characteristics for range edge populations dependent on their latitude. Likewise, niche conservatism and strong beta-diversity patterns seem to withhold at large macro-ecological scales[16–18]. This is generally true for plant species, with biomes being the most consistent classifier based on structural and functional similarity[19,20]. Recent efforts have been made to understand how accurate and substantial biome entities are at defining species distributions [20,21]. Nevertheless, it remains an open question if, and to what extent, the intersection between biomes is a source of species range edge hotspots. Deciphering such patterns will enable the proper understanding of how communities of species redistribute and are structured geographically, and if similar biomes in distinct geographic areas have similar effects on the distribution of species.

However, although there has been an increased interest in defining the biogeographical underpinnings of species distributions, most techniques have used species relatedness and diversity metrics to test for the existence of shared niche space between species and communities. The direct analysis of the ecological and climatic limitations to geographic space, although trivial, remain elusive. Here, we look at the universal set of climatic factors and geographical patterns of species distributions by focusing directly on species' range edge distributions.

We present the first-to-date global study of tree species range edges, applying a novel, simple, yet effective method of delineating range edges (REs) in order to: (1) identify deterministic patterns of REs, as seen by RE-dense areas (RE hotspots: REH); (2) determine whether the classification of biomes as distinct community-level patterns of biodiversity properly delineate the niche of species; (3) identify global-scale REH patterns, and; (4) discern the underlying niche factors responsible for RE formation. In particular, we investigate whether spatial heterogeneity of abiotic factors or a universal predictor for habitat quality are determinants of RE formation. We focus specifically on tree species, as they are an exemplary group of species in the study of the ecological changes predicted to occur at the peripheries of distributions, given their long-lived characteristics and fundamental role in many ecosystems[25,26], specifically biomes [18,27].

Although we do not study the interplay between climate and other ecological and evolutionary limiting factors to species distributions (e.g., seed dispersal, plasticity, and adaptation), discerning these patterns and the climatic components leading to such distributions will enable a better understanding of interplays between biotic and abiotic factors in future studies. A better understanding of the climatic factors affecting dispersal enable for better predictions of the success of species to track changing climates, and in turn, if they will be subject of migration lags[22–24].

Results

Global data set and REH distribution

We present the first report of the global distribution of tree range edge hotspots (REH) (Fig. 1), marked by the hexagons with significant clustering of REs (Supplementary Fig. 1). We did not identify any significant coldspots, given the baseline presence of REs around the globe. The visual patterns emerging from these distributions indicated that distributions stopped disproportionally more at the edges of biomes than within them. For example, northern REH occurred mostly at the intersection between a montane (Himalayas) and a desert biome (Gobi Desert) or at the edge of the tundra in North America. Southern REH in Africa and southern Eurasia tended to stop at the edges of desert biomes (e.g. Sahara Desert) but stopped mainly at the edges of temperate and montane grasslands in the Southern Hemisphere (pampas and Andes, respectively). Eastern and western REH were notably mostly in similar geographical locations, e.g. at the intersections between the Himalayas and central Asian deserts or at the edge of the Atacama Desert in South America.

In order to identify the underlying niche factors that define REs, we focused only on inland REs, since REs at the edge of a water source is probably due to an obvious geographical barrier rather than an ecological effect. The fraction of inland REs was globally similar between the continents (between 65 and 75%), except for Australia (~45%, Supplementary Fig. 2a). The large fraction of inland REs (Supplementary Fig. 2b) was mainly associated with the edges of biomes (e.g. northern REs in Africa, Asia, and North America were associated with the Sahara Desert, montane grasslands of the Himalayas, and the tundra biome, respectively; Fig. 1). A significant fraction of eastern inland REs were in Europe, but almost no significant REH were identified, implying the sparse distribution of REs throughout the continent, or as an effect from a smaller area for distribution compared to other continents.

Contribution of biome-biome intersections to REHs

We quantitatively identified the global patterns of arrest by analysing the fraction of REH that stopped at biome-biome intersections (14 central global ecological regions best distinguished by their climate, fauna, and flora obtained from World Wild Life (WWF, http://www.worldwildlife.org/; Fig. 1) and by delineating a buffer zone at points of intersection between two or more biomes. Although not significant, the number of REH was strongly associated with the intersection between biomes (Fig. 2a, first panel), indicating that climatic conditions were a probable cause for the REs of tree species at biome edges. Our results, however, identified an unequal contribution of the different biome edges to the fraction of REH (Fig 2a). We then analysed the individual biome-biome intersections normalised by a *global* permutation (i.e., from permutation of the global distribution of REH, see Materials and Methods). Here we identified (1) a strong contribution of REH at the intersections between temperate and desert biomes (Fig. 2b, left panel) in comparison to (2) a weaker contribution at the intersections of tropical and subtropical biomes (between themselves and with temperate biomes). The significance of the contribution of REH (p<0.05) (see Materials and Methods) was attributed almost exclusively to the intersections within temperate biomes (asterisks in Fig. 2b). This unequal contribution between the temperate vs tropical and subtropical biomes was globally consistent. Nevertheless, under a *per-biome* permutation (i.e., normalising each intersect by a selective permutation from the respective biome combinations separately; Fig 2b right panel), we observed a much weaker

contribution of range edges to the formation of biomes. Only a selected number of biome-biome intersections had a significant contribution to the formation of REH, yet no specific pattern between temperate and tropical biomes was observed. Indeed, the strong positive correlation between the number of REs between biomes with the number of REs within biomes (Fig. 2c; $R = 0.9$, linear regression, $p < 0.001$) reflects such discrepancy between panels in Fig 2b, as the number of REH at the edge of biomes had a strong linear association with the number of REH within that biome. We identified five biomes, however, outside the 95% confidence interval (CI) of the regression (Fig. 2c). The two biomes above the regression CI (desert; and boreal) were biomes where the number of REH at the edges was larger than predicted (Fig 2c). In contrast, temperate coniferous forest, mangrove and flooded grasslands fell below the CI of the regression, indicating a much larger number of REH within the biome compared to the edge.

A parallel analysis using the distribution of REs (rather than REH) was conducted to identify similarities and differences between the distributions. The results indicated a similar pattern of distribution (Supplementary Fig. 4a-b compared to Fig. 2b-c), suggesting global forces associated with REH.

Climatic predictors of RE formation

Intersections between biomes were a significant cause of RE formation, so we also investigated the dependence of RE formation on climate. We tested both the '*absolute climate*' (i.e. annual and seasonal average temperature and precipitation) and the '*spatial heterogeneity of climate*' (*SH*, i.e. spatial variability of absolute climate) at each of the global hexagons using 19 bioclimatic variables obtained from WorldClim (see Materials and Methods). Elevation (*absolute* and *SH*) and latitude were also accounted for at each hexagon resolution. Generalised linear models of regression between each climatic variable as an independent predictor indicated that all 40 climatic variables were significant predictors of RE formation ($p<0.05$) (Supplementary Fig5a). Interestingly, most *absolute* climatic variables were negatively associated with the global RE distribution (Fig. 3a and b), implying a general prediction of REs occurring in climates with low temperatures and low precipitation. A positive association was attributed to the four climatic variables that defined the temporal heterogeneity of temperature and precipitation (mean diurnal range, temperature

seasonality temperature annual range and precipitation seasonality; BioClim variables 2, 4, 7, and 15, respectively; Fig. 3a, Supplementary Fig. 5) and all of the *SH* climatic variables. A mixed model with biomes and continents as random factors gave a reduced number of variables that significantly predicted the formation of RE (Fig. 3a). A crossed model with both continents and biomes consistently gave a stronger fit (AIC values) than only considering either of these categorical random factors independently (Supplementary Fig 6b). In this more stringent global analysis, latitude, *absolute* temperature and *SH* temperature, precipitation and elevation were significant predictors of REH. The goodness of fit (R^2) of each model, as a measure of predictive strength, indicated that spatial heterogeneity accounted for RE formation better than did their absolute equivalents. A model selection was carried out to identify the most important factors associated with RE formation, followed by a model averaging of the models with a $\Delta AIC < 2$, (Figs. 3c). The *SH* climatic variables again defined RE better than the *absolute* climatic variables. Although *absolute* climatic variables such as isothermality (BioClim3), temperature of the wettest quarter (BioClim8) (both characteristic of tropical and subtropical climates), and annual precipitation (BioClim12) were strongly associated with REs in a generalised linear model, temperature of the warmest quarter (BioClim10) was the only predictor strongly associated with RE under the mixed model (Figs. 3a and S6). This difference in results can be visually seen when comparing between continents (Fig. 3b) e.g. in panel 3; with temperature of wettest quarter (BioClim8) having partial dissociations in Africa and South America, i.e. the two continents with the most tropical biomes. Furthermore, *SH* isothermality (BioClim3), temperature of coldest month (*SH* BioClim 6), precipitation (*SH* BioClim13 and 16) and elevation change (*SH* elevation) were all predictors of REs. Interestingly, the *absolute* climatic predictors (BioClim8 and BioClim10) were strongly negatively correlated with *spatial heterogeneity* at the minimum temperature of the coldest month (*SH* BioClim6) (Supplementary Fig. 7b), indicating its representation of a tropical biome climate. The weaker relative importance of *SH* precipitation (Fig. 3c, left panel) is due to its dissociation with the continents that most strongly represent temperate regions; Europe and North America (Fig 3b, panel 4).

In parallel, we ran models with the ENVIREM dataset, a dataset of environmental variables complementary to WorldClim that are more ecophysiologically meaningful for plant species[28]. Most ENVIREM variables associated with REH are spatially heterogeneous variables (Fig. 3c, Supplementary Fig. 7). *Absolute* potential evapotranspiration (PET) of the coldest quarter was strongly correlated with both *absolute* temperature of the warmest quarter

(BioClim10) and *SH* temperature of the coldest month (BioClim6); the two predictors indicative of a transition between tropical and temperate biomes (Supplementary Fig. 7b). *SH* EmbergerQ and *SH* PET of warmest quarter were strongly associated with *SH* precipitation (BioClim13 and 16) and *SH* temperature of the warmest quarter (BioClim10), respectively. Albeit the strong correlation of all *absolute* ENVIREM variables with *absolute* BioClim variables[28], we identified several *SH* ENVIREM variables to be weakly correlated to BioClim (e.g., *SH* moisture index and PET of driest month). All results in Fig. 3 were robust against spatial autocorrelation. (Supplementary Fig. 5b, see Materials and Methods for further information).

The REH distribution from our generated polygons was compared to a REH distribution from expert polygons (see Materials and Methods), in order to test for the accuracy of our generated dataset. Indeed, results of GLMs converge, indicating the robustness of our generated dataset to describe REs (Supplementary Fig. S8).

Discussion

The results support our hypothesis of a nonrandom distribution of REs. We were able to confirm that many tree species range edges were clustered rather than sparsely (stochastic) distributed, by obtaining a larger number of significant REH. This finding suggests the underlying presence of ecological and evolutionary forces governing REH formation. Similarly, the matching results obtained from the biome analyses when accounting for REs (Supplementary Fig. 4) or REH (Fig. 2) was also indicative of the deterministic clustering of species REs at these specific ecological barriers (biome edges). Nevertheless, we identified case-specific exceptions such as the scattered distribution of eastern REs throughout Europe, indicated by the strong identification of internal REs (Supplementary Fig. 2b) but not REH (Fig. 1c). This could potentially have occurred as a consequence of either the intense anthropogenic activity throughout[29,30] that prevented the distribution of tree species to reach their natural REs, or the effect of smaller land area compared to other continents, altering the effect of how species interact with abiotic factors, thus affecting their distribution and adaptation.

Our results also indicated a strong dependence of REH on the edges of biomes, strongly supporting the many efforts to determine whether the division of the planet into discrete geographical units has been appropriately delineated[21,31–34]. Our findings, however, were

unexpectedly biome-specific, identifying the biome borders that most defined tree distributions (Fig. 2). There is no seemingly obvious pattern of differential contribution to REH when analysing each biome independently (*per-biome* bootstrap; Fig 2a). Yet, a clear distinction between a tropical/temperate distribution of REH is strongly observed under a biome-biome pairs analysis (*global* bootstrap; Fig 2b). Specifically, REs were strongly dependent on desert, temperate, and montane edges compared to the weak dependence on tropical and subtropical biome edges. Despite the marginally significant association of range edges to biomes globally (Fig 2a), a tropical-temperate REH distribution is not specifically dependent on biome edges, but rather, it represents a global biome trend (Fig 2b, left panel; Fig. 2c). In addition, the importance of biomes in describing climate's association to REs, as seen from a mixed effects model (Fig. 3a, S6c), as well as the absence of latitude in predicting REs in a global model (Fig 3c) is further indicative of the importance of biomes (rather than a latitudinal effect) in defining REs.

The large differences in the relative number of REs between the tropical and temperate biomes can indicate adaptive mechanisms between species residing in either of these two types of biomes. These nontrivial results may have been related to the Latitude Diversity Gradient[35–37], a well-established pattern in which biodiversity is higher in the tropics than in temperate regions. Firstly, the notion that these differences surged through the effective evolutionary time hypothesis[36,38] suggests that genetic diversity is higher in the tropics due to the possible longer times needed for adaptation and expansion. Similarly, long-term climatic oscillations have been suggested to reduce cladogenesis at higher latitudes, consistent with the observation that tropical areas are centers of evolutionary novelty[36,39,40] and evolve faster than temperate regions[35], leading to the notion of tropics as centers for radial expansion of clades and species. Recent studies have identified mechanisms for this 'out of the tropics expansion' model[41,42] and have reported a higher fraction of bridge species (species violating niche conservatism) from the tropics compared to temperate regions. We conclude that the lower contribution to REs in tropical vs temperate regions is consistent with stronger radial expansions from the tropics than temperate regions.

However, other mechanisms could explain the distinctive REH patterns between the tropic and temperate regions. For example, based on several studies demonstrating the large physiological and evolutionary effects of forest fragmentation[29,43,44], the increased long-term anthropogenic activity, and consequently excessive fragmentation, in forests in temperate

regions compared to tropical and subtropical regions may have hindered the adaptation of species to novel climates. The lack of association of REH specifically to temperate grasslands, temperate conifers and Mediterranean biomes specifically (Fig 2a) could be indicative of such an effect, as these biomes have been historically subject to strong anthropogenic activity.

Our analyses confirmed the strong dependence of RE on climate. Lower temperature, higher climatic heterogeneity (both temporal and spatial), and elevation changes were the strongest climatic predictors of REs. The analysis of the best predictive variables indicated a noticeable weak correlation between most of the leading factors determining REs (Supplementary Fig. 7b), suggesting that these factors were site-specific predictors of REs (e.g. desert biome with low annual precipitation, montane grassland with high elevation, and tropical and subtropical biomes with temperature homogeneity (isothermality, temperature of the wettest quarter and spatial heterogeneity of low temperature – *SH6*)). We suggest that the strong negative correlation between these absolute climatic variables and the spatial heterogeneity of low temperature is an indication of different temperature patterns between tropical and temperate regions, i.e. a buffered heterogeneous temperature in the tropics in contrast to the latitudinal effect of decreasing temperatures in the temperate biomes. There was an overall positive trend for the effect of precipitation (BC12-BC19) when controlling for biome since the signal appeared only when biome is included as a random factor (Supplementary Fig. 6). This means there must be other non-accounted confounding factors altering the relationship when analysing all biomes together, suggesting that the effect is not biome dependent. Although this study does not make a note of the complex interplay between biotic and abiotic forces or the ecological traits of tree species (e.g. seed dispersal, phenology) in the formation of RE, we find a consistent global effect of temperature and spatial heterogeneity of temperature and precipitation to predict their formation throughout the different models. This is indicative of their principal role as universal predictors of species distributions.

The strong prediction of REH formation from PET of the coldest quarter and its covariance to the tropical-temperate transition variables (*SH* BioClim6 and *absolute* BioClim10), could reflect a possible mechanistic evolutionary constrain for the distribution of woody species' at temperate biomes. At higher latitudinal temperate biomes, where evapotranspiration is strongly reduced, especially during the coldest quarter, there is a strong limitation to photosynthesis and growth by the significant time reduction of stomatal conductance[45,46].

Similarly, given the strength of its prediction of REH, *SH* embergerQ (pluviothermic quotient) could indicate a more refined mechanism for RE formation than its covariates of precipitation (*SH* BioClim 13 and 16). This index describes mean annual precipitation in relation to annual changes in temperature. EmbergerQ thus increases the predictability of how precipitation also dictates the formation of REH in the more temperate biomes (Fig 3b, panel 4). The consistency of our results indicates spatial and temporal heterogeneity of climate and topography as overwhelmingly stronger predictors of RE formation than their absolute climatic counterparts (Fig. 3) and, in turn, frail evidence for a universal poor habitat quality. Even in the cases where *mean* climatic variables strongly predict REH, these were strongly associated with this transition between the spatial and temporal climatically homogenous tropical and subtropical biomes to the more heterogeneous temperate biomes. Nevertheless, we note the importance of a lack of evapotranspiration, particularly in cold climates, as a main predictor of RE formation.

These observations have substantial implications for the effects of climate change on tree distributions and its effects on tree migration. Although predictions for future steeper temperature gradient as a result of greenhouse gas emission and climate change has not been trivial[47–49], such an increase in temperature gradients could vastly affect the distribution of species. In particular, our results strengthen the growing understanding that the predicted poleward migration of tree species might not be as successful as previously predicted[23,24]. The increase of stronger spatial gradients (especially in the lower latitudes)[24,48,50] or extreme and spontaneous events might all be causes of migration lags, despite the suitable temperatures at higher latitudes and altitudes. Likewise, the importance of PET from temperate biomes on the formation of REH presented here could also suggest a possible migration lag or loss of adaptation due to the predicted reduction in PET at higher latitudes[51]. Our results thus highlight the importance of accounting for more precise spatial heterogeneity of climate as a critical feature in future models of species distribution and the development of more precise conservation efforts such as assisted migration.

Materials and Methods

Data acquisition and polygon formation

Supplementary Fig. 9 visually summarises the methodologies used to obtain global range edge hotspot (REH) distributions. We downloaded a data set of tree species from the open-source data set using R packages *rgbif* and *taxize*. Global Biodiversity Information Facility

(GBIF; 05 July 2021, https://doi.org/10.15468/dl.ajen6k) using the Botanic Gardens Conservation list of 60000 tree species. We downloaded occurrences with entries from 1980 onwards, removing any occurrence reported with a *geospatial issue*, species not belonging to the kingdom Plantae (in case of mismatched species names), and any occurrence marked as *unlikely*, *mismatched*, or *invalid*. We removed occurrences that had reported uncertainties of >100 km and records based on fossils and unknown sources. We then used the *CoordinateCleaner* R package[52] to remove any occurrences with zero coordinates, equal x and y coordinates, duplicates, occurrences at sea, coordinates at capitals, and centroids. To finalise, we again removed species with <300 occurrences. The data we used undoubtedly contained sampling bias[53], probably overrepresenting the number of REs in some regions with a reduced or negligible sampling effort. We tried to overcome this issue by basing our filtering steps on several previous studies[52–54]. The strength of the critical filtering steps applied in our analysis resembled those previously presented[53].

We converted the georeferenced species occurrences (x and y coordinates) into distributional polygons in parallel using two independent techniques; through concave-hull (Supplementary Methods) and multivariate kernel-density estimation (described below). Given the strong similarity between the two methods (Supplementary Fig. 8), we discuss the methods and results in detail only for the kernel-density estimated polygons. We created polygons using two-dimensional kernel density estimations. We first divided the extent of all the coordinates into 800 grid points in each dimension (longitude and latitude) in order to produce a matrix of 640,000 grid cells for each of the species. Subsequently, we selected for the grid cells with the highest 99% estimation of the species' occurrence and subsequently rasterised these. Polygons were then delineated around the contour of the rasters.

Polygon groupings

All polygons belonging to the same species were grouped based on absolute distance from one another. Polygons separated by ≤500 km were grouped together, with the assumption that fragmentation, gene flow, and unreported data could all warrant two nearby populations to be considered as one. We used the hclust function (package stats, agglomeration method: complete) to hierarchically cluster populations from a sequence of three or more populations by their distances, also using a cutoff of 500 km (cutree function, package stats) for determining the clusters. The final data set comprised >3600 tree species, ranging from one to

nine populations (polygons) per species, for a total of 8500 populations. All spatial data was analysed using R packages *sf* and *raster*.

RE determination

RE dense areas were determined by (1) defining distinct global units, (2) identifying the RE of each species and (3) map species' RE to the global units to calculate the density of REs/unit. In detail: (1) We rasterised the world map to spatially bin the density of REs. To overcome the problem of spatial distortion, we used hexagonal bins with the *dggridR* R package, developed using the ISEA Discrete Global Grids system, a repetition of polygons on the surface of an icosahedron, allowing for the projection of equal sized bins onto a 2D plane. We defined the size of each hexagon as ~23 000 km^2 (with an average spacing between center nodes of 165 km). (2) We used coordinates of the cardinal directions (north, south, east, and west) to represent species REs by subdividing each polygon cluster into four quartiles in the four cardinal directions (NE, NW, SE, and SW). The REs for each quartile were determined as the two most-outward coordinates of the corresponding cardinal directions (e.g. north and east cardinal coordinates for the NE quartile). Eight REs were thus determined for each population (two for each cardinal direction). As a filtering step, we accounted for both REs from the same cardinal direction if they were >20 arc-degrees apart, otherwise we only accounted for the farthest point from the centroid. (3) The total number of REs obtained using this method was normalised by the total number of species intersecting its respective hexagon. In parallel, we also defined REs by accounting for the perimeter of the polygon for each species (Supplementary Fig. 10). The 'perimeter' system may be a more realistic and complete system for identifying REs, but the 'cardinal coordinate' system, although more simplistic in nature, (1) provides a clearer visual representation of the distribution of REs, and (2) allows for the directionality of REs to be compared, essential farther along the pipeline by distinguishing between coastline and inland REs and identifying hotspots in the four cardinal directions.

We then classified coastline and inland REs, assuming that the arrest of species distribution at the edge of a water source was probably due to an obvious geographical barrier rather than an ecological effect. Coastline REs were determined by creating a semicircle (buffer of 3 arc-degrees) around each RE in the direction of its cardinal coordinate and measured the percentage overlap with water. Cardinal coordinates with >50% overlap were considered a "coastal RE".

In order to find the probability distribution of REs at the edge of water source (Supplementary Fig. 2), we permutated the global population of REs (except for the Australian population) and used the mean of this permutation to compare to the number of REs in each of the continents.

Hotspot analysis

Hotspots were identified using the Getis-Ord Gi* hotspot analysis[55,56] to find spatial correlations between hexagon (inland and normalised) RE densities. We initially compiled a list of neighbors between all hexagons using the poly2nb function and then obtained the local G statistic using the weighted density (normalised number of REs) of the global hexagons and their relative distance from each other. The G statistic calculates a Z-score (measure of standard deviation) for each hexagon. *P* values were then determined using the critical Z-scores at 95% confidence levels followed by a Bonferroni correction using the p.adjustSP function (using the number of neighbors between hexagons rather than the total number of hexagons). All analyses were carried out using the R *spdep* package[57].

We compared the analysis from the linear models with expert based polygons from three different sources – IUCN, BIEN and EUFORGEN. A randomised weighted sample of all of this dataset was used to generate a global distribution of REH by running this sample through our pipeline. GLMs were run on the global distribution of expert-based REH in the same way as with our generated polygons. Given the uneven distribution of expert-based polygons globally (Supplementary Fig. 8a), we ran models excluding Asia and Africa, in order to account for this bias. As seen by the strong similarity between the GLMs of expert polygons in a global and filtered model (Supplementary Fig 8b and Fig 8c, panel 3), we observed an overrepresentation of the expert polygons for these continents. Likewise, the GLMs from our generated polygons are much similar to those obtained from the filtered model. In this case, practically all variables showed the same relationship with range edges (either positive or negative β values) as well as similar magnitudes.

Statistical analyses

Contribution of biome edge to RE. We used the 14 biomes defined by the World Wildlife Fund (WWF) for our analyses. The distributions were downloaded from the WWF webpage (http://www.worldwildlife.org/). To identify the intersection between biomes, we reduced the

complexity of the polygon edge using the rmapshaper package, which can perform topologically aware polygon simplifications, thus maintaining the intersection between biomes upon reduction of "edginess" of the polygons. The intersections were delineated and subsequently enlarged (with a buffer distance of 0.1 arc-minute, ~185 m at the equator). A:

i. *Global* permutation assay was carried out by randomising (1000 iterations) the global distribution of hexagons with REH (absolute Z-score)
ii. *Per-biome* stratified permutation was carried out by randomising the distribution of hexagons with REH within each biome independently.

The averaged global bootstrap shown in panel 1 of Fig 2a, was calculated using a *per-biome* bootstrap to obtain the probability distribution of REH at biome edges (Supplementary Fig. 3a). Distributions that were not normally distributed as a result of their small size (flooded grassland, mangrove and tropical and subtropical coniferous forest; Supplementary Fig 3b), were removed from the analysis. A general trend for the probability of range edges falling at the intersection of biomes was therefore measured as a unified standardised z-distribution, and compared to the median z-score from the actual percent overlap for each biome.

The density of hotspots at the intersection between biomes was calculated using the sum of Z-scores of the hotspots at that intersection, and the percentage contribution was then calculated using this value over the total Z-score at all biome intersections. The *global* or *per-biome* 1000 permutation means were used as a normalising denominator for the values obtained from our data set (Fig 2b, Supplementary Fig 4a). The denominator could be either larger or smaller than the numerator, so we log-transformed the outcome to obtain a linear-like relationship. Contribution within a biome was calculated the same way as for the contribution at the edge, using the mean from a permutated assay to normalise for the absolute value.

Climatic dependency of RE. We tested the relationship between RE density and climatic features by assigning a set of environmental variables to each hexagon. We used the bioclimatic attributes downloaded from WorldClim Global Climate Data[58] at a resolution of 5 arc-minutes. The 19 BIOCLIM variables and elevations for each hexagon were extracted using the R raster package. We also used the 16 ENVIREM variables described by [28], downloaded from their website, at a resolution of 2.5 arc-minutes. *Absolute* climate for each

hexagon was obtained using the mean over all pixels. Climatic *spatial heterogeneity* (*SH*) was calculated using the proportional variability index[59,60] (PV) over all pixels in each hexagon.

We used an array of linear mixed models (Fig 3, Supplementary Fig 5 and 6) to test for the dependence of REs to climate and the robustness of the results. Linear mixed models were carried out to account for biomes and continents. Both of these variables were introduced as random effects in random intercept models. A model selection analysis was used to determine the models that best predicted the formation of RE. Random intercept models using both continent and biome as random variables were run, and models with $\Delta AIC < 2$ were selected for. Given the strong correlation between different predictor variables, we ran models only with variable combinations that had a Pearson's correlation value $r < 0.7$. The relative contribution of the variable included in the model were calculated from the selected models. Analyses were run using R package *MuMIn*[61].

All statistical analyses (individual GLMs, and multiple-predictor GLMs) were tested for their robustness to spatial autocorrelation by creating a spatial autocovariate (autocov_dist function, spdep package), calculated as the distance-weighted average of neighboring dependent variables[62], so hexagons in proximity were averaged and those farther away received a lower weighting average. We set the predetermined distance to 200 km based on the average distance between cells. The spatial autocovariate was then included in the regression model as a dependent variable.

Data Availability

The occurrence points used from GBIF can be found in the GBIF webpage (https://doi.org/10.15468/dl.ajen6k). Polygons generated from occurrence points are provided in the public GitHub repository https://github.com/dlernerg/Global-Range-edges. Biome polygons were obtained from the WWF webpage (http://www.worldwildlife.org/). Bioclimatic attributes were downloaded from WorldClim Global Climate Data[58]. ENVIREM variables were downloaded from their webpage (https://envirem.github.io/).

Code Availability

Custom codes related to this paper can be found in a GitHub repository at https://github.com/dlernerg/Global-Range-edges

Acknowledgments

TK wishes to thank Edith and Nathan Goldenberg Career Development Chair; Mary and Tom Beck-Canadian Center for Alternative Energy Research; Larson Charitable Foundation New Scientist Fund; Yotam Project; Dana and Yossie Hollander; Estate of Emile Mimran; and the Estate of Helen Nichunsky. DL was supported by the Sustainability and Energy Research Initiative PhD grant. MF-M and JP were supported by the PID2019-110521GB-I00 and TED2021-132627B-I00 grants funded by MCIN, AEI/10.13039/501100011033 and the NextGeneration EU/PRTR. MF-M. was supported by a postdoctoral fellowship from "la Caixa" Foundation (ID 100010434), code: LCF/BQ/PI21/11830010.

Author Contributions: D.L., J.B. and T.K. designed the research; D.L. performed the research and analysed the data; M.F.M, J.P., T.K., J.B., and S.L.L. provided scientific advice; M.F.M. and J.B. advised on statistical aspects; D.L. wrote the paper with special contribution from J.P., J.B., M.F.M. and T.K.

Competing Interest Statement: The authors declare no conflict of interest.

Figure 1. Range-edge hotspots. Hotspots were identified using Getis-Ord Gi* analysis, which returns a Z-score for each hexagon in the world. Only hexagons with $p<0.05$ are considered hotspots and subsequently shown here. The Z-score of each hexagon is represented by the color gradient. Biomes (as defined by the World Wildlife Fund) are marked by colors. 'T&sT' and 'Temp' stand for 'Tropic and Subtropic' and 'Temperate', respectively.

Figure 2. Range-edge hotspots at intersections between biomes. (**a**) Modeled distribution of the percentage of hotspots at the edges of biomes from a permutated (randomised) *per-biome* distribution of hotspots. The first panel represented the median value of all the other biomes in the figure over a standardised z-distribution of biomes. The arrow marks the percentage of RE hotspots at biome intersections in the data set (one-sided *p-values*) (**b**) Heat maps of the percentage of RE hotspots (relative to the total number of hotspots) at the intersection between two biomes. A biome-pair intersection with a significant number of hotspots (*p-value* < 0.08) is marked with an asterisk. Biome-pairs that have no intersections are gray. Panel 1 is normalised over a *global* bootstrap and panel 2 over a *per-biome*

bootstrap. (**c**) Correlations (and regression lines) of the relationships between the number of REs at the edges of biomes and the number of REs within the biomes. The shaded area represents the 95% confidence interval around the regression line. *P-values* are calculated using a two-sided Student's T-test (*degrees of freedom* = 12). See Methods and Materials for further information on the methodology for obtaining significance levels (for (**b**)) and hotspot permutations. 'T&sT' and 'Temp' stand for 'Tropic and Subtropic' and 'Temperate', respectively.

Figure 3. Climatic predictors of range-edge formation. Models of RE formation using the absolute climate (mean) and the spatial heterogeneity (PV index) of the 19 WorldClim variables and elevation. (**a**) Individual (binomial) mixed regression models between each of the predictor variables and number of REs (1-19 are BioClim variables) and accounting for continents and biomes as random effects. The estimated coefficients of the explanatory variables (β) are represented by the color gradient. '*' and '**' represent *p-values* <0.05 and <0.01, respectively (two-sided Student T-test).
BioClim1 – Annual Mean Temperature, BioClim2 - Mean Diurnal Range (Mean of monthly (max temp - min temp), BioClim3 – Isothermality, BioClim4 - Temperature Seasonality, BioClim5 - Max Temperature of Warmest Month, BioClim6 - Min Temperature of Coldest Month, BioClim7 - Temperature Annual Range, BioClim8 - Mean Temperature of Wettest Quarter, BioClim9 - Mean Temperature of Driest Quarter, BioClim10 - Mean Temperature of Warmest Quarter, BioClim11 - Mean Temperature of Coldest Quarter, BioClim12 - Annual Precipitation, BioClim13 – Precipitation of Wettest Month, BioClim14 – Precipitation of Driest Month, BioClim15 – Precipitation Seasonality (PV), BioClim16 – Precipitation of Wettest Quarter, BioClim17 – Precipitation of Driest Quarter, BioClim18 – Precipitation of Driest Quarter, BioClim19 – Precipitation of Coldest Quarter.
(**b**) Violin plots depicting the results from (**a**) for four predictor variables. The distribution of continental climates is shown in gray, contrasted with the climatic distribution specific to the RE hotspots (scaled to the intensity of the hotspot, i.e. the Z- score). (**c**) Forest plot of a model average from the highest predicting LMM with both BioClim and ENVRIEM variables (identified with a model selection). Beta values (log-odds) are shown for each predictor. Absolute and SH climate was obtained from all of the inland global hexagonal units (*n = 5851*). Error bars represent 95% confidence interval around the average effect.

References

1. Soule, M. The epistasis cycle: A theory of marginal populations. *Annu Rev Ecol Syst* **4**, 165–187 (1973).

2. Brown, J. H. On the Relationship between Abundance and Distribution of Species. *Am Nat* **124**, 255–279 (1984).

3. Gaston, K. J. The Structure and Dynamics of Geographic Ranges. *Oxford University Press* (2003).

4. Sexton, J. P., McIntyre, P. J., Angert, A. L. & Rice, K. J. Evolution and ecology of species range limits. *Annu Rev Ecol Evol Syst* **40**, 415–436 (2009).

5. Gaston, K. J. Geographic range limits: Achieving synthesis. *Proceedings of the Royal Society B: Biological Sciences* **276**, 1395–1406 (2009).

6. Goldberg, E. E. & Lande, R. *Notes and Comments Species' Borders and Dispersal Barriers*. *Am. Nat* vol. 170 (2007).

7. Bachmann, J. C., Rensburg, A. J. van, Cortazar-Chinarro, M., Laurila, A. & Buskirk, J. van. Gene flow limits adaptation along steep environmental gradients. *American Naturalist* **195**, E67–E86 (2020).

8. Hargreaves, A. L., Samis, K. E. & Eckert, C. G. Are species' range limits simply niche limits writ large? A review of transplant experiments beyond the range. *American Naturalist* **183**, 157–173 (2014).

9. Henry, R. C., Bartoń, K. A. & Travis, J. M. J. Mutation accumulation and the formation of range limits. *Biol Lett* **11**, 11DUUMY (2015).

10. Perrier, A., Sánchez-Castro, D. & Willi, Y. Environment dependence of the expression of mutational load and species' range limits. *J Evol Biol* **35**, 731–741 (2022).

11. Bontrager, M. *et al.* Adaptation across geographic ranges is consistent with strong selection in marginal climates and legacies of range expansion. *Evolution (N Y)* 1–18 (2021) doi:10.1111/evo.14231.

12. Santini, L., Pironon, S., Maiorano, L. & Thuiller, W. Addressing common pitfalls does not provide more support to geographical and ecological abundant-centre hypotheses. *Ecography* **42**, 696–705 (2019).

13. Oldfather, M. F., Kling, M. M., Sheth, S. N., Emery, N. C. & Ackerly, D. D. Range edges in heterogeneous landscapes: Integrating geographic scale and climate complexity into range dynamics. *Glob Chang Biol* **26**, 1055–1067 (2020).

14. Janzen, D. H. *Why Mountain Passes are Higher in the Tropics WHY MOUNTAIN PASSES ARE HIGHER IN THE TROPICS*. Source: The American Naturalist* vol. 101.

15. Maxwell, M. F., Leprieur, F., Quimbayo, J. P., Floeter, S. R. & Bender, M. G. Global patterns and drivers of beta diversity facets of reef fish faunas. *J Biogeogr* **49**, 954–967 (2022).

16. Roy, K., Hunt, G., Jablonski, D., Krug, A. Z. & Valentine, J. W. A macroevolutionary perspective on species range limits. *Proceedings of the Royal Society B: Biological Sciences* **276**, 1485–1493 (2009).

17. Loiseau, N. *et al.* Global distribution and conservation status of ecologically rare mammal and bird species. *Nat Commun* **11**, (2020).

18. Kerkhoff, A. J., Moriarty, P. E. & Weiser, M. D. The latitudinal species richness gradient in New World woody angiosperms is consistent with the tropical conservatism hypothesis. *Proc Natl Acad Sci U S A* **111**, 8125–8130 (2014).

19. Donoghue, M. J. & Edwards, E. J. Biome shifts and niche evolution in plants. *Annu Rev Ecol Evol Syst* **45**, 547–572 (2014).

20. Ringelberg, J. J., Zimmermann, N. E., Weeks, A., Lavin, M. & Hughes, C. E. Biomes as evolutionary arenas: Convergence and conservatism in the trans-continental succulent biome. *Global Ecology and Biogeography* **29**, 1100–1113 (2020).

21. Smith, J. R. *et al.* A global test of ecoregions. *Nat Ecol Evol* **2**, 1889–1896 (2018).

22. Zhu, K., Woodall, C. W. & Clark, J. S. Failure to migrate: Lack of tree range expansion in response to climate change. *Glob Chang Biol* **18**, 1042–1052 (2012).

23. Corlett, R. T. & Westcott, D. A. Will plant movements keep up with climate change? *Trends in Ecology and Evolution* vol. 28 482–488 Preprint at https://doi.org/10.1016/j.tree.2013.04.003 (2013).

24. la Sorte, F. A., Butchart, S. H. M., Jetz, W. & Böhning-Gaese, K. Range-Wide Latitudinal and Elevational Temperature Gradients for the World's Terrestrial Birds: Implications under Global Climate Change. *PLoS One* **9**, e98361 (2014).

25. Paquette, A. & Messier, C. The effect of biodiversity on tree productivity: From temperate to boreal forests. *Global Ecology and Biogeography* **20**, 170–180 (2011).

26. Pichancourt, J. B., Firn, J., Chadès, I. & Martin, T. G. Growing biodiverse carbon-rich forests. *Glob Chang Biol* **20**, 382–393 (2014).

27. Pennington, R. T., Lavin, M. & Oliveira-Filho, A. Woody plant diversity, evolution, and ecology in the tropics: Perspectives from seasonally dry tropical forests. *Annu Rev Ecol Evol Syst* **40**, 437–457 (2009).

28. Title, P. O. & Bemmels, J. B. ENVIREM: an expanded set of bioclimatic and topographic variables increases flexibility and improves performance of ecological niche modeling. *Ecography* **41**, 291–307 (2018).

29. Veresoglou, S. D. & Peñuelas, J. Variance in biomass-allocation fractions is explained by distribution in European trees. *New Phytologist* **222**, 1352–1363 (2019).

30. Grantham, H. S. *et al.* Anthropogenic modification of forests means only 40% of remaining forests have high ecosystem integrity. *Nat Commun* **11**, (2020).

31. Holdridge, L. R. Determination of world plant formations from simple climatic data. *Science (1979)* **105**, 367–368 (1947).

32. Whittaker, R. H. Classification of Natural Communities. *Botanical Review* **28**, 1–239 (1962).

33. McDonald, R. *et al.* Species compositional similarity and ecoregions: Do ecoregion boundaries represent zones of high species turnover? *Biol Conserv* **126**, 24–40 (2005).

34. von Humboldt, A. & Bonpland, A. *Essay on the Geography of Plants*. (2013).

35. Cardillo, M. Latitude and rates of diversifcation in birds and butterfies. *Proc. R. Soc. Lond. B* **266**, 1221–1225 (1999).

36. Hillebrand, H. *On the Generality of the Latitudinal Diversity Gradient*. *Am. Nat* vol. 163 (2004).

37. Mittelbach, G. G. *et al.* Evolution and the latitudinal diversity gradient: Speciation, extinction and biogeography. *Ecology Letters* vol. 10 315–331 Preprint at https://doi.org/10.1111/j.1461-0248.2007.01020.x (2007).

38. Hewitt, G. M. Genetic consequences of climatic oscillations in the Quaternary. in *Philosophical Transactions of the Royal Society B: Biological Sciences* vol. 359 183–195 (2004).

39. Crane, P. & Scott, L. Angiosperm Diversification and Paleolatitudinal Gradients in Cretaceous Floristic Diversity. *Science (1979)* **246**, 675–678 (1989).

40. Jablonski, D. The tropics as a source of evolutionary novelty through geological time. *P. J. IEEE Trans. Geosci. Remote Sensing* **361**, 180–191 (1993).

41. Jablonski, D. *et al.* Out of the tropics, but how? Fossils, bridge species, and thermal ranges in the dynamics of the marine latitudinal diversity gradient. *Proc Natl Acad Sci U S A* **110**, 10487–10494 (2013).

42. Antonelli, A. *et al.* An engine for global plant diversity: Highest evolutionary turnover and emigration in the American tropics. *Front Genet* **6**, (2015).

43. Jump, A. S. & Peñuelas, J. Running to stand still: Adaptation and the response of plants to rapid climate change. *Ecol Lett* **8**, 1010–1020 (2005).

44. Morreale, L. L., Thompson, J. R., Tang, X., Reinmann, A. B. & Hutyra, L. R. Elevated growth and biomass along temperate forest edges. *Nat Commun* **12**, 7181 (2021).

45. Wilkinson, S., Clephan, A. L. & Davies, W. J. Rapid Low Temperature-Induced Stomatal Closure Occurs in Cold-Tolerant Commelina communis Leaves But Not in Cold-Sensitive Tobacco Leaves, via a Mechanism That Involves Apoplastic Calcium But Not Abscisic Acid. *Plant Physiol* **126**, 1566–1578 (2001).

46. Brodribb, T. J. & Holbrook, N. M. Stomatal protection against hydraulic failure: a comparison of coexisting ferns and angiosperms. *New Phytologist* **162**, 663–670 (2004).

47. Davis, B. A. S. & Brewer, S. Orbital forcing and role of the latitudinal insolation/temperature gradient. *Clim Dyn* **32**, 143–165 (2009).

48. Seager, R. *et al.* Strengthening tropical Pacific zonal sea surface temperature gradient consistent with rising greenhouse gases. *Nat Clim Chang* **9**, 517–522 (2019).

49. Xu, Y. & Ramanathan, V. Latitudinally asymmetric response of global surface temperature: Implications for regional climate change. *Geophys Res Lett* **39**, n/a-n/a (2012).

50. Colwell, R. K., Brehm, G., Cardelús, C. L., Gilman, A. C. & Longino, J. T. *Global Warming, Elevational Range Shifts, and Lowland Biotic Attrition in the Wet Tropics*. vol. 322 www.sciencemag.org (2008).

51. Basso, B., Martinez-Feria, R. A., Rill, L. & Ritchie, J. T. Contrasting long-term temperature trends reveal minor changes in projected potential evapotranspiration in the US Midwest. *Nat Commun* **12**, 1476 (2021).

52. Zizka, A. *et al.* CoordinateCleaner: Standardized cleaning of occurrence records from biological collection databases. *Methods Ecol Evol* **10**, 744–751 (2019).

53. Serra-diaz, J. M., Enquist, B. J., Maitner, B., Merow, C. & Svenning, J. Big data of tree species distributions : how big and how good ? (2018) doi:10.1186/s40663-017-0120-0.

54. Zizka, A. *et al.* No one-size-fits-all solution to clean GBIF. *PeerJ* **8**, (2020).

55. Getis, A. & Ord, J. K. The Analysis of Spatial Association by Use of Distance Statistics. *Geogr Anal* (1992).

56. Mendez, C. Spatial autocorrelation analysis in R. R Studio/RPubs. *https://rpubs.com/quarcs-lab/spatial-autocorrelation* (2020).

57. Bivand, R. S., Pebesma, E. & Gómez-Rubio, V. *Applied Spatial Data Analysis with R*. (2013).

58. Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G. & Jarvis, A. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* **25**, 1965–1978 (2005).

59. Heath, J. P. Quantifying temporal variability in population abundances. *Oikos* vol. 115 573–581 Preprint at https://doi.org/10.1111/j.2006.0030-1299.15067.x (2006).

60. Fernández-Martínez, M. *et al.* The consecutive disparity index, D: a measure of temporal variability in ecological studies. *Ecosphere* **9**, (2018).

61. Bartoń, K. MuMIn: Multi-model inference. R package version 1.10.0.. 1. . (2013).

62. F. Dormann, C. *et al.* Methods to account for spatial autocorrelation in the analysis of species distributional data: A review. *Ecography* vol. 30 609–628 Preprint at https://doi.org/10.1111/j.2007.0906-7590.05171.x (2007).