

Integrating global patterns and drivers of tree diversity across a continuum of spatial grains

Petr Keil^{1} & Jonathan M. Chase^{1,2}*

¹German Centre for Integrative Biodiversity Research (iDiv), Halle-Jena-Leipzig, Deutscher Platz 5e, 04103 Leipzig, Germany

²Institute of Computer Science, Martin-Luther University Halle-Wittenberg, 06099 Halle (Saale), Germany

1 **What drives biodiversity and where are the most biodiverse places on Earth? The answer**
2 **critically depends on spatial scale (grain), and is obscured by lack of data and mismatches in**
3 **their grain. We resolve this with cross-scale models integrating global data on tree species**
4 **richness (S) from 1338 local forest surveys and 287 regional checklists, enabling estimation of**
5 **drivers and patterns of biodiversity at any desired grain. We uncover grain-dependent effects of**
6 **both environment and biogeographic regions on S, with a positive regional effect of Southeast**
7 **Asia at coarse grain that disappears at fine grains. We show that, globally, biodiversity cannot**
8 **be attributed to purely environmental or regional drivers, since regions are environmentally**
9 **distinct. Finally, we predict global maps of biodiversity at two grains, identifying areas of**
10 **exceptional species turnover in China, East Africa, and North America. Our cross-scale**
11 **approach unifies disparate results from previous studies regarding environmental versus**
12 **biogeographic predictors of biodiversity, and enables efficient integration of heterogeneous data.**

13 What drives global variation in the numbers of species that live from place to place? For example, why
14 are there fewer than 100 species of trees that live in millions of km² in the boreal forests of Eurasia or
15 North America ¹, while there can be hundreds of species co-occurring in as little as 50 ha in tropical
16 forests of South America and Asia ²?

17 The most important obstacle to answering these fundamental questions is a lack of data, especially in
18 places with the highest biodiversity ^{3,4}. But even in the regions and taxa which have been well-
19 sampled, the data are a heterogeneous mixture of point observations, survey plots, and regional
20 checklists, all with varying area and sampling protocol ⁴. For example, for trees, there are hundreds of
21 0.1 ha Gentry forest plots mostly in the New World ⁵, hundreds of 1 ha ForestPlots.net plots
22 throughout tropical forests ⁶, dozens of > 2 ha CTFS-ForestGEO plots (www.forestgeo.si.edu),
23 hundreds of published regional checklists ⁷, and hundreds to thousands of other published surveys and
24 checklists scattered throughout the published and grey literature. These together hold key information

25 on global distribution of tree biodiversity, yet the lack of methods to address differences in sampling
26 have so far prevented their integration.

27 Further, as could be said for many problems in ecology, attempts to map global biodiversity and to
28 assess its potential drivers are severely complicated by the issues of spatial scale^{8–11}: The most
29 straightforward issue is the non-linear increase of number of species (S) with area¹², which is why
30 patterns of biodiversity cannot be readily inferred from sampling locations of varying area. The second
31 issue concerns sets of sampling locations that do have a constant area (hereafter grain); even then a
32 spatial pattern of S observed at a small grain may differ from a pattern at large grain^{13–15} – an example
33 is grain-dependence of latitudinal diversity gradient¹⁶ [but see ref¹⁷]. The reason is that beta diversity
34 (the ratio between fine-grain alpha diversity and coarse-grain gamma diversity) varies over large
35 geographic extents¹⁸. Finally drivers and predictors of diversity have different associations with S at
36 different grains^{19–22}. For example, at global and continental extents, the association of S with
37 topography increases with grain in Neotropical birds²² and the association with temperature increases
38 with grain in global vertebrates²¹ and eastern Asian and North American trees²³. Thus, biodiversity
39 should ideally be studied, mapped, and explained at multiple grains¹⁴.

40 Although the abovementioned scaling issues are well-known^{13,19,24,25}, methods are lacking that
41 explicitly incorporate grain-dependence within a single model, allowing cross-grain inference and
42 predictions. Furthermore, it is still common to report patterns and drivers of biodiversity at a single
43 grain, resulting in pronounced mismatches of spatial grain among studies, and hindering synthesis. An
44 example is the debate over whether biodiversity is more associated with regional proxy variables for
45 macroevolutionary diversification and historical dispersal limitation, or with ecological drivers that
46 include climatic and other environmental drivers, as well as biotic interactions^{25–29}. While climate and
47 other ecological factors usually play a strong role [but see ref³⁰], studies differ in whether they view
48 residual regional forces being weak^{31–33} or strong^{34–36}. Even within the same group of organisms –
49 trees – there is debate regarding whether environment^{23,37–40} or regional history^{41–44} drive global
50 patterns. And yet, these studies are rarely done at a comparable spatial grain, and perhaps not
51 surprisingly, studies from smaller plot-scale analyses^{39,40} typically conclude a strong role for
52 environmental variation, whereas large-grain analyses^{43,45} show a strong role of historical
53 biogeographic processes.

54 Here, we propose a cross-grain approach that allows estimation of contemporary environmental and
55 regional predictors, as well as global patterns, of tree species richness across a continuum of grains,
56 from plots of 10 x 10 m² up to the entire continents. Our study has three main goals: (i) by explicitly
57 considering spatial grain as a modifier of the influence of ecology versus regional biogeography, we
58 aim to synthesize results among studies, and illustrate how the importance of these processes varies
59 with grain. Apart from the well-known grain-dependent effects of environment, we also focus on the

60 so far overlooked grain-dependent effects of biogeographic regions. (ii) The novelty of the approach is
61 to model grain-dependence of every predictor (spatial, regional, or ecological) within a single model
62 as having a statistical interaction with area, which enables integration of an unprecedented volume of
63 heterogeneous data from local surveys and country-wide checklists – although such interaction has
64 been occasionally tested^{16,17,36}, to our knowledge it has not been applied to both spatial and
65 environmental effects, nor for data integration and cross-grain predictions. (iii) We take the advantage
66 of being able to predict biodiversity patterns at any arbitrarily chosen grain and we map the estimates
67 of alpha, beta, and gamma diversity of trees across the entire planet.

Results and Discussion

68 **Macroecological patterns.** To explain the observed global variation of tree diversity (Fig. 1), we
69 specified two models that predict S by grain-dependent effects of environmental variables
70 (Supplementary Table 1), but differ in the way they model the grain-dependent regional component of
71 biodiversity: model REALM attributes residual variation of S to location's membership within a pre-
72 defined biogeographic realm [as in ref⁴⁶], while model SMOOTH estimates the regional imprints in S
73 directly from the data using smooth autocorrelated surfaces. Both models explain more than 90% of
74 deviance of the data (Supplementary Table 2) and both predict S that matches the observed S
75 (Supplementary Fig. 1). This is in line with other studies from large geographical extents, where 70-
76 90% model fits are common even for relatively simple climate-based models^{23,40,46–48}.

77 Next, we used model SMOOTH to predict patterns of S and beta diversity over the entire mainland, at
78 a regular grid of large hexagons of 209,903 km² and at a grid of local plots of 1 ha (Fig. 2A-C). We
79 predict latitudinal gradient of S at both grains (Fig. 2A, B, Supplementary Fig. 2), which matches the
80 traditional narrative in trees⁴⁹ and other groups [ref⁵⁰, p. 662-667]. However there are also differences
81 between the patterns at the two grains, particularly in China, East Africa, and southern North America
82 (Fig. 2C), where the plot-grain S is disproportionately lower than what would be expected from the
83 coarse-grain S . These are regions with exceptionally high beta diversity and are in the dry tropics and
84 sub-tropics with high topographic heterogeneity – examples are Ethiopian Highlands and Mexican
85 Sierra Madre ranges, which have sharp environmental gradients and patchy forests, resulting in
86 relatively low local alpha diversity but high regional gamma diversity. The exception is the predicted
87 high beta diversity in China, where the historical component of beta diversity dominates the effect of
88 environmental gradients (Fig. 2C vs F), as also suggested by refs^{23,41,51,52}, and as discussed below.

89 **Grain-dependent effects of region.** Although model REALM treats the regional biogeographic
90 effects on S as discrete, while model SMOOTH treats them as continuous, both models reveal similar
91 grain-dependence of these regional effects. At the coarse grains (i.e. in larger regions), model REALM
92 shows that the anomaly of S that is independent of environment (and thus attributed to the effect of

93 regions) is highest in the Indo-Malay region, followed by parts the Neotropics, Australasia, and
94 Eastern Palaearctic (Fig. 3). Similar pattern emerges at the coarse grain from model SMOOTH, where
95 particularly China, and Central America to some degree, are hotspots of environmentally-independent
96 S (i.e., strong effects of biogeographic regions) (Fig. 2D). This follows the existing narrative^{44,46}
97 where tree diversity is typically highest, and anomalous from the climate-driven expectation, in eastern
98 Asia. However, at the smaller plot grain, a different pattern emerges in both the REALM (Fig. 3) and
99 SMOOTH (Fig. 2E) models: the regional biogeographic effects are present, but weaker. Further, they
100 shift away from the Indo-Malay and the Neotropical regions (REALM model) or China and Central
101 America (SMOOTH model) at the coarse grains towards the equator, particularly to Australasia, at the
102 plot grain (Fig. 2F, 3).

103 These results can be viewed through the logic of species-area relationship (SAR), and its link to alpha,
104 beta, and gamma diversity^{12,53}: If environmental conditions are constant, or statistically controlled for,
105 then S depends only on area and on specific regional history. Since these interact, what emerges are
106 region-dependent SARs (in model REALM; Fig. 3), which are equivalent to grain-dependent effects of
107 regions (in model SMOOTH; Fig. 2). In both, what geographically varies is the environmentally-
108 independent local (RE_{plot} in Fig. 2E) S and regional (RE_{hex} , Fig. 2D) S , and their ratio (i.e. difference in
109 log space in Fig. 3F), which directly links to the slopes of relationships in Fig. 3. We can explain this
110 through different range dynamics in different parts of the world. Areas with high levels of
111 environmentally-independent S at large grains, such as China and Central America, could have
112 historically accumulated species that are spatially segregated with relatively small ranges, for example
113 due to allopatric speciation⁴⁴, climate refugia [as in Europe⁵⁴], or due to dispersal barriers and/or
114 large-scale habitat heterogeneity⁴⁴. This would lead to increased regional richness but contribute less
115 to local richness, leading to stronger regional effects at larger than smaller grains, as we observed.

116 We also found pronounced autocorrelation in the residuals of the REALM model at the country grain,
117 but low autocorrelation at both grains in the residuals of model SMOOTH (Supplementary Fig. 5).
118 Residual autocorrelation in S is the spatial structure that was not accounted for by environmental
119 predictors; it can emerge as a result of dispersal barriers or particular evolutionary history in a given
120 location or region^{55,56}. The autocorrelation in REALM residuals thus indicates that the discrete
121 biogeographical regions (Fig. 3A) fail to delineate areas with unique effects on S ; these are better
122 derived directly from the data, for example using the splines in model SMOOTH (Fig. 2D, E). As
123 such, the smoothing not only addresses a prevalent nuisance [i.e. biased parameter estimates due to
124 autocorrelation⁵⁷], but can also be used to delineate the regions relevant for biodiversity more
125 accurately than the use of *á priori* defined regions.

126 **Grain-dependent effects of environment.** Generally, the signs and magnitudes of the standardized
127 coefficients of environmental predictors (Fig. 4) at the plot grain are in line with those observed⁴⁶.

128 However, as far as we are aware, only Kreft & Jetz³⁶ modelled richness-environment associations as
129 grain-dependent by using the statistical interactions between an environment and area. In our analyses,
130 several of these interactions terms were significant in both models REALM and SMOOTH (Fig. 4);
131 this in line with refs²¹⁻²³, but it contrasts with Kreft & Jetz³⁶ who detected no interaction between area
132 and environment at the global extent in plants. However, the latter study lacked data from local plots
133 (i.e. had a limited range of areas). We detected the clearest grain dependence in the effect of Gross
134 Primary Productivity (GPP, a proxy for energy input) and Tree density (Fig. 4); both effects decrease
135 with area. The reason is that, as area increases, large parts of barren, arid, and forest-free land are
136 included in the large countries such as Russia, Mongolia, Saudi Arabia, or Sudan, diluting the
137 importance of the total tree density at large grains.

138 Further, we failed to detect an effect of elevation span at fine grain, but it emerged at coarse grains
139 (Fig. 4). This is in line with other studies^{21,22}, and it shows that topographic heterogeneity is most
140 important over large areas where clear barriers (mountain ranges and deep valleys) limit colonization
141 and promote diversification⁵⁸. Also note the wide credible intervals (i.e. high uncertainty) around the
142 effects of islands and most of the climate-related variables across grains (Fig. 4). A likely source of
143 this uncertainty is the collinearity between environmental and regional predictors (see below). This
144 prevented us from detecting grain-dependency of the effect of temperature, although we expected it
145 based on previous studies^{21,23}. Finally, we detected a consistently negative effect of islands on S, but
146 with broad credible intervals across all grains (Fig. 4); this uncertainty is likely caused by our binomial
147 definition of islands. We suggest that inclusion of proximity to mainland or island history (Britain or
148 the Sunda Shelf islands used to be mainland), and inclusion of remote oceanic islands, could reduce
149 the uncertainty.

150 **Regions vs environment.** We used deviance partitioning^{59,60} to assess the relative importance of
151 biogeographic regions versus environmental conditions in explaining the variation of S across grains.
152 At the global extent, the independent effects of biogeographic realms strengthened towards coarse
153 grain, from 5% at the plot grain to 20% for country grain in model REALM (Fig. 5A). In contrast, the
154 variation of S explained uniquely by environmental conditions (around 14%, Fig. 5A) showed little
155 grain dependence. However and importantly, at both grains, roughly 50% of the variation of S is
156 explained by an overlap between biogeographic realms and environment, and it is impossible to tease
157 these apart due to the collinearity between them. In other words, biogeographic realms also tend to be
158 environmentally distinct (Supplementary Figs. 6-7), i.e. they are not environmentally similar replicates
159 in different parts of the world [see also ref⁴⁶ for similar conclusion]. The same problem prevails when
160 the World is split into two halves and when the partitioning is done in each half separately (Fig. 5B,
161 C). This climate-realm collinearity at the global extent weakens our ability to draw conclusions about
162 the relative importance of contemporary environment versus historical biogeography, since by
163 accounting for environment, we inevitably throw away a large portion of the regional signal, and vice

164 versa. Thus, we caution interpretations of analyses such as ours and others^{30,31,33,46,61} inferring the
165 relative magnitude biogeographic versus environmental effects merely from contemporary
166 observational data.

167 Given this covariation, we cannot clearly say whether environment or regional effect are more
168 important in driving patterns of richness. We can, however, make statements about the grain
169 dependence of both environment and region, as above. The climate-realm collinearity is likely
170 responsible for the inflated uncertainty [as expected from ref⁶²] around the effects of environmental
171 predictors (Fig. 4) and biogeographic realms (Fig. 3), but there remains enough certainty about the
172 effects of some predictors, such as tree density or GPP (Fig. 4), which are more orthogonal to climate
173 and regions.

174 To overcome the global collinearity problem and to better answer the classical question of whether
175 diversity is more influenced by historical or contemporary processes, we suggest the following
176 alternative strategies: (i) analyze smaller subsets of data where environmental and regional data are
177 less collinear, e.g. across islands⁶³ or biogeographic boundaries^{44,64} with similar environment but
178 different history, (ii) use historical data from fossil or pollen records⁶⁵, (iii) use long-term range
179 dynamics or other patterns reconstructed from phylogenies^{66,67}, and (iv) use historical data on past
180 environmental conditions⁶⁸. Finally, (v) we see a promise in the emerging use of process-based and
181 mechanistic models in macroecology^{69,70} which can predict multiple patterns, ideally at multiple
182 grains, and as such can offer a strong tests⁷¹ of the relative importance of historical biogeography
183 versus contemporary environment in generating biodiversity, irrespectively on the mutual arrangement
184 of geography and environment.

185 **Implications.** We have compiled a global dataset on tree species richness, and used it to integrate
186 highly heterogeneous data in a model that contains grain-dependence as well as spatial autocorrelation,
187 and predicts hotspots of biodiversity across grains that span 11 orders of magnitude, from local plots
188 to the entire continents. This is an improvement of data, methods, and concepts, and importantly, we
189 reveal a critical grain-dependence in the both regional and environmental predictors. We propose that
190 this grain-dependence, together with the confounding collinearity between environment and
191 geography, is the reason why studies comparing the importance of environmental versus historical
192 biogeographic predictors of global diversity patterns have come to disparate conclusions. Studies using
193 smaller-grained data tend to find strong influence of environment^{39,40}, whereas those that use larger-
194 grained data find strong effect historical biogeography^{43,45}. We reconcile this with a grain-explicit
195 analysis and show that smaller-grain (alpha-diversity) patterns are less strongly influenced by regional
196 biogeography than larger-grained (gamma-diversity) patterns. Finally, we suggest that the advantages
197 of having a formal statistical way to directly embrace grain dependence are twofold: Not only it will
198 allow ecologists to test grain-explicit theories, but it is precisely the same grain dependence that will

199 also allow integration of heterogeneous, messy, and haphazard data from various taxonomic groups,
200 especially the data deficient ones. This is desperately needed in the field that has restricted its global
201 focus to a small number of well-surveyed taxa.

Methods

202 The complete data and R codes used for all analyses are available under CC-BY license in a GitHub
203 repository at https://github.com/petrkeil/global_tree_S. Extended description of methods is in SI Text.

204 **Data on S at the plot grain.** We compiled a global database of tree species richness in 1932 forest
205 plots, from which we selected only plots with unique coordinates, and with data on number of
206 individual trees, minimum diameter at breast height (DBH), and area of the plot. We included only
207 plots that covered a contiguous area and in which all trees within the plot above the minimum DBH
208 were determined. In case there were several plots with the exactly same geographic coordinates, we
209 chose one plot with the largest area. If areas were the same, we chose one plot randomly. This left us
210 with 1338 forest plots for our main analyses. Although all of these plots are in forests, the authors of
211 the primary studies still differ in which individuals are actually determined. For instance, authors may
212 include or exclude lianas. Thus, in the main analyses we included all plots that have the following
213 morphological scope: “trees”, “woody species”, “trees and palms”, “trees and shrubs”, “trees and
214 lianas”, “all living stems”. In a parallel sensitivity analysis we used a more stringent selection criteria
215 to create a subset of the data (see below). The data are available at
216 https://github.com/petrkeil/global_tree_S. The list of references used for data extraction is in
217 Supplementary Information.

218 **Data on S at the country grain.** We compiled data on tree species richness of 287 countries and other
219 administrative units (US and Brazilian states, Chinese provinces). We downloaded the data from
220 BONAP taxonomic data center at <http://bonap.net/tdc> for the United States ⁷², from ⁷³ for the provinces
221 of China, from Flora do Brasil 2020 at <http://floradobrasil.jbrj.gov.br> ⁷⁴, and from Botanic Gardens
222 Conservation International database GlobalTreeSearch ⁷ (accessed 18 Aug 2017) for the rest of the
223 world. To download the data from GlobalTreeSearch we used Selenium software interfaced through a
224 custom R script.

225 **Predictors of species richness.** For each plot and each country we calculated its latitude, longitude,
226 and area, and we extracted environmental variables (Supplementary Table 1) related to energy
227 availability, climate seasonality, climatic limits, topographic heterogeneity, insularity, tree density, and
228 productivity, all of which are known to predict plant and tree species richness ^{36,38,40,46,75}
229 (Supplementary Table 1). For each plot we also noted minimum DBH that was used as a criterion to

230 include tree individuals in a study. All continuous predictors were standardized to 0 mean and unit
231 variance prior to the statistical modelling.

232 **Cross-grain models.** Our core approach is that ‘grain dependence’ of an effect of a predictor can be
233 modelled as a *statistical interaction* between the predictor and area. For example, imagine a linear
234 relationship between species richness S and temperature T , defined as $S = a + bT$. Now let us assume
235 that the coefficient b also depends linearly on area (grain) A as $b = \alpha + \beta A$; by substitution we get
236 $S = a + \alpha + \beta AT$, where βAT is the interaction term. Following this logic, we built statistical models
237 that treat environmental and historical predictors of S as grain-dependent. Specifically, we built two
238 models (REALM and SMOOTH) representing the same general idea of grain-dependency, but each
239 implementing it in a somewhat different way. These models are not mutually exclusive, but are
240 complementary approaches to the same problem.

241 **Model REALM.** This model follows the traditional approach to assess regional effects on S , that is,
242 variation of S that is not accounted for by environmental predictors can be accounted for by
243 membership in pre-defined discrete geographic regions [as in ⁴⁶], a.k.a. realms. We extend this idea by
244 assuming that the effect of biogeographic regions interacts with area (i.e. grain), i.e. there is a different
245 species-area relationship at work in each region. Formally, species richness S_i in i th plot or country is
246 a negative binomial random variable $S_i \sim NegBin(\mu_i, \theta)$, where

$$247 \quad \log \mu_i = \alpha_j + \sum_{k=1}^3 A_i^k \beta_{j,k} + X_i \gamma + A_i X_i \delta, \quad (1)$$

248 and where α_j are the intercepts for each j th region, $\sum_{k=1}^3 A_i^k \beta_{j,k}$ is the interaction between a third-
249 order polynomial of area A and the j th region; we have chosen the third-order polynomial to ensure an
250 ability to produce a tri-phasic effect of area ¹² in each region. $X_i \gamma$ is the term for area-independent
251 effects of environmental predictors in a matrix X , and $A_i X_i \delta$ is the interaction term between area A
252 and X . Parameters to be estimated are the vectors α , β , γ , δ , and the constant θ . The model can be
253 specified in R package mgcv ⁷⁶ as `gam(S ~ REALM + poly(A,3):REALM + X + X:A, family='nb')`,
254 where REALM is a factor identifying the regions.

255 **Model SMOOTH.** In this model we avoid using discrete biogeographic regions; instead, we use thin-
256 plate spline functions (a.k.a. splines) ⁷⁶ of geographic coordinates. This allows us (i) to identify the
257 areas of historically accumulated S directly from the data, and (ii) to account for spatial
258 autocorrelation in model residuals ⁵⁷ at the same time. As above, $S_i \sim NegBin(\mu_i, \theta)$, but

$$259 \quad \log \mu_i = \alpha + \sum_{k=1}^3 A_i^k \beta_k + X_i \gamma + A_i X_i \delta + s_1(Lat, Lon)Plt_i + s_2(Lat, Lon)Cntr_i. \quad (2)$$

260 The notation is the same as in the previous model, with the exception of α and β now being constant,
261 and with the spline functions s_1 and s_2 (each with 14 spline bases), and with Plt_i and $Cntr_i$ as binary

262 (0 or 1) variables specifying if an observation i is a country or a plot. In R package mgcv the model
263 writes as $\text{gam}(S \sim s(\text{Lat}, \text{Lon}, \text{by}=\text{Plt.or.Cntr}, \text{bs}='s', k=14) + \text{poly}(A, 3) + X + X:A, \text{family}='nb')$,
264 where Plt.or.Cntr is a factor identifying if an observation is a plot or a country.

265 **Model fitting, inference, predictions, and sensitivity analysis.** We used a combination of maximum
266 likelihood (fast, easy to work with) and Hamiltonian Monte Carlo (slow, but handles uncertainty well)
267 to optimize and fit the models. To compare the effects of contemporary environment vs biogeographic
268 regions, we used partitioning of deviance. We used model SMOOTH to generate the global
269 predictions (Fig. 2) in a set of artificially generated plots (each with an area of 1 ha) and hexagons
270 (each with an area of 209,903 km²). We additionally tested if our results are sensitive to data sources
271 and definition of what a ‘tree’ is. All these steps are described in detail in SI Text.

References

- 272 1. Fine, P. V. A. & Ree, R. H. Evidence for a time-integrated species-area effect on the latitudinal
273 gradient in tree diversity. *Am. Nat.* **168**, 796–804 (2006).
- 274 2. Losos, E. & Leigh, E. G. *Tropical forest diversity and dynamism*. (University of Chicago Press,
275 2004).
- 276 3. Meyer, C., Kreft, H., Guralnick, R. & Jetz, W. Global priorities for an effective information basis
277 of biodiversity distributions. *Nat. Commun.* **6**, 8221 (2015).
- 278 4. Jetz, W., McPherson, J. M. & Guralnick, R. P. Integrating biodiversity distribution knowledge:
279 toward a global map of life. *TREE* **27**, 151–159 (2012).
- 280 5. Phillips, O. L. & Miller, J. S. *Global patterns of plant diversity: Alwyn H. Gentry’s forest transect*
281 *data set*. (Missouri Botanical Garden Press, 2002).
- 282 6. Sullivan, M. *et al.* Diversity and carbon storage across the tropical forest biome. *Scientific Reports*
283 **7**, (2017).
- 284 7. BGCI. GlobalTreeSearch online database. (2017).
- 285 8. Levin, S. A. Multiple scales and the maintenance of biodiversity. *Ecosystems* **3**, 498–506 (2000).
- 286 9. Chave, J. The problem of pattern and scale in ecology: what have we learned in 20 years? *Ecol*
287 *Lett* **16**, 4–16 (2013).
- 288 10. Chase, J. M. Spatial scale resolves the niche versus neutral theory debate. *J Veg Sci* **25**, 319–322
289 (2014).

- 290 11. Leibold, M. A. & Chase, J. M. *Metacommunity Ecology*. (Princeton University Press, 2017).
- 291 12. Storch, D. The theory of the nested species–area relationship: geometric foundations of
292 biodiversity scaling. *J Veg Sci* **27**, 880–891 (2016).
- 293 13. Rahbek, C. The role of spatial scale and the perception of large-scale species-richness patterns.
294 *Ecology Letters* **8**, 224 (2005).
- 295 14. Rahbek, C. & Graves, G. R. Detection of macro-ecological patterns in South American
296 hummingbirds is affected by spatial scale. *Proceedings of the Royal Society B: Biological*
297 *Sciences* **267**, 2259–2265 (2000).
- 298 15. Chase, J. M. & Knight, T. M. Scale-dependent effect sizes of ecological drivers on biodiversity:
299 why standardised sampling is not enough. *Ecology Letters* **16**, 17–26 (2013).
- 300 16. Blowes, S. A., Belmaker, J. & Chase, J. M. Global reef fish richness gradients emerge from
301 divergent and scale-dependent component changes. *Proceedings of the Royal Society B:*
302 *Biological Sciences* **284**, 20170947 (2017).
- 303 17. Lyons, S. K. & Willig, M. R. A hemispheric assessment of scale dependence in latitudinal
304 gradients of species richness. *Ecology* **80**, 2483–2491 (1999).
- 305 18. Buckley, L. B. & Jetz, W. Linking global turnover of species and environments. *PNAS* **105**,
306 17836–17841 (2008).
- 307 19. Shmida, A. & Wilson, M. V. Biological determinants of species diversity. *Journal of*
308 *Biogeography* **12**, 1–20 (1985).
- 309 20. Böhning-Gaese, K. Determinants of avian species richness at different spatial scales. *Journal of*
310 *Biogeography* **24**, 49–60 (1997).
- 311 21. Belmaker, J. & Jetz, W. Cross-scale variation in species richness–environment associations. *Glob.*
312 *Ecol. Biogeogr.* **20**, 464–474 (2011).
- 313 22. Rahbek, C. & Graves, G. R. Multiscale assessment of patterns of avian species richness. *PNAS* **98**,
314 4534–4539 (2001).
- 315 23. Wang, Z., Brown, J. H., Tang, Z. & Fang, J. Temperature dependence, spatial scale, and tree
316 species diversity in eastern Asia and North America. *PNAS* **106**, 13388–13392 (2009).

- 317 24. Whittaker, R. J., Willis, K. J. & Field, R. Scale and species richness: towards a general,
318 hierarchical theory of species diversity. *Journal of Biogeography* **28**, 453–470 (2001).
- 319 25. Ricklefs, R. E. Intrinsic dynamics of the regional community. *Ecol Lett* **18**, 497–503 (2015).
- 320 26. Vázquez-Rivera, H. & Currie, D. J. Contemporaneous climate directly controls broad-scale
321 patterns of woody plant diversity: a test by a natural experiment over 14,000 years. *Global*
322 *Ecology and Biogeography* **24**, 97–106 (2015).
- 323 27. Fine, P. V. A. Ecological and evolutionary drivers of geographic variation in species diversity.
324 *Annual Review of Ecology, Evolution, and Systematics* **46**, 369–392 (2015).
- 325 28. Harmon, L. J. & Harrison, S. Species diversity is dynamic and unbounded at local and continental
326 scales. *Am. Nat.* **185**, 584–593 (2015).
- 327 29. Rabosky, D. L. & Hurlbert, A. H. Species richness at continental scales is dominated by
328 ecological limits. *Am. Nat.* **185**, 572–583 (2015).
- 329 30. Wiens, J. J., Pyron, R. A. & Moen, D. S. Phylogenetic origins of local-scale diversity patterns and
330 the causes of Amazonian megadiversity. *Ecol. Lett.* **14**, 643–652 (2011).
- 331 31. Hawkins, B. A., Porter, E. E. & Diniz-Filho, J. A. F. Productivity and history as predictors of the
332 latitudinal diversity gradient of terrestrial birds. *Ecology* **84**, 1608–1623 (2003).
- 333 32. Algar, A. C., Kerr, J. T. & Currie, D. J. Evolutionary constraints on regional faunas: whom, but
334 not how many. *Ecology Letters* **12**, 57–65 (2009).
- 335 33. Dunn, R. R. *et al.* Climatic drivers of hemispheric asymmetry in global patterns of ant species
336 richness. *Ecol. Lett.* **12**, 324–333 (2009).
- 337 34. Araújo, M. B. *et al.* Quaternary climate changes explain diversity among reptiles and amphibians.
338 *Ecography* **31**, 8–15 (2008).
- 339 35. Belmaker, J. & Jetz, W. Relative roles of ecological and energetic constraints, diversification rates
340 and region history on global species richness gradients. *Ecol. Lett.* **18**, 563–571 (2015).
- 341 36. Kreft, H. & Jetz, W. Global patterns and determinants of vascular plant diversity. *PNAS* **104**,
342 5925–5930 (2007).
- 343 37. Currie, D. J. & Paquin, V. Large-scale biogeographical patterns of species richness of trees.
344 *Nature* **329**, 326 (1987).

- 345 38. Francis, A. P. & Currie, D. J. Global patterns of tree species richness in moist forests: another
346 look. *Oikos* **81**, 598–602 (1998).
- 347 39. Currie, D. J. *et al.* Predictions and tests of climate-based hypotheses of broad-scale variation in
348 taxonomic richness. *Ecology Letters* **7**, 1121–1134 (2004).
- 349 40. Šímová, I. *et al.* Global species–energy relationship in forest plots: role of abundance, temperature
350 and species climatic tolerances. *Global Ecology and Biogeography* **20**, 842–856 (2011).
- 351 41. Latham, R. & Ricklefs, R. E. Global patterns of tree species richness in moist forests: energy-
352 diversity theory does not account for variation in species richness. *Oikos* **67**, 325–333 (1993).
- 353 42. Ricklefs, R. E., Latham, R. E. & Qian, H. Global patterns of tree species richness in moist forests:
354 distinguishing ecological influences and historical contingency. *Oikos* **86**, 369–373 (1999).
- 355 43. Qian, H., Wiens, J. J., Zhang, J. & Zhang, Y. Evolutionary and ecological causes of species
356 richness patterns in North American angiosperm trees. *Ecography* **38**, 241–250 (2015).
- 357 44. Qian, H. & Ricklefs, R. E. Large-scale processes and the Asian bias in species diversity of
358 temperate plants. *Nature* **407**, 180–182 (2000).
- 359 45. Ricklefs, R. E., Qian, H. & White, P. S. The region effect on mesoscale plant species richness
360 between eastern Asia and eastern North America. *Ecography* **27**, 129–136 (2004).
- 361 46. Ricklefs, R. E. & He, F. Region effects influence local tree species diversity. *PNAS* **113**, 674–679
362 (2016).
- 363 47. O’Brien, E. M., Field, R. & Whittaker, R. J. Climatic gradients in woody plant (tree and shrub)
364 diversity: water-energy dynamics, residual variation, and topography. *Oikos* **89**, 588–600 (2000).
- 365 48. Field, R., O’Brien, E. M. & Whittaker, R. J. Global models for predicting woody plant richness
366 from climate: development and evaluation. *Ecology* **86**, 2263–2277 (2005).
- 367 49. Brown, J. H. *Macroecology*. (University of Chicago Press, 1995).
- 368 50. Lomolino, M. V., Riddle, B. R., Whittaker, R. J. & Brown, J. H. *Biogeography*. (Sinauer
369 Associates, 2010).
- 370 51. Qian, H. A comparison of the taxonomic richness of temperate plants in East Asia and North
371 America. *American Journal of Botany* **89**, 1818–1825 (2002).

- 372 52. Qian, H. & Ricklefs, R. E. Large-scale processes and the Asian bias in species diversity of
373 temperate plants. *Nature* **407**, 180 (2000).
- 374 53. Crist, T. O. & Veech, J. A. Additive partitioning of rarefaction curves and species-area
375 relationships: unifying alpha-, beta- and gamma-diversity with sample size and habitat area. *Ecol.*
376 *Lett.* **9**, 923–932 (2006).
- 377 54. Svenning, J.-C. & Skov, F. Limited filling of the potential range in European tree species: Limited
378 range filling in European trees. *Ecology Letters* **7**, 565–573 (2004).
- 379 55. Legendre, P. Spatial autocorrelation: trouble or new paradigm? *Ecology* **74**, 1659–1673 (1993).
- 380 56. Dormann, C. F. *et al.* Methods to account for spatial autocorrelation in the analysis of species
381 distributional data: a review. *Ecography* **30**, 609–628 (2007).
- 382 57. Dormann, C. F. Effects of incorporating spatial autocorrelation into the analysis of species
383 distribution data. *Global Ecology and Biogeography* **16**, 129–138 (2007).
- 384 58. Quintero, I. & Jetz, W. Global elevational diversity and diversification of birds. *Nature* **555**, 246–
385 250 (2018).
- 386 59. Carrete, M. *et al.* Habitat, human pressure, and social behavior: Partialling out factors affecting
387 large-scale territory extinction in an endangered vulture. *Biological Conservation* **136**, 143–154
388 (2007).
- 389 60. Randin, C. F. *et al.* Climate change and plant distribution: local models predict high-elevation
390 persistence. *Global Change Biology* **15**, 1557–1569 (2009).
- 391 61. White, E. P. & Hurlbert, A. H. The combined influence of the local environment and regional
392 enrichment on bird species richness. *The American Naturalist* **175**, E35–E43 (2010).
- 393 62. Dormann, C. F. *et al.* Collinearity: a review of methods to deal with it and a simulation study
394 evaluating their performance. *Ecography* **36**, 27–46 (2013).
- 395 63. Rominger, A. J. *et al.* Community assembly on isolated islands: macroecology meets evolution.
396 *Glob. Ecol. Biogeogr.* **25**, 769–780 (2016).
- 397 64. Swenson, N. G. *et al.* Constancy in functional space across a species richness anomaly. *The*
398 *American Naturalist* **187**, E83–E92 (2016).

- 399 65. Šizling, A. L. *et al.* Can people change the ecological rules that appear general across space?
400 *Global Ecology and Biogeography* **25**, 1072–1084 (2016).
- 401 66. Quintero, I., Keil, P., Jetz, W. & Crawford, F. W. Historical biogeography using species
402 geographical ranges. *Syst Biol* **64**, 1059–1073 (2015).
- 403 67. Arias, J. S. An event model for phylogenetic biogeography using explicitly geographical ranges.
404 *Journal of Biogeography* **44**, 2225–2235 (2017).
- 405 68. Jetz, W. & Fine, P. V. A. Global gradients in vertebrate diversity predicted by historical area-
406 productivity dynamics and contemporary environment. *PLOS Biology* **10**, e1001292 (2012).
- 407 69. Cabral, J. S., Valente, L. & Hartig, F. Mechanistic simulation models in macroecology and
408 biogeography: state-of-art and prospects. *Ecography* **40**, 267–280 (2017).
- 409 70. Connolly, S. R., Keith, S. A., Colwell, R. K. & Rahbek, C. Process, mechanism, and modeling in
410 macroecology. *Trends in Ecology & Evolution* **32**, 835–844 (2017).
- 411 71. McGill, B. Strong and weak tests of macroecological theory. *Oikos* **102**, 679–685 (2003).
- 412 72. Kartesz, J. T. *The Biota of North America Program (BONAP)*. (North American Plant Atlas,
413 2015).
- 414 73. Qian, H. Environmental determinants of woody plant diversity at a regional scale in China. *PLoS*
415 *ONE* **8**, e75832 (2013).
- 416 74. Anonymous. *Flora do Brazil 2020 under construction*. (2017).
- 417 75. Liang, J. *et al.* Positive biodiversity-productivity relationship predominant in global forests.
418 *Science* **354**, aaf8957–aaf8957 (2016).
- 419 76. Wood, S. N. *Generalized additive models: an introduction with R*. (CRC Press/Taylor & Francis
420 Group, 2017).
- 421

Acknowledgements

422 We thank to Dylan Craven and Irena Šimová for valuable advice, and to Robert Ricklefs, Shane
423 Blowes and two anonymous referees for critical comments that greatly improved early versions of the
424 manuscript. We acknowledge the support of the German Centre for Integrative Biodiversity Research
425 (iDiv) Halle-Jena-Leipzig funded by the German Research Foundation (FZT 118).

Author contributions

426 P.K. formalized the ideas, collated the data, performed the analyses, and led the writing. J.M.C.
427 proposed the initial idea, contributed to its development, discussed the results, and contributed to the
428 writing.

Competing interests

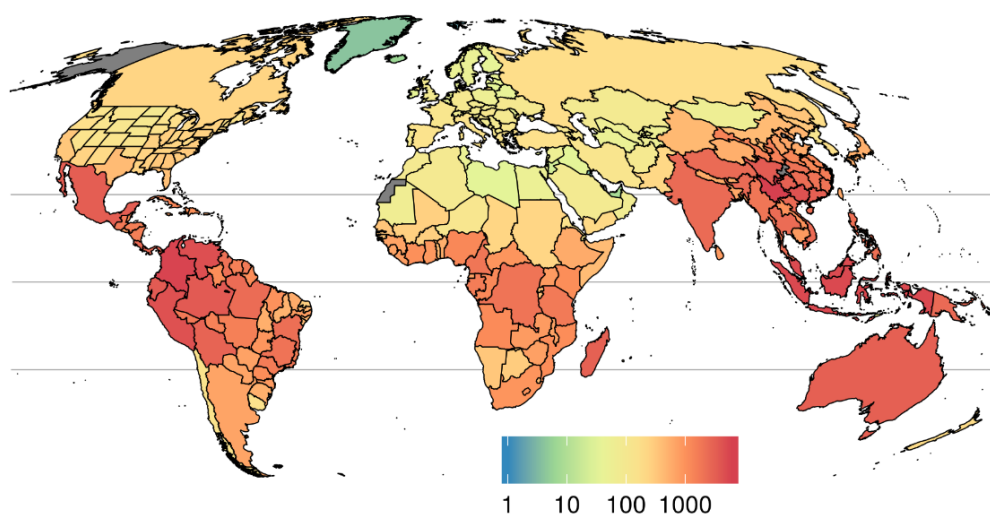
429 The authors declare no competing interests

Corresponding author

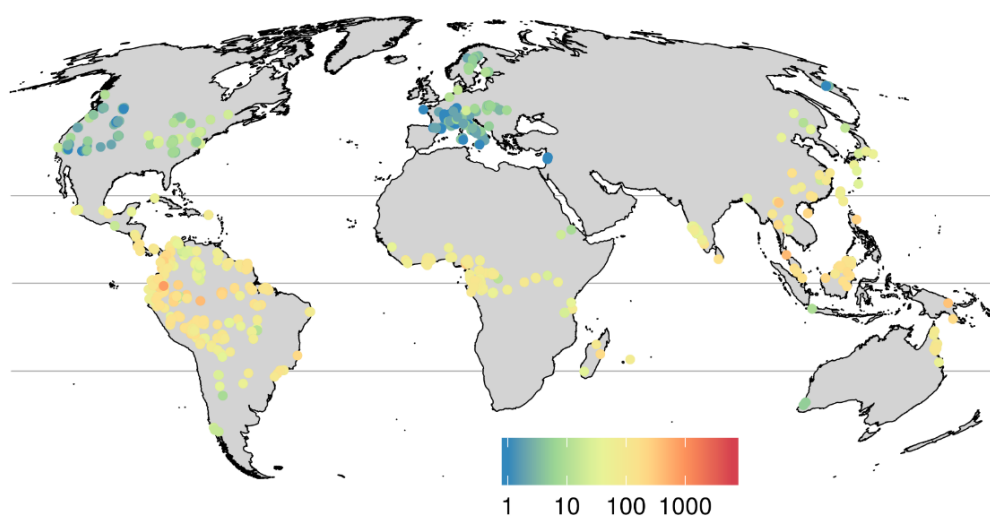
430 Petr Keil, German Centre for Integrative Biodiversity Research (iDiv), Halle-Jena-Leipzig, Deutscher
431 Platz 5e, 04103 Leipzig, Germany. Phone: 00420 776 371 840. Email: pkeil@seznam.cz

Figures

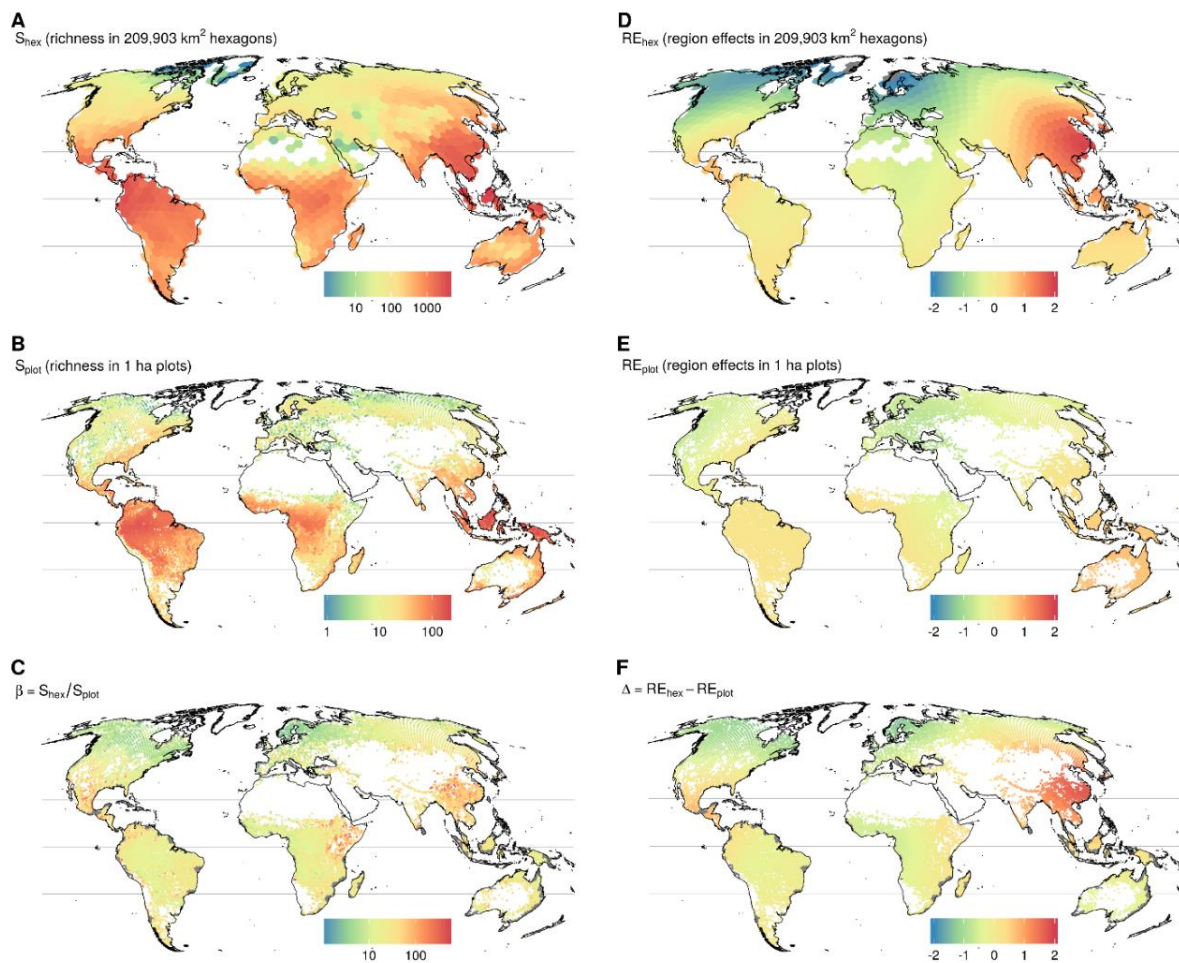
A



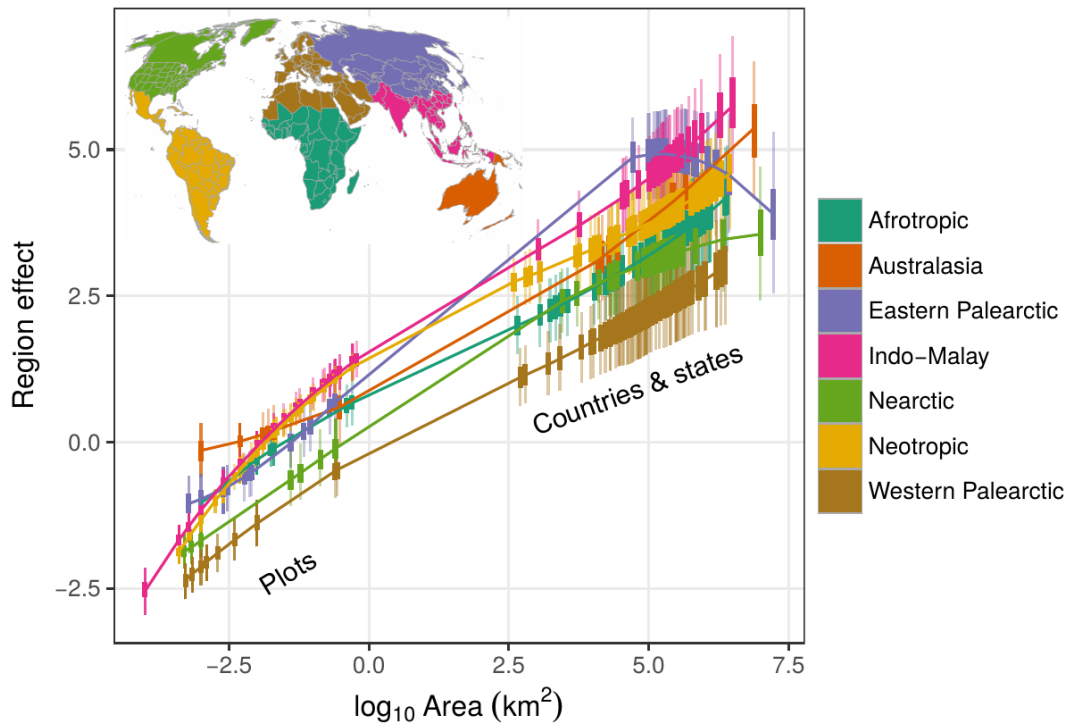
B



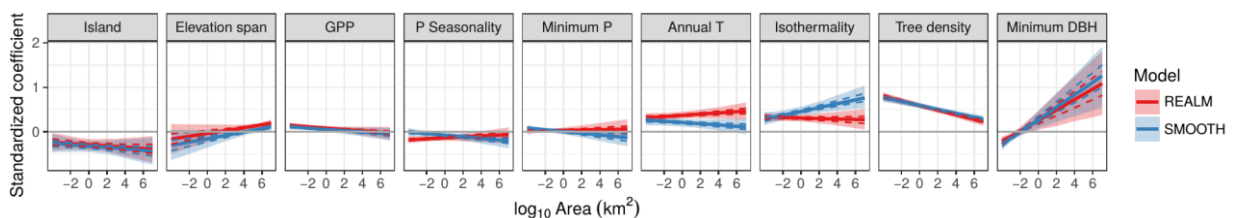
432 **Figure 1** | Raw data on observed tree species richness S (\log_{10} scale) (A) at the country/states grain
433 with 287 spatial units, and (B) at the plot grain with 1338 plots. Maps use Mollweide projection.



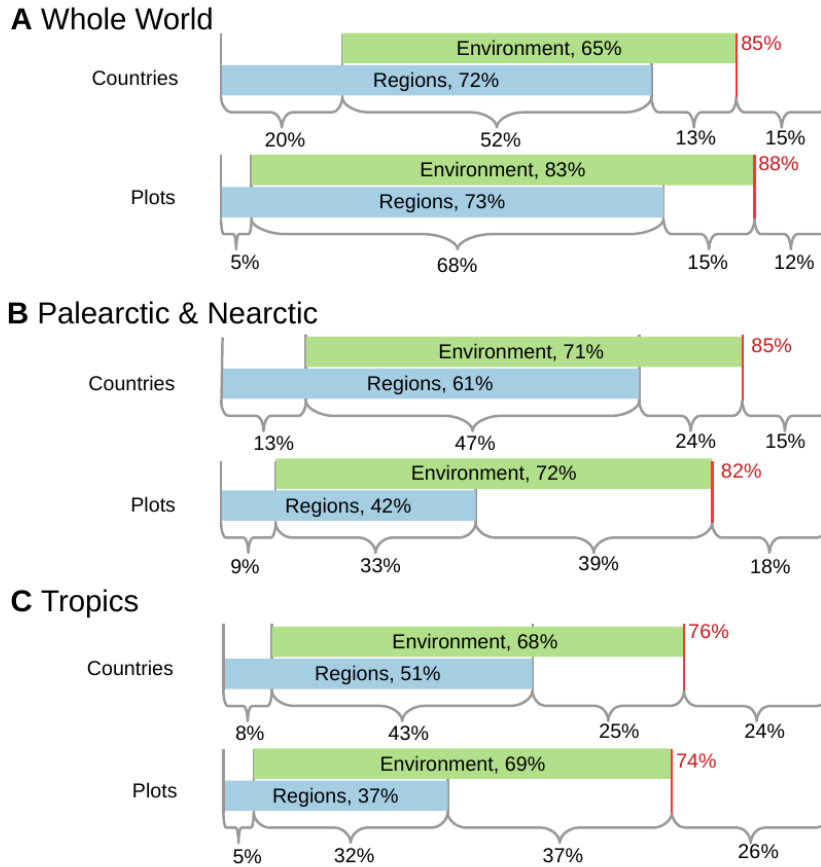
434 **Figure 2** | Predicted patterns of species richness, beta diversity, and continuous region effects from
435 model SMOOTH at two spatial grains. Species richness S_{hex} is regional gamma diversity (A), S_{plot} is
436 alpha diversity (B), and their ratio is beta diversity (C). Regional effects RE (D-E) are smooth splines
437 representing anomaly of S (on natural log scale) from expectation based purely on environmental
438 conditions. White mainland areas are those for which we lacked data on at least one predictor. Panels
439 A-C use \log_{10} scale.



440 **Figure 3** | Grain-dependent effects of discrete biogeographic regions on species richness, estimated by
 441 model REALM across a continuum of areas (grains). A region effect is what remains after accounting
 442 for effects of all other predictors at a given area. Vertical bars show 2.5, 25, 50, 75 and 95.5
 443 percentiles of posterior densities.



444 **Figure 4** | Grain-dependent standardized model coefficients for environmental predictors. The y-axis
 445 is the value of model coefficient, not S -- although the lines for tree density are declining, its effect on
 446 S is positive at all values of area. Lines and shadings are 2.5, 25, 50, 75 and 95.5 percentiles of
 447 posterior densities.



448 **Figure 5** | Partitioning of deviance of species richness (S) to components explained by environment vs
 449 biogeographic regions, at two spatial grains, using the REALM model. In panel A the extent is the
 450 whole World, panel B uses data from the Palearctic and Nearctic regions. Panel C uses data from the
 451 Neotropic, Afrotropic, Indo-Malay and Australasian regions. Red lines are the total deviance explained
 452 by the full model with both the environment and regions.