

# Global patterns and drivers of tree diversity integrated across a continuum of spatial grains

Petr Keil <sup>1\*</sup> and Jonathan M. Chase <sup>1,2</sup>

**Controversy remains over what drives patterns in the variation of biodiversity across the planet. The resolution is obscured by lack of data and mismatches in their spatial grain (scale), and by grain-dependent effects of the drivers. Here we introduce cross-scale models integrating global data on tree-species richness from 1,336 local forest surveys and 282 regional checklists, enabling the estimation of drivers and patterns of biodiversity across spatial grains. We uncover grain-dependent effects of both environment and biogeographic regions on species richness, with a striking positive effect of Southeast Asia at coarse grain that disappears at fine grains. We show that, globally, biodiversity cannot be attributed purely to environmental or regional drivers, as the regions are environmentally distinct even within a single latitudinal band. Finally, we predict global maps of biodiversity at local (plot-based) and regional grains, identifying areas of exceptional beta-diversity in China, East Africa and North America. By allowing the importance of drivers of diversity to vary with grain in a single model, our approach unifies disparate results from previous studies regarding environmental versus biogeographic predictors of biodiversity, and enables efficient integration of heterogeneous data.**

Why are there fewer than 100 species of trees that live in the millions of km<sup>2</sup> of boreal forests in Eurasia and North America<sup>1,2</sup>, while there can be hundreds of species co-occurring in as little as 50 ha in tropical forests of South America and Asia<sup>3</sup>? What drives global variation in the numbers of species that live in different places, and where exactly are the places of highest biodiversity? The fundamental scientific appeal of these questions can be traced back at least to Humboldt<sup>4</sup>, yet understanding biological diversity has taken on new urgency as it faces threats from increasing human pressure. However, despite decades of research and hypotheses proposed<sup>5–9</sup>, there has been lack of consensus on the determinants of global variations in diversity, and for most taxa the global map of biodiversity is still largely incomplete.

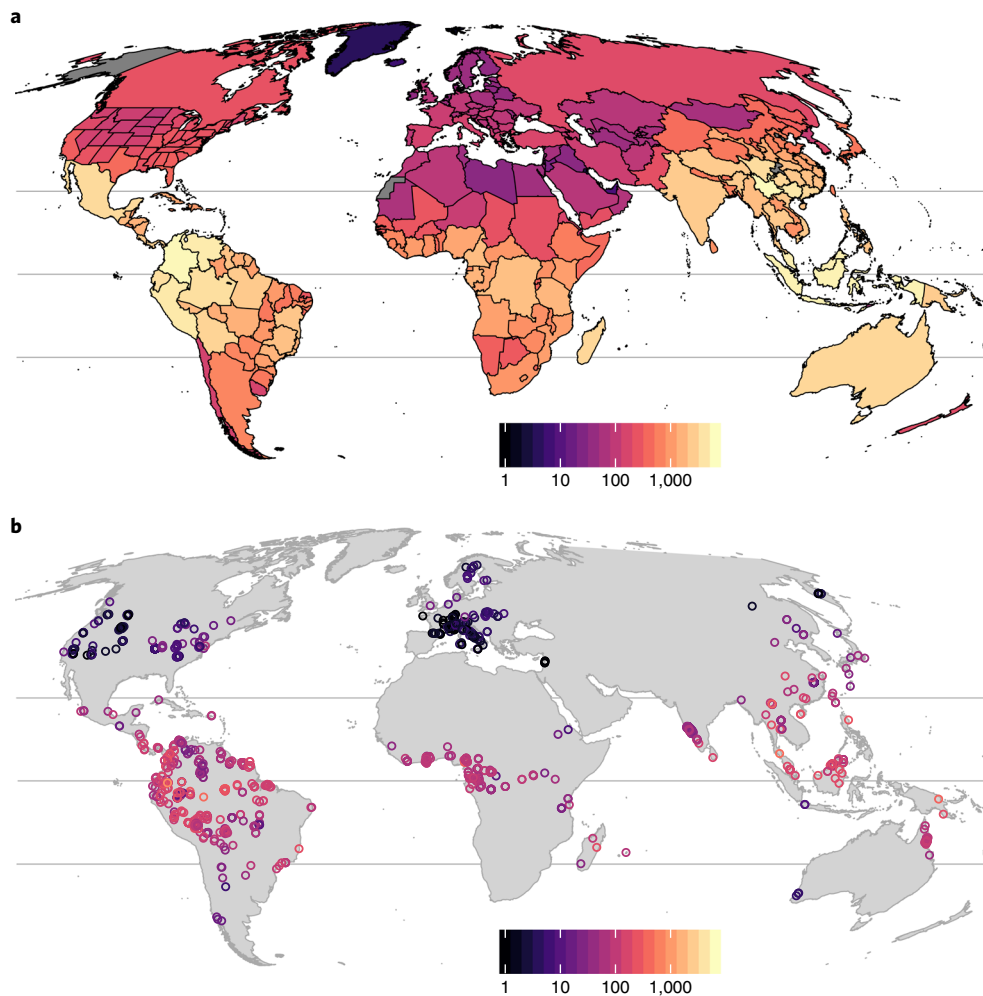
The most important obstacle to answering the fundamental questions about drivers and patterns of biodiversity is a lack of data, especially in places where diversity is thought to be highest<sup>10,11</sup>. But even in regions and taxa that have been well sampled, the data are a heterogeneous mixture of point observations, survey plots, and regional checklists, all with varying area and sampling protocol<sup>11</sup>. For example, for trees, there are hundreds of 0.1 ha Gentry forest plots mostly in the New World<sup>12</sup>, hundreds of 1 ha ForestPlots.net plots throughout tropical forests<sup>13</sup>, dozens of CTFS-ForestGEO plots of more than 2 ha ([www.forestgeo.si.edu](http://www.forestgeo.si.edu)), hundreds of published regional checklists<sup>14</sup>, and hundreds to thousands of other published surveys and checklists scattered throughout the published and grey literature. These together hold key information on the global distribution of tree biodiversity, and there are initiatives that mobilize this information over large scale<sup>15</sup>, yet the lack of methods to address differences in sampling has so far prevented their integration for the purpose of model-based prediction and inference.

Further, as could be said for many problems in ecology, attempts to map global biodiversity and to assess its potential drivers are severely complicated by the issues of spatial scale<sup>16–19</sup>: The most straightforward, but fundamental, issue is that the number of species ( $S$ ) increases nonlinearly with area<sup>20</sup>. This is why patterns in

the variation of biodiversity from place to place cannot be readily inferred from sampling locations of varying area. However, even when sets of sampling locations do have a constant area (hereafter grain), a spatial pattern of  $S$  observed at small grains will usually differ from a pattern observed at large grains<sup>21–23</sup>. Examples include the grain dependence of altitudinal<sup>21</sup> and latitudinal<sup>24,25</sup> diversity gradients. The reason for this is that beta-diversity (the ratio between fine-grain alpha-diversity and coarse-grain gamma-diversity) typically varies over large geographic extents<sup>26</sup>. Finally drivers and predictors of diversity have different associations with  $S$  at different grains<sup>27–30</sup>. For example, at global and continental extents, the association of  $S$  with topography increases with grain in Neotropical birds<sup>30</sup> and the association with temperature increases with grain in global vertebrates<sup>29</sup> and eastern Asian and North American trees<sup>31</sup>. Thus, biodiversity should ideally be studied, mapped, and explained at multiple grains<sup>22</sup>.

Although the abovementioned scaling issues are well known<sup>21,27,32,33</sup>, methods that explicitly incorporate grain-dependence within statistical models of biodiversity, which would allow cross-grain inference and predictions, are lacking. Furthermore, it has been common to report patterns and drivers of biodiversity at a single grain, resulting in pronounced mismatches of spatial grain among studies, but also offering an opportunity for synthesis. An example is the debate over whether biodiversity is more associated with regional proxy variables for macroevolutionary diversification and historical dispersal limitation, or with ecological drivers that include climatic and other environmental drivers, as well as biotic interactions<sup>9,33–36</sup>. Although climate and other ecological factors usually play a strong role (but see ref. <sup>37</sup>), studies differ in whether they view residual effects of biogeographic regions on diversity (after accounting for climate and environment), as being weak<sup>38–40</sup> or strong<sup>41–43</sup>. Even within the same growth form of organisms—trees—there is debate regarding whether environment<sup>5,31,44–46</sup> or regional history<sup>47–50</sup> are more important in driving global patterns. And yet, these studies are rarely done at a comparable spatial grain,

<sup>1</sup>German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Leipzig, Germany. <sup>2</sup>Institute of Computer Science, Martin-Luther University Halle-Wittenberg, Halle (Saale), Germany. \*e-mail: [pkeil@seznam.cz](mailto:pkeil@seznam.cz)



**Fig. 1 | Raw data on observed tree species richness  $S$  ( $\log_{10}$  scale).** **a**, Country/states grain with 282 spatial units. **b**, Plot grain with 1,336 plots. Maps use Mollweide projection. Horizontal lines are the Equator and tropics.

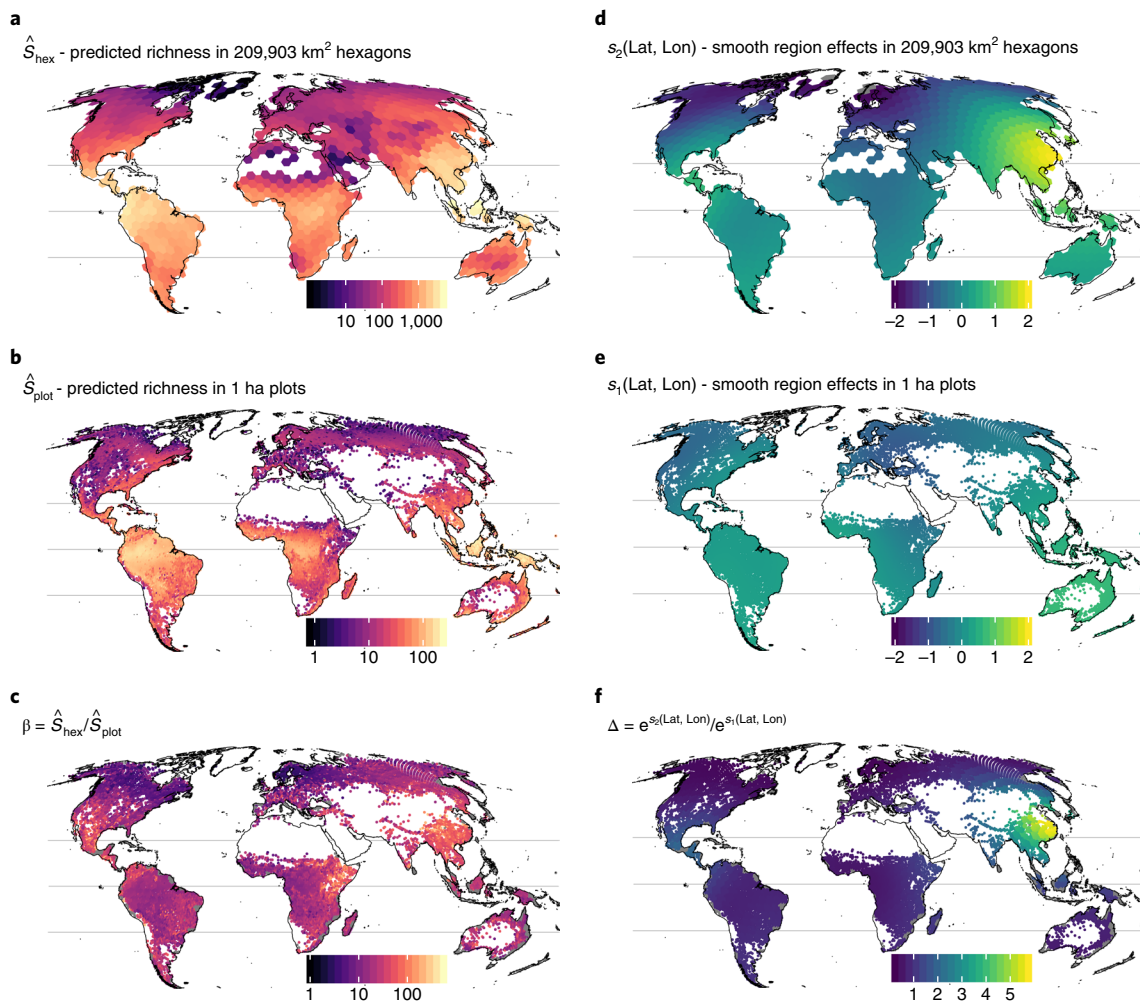
and perhaps not surprisingly, studies from smaller plot-scale analyses<sup>6,46</sup> typically conclude a strong role for environmental variation in driving patterns of biodiversity, whereas large-grain analyses<sup>49,51</sup> demonstrate a strong role of historical biogeographic processes.

Here we propose a cross-grain approach that allows estimation of the role of contemporary environmental and regional predictors of, and prediction of global patterns in, tree species richness across a continuum of spatial grains, from small forest plots (for example, 0.01 or 0.1 ha) up to entire continents. Our study has three main goals: (1) by explicitly considering spatial grain as a modifier of the influence of ecology versus regional biogeography, we aim to synthesize results among studies, and illustrate how the importance of these processes varies with grain. Apart from the well-known grain-dependent effects of environment, we also focus on the so-far-overlooked grain-dependent effects of biogeographic regions. (2) The novelty of the approach is to model grain-dependence of every predictor (spatial, regional, or ecological) within a single model as having a statistical interaction with area, which enables the integration of an unprecedented volume of heterogeneous data from local surveys and country-wide checklists. Although such interaction has been tested occasionally<sup>24,43,52</sup>, it has not been applied to both spatial and environmental effects, nor for data integration and cross-grain predictions. (3) Finally, we take the advantage of being able to predict biodiversity patterns at any desired grain and we map the estimates of alpha-, beta-, and gamma-diversity of trees across Earth.

## Results and Discussion

**Macroecological patterns.** To explain the observed global variation of tree diversity (Fig. 1), we specified two models that predict  $S$  by grain-dependent effects of environmental variables, but differ in the way they model the grain-dependent regional component of biodiversity: model REALM attributes residual variation of  $S$  to locations' membership within a pre-defined biogeographic realm (as in ref. <sup>7</sup>), while model SMOOTH estimates the regional imprints in  $S$  directly from the data using smooth autocorrelated surfaces. Both models explain more than 90% of the deviance of the data (Supplementary Table 1), both predict a value for  $S$  that matches the observed  $S$  (Supplementary Fig. 1), and they give good out-of-sample predictive performance (Supplementary Figs. 2 and 3). This is in line with other studies from large geographical extents, where 70–90% model fits are common even for relatively simple climate-based models<sup>7,31,46,53,54</sup>.

Next, we used model SMOOTH to predict patterns of  $S$  and beta-diversity over the entire mainland, on a regular grid of large hexagons of 209,903 km<sup>2</sup> (Fig. 2a) and on a grid of local plots of 1 ha (Fig. 2b). On average, for a given 1 ha plot or hexagon, the 95% prediction interval spans 1 order of magnitude around the median predicted  $S$  (Supplementary Fig. 4h,j), with highest prediction uncertainty in areas with extreme environments or with no plot data, such as deserts and arctic regions (Supplementary Fig. 4k,l). We predict a pronounced latitudinal gradient of  $S$  at both grains (Fig. 2a,b and



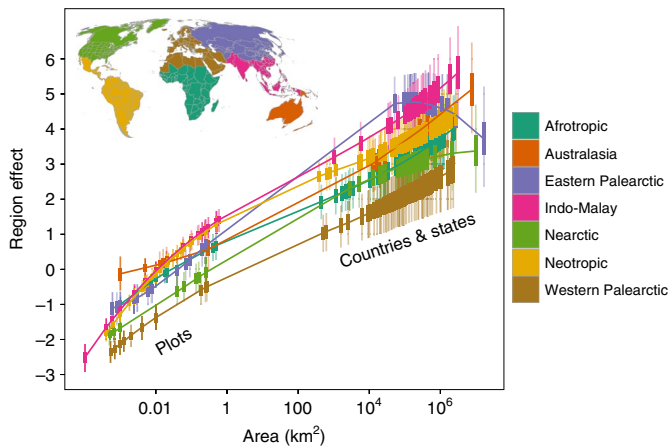
**Fig. 2 | Predicted patterns of species richness, beta-diversity, and continuous region effects at two spatial grains. a–c.** All maps are derived from model SMOOTH. Species richness  $\hat{S}_{\text{hex}}$  (**a**) is regional gamma-diversity,  $\hat{S}_{\text{plot}}$  (**b**) is alpha-diversity, and their ratio (**c**) is beta-diversity. In **a–c**, the  $\log_{10}$  scale is used, and all maps use Mollweide projection. In **d,e**, Region effects are smooth splines representing an anomaly of  $S$  (on natural log scale) from expectation based purely on environmental conditions; they are the  $s_1(\text{Lat, Lon})$  and  $s_2(\text{Lat, Lon})$  terms from equation (3) (see Methods). For example, being in China increases the richness of a hexagon (**d**) roughly  $e^{1.8}/e^{1.2} = 5$  times, compared to a hexagon in Mexico at a similar latitude. **f**, The degree to which the region effect is stronger at the hexagon grain, relative to the 1 ha plot grain. White mainland areas are those for which we lacked data on at least one predictor.

Supplementary Fig. 5), which matches other empirical studies of trees<sup>55</sup>, all vascular plants<sup>56</sup>, and other groups (ref. <sup>57</sup>, pages 662–667). However there are also differences between the patterns at the two grains, particularly in China, East Africa, and southern North America (Fig. 2c). These are regions with exceptionally high beta-diversity and are in the dry tropics and sub-tropics with high topographic heterogeneity—examples include the Ethiopian Highlands and Mexican Sierra Madre ranges, which have sharp environmental gradients and patchy forests, resulting in relatively low local alpha-diversity but high regional gamma-diversity. The exception is the predicted high beta-diversity in China, where the historical component of beta-diversity dominates the effect of environmental gradients (compare Fig. 2c and Fig. 2f). This exception has also been suggested previously<sup>31,47,50,58</sup>, and is discussed below.

**Grain-dependent effects of region.** Any geographic pattern (for example, a gradient or a regionally elevated richness) of  $S$  that remains after accounting for the effect of environmental drivers can be seen as a ‘region effect’, potentially reflecting unique diversification history and dispersal limitations of a given region. Although model REALM treats the region effects on  $S$  as discrete, while model

SMOOTH treats them as continuous, both models reveal similar grain-dependence of these regional effects. At coarse grains (that is area  $>100 \text{ km}^2$ ), model REALM shows that the regional anomaly of  $S$  that is independent of environment is highest in the Indo-Malay region, followed by parts of the Neotropics, Australasia, and Eastern Palearctic (Fig. 3 and Supplementary Fig. 6). A similar pattern emerges at coarse grains from model SMOOTH, in which particularly China and Central America are hotspots of environmentally independent  $S$  (that is, there are strong effects of biogeographic regions) (Fig. 2d). This follows the existing narrative<sup>7,50</sup> where tree diversity is typically highest, and anomalous from the climate-driven expectation, in eastern Asia. However, at the smaller plot grain, the regional biogeographic effects are present, but weaker in both the REALM (Fig. 3) and SMOOTH (Fig. 2e) models. Further, the regional effects shift away from the Indo-Malay and the Neotropical regions (REALM model) or China and Central America (SMOOTH model) at the coarse grains towards the Equator, particularly to Australasia, at the plot grain (Figs. 2e and 3).

These results can be viewed through the logic of species–area relationship (SAR), and its link to alpha-, beta-, and gamma-diversity<sup>20,59</sup>: if environmental conditions are constant (or statistically controlled



**Fig. 3 | Grain-dependent effects of discrete pre-defined biogeographic regions on species richness.** A region effect is a contribution of a given region to expected richness in a given plot or country (on a natural log scale) that remains after accounting for the grain-dependent effects of all environmental predictors in model REALM. Even more technically, the region effects here are the  $\alpha_j + \sum_{k=1}^3 \log(A_i)^k \beta_{j,k}$  terms from equation (2), excluding the intercept. These effects are on a natural log scale and are relative to the intercept (baseline) of 3.95. For example, in an area of  $10^3 \text{ km}^2$ , Nearctic region contributes  $e^{3.95+2}/e^{3.95+1.3} = 2$  times more species compared to Western Palearctic region, which is a difference of 193 species. Also, note the data gap between the plots and the countries (at grains from 1 to  $100 \text{ km}^2$ ). Vertical bars show 2.5, 25, 50, 75 and 95.5 percentiles of posterior densities.

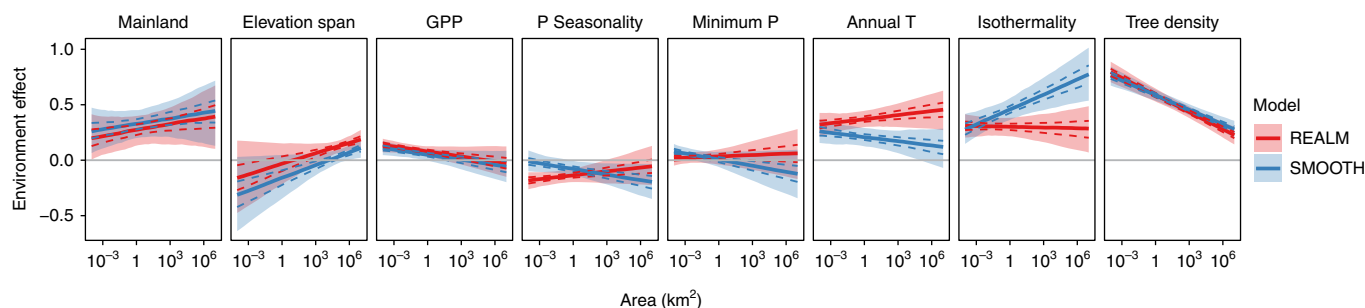
for) then  $S$  depends only on area and on specific regional history. Since these interact, what emerge are region-dependent SARs in model REALM (Fig. 3), which are equivalent to grain-dependent effects of regions in model SMOOTH (Fig. 2). In both models, what geographically varies is the environmentally-independent local  $S$  (Fig. 2e) and regional  $S$  (Fig. 2d), as well as their ratio ( $\Delta$  in Fig. 2f), which directly links to the slopes of relationships in Fig. 3. One way to explain this through different range dynamics in different parts of the world. Areas with high levels of environmentally independent  $S$  at large grains, such as China and Central America, could have historically accumulated species that are spatially segregated with relatively small ranges, for example, by being climate refuges (as in Europe<sup>60</sup>), or owing to dispersal barriers and/or large-scale habitat heterogeneity<sup>50</sup>. This would lead to increased regional richness but contribute less to local richness, leading to stronger regional effects at larger grains than at smaller grains, as we observed. An alternative explanation of the pattern would be elevated diversification rates at large grains in China and Central America; however, we think this is unlikely, given that these areas do not exhibit elevated diversification rates in other groups<sup>42,61</sup>.

We also found pronounced autocorrelation in the residuals of the REALM model at the country grain, but low autocorrelation at both grains in the residuals of model SMOOTH (Supplementary Fig. 8). Residual autocorrelation in  $S$  is the spatial structure that was not accounted for by environmental predictors; it can emerge as a result of dispersal barriers or a particular evolutionary history in a given location or region<sup>62,63</sup>. The autocorrelation in REALM residuals thus indicates that the discrete biogeographical regions (Fig. 3) fail to delineate areas with unique effects on  $S$ . These are better derived directly from the data, for example, using the splines in model SMOOTH (Fig. 2d,e). As such, the smoothing not only addresses a prevalent nuisance (that is, biased parameter estimates due to autocorrelation<sup>64</sup>), but can also be used to delineate the regions relevant for biodiversity more accurately than the use of a priori defined regions.

**Grain-dependent effects of environment.** Generally, the signs and magnitudes of the coefficients of environmental predictors (Fig. 4) at the plot grain are in line with those observed elsewhere<sup>7</sup>. However, as far as we are aware, only Kreft and Jetz<sup>43</sup> modelled richness–environment associations as grain-dependent by using the statistical interactions between an environment and area. In our analyses, several of these interaction terms were significant in both models REALM and SMOOTH (Fig. 4). This is in agreement with some previous work<sup>29–31</sup>, but contrasts with Kreft and Jetz<sup>43</sup> who detected no interaction between area and environment at the global extent in plants. However, the lack of area by environment interaction in their study might have been due to a limited range of areas (grains) examined. We detected clear grain dependence, supported by both models, in the effects of tree density and gross primary productivity (GPP, a proxy for energy input); both effects decrease with area (Fig. 4). The reason for this is that, as area increases, large parts of barren, arid, and forest-free land are included in the large countries such as Russia, Mongolia, Saudi Arabia, or Sudan, diluting the importance of the total tree density at large grains.

We failed to detect an effect of elevation span at fine grains (probably because the elevation data themselves were coarse-grained; see Supplementary Discussion), but it emerged at coarse grains (Fig. 4), in line with other studies<sup>29,30</sup>. This suggests that topographic heterogeneity is important over large areas in which clear barriers (mountain ranges and deep valleys) limit colonization and promote diversification<sup>65</sup>, or that it creates refuges in which species can persist during adverse environmental conditions<sup>66</sup>. Also note the high uncertainty around the effects in the climate-related variables across grains (Fig. 4). A probable source of this uncertainty is the co-linearity between environmental and regional predictors (see below in ‘Regions versus environment’). This prevented us from detecting the grain-dependency of the effect of temperature, although we expected it on the basis of previous studies<sup>29,31</sup>. Finally, we detected a consistent positive effect of mainland as compared to islands, which is expected<sup>67</sup>. However, the effect had broad credible intervals across all grains (Fig. 4); this uncertainty is likely caused by our binomial definition of islands, by the lack of consideration of distance from mainland, and by the classification of some of countries as mainland, although they also overlap islands (see Supplementary Methods and Supplementary Discussion).

**Regions versus environment.** We used deviance partitioning<sup>68,69</sup> to assess the relative importance of biogeographic regions versus environmental conditions in explaining the variation of  $S$  across grains. At the global extent, the independent effects of biogeographic realms strengthened towards coarse grain, from 5% at the plot grain to 20% for country grain in model REALM (Fig. 5a). In contrast, the variation of  $S$  explained uniquely by environmental conditions (around 14%, Fig. 5a) showed little grain dependence. However, and importantly, at both grains, roughly half of the variation of  $S$  is explained by an overlap between biogeographic realms and environment, and it is impossible to tease these apart owing to the co-linearity between them. In other words, biogeographic realms also tend to be environmentally distinct (Supplementary Figs. 9 and 10); that is, they are not environmentally similar replicates in different parts of the world (see also ref. <sup>7</sup> for a similar conclusion). The same problem prevails when the Earth is split into two halves and when the partitioning is done in each half separately (Fig. 5b,c). This climate–realm co-linearity at the global extent weakens our ability to draw conclusions about the relative importance of contemporary environment versus historical biogeography, as by accounting for environment, we inevitably throw away a large portion of the regional signal, and vice versa. Thus, we caution interpretations of analyses such as ours and others<sup>7,37,38,40,70</sup> inferring the relative magnitude biogeographic versus environmental effects merely from contemporary observational data.



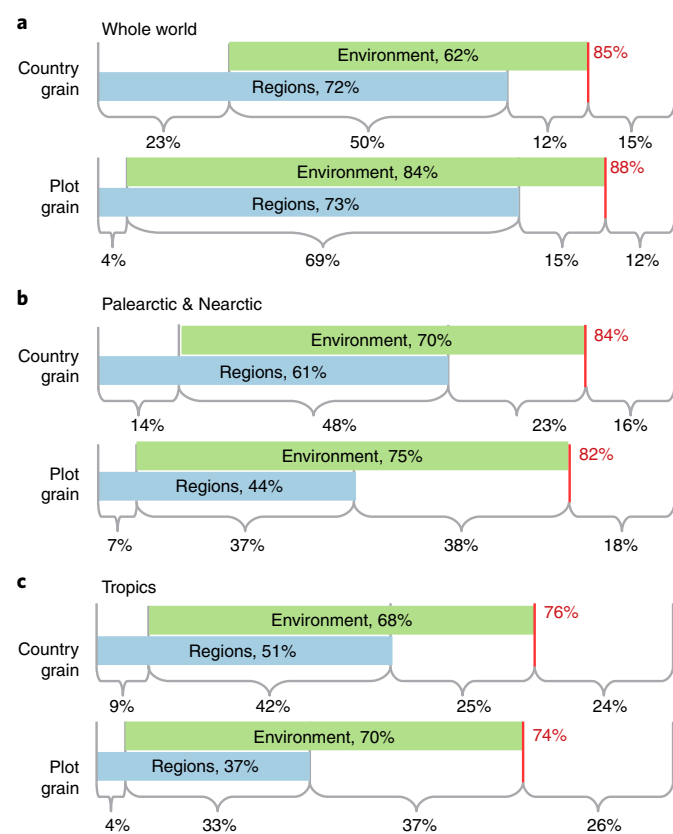
**Fig. 4 | Grain-dependent effects of environmental predictors on species richness.** These are the effects of the predictors on expected species richness that remain after accounting for the grain-dependent effects of regions. Specifically, the environment effects are the  $X_i\gamma + X_i\log(A_i)\delta$  terms from equations (2) and (3) (see Methods). The y-axis is the value of model coefficient, not  $S$ ; thus, although the lines for tree density are declining, its effect on  $S$  is positive at all values of area. Also, these effects are derived from predictors standardized to zero mean and variance of 1 (with the exception of the mainland effect), are on a natural log scale, since the models use log link. Lines and shadings are 2.5, 25, 50, 75 and 95.5 percentiles of posterior densities.

Given this covariation, we cannot clearly say whether environment or regional effect are more important in driving patterns of richness. We can, however, make statements about the grain dependence of both environment and region, as above. The climate–realm co-linearity is likely responsible for the inflated uncertainty<sup>71</sup> around the effects of environmental predictors (Fig. 4) and biogeographic realms (Fig. 3), but there remains enough certainty about the effects of some predictors, such as tree density or GPP (Fig. 4), which are more orthogonal to climate and regions.

To overcome the global co-linearity problem and to better answer the classical question of whether diversity is more influenced by historical or contemporary processes, we suggest the following alternative strategies: (1) analyse smaller subsets of data in which environmental and regional data are less collinear, such as across islands<sup>72</sup> or biogeographic boundaries<sup>50,73</sup> with similar environments, but different history; (2) use historical data from fossil or pollen records<sup>74</sup>; (3) use long-term range dynamics or other patterns reconstructed from phylogenies<sup>75,76</sup>; (4) use predictors reflecting past environmental conditions<sup>77,78</sup> or predictors that statistically interact with time<sup>79</sup>; finally, (5) we see a promise in the emerging use of process-based and mechanistic models in macroecology<sup>80,81</sup>, which can predict multiple patterns, ideally at multiple grains, and as such can offer a strong test<sup>82</sup> of the relative importance of historical biogeography versus contemporary environment in generating biodiversity.

**Conclusions.** We have compiled a global dataset on tree species richness, and used it to integrate highly heterogeneous data in a model that contains grain-dependence as well as spatial autocorrelation, and predicts patterns of biodiversity across grains that span 11 orders of magnitude, from local plots to the entire continents. This is an improvement of data, methods, and concepts, and importantly, we reveal a critical grain-dependence in both regional and environmental predictors. We propose that this grain-dependence, together with the confounding co-linearity between environment and geography, is the reason why studies comparing the importance of environmental versus historical biogeographic predictors of global diversity patterns have come to disparate conclusions. Studies using smaller-grained data tend to find strong influence of environment<sup>5,46</sup>, whereas those that use larger-grained data find strong effect historical biogeography<sup>49,51</sup>. We reconcile this with a grain-explicit analysis and show that smaller-grain (alpha-diversity) patterns are less influenced by regional biogeography than larger-grained (gamma-diversity) patterns. Finally, we suggest that the advantages of having a formal statistical way to directly embrace grain dependence are twofold: not only will it allow ecologists to test grain-explicit theories, but it is precisely the same grain dependence

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



**Fig. 5 | Partitioning of deviance of species richness (S) to fraction explained by environment versus fraction explained by biogeographic regions.** The fractions were obtained using the REALM model, fitted at two grains of the data. Specifically, the environmental fraction is the deviance explained by the  $X_i\gamma + X_i\log(A_i)\delta$  term from equation (2) plus an intercept, and the regional fraction by the  $\alpha_j + \sum_{k=1}^3 \log(A_i)^k \beta_{j,k}$  term. **a–c**, Data are from the whole world (**a**), the Palearctic and Nearctic (**b**) and from the Neotropical, Afrotropical, Indo-Malayan and Australasian regions (**c**). Red lines represent the total deviance explained by the full model with both the environment and regions.

## Methods

**Data on *S* at the plot grain.** We compiled a global database of tree species richness from 1,933 forest plots; these were taken from published database compilations<sup>7,12,13,83–86</sup>, from national forest inventory surveys<sup>87–89</sup>, others were extracted manually from primary sources<sup>90–132</sup>. From this set of plots we then selected only those with unique geographic coordinates, and with data on the number of individual trees, minimum diameter at breast height (DBH), and area of the plot. We made the effort to include only plots that spanned a contiguous area and in which all trees within the plot above the minimum DBH were determined. In cases where there were several plots with the exact same geographic coordinates, we chose the plot with the largest area. If areas were the same, we chose one plot randomly. This left us with 1,336 forest plots for our main analyses. Although all of these plots are in forests, the authors of the primary studies still differ in which individuals are actually determined. For instance, authors may include or exclude lianas. Thus, in the main analyses we included all plots that have the following morphological scope: ‘trees’, ‘woody species’, ‘trees and palms’, ‘trees and shrubs’, ‘trees and lianas’, ‘all living stems’. In a parallel sensitivity analysis we used a more stringent selection criteria to create a subset of the data (see below).

**Data on *S* at the country grain.** We compiled data on tree species richness of 282 countries and other administrative units (US and Brazilian states, Chinese provinces). We downloaded the data from BONAP taxonomic data center at <http://bonap.net/tdc> for the United States<sup>133</sup>, from ref.<sup>134</sup> for the provinces of China, from Flora do Brasil 2020 at <http://floradobrasil.jbrj.gov.br><sup>135</sup>, and from Botanic Gardens Conservation International database GlobalTreeSearch<sup>14</sup> (accessed 18 August 2017) for the rest of the world. To download the data from GlobalTreeSearch, we used the Selenium software interfaced through a custom R script. We note that there are more potential data sources that could have been further leveraged to make our dataset even larger, both at the country and the plot grain, and perhaps also at the intermediate grain. However, our priority has been to make the data for this paper open, and thus we use only the easily available open databases and primary published sources.

**Sensitivity to data sources and tree definition.** Data sources vary in their definition of what a tree is, and for the main analyses presented here we used an inclusive and broad definition, which gave us the advantage of larger *N*. To be sure that our results are robust to this definition, and also robust to a potential co-linearity between data sources and biogeographic regions, we performed an analysis on a subset of the data, selected using the following rules: (1) at the plot grain, we used only plots with trees defined as ‘trees’ or with DBH ≥ 1 cm, (2) at the country grain we used USA, Brazil and China as complete spatial units and we did not ‘disaggregate’ them to the smaller administrative regions. This gave us a subset of 1,166 plots and 183 countries. The results obtained from this subset were similar to the results obtained from the full dataset (see figures in [https://github.com/petrkeil/global\\_tree\\_S/tree/master/Figures/Subset\\_data\\_sensitivity\\_analysis](https://github.com/petrkeil/global_tree_S/tree/master/Figures/Subset_data_sensitivity_analysis)), and thus we consider our results to be robust to data source and tree definition.

**Predictors of species richness.** For each plot and each country, we extracted characteristics that are proxies for environmental heterogeneity, energy availability, productivity, climatic limits, climatic stability, insularity, and regional or spatial variables. These are known, or have been hypothesized, to be associated with plant species richness<sup>7,43,45,46,136</sup> (Supplementary Methods). Specifically, we calculated the following predictors of species richness: area, latitude and longitude of its centroid, membership in a discrete biogeographical realm, its location on mainland or island, difference between highest and lowest altitude, mean gross primary productivity, mean annual temperature, mean isothermality, precipitation in the driest quarter of the year, and mean precipitation seasonality. For each plot we also noted minimum DBH that was used as a criterion to include tree individuals in a study. All continuous predictors were standardized to 0 mean and unit variance before statistical modelling. Area and tree density were log transformed. See Supplementary Methods for a detailed description of each predictor, its source reference, hypothesized effect on *S*, and original spatial grain.

**Cross-grain models and grain-dependent effects.** Our core approach is that ‘grain dependence’ of an effect of a predictor can be modelled using a statistical interaction between the predictor and area. Specifically, imagine a log–linear relationship between expected mean species richness  $\hat{S}_i$  and an environmental predictor  $x_i$  at site *i*, defined as  $\log \hat{S}_i = \alpha + \beta x_i$ , where  $\alpha$  is the intercept and  $\beta$  is the slope (effect) that linearly depends on logarithm of area  $A_i$  of site *i* as  $\beta_i = \gamma + \delta \log(A_i)$ . By substitution we get

$$\log \hat{S}_i = \alpha + x_i \gamma + x_i \log(A_i) \delta \quad (1)$$

where  $\gamma$  is grain-independent effect of predictor  $x_i$  and  $\delta$  is the effect of the statistical interaction between  $x_i$  and  $\log(A_i)$ . By estimating the  $\gamma$  and  $\delta$  coefficients we can then plot the overall effect  $\beta_i$  as a function of area (for example, in Fig. 4). Extending this logic, we built statistical models that treat environmental and regional predictors of species richness as grain-dependent. Specifically, we built two models (REALM and SMOOTH) representing the same general idea of grain-

dependency, but each implementing it in a somewhat different way. These models are not mutually exclusive, but are complementary approaches to the same problem.

**Model REALM.** This model follows the traditional approach to assess regional effects on *S*, that is, variation of *S* that is not accounted for by environmental predictors can be accounted for by membership in pre-defined discrete geographic regions (as in ref. 7), also known as realms. We extend this idea by assuming that the effect of biogeographic regions interacts with area (that is, grain). That is, there are different SARs at work in each region. These SARs set the mean richness at a given grain, and the environmental variables then predict variation around that mean. Formally, observed species richness  $S_i$  in *i*th plot or country is a negative binomial random variable  $S_i \sim \text{NegBin}(\hat{S}_i, \theta)$ , where

$$\log \hat{S}_i = \alpha_j + \sum_{k=1}^3 \log(A_i)^k \beta_{j,k} + X_i \gamma + X_i \log(A_i) \delta \quad (2)$$

and where  $\alpha_j$  are the area-independent effects of *j*th region (one of them is the intercept),  $\sum_{k=1}^3 \log(A_i)^k \beta_{j,k}$  is the interaction between a third-order polynomial of area *A* and the *j*th region; we have chosen the third-order polynomial to ensure an ability to produce the well-known tri-phasic effect of area<sup>20</sup>.  $X_i \gamma$  is the term for area-independent effects of environmental predictors in a matrix *X*, and  $A_i X_i \delta$  is the interaction term between area *A* and *X*. Parameters to be estimated are the vectors  $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta$ , and the dispersion parameter  $\theta$ . If we only had a single predictor *x*, the model would be specified in R package ‘mgcv’<sup>137</sup> as `gam(S ~ REALM + poly(A,3):REALM + x + x:A, family = ‘nb’)`, where REALM is a factor identifying the regions. We use the negative binomial distribution (specifically, its mean and dispersion parametrization) since it can deal with over-dispersion of the response, it was used in a key single-grain study<sup>7</sup> that we wish to contrast with ours, and it allows calculation of Akaike information criterion (AIC) and Bayesian information criterion (BIC). Also note that the interaction terms with  $\log(A_i)$  are linear—this is an intentional simplification to make the idea presented here clearer, but we suggest that future studies may consider non-linear interaction terms.

**Model SMOOTH.** In this model we avoid using discrete biogeographic regions; instead, we use thin-plate spline functions (hereafter, splines)<sup>137</sup> of geographic coordinates. This allows us (1) to identify the areas of historically accumulated *S* directly from the data, freeing us from the need to use pre-defined geographic realms, and (2) it accounts for spatial autocorrelation in model residuals at the same time<sup>64</sup>. As above,  $S_i \sim \text{NegBin}(\hat{S}_i, \theta)$ , but now

$$\log S_i = \alpha + \sum_{k=1}^3 \log(A_i)^k \beta_k + X_i \gamma + X_i \log(A_i) \delta + s_1(\text{Lat}, \text{Lon}) \text{Plt}_i + s_2(\text{Lat}, \text{Lon}) \text{Cntr}_i \quad (3)$$

The first difference from the REALM model is that  $\alpha$  and  $\beta$  do not vary geographically, but there is a single global species–area relationship (see also Supplementary Fig. 7). The second difference is the spline functions  $s_1$  and  $s_2$  (each with 14 spline bases), and with Plt<sub>*i*</sub> and Cntr<sub>*i*</sub> as binary (0 or 1) variables specifying if an observation *i* is a plot or a country. If we only had a single predictor *x*, the model would be specified in R package ‘mgcv’ as `gam(S ~ s(Lat, Lon, by = Plt or Cntr, bs = ‘sos’, k = 14) + poly(A, 3) + x + x:A, family = ‘nb’)`, where Plt or Cntr is a factor identifying if an observation is a plot or a country.

**Null model.** To set a baseline for the performance of models REALM and SMOOTH, we also fitted a ‘null’ model with only the intercept  $\alpha$  and the dispersion parameter  $\theta$ . The model writes as  $S_i \sim \text{NegBin}(\alpha, \theta)$ . The performance ( $R^2$ , AIC, BIC) of models REALM and SMOOTH was then judged relative to this null model.

**Model fit, diagnostics, and inference.** For the initial model assessment, optimizing the number of spline nodes, extraction of the splines, extraction of residual autocorrelation, and for AIC and BIC calculations, we fitted the models using maximum likelihood (gam function in R package ‘mgcv’<sup>137</sup>). For Bayesian inference and for assessment of uncertainty about model parameters and predictions, we fitted the models using Hamiltonian Monte Carlo (HMC) sampler Stan<sup>138</sup>, interfaced through R function ‘brm’ (package ‘brms’<sup>139</sup>) with 3 chains, 3,000 iterations with 1,000 as a warmup, and every 10th iteration kept for inference. For all parameters we used uninformative prior distributions that are the default setting in the ‘brm’ function. Visual check of the HMC chains showed excellent convergence. To measure model fit, we used plots of observed versus predicted values of *S*, and we also calculated AIC and BIC, which we additionally compared with AIC and BIC of the ‘null’ model with only the constant intercept  $\alpha$  (Supplementary Table 1). To assess spatial autocorrelation in species richness and in the residuals of both models, we used spatial correlograms with Moran’s *I* as a function of geographic distance (Supplementary Fig. 8), with distance bins of 200 km, using corlog function in R package ‘ncf’.

**Global predictions.** To demonstrate the ability of our statistical approach to predict patterns of  $S$  at any arbitrarily chosen grain, we used model SMOOTH to make predictions in a set of artificially generated plots (each with an area of 1 ha) and hexagons (each with an area of 209,903 km<sup>2</sup>) distributed at regular distances across the global mainland. We used R package 'dggidR' to generate both. We used hexagons since they suffer almost no geometrical distortion of their shape due to the geographic projection of Earth. We further eliminated all plots for which at least one environmental variable was unavailable, and hexagons with less than 50% of mainland area, which left us with 9,761 local plots and 620 hexagons (Fig. 2). For each plot and hexagon we extracted the same predictors as for the empirical data, using exactly the same procedures. We then plugged these predictors in to the SMOOTH model, generated the expected  $\hat{S}$  (see equation (3)), and mapped it across the 1 ha plots (hereafter  $\hat{S}_{\text{plot}}$  or alpha-diversity) and hexagons ( $\hat{S}_{\text{hex}}$  or gamma-diversity); we also mapped the ratio gamma/alpha, which is beta-diversity. Finally, we extracted the smooth region effects  $s_2(\text{Lat}, \text{Lon})$  and  $s_1(\text{Lat}, \text{Lon})$  in the hexagons and 1 ha plots respectively, these are the spline functions from equation (3) using the geographic coordinates of the centroids of the hexagons or the 1 ha plots.

**Cross-validation and external validation of the predictions.** To assess predictive performance of the models, we employed two approaches: first, we used fourfold cross-validation in which the original dataset was split to 4 folds (fractions) with approximately equal  $N$ ; each of these folds then served as a test dataset which was compared with predictions of a model fitted using the other 3/4 of the data (the training dataset). Instead of doing a computationally intensive Bayesian cross-validation, we performed the cross-validation using the maximum likelihood model fitting, and thus we report no prediction intervals, and we report results of this exercise as plots of observed versus mean predicted richness (Supplementary Fig. 2).

Second, we performed an external validation of the coarse-grained predictions against an independently assembled dataset that was not used in model training, and comes as a fundamentally different data type: point observations. Specifically, we amassed data on point observations from three databases: (1) The RAINBIO database (<http://rainbio.cesab.org/>) of African vascular plants distributions<sup>140</sup>, (2) the BIEN 3+ database<sup>87,140–149</sup> (<http://bien.nceas.ucsb.edu/bien/>) for the New World plant observations accessed through 'BIEN' R package<sup>150</sup>, and (iii) the high-resolution EU-Forest database of tree occurrences in Europe<sup>151</sup>, although records in the latter come from standardized surveys, rather than haphazard observations. We restricted the RAINBIO records to only those with habit = 'tree', and BIEN records to those with whole plant woodiness = 'woody'. Based on these records, we then calculated the number of observations (records) and species richness in the 209,903 km<sup>2</sup> hexagons. We then excluded all under-sampled grid cells, as those with at least 4,000, 10,000, and 1,000 records per hexagon in RAINBIO, BIEN, and EU-Forest respectively. The observed richness in these hexagons was then plotted against predictions (and their full Bayesian prediction intervals) of model SMOOTH.

As requested by the BIEN data use policy, we also acknowledge the herbaria that contributed data to this work: FCO, UNEX, LPB, AD, CVRD, FURB, IAC, IB, INPA, IPA, MBML, UBC, UESC, UFMA, UFRJ, UFRN, UFS, ULS, US, USP, RB, TRH, ZMT, BRIT, MO, NCU, NY, TEX, U, UNCC, A, AAU, GH, AS, ASU, BAI, B, BA, BAA, BAB, BACP, BAF, BC, BCRU, BG, BH, SEV, BM, MJG, BOUM, BR, C, CANB, CAS, CAY, CEN, CHR, CICY, CIMI, COA, COAH, CP, COL, CONC, CORD, CRAI, CU, CS, CTES, CTESN, DAO, DAV, DS, E, ENCB, ESA, F, UVIC, FLAS, FR, FTG, FUEL, G, GB, GLM, K, GZU, HAL, HAMAB, HAST, HBG, HBR, HO, HRP, HSS, HU, HUSA, IBUG, ICN, IEB, ILL, FCQ, ABH, INEGI, UCSB, ISU, SD, JUA, ECON, USE, TALL, CATA, KSTC, LAGU, KU, LA, GMDRC, LD, LEB, LI, LIL, CNH, MACF, LL, LOJA, LP, LPAG, MGC, LPS, IRVC, JOTR, LSU, DBG, HSC, MELU, NZFRI, M, MA, CSUB, MB, MBM, UCSC, UCS, JBG, OBI, MCNS, ICESI, MEL, MEN, TUB, MERL, MEXU, FSU, MG, MICH, BABY, SCFS, SACT, JROH, SBBG, SJSU, MNHM, MNHN, SDSU, MOR, MSC, SFV, CNS, JEPS, CIB, VIT, MU, PGM, MVM, PASA, BOON, ND, NE, NHM, NMB, NMSU, NSW, O, CHSC, CHAS, CDA, OSC, P, UPS, SGO, PH, SI, POM, PY, QMEX, TROM, RM, RSA, S, SALA, SANT, SNM, SP, SRA, TAIE, TU, UADY, UAM, UAS, UB, UC, UCR, UEC, UFG, UFMT, UJAT, ULM, UNM, UNR, UT, UTEP, VAL, VEN, W, WAG, WELT, WIS, WTU, WU, ZT, CUV, AAS, BHCB, PERTH.

**Partitioning of deviance.** To estimate the relative effects of contemporary environment versus biogeographic regions, we used partitioning of deviance<sup>65,69</sup>, an approach related to variance partitioning<sup>152</sup>. Specifically, the deviance from the null model with no predictors is partitioned to (1) a fraction explained by environmental variables and their interaction with area, (2) to region effects represented by biogeographic realms and their interaction with area, (3) to their overlap (caused by co-linearity between environment and realms), and (4) to their independent effects. We used only model REALM to do the partitioning, since it does not contain area (grain) as a standalone term, which makes the partitioning easier to interpret in terms of the purely environmental versus regional fraction. We did the partitioning at the global extent (using data from all biogeographic realms), but also for two hemispheric subsets in an attempt to reduce the co-linearity between realms and environment: (1) the Nearctic and Palaearctic realms, which represent the boreal, temperate and sub-tropical realms of the northern hemisphere, (2) the Neotropic, Afrotropic, Indo-Malay and Australasian realms that represent the sub-tropics and tropics around the equator.

**Reporting Summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

All data and R codes used for the analyses are available under CC-BY 4.0 license in a GitHub repository at [https://github.com/petrkeil/global\\_tree\\_S](https://github.com/petrkeil/global_tree_S), which is also mirrored at figshare at [https://figshare.com/articles/global\\_tree\\_S/7461509](https://figshare.com/articles/global_tree_S/7461509). Please note that if the data on species richness are reused, the original data sources should be credited.

Received: 28 July 2018; Accepted: 7 January 2019;

Published online: 18 February 2019

## References

1. Fine, P. V. A. & Ree, R. H. Evidence for a time-integrated species-area effect on the latitudinal gradient in tree diversity. *Am. Nat.* **168**, 796–804 (2006).
2. Frodin, D. G. *Guide to Standard Floras of the World* (Cambridge Univ. Press, Cambridge, 2001).
3. Losos, E. & Leigh, E. G. *Tropical Forest Diversity and Dynamism* (Univ. of Chicago Press, Chicago, 2004).
4. Hawkins, B. A. Ecology's oldest pattern? *Trends Ecol. Evol.* **16**, 470 (2001).
5. Storch, D., Bohdalková, E. & Okie, J. The more-individuals hypothesis revisited: the role of community abundance in species richness regulation and the productivity-diversity relationship. *Ecol. Lett.* **21**, 920–937 (2018).
6. Currie, D. J. et al. Predictions and tests of climate-based hypotheses of broad-scale variation in taxonomic richness. *Ecol. Lett.* **7**, 1121–1134 (2004).
7. Ricklefs, R. E. & He, F. Region effects influence local tree species diversity. *Proc. Natl Acad. Sci. USA* **113**, 674–679 (2016).
8. Wiens, J. J. et al. Niche conservatism as an emerging principle in ecology and conservation biology. *Ecol. Lett.* **13**, 1310–1324 (2010).
9. Rabosky, D. L. & Hurlbert, A. H. Species richness at continental scales is dominated by ecological limits. *Am. Nat.* **185**, 572–583 (2015).
10. Meyer, C., Kreft, H., Guralnick, R. & Jetz, W. Global priorities for an effective information basis of biodiversity distributions. *Nat. Commun.* **6**, 8221 (2015).
11. Jetz, W., McPherson, J. M. & Guralnick, R. P. Integrating biodiversity distribution knowledge: toward a global map of life. *Trends Ecol. Evol.* **27**, 151–159 (2012).
12. Phillips, O. L. & Miller, J. S. *Global Patterns of Plant Diversity: Alwyn H. Gentry's Forest Transect Data Set* (Missouri Botanical Garden Press, St. Louis, 2002).
13. Sullivan, M. et al. Diversity and carbon storage across the tropical forest biome. *Sci. Rep.* **7**, 39102 (2017).
14. *GlobalTreeSearch Online Database* (BCGI, 2017); [https://www.bcgi.org/global\\_tree\\_search.php](https://www.bcgi.org/global_tree_search.php)
15. Enquist, B. J., Condit, R., Peet, R. K., Schildhauer, M. & Thiers, B. M. Cyberinfrastructure for an integrated botanical information network to investigate the ecological impacts of global climate change on plant biodiversity. *PeerJ Preprints* **4**, e2615v2 (2016).
16. Levin, S. A. Multiple scales and the maintenance of biodiversity. *Ecosystems* **3**, 498–506 (2000).
17. Chave, J. The problem of pattern and scale in ecology: what have we learned in 20 years? *Ecol. Lett.* **16**, 4–16 (2013).
18. Chase, J. M. Spatial scale resolves the niche versus neutral theory debate. *J. Veg. Sci.* **25**, 319–322 (2014).
19. Leibold, M. A. & Chase, J. M. *Metacommunity Ecology* (Princeton Univ. Press, Princeton, 2017).
20. Storch, D. The theory of the nested species–area relationship: geometric foundations of biodiversity scaling. *J. Veg. Sci.* **27**, 880–891 (2016).
21. Rahbek, C. The role of spatial scale and the perception of large-scale species-richness patterns. *Ecol. Lett.* **8**, 224 (2005).
22. Rahbek, C. & Graves, G. R. Detection of macro-ecological patterns in South American hummingbirds is affected by spatial scale. *Proc. R. Soc. B* **267**, 2259–2265 (2000).
23. Chase, J. M. & Knight, T. M. Scale-dependent effect sizes of ecological drivers on biodiversity: why standardised sampling is not enough. *Ecol. Lett.* **16**, 17–26 (2013).
24. Blowes, S. A., Belmaker, J. & Chase, J. M. Global reef fish richness gradients emerge from divergent and scale-dependent component changes. *Proc. R. Soc. B* **284**, 20170947 (2017).
25. Kraft, N. J. B. et al. Disentangling the drivers of  $\beta$  diversity along latitudinal and elevational gradients. *Science* **333**, 1755–1758 (2011).
26. Buckley, L. B. & Jetz, W. Linking global turnover of species and environments. *Proc. Natl Acad. Sci. USA* **105**, 17836–17841 (2008).
27. Shmida, A. & Wilson, M. V. Biological determinants of species diversity. *J. Biogeogr.* **12**, 1–20 (1985).

28. Böhning-Gaese, K. Determinants of avian species richness at different spatial scales. *J. Biogeogr.* **24**, 49–60 (1997).
29. Belmaker, J. & Jetz, W. Cross-scale variation in species richness–environment associations. *Glob. Ecol. Biogeogr.* **20**, 464–474 (2011).
30. Rahbek, C. & Graves, G. R. Multiscale assessment of patterns of avian species richness. *Proc. Natl Acad. Sci. USA* **98**, 4534–4539 (2001).
31. Wang, Z., Brown, J. H., Tang, Z. & Fang, J. Temperature dependence, spatial scale, and tree species diversity in eastern Asia and North America. *Proc. Natl Acad. Sci. USA* **106**, 13388–13392 (2009).
32. Whittaker, R. J., Willis, K. J. & Field, R. Scale and species richness: towards a general, hierarchical theory of species diversity. *J. Biogeogr.* **28**, 453–470 (2001).
33. Ricklefs, R. E. Intrinsic dynamics of the regional community. *Ecol. Lett.* **18**, 497–503 (2015).
34. Vázquez-Rivera, H. & Currie, D. J. Contemporaneous climate directly controls broad-scale patterns of woody plant diversity: a test by a natural experiment over 14,000 years. *Glob. Ecol. Biogeogr.* **24**, 97–106 (2015).
35. Fine, P. V. A. Ecological and evolutionary drivers of geographic variation in species diversity. *Annu. Rev. Ecol. Syst.* **46**, 369–392 (2015).
36. Harmon, L. J. & Harrison, S. Species diversity is dynamic and unbounded at local and continental scales. *Am. Nat.* **185**, 584–593 (2015).
37. Wiens, J. J., Pyron, R. A. & Moen, D. S. Phylogenetic origins of local-scale diversity patterns and the causes of Amazonian megadiversity. *Ecol. Lett.* **14**, 643–652 (2011).
38. Hawkins, B. A., Porter, E. E. & Diniz-Filho, J. A. F. Productivity and history as predictors of the latitudinal diversity gradient of terrestrial birds. *Ecology* **84**, 1608–1623 (2003).
39. Algar, A. C., Kerr, J. T. & Currie, D. J. Evolutionary constraints on regional faunas: whom, but not how many. *Ecol. Lett.* **12**, 57–65 (2009).
40. Dunn, R. R. et al. Climatic drivers of hemispheric asymmetry in global patterns of ant species richness. *Ecol. Lett.* **12**, 324–333 (2009).
41. Araújo, M. B. et al. Quaternary climate changes explain diversity among reptiles and amphibians. *Ecography* **31**, 8–15 (2008).
42. Belmaker, J. & Jetz, W. Relative roles of ecological and energetic constraints, diversification rates and region history on global species richness gradients. *Ecol. Lett.* **18**, 563–571 (2015).
43. Kreft, H. & Jetz, W. Global patterns and determinants of vascular plant diversity. *Proc. Natl Acad. Sci. U.S.A.* **104**, 5925–5930 (2007).
44. Currie, D. J. & Paquin, V. Large-scale biogeographical patterns of species richness of trees. *Nature* **329**, 326 (1987).
45. Francis, A. P. & Currie, D. J. Global patterns of tree species richness in moist forests: another look. *Oikos* **81**, 598–602 (1998).
46. Šimová, I. et al. Global species–energy relationship in forest plots: role of abundance, temperature and species climatic tolerances. *Glob. Ecol. Biogeogr.* **20**, 842–856 (2011).
47. Latham, R. & Ricklefs, R. E. Global patterns of tree species richness in moist forests: energy-diversity theory does not account for variation in species richness. *Oikos* **67**, 325–333 (1993).
48. Ricklefs, R. E., Latham, R. E. & Qian, H. Global patterns of tree species richness in moist forests: distinguishing ecological influences and historical contingency. *Oikos* **86**, 369–373 (1999).
49. Qian, H., Wiens, J. J., Zhang, J. & Zhang, Y. Evolutionary and ecological causes of species richness patterns in North American angiosperm trees. *Ecography* **38**, 241–250 (2015).
50. Qian, H. & Ricklefs, R. E. Large-scale processes and the Asian bias in species diversity of temperate plants. *Nature* **407**, 180–182 (2000).
51. Ricklefs, R. E., Qian, H. & White, P. S. The region effect on mesoscale plant species richness between eastern Asia and eastern North America. *Ecography* **27**, 129–136 (2004).
52. Lyons, S. K. & Willig, M. R. A hemispheric assessment of scale dependence in latitudinal gradients of species richness. *Ecology* **80**, 2483–2491 (1999).
53. O'Brien, E. M., Field, R. & Whittaker, R. J. Climatic gradients in woody plant (tree and shrub) diversity: water-energy dynamics, residual variation, and topography. *Oikos* **89**, 588–600 (2000).
54. Field, R., O'Brien, E. M. & Whittaker, R. J. Global models for predicting woody plant richness from climate: development and evaluation. *Ecology* **86**, 2263–2277 (2005).
55. Brown, J. H. *Macroecology* (Univ. of Chicago Press, Chicago, 1995).
56. Mutke, J. & Barthlott, W. Patterns of vascular plant diversity at continental to global scale. *Biol. Skrift.* **55**, 521–538 (2005).
57. Lomolino, M. V., Riddle, B. R., Whittaker, R. J. & Brown, J. H. *Biogeography* (Sinauer Associates, Sunderland, 2010).
58. Qian, H. A comparison of the taxonomic richness of temperate plants in East Asia and North America. *Am. J. Bot.* **89**, 1818–1825 (2002).
59. Crist, T. O. & Veech, J. A. Additive partitioning of rarefaction curves and species-area relationships: unifying alpha-, beta- and gamma-diversity with sample size and habitat area. *Ecol. Lett.* **9**, 923–932 (2006).
60. Svenning, J.-C. & Skov, F. Limited filling of the potential range in European tree species: limited range filling in European trees. *Ecol. Lett.* **7**, 565–573 (2004).
61. Jansson, R. & Davies, T. J. Global variation in diversification rates of flowering plants: energy vs. climate change. *Ecol. Lett.* **11**, 173–183 (2007).
62. Legendre, P. Spatial autocorrelation: trouble or new paradigm? *Ecology* **74**, 1659–1673 (1993).
63. Dormann, C. F. et al. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography* **30**, 609–628 (2007).
64. Dormann, C. F. Effects of incorporating spatial autocorrelation into the analysis of species distribution data. *Glob. Ecol. Biogeogr.* **16**, 129–138 (2007).
65. Quintero, I. & Jetz, W. Global elevational diversity and diversification of birds. *Nature* **555**, 246–250 (2018).
66. Stein, A., Gerstner, K. & Kreft, H. Environmental heterogeneity as a universal driver of species richness across taxa, biomes and spatial scales. *Ecol. Lett.* **17**, 866–880 (2014).
67. MacArthur, R. H. & Wilson, E. O. *The Theory of Island Biogeography* (Princeton Univ. Press, Princeton, 1967).
68. Carrete, M. et al. Habitat, human pressure, and social behavior: Partiaing out factors affecting large-scale territory extinction in an endangered vulture. *Biol. Conserv.* **136**, 143–154 (2007).
69. Randin, C. F. et al. Climate change and plant distribution: local models predict high-elevation persistence. *Glob. Change Biol.* **15**, 1557–1569 (2009).
70. White, E. P. & Hurlbert, A. H. The combined influence of the local environment and regional enrichment on bird species richness. *Am. Nat.* **175**, E35–E43 (2010).
71. Dormann, C. F. et al. co-linearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography* **36**, 27–46 (2013).
72. Rominger, A. J. et al. Community assembly on isolated islands: macroecology meets evolution. *Glob. Ecol. Biogeogr.* **25**, 769–780 (2016).
73. Swenson, N. G. et al. Constancy in functional space across a species richness anomaly. *Am. Nat.* **187**, E83–E92 (2016).
74. Şizling, A. L. et al. Can people change the ecological rules that appear general across space? *Glob. Ecol. Biogeogr.* **25**, 1072–1084 (2016).
75. Quintero, I., Keil, P., Jetz, W. & Crawford, F. W. Historical biogeography using species geographical ranges. *Syst. Biol.* **64**, 1059–1073 (2015).
76. Arias, J. S. An event model for phylogenetic biogeography using explicitly geographical ranges. *J. Biogeogr.* **44**, 2225–2235 (2017).
77. Hawkins, B. A. & Porter, E. E. Relative influences of current and historical factors on mammal and bird diversity patterns in deglaciated North America: climate, ice and diversity. *Glob. Ecol. Biogeogr.* **12**, 475–481 (2003).
78. Sandel, B. et al. The influence of late quaternary climate-change velocity on species endemism. *Science* **334**, 660–664 (2011).
79. Jetz, W. & Fine, P. V. A. Global gradients in vertebrate diversity predicted by historical area-productivity dynamics and contemporary environment. *PLoS Biol.* **10**, e1001292 (2012).
80. Cabral, J. S., Valente, L. & Hartig, F. Mechanistic simulation models in macroecology and biogeography: state-of-art and prospects. *Ecography* **40**, 267–280 (2017).
81. Connolly, S. R., Keith, S. A., Colwell, R. K. & Rahbek, C. Process, mechanism, and modeling in macroecology. *Trends Ecol. Evol.* **32**, 835–844 (2017).
82. McGill, B. Strong and weak tests of macroecological theory. *Oikos* **102**, 679–685 (2003).
83. Coelho de Souza, F. et al. Evolutionary heritage influences Amazon tree ecology. *Proc. R. Soc. B* **283**, 20161587 (2016).
84. Phillips, O. L. et al. Efficient plot-based floristic assessment of tropical forests. *J. Trop. Ecol.* **19**, 629–645 (2003).
85. Ramesh, B. R. et al. Forest stand structure and composition in 96 sites along environmental gradients in the central Western Ghats of India. *Ecology* **91**, 3118–3118 (2010).
86. Myers, J. A., Chase, J. M., Crandall, R. M. & Jiménez, I. Disturbance alters beta-diversity but not the relative importance of community assembly mechanisms. *J. Ecol.* **103**, 1291–1299 (2015).
87. US Department of Agriculture. *Forest Inventory and Analysis – Fiscal Year 2016 Business Report* (US Department of Agriculture, Washington, D.C., 2016).
88. De Natale, F. et al. *Inventario Nazionale delle Foreste e dei Serbatoi Forestali di Carbonio* (Ispettorato Generale del Corpo Forestale dello Stato, CRA-ISAF, Trento, 2005).
89. Institut national de l'information géographique et forestière. *French National Forest Inventory (FNFI)* (IGN, Saint-Mandé, 2017); <http://inventaire-forestier.ign.fr/>
90. Abbott, I. Comparisons of spatial pattern, structure, and tree composition between virgin and cut-over jarrah forest in Western Australia. *For. Ecol. Manag.* **9**, 101–126 (1984).



91. Adam, J. H. Changes in forest community structures of tropical montane rain forest on the slope of Mt. Trus Madi in Sabah, Malaysia. *J. Trop. For. Sci.* **13**, 76–92 (2001).
92. Addo-Fordjour, P., Obeng, S., Anning, A. & Addo, M. Floristic composition, structure and natural regeneration in a moist semi-deciduous forest following anthropogenic disturbances and plant invasion. *Int. J. Biodiv. Conserv.* **1**, 21–37 (2009).
93. Adekunle, V. A. J. Conservation of tree species diversity in tropical rainforest ecosystem of South-West Nigeria. *J. Trop. For. Sci.* **18**, 91–101 (2006).
94. Ansley, S. J.-A. & Battles, J. J. Forest composition, structure, and change in an old-growth mixed conifer forest in the northern Sierra Nevada. *J. Torrey Bot. Soc.* **125**, 297–308 (1998).
95. Beals, E. W. The remnant cedar forests of Lebanon. *J. Ecol.* **53**, 679–694 (1965).
96. Bonino, E. E. & Araujo, P. Structural differences between a primary and a secondary forest in the Argentine Dry Chaco and management implications. *For. Ecol. Manag.* **206**, 407–412 (2005).
97. Cairns, M. A., Olmsted, I., Granados, J. & Argaez, J. Composition and aboveground tree biomass of a dry semi-evergreen forest on Mexico's Yucatan Peninsula. *For. Ecol. Manag.* **186**, 125–132 (2003).
98. Cao, M. & Zhang, J. Tree species diversity of tropical forest vegetation in Xishuangbanna, SW China. *Biodivers. Conserv.* **6**, 995–1006 (1997).
99. Cheng-Yang, Z., Zeng-Li, L. I. U. & Jing-Yun, F. Tree species diversity along latitudinal gradient on southeastern and northwestern slopes of Mt. Huanggang, Wuyi Mountains, Fujian, China. *Biodivers. Sci.* **12**, 63–74 (2004).
100. Davis, M. A., Curran, C., Tietmeyer, A. & Miller, A. Dynamic tree aggregation patterns in a species-poor temperate woodland disturbed by fire. *J. Veg. Sci.* **16**, 167–174 (2005).
101. Do, T. V. et al. Effects of micro-topographies on stand structure and tree species diversity in an old-growth evergreen broad-leaved forest, southwestern Japan. *Glob. Ecol. Conserv.* **4**, 185–196 (2015).
102. Eichhorn, M. Boreal forests of Kamchatka: structure and composition. *Forests* **1**, 154–176 (2010).
103. Enoki, T. Microtopography and distribution of canopy trees in a subtropical evergreen broad-leaved forest in the northern part of Okinawa Island, Japan. *Ecol. Res.* **18**, 103–113 (2003).
104. Eshete, A., Sterck, F. & Bongers, F. Diversity and production of Ethiopian dry woodlands explained by climate- and soil-stress gradients. *For. Ecol. Manag.* **261**, 1499–1509 (2011).
105. Fashing, P. J., Forrester, A., Scully, C. & Cords, M. Long-term tree population dynamics and their implications for the conservation of the Kakamega Forest, Kenya. *Biodivers. Conserv.* **13**, 753–771 (2004).
106. Graham, A. W. *The CSIRO Rainforest Permanent Plots of North Queensland - Site, Structural, Floristic and Edaphic Descriptions* (CSIRO and the Cooperative Research Centre for Tropical Rainforest Ecology and Management, Rainforest CRC, Cairns, 2006).
107. Hirayama, K. & Sakimoto, M. Spatial distribution of canopy and subcanopy species along a sloping topography in a cool-temperate conifer-hardwood forest in the snowy region of Japan. *Ecol. Res.* **18**, 443–454 (2003).
108. Jing-Yun, F., Yi-De, L. I., Biao, Z. H. U., Guo-Hua, L. I. U. & Guang-Yi, Z. Community structures and species richness in the montane rain forest of Jianfengling, Hainan Island, China. *Biodivers. Sci.* **12**, 29–43 (2004).
109. Kohira, M., Ninomiya, I., Ibrahim, A. Z. & Latiff, A. Diversity, diameter structure and spatial pattern of trees in semi-evergreen rain forest of Langkawi island, Malaysia. *J. Trop. For. Sci.* **13**, 460–476 (2001).
110. Kohyama, T. Tree size structure of stands and each species in primary warm-temperate rain forests of Southern Japan. *Bot. Mag. Tokyo* **99**, 267–279 (1986).
111. Krishnamurthy, Y. L. et al. Vegetation structure and floristic composition of a tropical dry deciduous forest in Bhadra Wildlife Sanctuary, Karnataka, India. *Trop. Ecol.* **51**, 235–246 (2010).
112. Lalfakawma, Sahoo, U., Roy, S., Vanlalhratpuia, K. & Vanlalhluna, P. C. Community composition and tree population structure in undisturbed and disturbed tropical semi-evergreen forest stands of North-East India. *Appl. Ecol. Environ. Res.* **7**, 303–318 (2010).
113. Linder, P., Elfving, B. & Zackrisson, O. Stand structure and successional trends in virgin boreal forest reserves in Sweden. *For. Ecol. Manag.* **98**, 17–33 (1997).
114. Lopes, C. G. R., Ferraz, E. M. N. & Araújo, EdeL. Physiognomic-structural characterization of dry- and humid-forest fragments (Atlantic Coastal Forest) in Pernambuco State, NE Brazil. *Plant Ecol.* **198**, 1–18 (2008).
115. Lü, X.-T., Yin, J. & Tang, J.-W. Structure, tree species diversity and composition of tropical seasonal rainforests in Xishuangbanna, South-West China. *J. Trop. For. Sci.* **22**, 260–270 (2010).
116. Malizia, A. & Grau, R. Liana-host tree associations in a subtropical montane forest of north-western Argentina. *J. Trop. Ecol.* **22**, 331–339 (2006).
117. Maycock, F. P., Guzik, J., Jankovic, J., Shevera, M. & Carleton, J. T. Composition, structure and ecological aspects of mesic old growth Carpathian deciduous forests of Slovakia, southern Poland and the western Ukraine. *Fragm. Flor. Geobot.* **45**, 281–321 (2000).
118. Nagel, A. T., Svoboda, M., Rugani, T. & Diaci, J. Gap regeneration and replacement patterns in an old-growth Fagus–Abies forest of Bosnia–Herzegovina. *Plant. Ecol.* **208**, 307–318 (2010).
119. Namikawa, K., Matsui, T., Kobayashi, M., Goto, R. & Kuramoto, S. Initial establishment and regeneration processes of an outlying isolated Fagus crenata Blume forest stand in the northernmost boundary of its range in Hokkaido, northern Japan. *Plant. Ecol.* **207**, 161–174 (2010).
120. Narayanan, A. & Parthasarathy, N. Biodiversity inventory of trees in a large scale permanent plot of tropical evergreen forest at Varagaliar, Anamalais, Western Ghats, India. *Biodivers. Conserv.* **8**, 1533–1554 (1999).
121. Popradit, A. et al. Anthropogenic effects on a tropical forest according to the distance from human settlements. *Sci. Rep.* **5**, 14689 (2015).
122. Round, P., Pierce, A., Sankamethawee, W. & Gale, G. The Mo Singto forest dynamics plot, Khao Yai National Park, Thailand. *Nat. Hist. Bull. Siam Soc.* **57**, 57–80 (2011).
123. Sanchez, M., Pedroni, F., Eisenlohr, P. V. & Oliveira-Filho, A. T. Changes in tree community composition and structure of Atlantic rain forest on a slope of the Serra do Mar range, southeastern Brazil, from near sea level to 1000m of altitude. *Flora* **208**, 184–196 (2013).
124. Sawada, H., Ohkubo, T., Kaji, M. & Oomura, K. Spatial distribution and topographic dependence of vegetation types and tree populations of natural forests in the Chichibu Mountains, central Japan. *J. Japan. Forest Soc.* **87**, 293–303 (2005).
125. Sheil, D. & Salim, A. Forest tree persistence, elephants, and stem scars. *Biotropica* **36**, 505–521 (2004).
126. Shu-Qing, Z. et al. Structure and species diversity of boreal forests in Mt. baikal, huzhong area, daxing'an mountains, northeast china. *Biodivers. Sci.* **12**, 182–189 (2004).
127. Splachtna, B. E., Gratzler, G. & Black, B. A. Disturbance history of a European old-growth mixed-species forest—A spatial dendro-ecological analysis. *J. Veg. Sci.* **16**, 511–522 (2005).
128. Szwagrzyk, J. & Gazda, A. Above-ground standing biomass and tree species diversity in natural stands of Central Europe. *J. Veg. Sci.* **18**, 555–562 (2007).
129. Wu, X.-P., Zhu, B. & Zhao, S.-Q. Comparison of community structure and species diversity of mixed forests of deciduous broad-leaved tree and Korean pine in Northeast China. *Biodivers. Sci.* **12**, 174–181 (2004).
130. Wusheng, X., Tao, D., Shihong, L. & Li, X. A comparison of tree species diversity in two subtropical forests, Guangxi, Southwest China. *J. Res. Ecol.* **6**, 208–216 (2015).
131. Yamada, I. Forest ecological studies of the montane forest of Mt. Pangrango, West Java: II. Stratification and floristic composition of the forest vegetation of the higher part of Mt. Pangrango. *South East Asian Studies* **13**, 513–534 (1976).
132. Yasuoka, H. The variety of forest vegetations in south-eastern Cameroon, with special reference to the availability of wild yams for the forest hunter-gatherers. *Afr. Study Monogr.* **30**, 89–119 (2008).
133. Kartesz, J. T. *The Biota of North America Program (BONAP)* (North American Plant Atlas, Chapel Hill, 2015).
134. Qian, H. Environmental determinants of woody plant diversity at a regional scale in China. *PLoS ONE* **8**, e75832 (2013).
135. Forzza, R. C. et al. *Flora do Brasil 2020* (Jardim Botânico do Rio de Janeiro, Rio de Janeiro, 2017); <http://floradobrasil.jbrj.gov.br/>
136. Liang, J. et al. Positive biodiversity-productivity relationship predominant in global forests. *Science* **354**, aaf8957 (2016).
137. Wood, S. N. *Generalized Additive Models: An Introduction with R* (CRC Press/Taylor & Francis Group, 2017).
138. Carpenter, B. et al. Stan: a probabilistic programming language. *J. Stat. Softw.* **76**, <https://doi.org/10.18637/jss.v076.i01> (2017).
139. Bürkner, P.-C. brms: an R package for Bayesian multilevel models using Stan. *J. Stat. Softw.* **80**, <https://doi.org/10.18637/jss.v080.i01> (2017).
140. Dauby, G. et al. RAINBIO: a mega-database of tropical African vascular plants distributions. *PhytoKeys* **74**, 1–18 (2016).
141. Anderson-Teixeira, K. J. et al. CTFS-ForestGEO: a worldwide network monitoring forests in an era of global change. *Glob. Change Biol.* **21**, 528–549 (2015).
142. DeWalt, S. J., Bourdy, G., Chavez de Michel, L. R. & Quenevo, C. Ethnobotany of the Tacana: Quantitative inventories of two permanent plots of Northwestern Bolivia. *Econ. Bot.* **53**, 237–260 (1999).
143. Enquist, B. & Boyle, B. SALVIAS—the SALVIAS vegetation inventory database. *Biodivers. Ecol.* **4**, 288–288 (2012).
144. Fegraus, E. Tropical ecology assessment and monitoring network (TEAMNetwork). *Biodivers. Ecol.* **4**, 287–287 (2012).

145. Oliveira-filho, A. T. *NeoTropTree, Flora Arbórea da Região Neotropical: um Banco de Dados Envolvendo Biogeografia, Diversidade e Conservação* (Universidade Federal de Minas Gerais, Belo Horizonte, 2017).
146. Peet, R. K. et al. Vegetation-plot database of the Carolina Vegetation Survey. *Biodivers. Ecol.* **4**, 243–253 (2012).
147. Peet, R. K., Lee, M. T., Jennings, M. D. & Faber-Langendoen, D. VegBank: a permanent, open-access archive for vegetation plot data. *Biodivers. Ecol.* **4**, 233–241 (2012).
148. Sosef, M. S. M. et al. Exploring the floristic diversity of tropical Africa. *BMC Biol.* **15**, 15 (2017).
149. Canhos, V. P. et al. *Rede speciesLink: avaliação 2006* (Fapesp, São Paulo, 2006); <http://splink.cria.org.br>
150. Maitner, B. S. et al. The BIEN R package: A tool to access the Botanical Information and Ecology Network (BIEN) database. *Meth. Eco. Evo.* **9**, 373–379 (2018).
151. Mauri, A., Strona, G. & San-Miguel-Ayanz, J. EU-Forest, a high-resolution tree occurrence dataset for Europe. *Sci. Data* **4**, 160123 (2017).
152. Borcard, D., Legendre, P. & Drapeau, P. Partialling out the spatial component of ecological variation. *Ecology* **73**, 1045–1055 (1992).

### Acknowledgements

We thank D. Craven and I. Šimová for valuable advice, and H. Kreft, J. Coyle, R. Ricklefs, and S. Blowes for critical comments that greatly improved the manuscript.

We acknowledge the support of the German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig funded by the German Research Foundation (FZT 118).

### Author contributions

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

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41559-019-0799-0>.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Correspondence and requests for materials** should be addressed to P.K.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2019

## Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided  
*Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g.  $F$ ,  $t$ ,  $r$ ) with confidence intervals, effect sizes, degrees of freedom and  $P$  value noted  
*Give  $P$  values as exact values whenever suitable.*
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's  $d$ , Pearson's  $r$ ), indicating how they were calculated

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

#### Data collection

Data on species richness were manually collated from primary sources or extracted from existing databases. Large fraction of the data on species richness in countries were downloaded using free Selenium software, interfaced through free software R, version 3.5.1. Data on environmental conditions in plots and countries were extracted from high-resolution global raster layers using R, mainly using packages 'raster', 'sp', 'rgdal' and 'rgeos'. All data preparation, cleaning, and harmonization were done in R, with heavy reliance on packages from the 'tidyverse' family. The R code is available under CC-BY 4.0 license in a GitHub repository at [https://github.com/petrkeil/global\\_tree\\_S](https://github.com/petrkeil/global_tree_S), which is also mirrored at figshare at [https://figshare.com/articles/global\\_tree\\_S/7461509](https://figshare.com/articles/global_tree_S/7461509).

#### Data analysis

Data processing and analyses for this study were done using free software R, version 3.5.1. The most important R packages that were used to fit the statistical models were 'mgcv', 'rstan', and 'brms'. The main R package used for most visualization purposes was 'ggplot2'. The R code is available under CC-BY 4.0 license in a GitHub repository at [https://github.com/petrkeil/global\\_tree\\_S](https://github.com/petrkeil/global_tree_S), which is also mirrored at figshare at [https://figshare.com/articles/global\\_tree\\_S/7461509](https://figshare.com/articles/global_tree_S/7461509).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

### Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All data used for the analyses are available under CC-BY 4.0 license in a GitHub repository at [https://github.com/petrkeil/global\\_tree\\_S](https://github.com/petrkeil/global_tree_S), which is also mirrored at figshare at [https://figshare.com/articles/global\\_tree\\_S/7461509](https://figshare.com/articles/global_tree_S/7461509).

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences  Behavioural & social sciences  Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://nature.com/documents/nr-reporting-summary-flat.pdf)

## Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description

In the core of our study are statistical models that explain species richness by predictors related to environmental conditions and regional biogeography. The study is purely observational and correlative, there is no direct examination of causality (e.g. using a randomized experiment). The main purpose of the models is estimation of parameters and inductive predictions to new locations, rather than hypothesis testing. All models are variants of generalized linear models or generalized additive models, with only fixed effects and no random effects, and are fitted in a Bayesian setting. All predictors used in the models are assumed to have statistical interaction with area. Performance of the models is evaluated using information criteria and cross-validation. The total number of observations is 1618, and they represent local forest plots (N = 1336) or country-like administrative units (N = 282).

Research sample

Our focal group of organisms is trees, which is a morphological and functional distinction, rather than taxonomic or phylogenetic one. Our sample is a mixture of local forest surveys (plots) and country-like administrative units (countries), all of which vary in their area. For each sampling unit we collected data on species richness of trees, total number of trees, geographic coordinates, and a set of environmental and spatial predictor variables. The data on species richness were extracted from existing and openly accessible databases and primary sources, data on environmental conditions were extracted from openly accessible global high resolution layers.

Sampling strategy

The study uses a compilation of data on tree species richness extracted from published sources. These were identified by an extensive search on Web of Science, Google Scholar, by examining reference lists of major publications on the issue of global tree biodiversity, and by contacting colleagues that work in the field of tree macroecology.

Data collection

All data collection and extraction was done by the corresponding author.

Timing and spatial scale

All the original forest plot data were collected in the field in the period between 1960 and present. The country-level data have no exact temporal period specified, but can be assumed to be collected over the second half of the 20th century. The spatial extent of our study is global, spatial resolution (grain) varies from local plots (< 1 ha) up to countries the size of Russia, Canada, or Australia

Data exclusions

We excluded all data that were not openly available, which reduced our sample size, but ensured full reproducibility of our analyses.

Reproducibility

The study uses a compilation of existing data. This compilation is openly accessible, together with all the code used for the data extraction and analyses, and this part of the study is thus fully reproducible. We also provide geographic coordinates of all local forest plots used in our analyses, which means that these locations (or locations in a close proximity of the original ones) can be re-visited and the ecological variables of interest (e.g. number of individuals or species richness) can be re-examined.

Randomization

Our study is observational and does not rely on randomized experiments. It uses data that were compiled from the literature.

Blinding

Blinding was not relevant to our observational study.

Did the study involve field work?  Yes  No

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

n/a	Involvement in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data

### Methods

n/a	Involvement in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging