ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/362567425

Co-limitation towards lower latitudes shapes global forest diversity gradients

Article *in* Nature Ecology & Evolution · August 2022

DOI: 10.1038/s41559-022-01831-x

CITATIONS
0

READS
39

249 authors, including:



Sergio de Miguel
Universitat de Lleida
186 PUBLICATIONS **3,097** CITATIONS
SEE PROFILE



Robert Bitariho
Mbarara University of Science & Technology (MUST)
77 PUBLICATIONS **2,134** CITATIONS
SEE PROFILE



Sebastian kepfer rojas
University of Copenhagen
53 PUBLICATIONS **744** CITATIONS
SEE PROFILE



Sharif A. Mukul
University of the Sunshine Coast
123 PUBLICATIONS **2,376** CITATIONS
SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Water Treatment View project

StarTree - Multipurpose trees and non-wood forest products: a challenge and opportunity View project

All content following this page was uploaded by Robert Bitariho on 09 August 2022.

The user has requested enhancement of the downloaded file.

nature ecology & evolution

ARTICLES

https://doi.org/10.1038/s41559-022-01831-x

Check for updates

Co-limitation towards lower latitudes shapes global forest diversity gradients

The latitudinal diversity gradient (LDG) is one of the most recognized global patterns of species richness exhibited across a wide range of taxa. Numerous hypotheses have been proposed in the past two centuries to explain LDG, but rigorous tests of the drivers of LDGs have been limited by a lack of high-quality global species richness data. Here we produce a high-resolution (0.025° × 0.025°) map of local tree species richness using a global forest inventory database with individual tree information and local biophysical characteristics from ~1.3 million sample plots. We then quantify drivers of local tree species richness patterns across latitudes. Generally, annual mean temperature was a dominant predictor of tree species richness, which is most consistent with the metabolic theory of biodiversity (MTB). However, MTB underestimated LDG in the tropics, where high species richness was also moderated by topographic, soil and anthropogenic factors operating at local scales. Given that local landscape variables operate synergistically with bioclimatic factors in shaping the global LDG pattern, we suggest that MTB be extended to account for co-limitation by subordinate drivers.

Identifying which mechanisms moderate global biodiversity patterns[1,2] has perplexed the scientific community for more than two centuries[3,4]. The most noticeable pattern, the latitudinal diversity gradient (LDG), is a trend of declining local species richness (alpha diversity) from low to high latitudes. This trend has been observed for many taxonomic groups and across land, freshwater and marine environments[5,6]. More than 30 hypotheses have been proposed[3,4,7,8] to explain LDG[9], but few can be reconciled with existing observational data for predicting biodiversity decline towards the poles. To test these varied hypotheses, biodiversity data must be assembled that are global in scope with sufficient sample coverage across all ecoregions and biomes.

In addition to biodiversity data, testing these varied hypotheses also requires data on a wide spectrum of potential drivers that may moderate biodiversity at local scales[9,10], such as climate, soil and land features and anthropogenic factors. For instance, environmental temperature (that is, ambient temperature of the air, represented by annual mean temperature) is largely responsible for the generation and maintenance of biodiversity through the effects of solar radiation on demographic rates (for example, growth and mortality), ecological interactions (for example, predation and competition) and evolutionary rates of change (for example, speciation and extinction)[11,12]. Soil and topographic heterogeneity facilitate niche partitioning via inducing microclimatic variation, contributing to compositional variation[13] and biodiversity maintenance[14,15]. Furthermore, humans have a long history of reshaping biodiversity through the selective use of natural resources and the modification of native species composition[16]. In addition, multiple subordinate factors jointly affecting biodiversity could potentially increase the diversity of niche opportunities, thereby resulting in species-rich assemblages.

Here we quantified the relative contribution of a wide range of environmental factors across space on local tree species richness in forested areas around the world. To accomplish this, we standardized a global tree species richness (that is, as alpha diversity) database (Fig. 1) and quantified the relative contribution of 47 explanatory variables including bioclimatic conditions (for example, annual mean temperature), vegetation and survey attributes (for example, sample plot size), topographic covariates (for example, terrain roughness), soil covariates (for example, bulk density) and anthropogenic spatial features (for example, size of roadless areas) in an attempt to test whether local co-limitation exists when multiple subordinate drivers co-dominate (Figs. 2 and 3). We conducted a three-stage analysis (Fig. 1 and Methods in Supplementary Information for details) based on two independent ground-sourced forest inventory datasets (Phase I and Phase II; Extended Data Fig. 1). The main dataset (Phase I) consisted of 1,255,444 sample plots, while the validating dataset (Phase II) consisted of 22,131 sample plots, most of which are located in unsampled and under-sampled regions of the Phase I dataset. Together, our sample data covered 424 of the 435 (97%) forested ecoregions worldwide (Extended Data Fig. 1), with a total of ~55 million sample trees representing more than 32,000 species.

Results and discussion

Global patterns of local tree species richness and LDG. Our analyses confirmed, with a high level of accuracy, one general spatial trend in local tree species richness worldwide that has led us to three conclusions regarding the mechanisms underlying patterns of tree species richness. We found that LDG for tree species richness was consistent with that of most other groups of organisms, with a decline from the tropics to the poles (Figs. 2 and 4). In the Northern Hemisphere, tree species richness dropped sharply from the equator (98 species per ha) to 10° N with an average rate of decline of 6 species per ha per 1° increase in latitude, after which the decline diminished and stabilized at 4 species per ha at 50° N. In the Southern Hemisphere, tree species richness declined from the equator to 25° S on average by 3 species per ha per 1° increase in latitude, after which tree species richness fluctuated before another steep drop from 25 species per ha (43° S) to 4 species per ha (50° S). We were able to detect and map regional patterns and global peaks of tree species diversity, with a high spatial resolution (0.025° × 0.025°). The Amazonian, Southeast Asian and Melanesian rainforests are the regions with the greatest local tree species richness worldwide, containing >200 tree species per ha above the 5 cm diameter-at-breast-height (DBH) threshold, confirming previous findings[17,18]. Tropical African rainforests generally contain 50% fewer tree species per hectare than Amazonian rainforests. In the temperate forests of the Northern Hemisphere, the Changbai Mountains in Northeast Asia (up to ~28 species per ha) and the

A full list of authors and affiliations appears at the end of the paper.

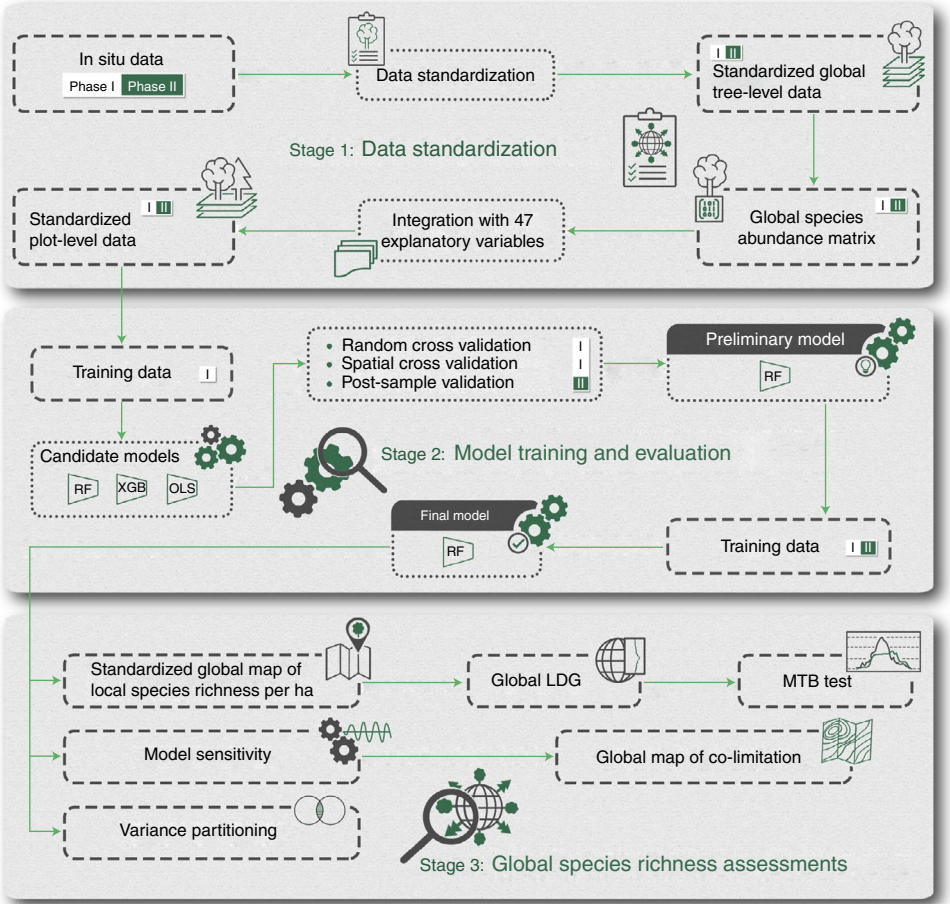


Fig. 1 | A conceptual diagram of the three-stage process employed in the study. Stage 1: Two independent global forest biodiversity individual-based (GFBi) datasets (Phase I and Phase II; Extended Data Fig. 1) were standardized into a global tree-level dataframe and aggregated into a global species abundance matrix. On the basis of plot locations, we merged the abundance matrix with 47 explanatory variables (Fig. 3) into a standardized plot-level dataframe. Stage 2: We compared three candidate models (RF, random forests; XGB, XGBoost; OLS, ordinary least squares) trained from the Phase I plot-level dataframe, using random and spatial cross validation based on Phase I data and post-sample validation based on Phase II data (Extended Data Fig. 2). The final model was then selected and re-calibrated with both Phase I and Phase II data. Stage 3: Using the final model, we standardized and mapped local tree species richness per hectare across the global forest range. On the basis of this globally continuous map, we quantified the associated LDG (Fig. 4a) and tested for the MTB (Fig. 2). We further developed the global map of co-limitation (Fig. 5a) based on model sensitivity analysis and quantified the contribution of key factors to local species richness patterns using variance partitioning (Fig. 6). Dotted boxes represent processes or models and dashed ones represent data or results. Methods provide details.

Central Appalachian forests in the Eastern United States (up to ~20 species per ha) display high local species richness. In the Southern Hemisphere, the sclerophyllous and *Nothofagus*-dominated forests in south-central Chile are among the most species-rich temperate communities (up to 50 species per ha). Boreal forest communities are consistently low in local tree species richness, with typically five or fewer tree species per hectare.

The above LDG pattern of tree species richness was generally consistent with the metabolic theory of biodiversity (MTB)[19,20], except at low latitudes (Fig. 5). According to MTB, environmental temperature is largely responsible for the generation and maintenance of biodiversity[12,21,22], and the natural logarithm of species richness is linearly associated with $1{,}000\,T^{-1}$, where T is the absolute environmental temperature in Kelvin (mean annual temperature +273.15 K), with a slope ranging from −7.5 K to −9.0 K. Our global tree species richness gradient was largely consistent with MTB, with a slope of −8.0 K ($P < 0.001$) and a coefficient of determination of 0.82 (Metabolic Theory of Biodiversity in Methods), indicating that environmental temperature is generally a good predictor of LDG. However, at low latitudes, MTB substantially underestimated LDG. In fact, near the equator where the actual LDG peaked (98 species per ha), observed tree species richness was almost twice as high as predicted by MTB (56 species per ha) (Fig. 4a). Our results suggest that within this low latitudinal range, other factors are also important to the maintenance of biodiversity.

The under-estimation of local tree species richness by MTB at low latitudes is attributable, in part, to the lack of a definite dominant environmental factor, suggesting a co-limitation of multiple subordinate drivers at low latitudes (Fig. 5). In general, bioclimatic factors predominantly determined species richness in 82.6% of the forested areas, while co-limitation (that is, absence of any dominating factor) occurred in 11.7% of forested areas globally. However, in the low-latitude range between 5° N and 15° S, the percentage area of co-limitation increased to 37.1%, more than three times the global average. Furthermore, forested areas under co-limitation contained on average 81.1 ± 0.1 species per ha, much higher than the average local tree species richness of forested areas predominantly determined by topographic (43.9 ± 0.1), anthropogenic (35.6 ± 0.2), soil (33.9 ± 0.2) and bioclimatic (19.4 ± 0.02) factors (Fig. 5b). This suggests that the pattern of co-limitation is pervasive

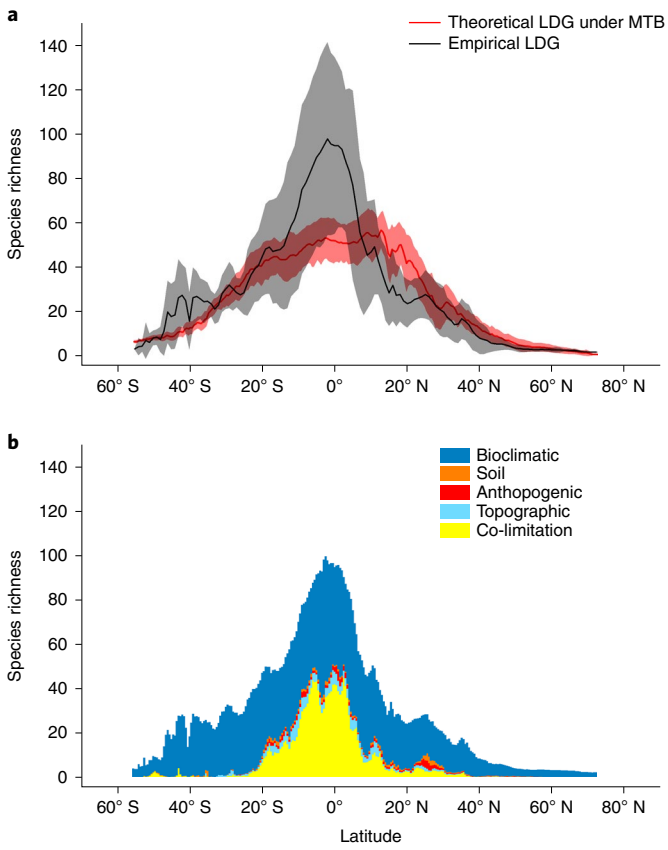


Fig. 2 | Latitudinal gradients of estimated tree species richness and co-limitation of drivers. a, The LDG of tree species richness per hectare was first empirically derived for all 0.025° pixels within the global forest range and aggregated by latitude (Methods). Data are presented as mean values (solid lines) ± standard deviation (shaded areas) and then compared to LDG predicted by the MTB based on local mean annual temperatures. **b**, The co-limitation illustrated here was the product of LDG and the percentage prevalence of dominant drivers by latitude (Fig. 5).

in species-rich tropical forests. In South America, transitional areas between Amazonia and savanna formations nearby are subject to co-limitation that is partly attributable to a dynamic equilibrium between closed forest and savanna[23], edaphic conditions and natural fire regimes[24]. In Africa, anthropogenic influences such as selective timber extraction and fuelwood collection, together with large-scale degradation[25] affect local tree species richness (Fig. 5 and Extended Data Fig. 7). In Central Africa, the evolution of anthropogenic influences from prehistoric to present times has imposed a substantial effect on species diversity[26] and resulted in the development of a complex system of mixes with light-demanding and old-growth tree species.

Bioclimatic dominance and co-limitation. In addition to an overall positive response of local tree species richness to the rise of annual mean temperature (partial dependence plot of C1 in Fig. 3 and Extended Data Fig. 3), the importance of environmental temperature (2.7%) was topped by the total annual precipitation (C12, 7.6%) (Fig. 3). Our findings are consistent with previous discoveries of a joint role of water and temperature/energy—as a proxy for net primary productivity[27]—on plant species richness, with water dominating particularly at warmer, lower latitudes[22,28]. Predicted tree species richness accelerated exponentially with temperature and rainfall, although independently, as shown in the cold–dry quadrant and the convex contours of the 2D partial dependence plot (Extended Data Fig. 3), until each has reached its respective threshold (1,500 mm for total annual precipitation and 10 °C for annual mean temperature). Beyond one of these thresholds, species richness is limited only by the predictor below its threshold (that is, by annual mean temperature in the cold–wet quadrant or by annual precipitation in the hot–dry quadrant). When both predictors have reached their thresholds, that is, in the hot–wet quadrant, co-limitation predominates in most tropical forests. Net primary productivity in the tropics, thus, requires co-limitation of other factors besides only temperature and rainfall[29]. As the response of carbon flux mirror the low-latitude co-limitation pattern for tree species diversity, the matching determinants for both diversity and productivity may explain the similar latitudinal gradient in productivity and the positive diversity–productivity relationship[30,31]. Our findings also indicate that under climate change, intensified droughts coupled with increased annual mean temperature[32] can potentially trigger declines of tree species richness, although possible increases in water-use efficiency from elevated CO_2 and the dominance of highly contingent co-limiting factors may partially buffer this effect in the tropics[33].

Here we articulate evidence for co-limitation in LDG. Resource co-limitation is a common concept in ecology (for example, refs. [34,35]), often used to describe how the synergistic interactions of two or more factors limit ecological productivity[36]. Our use of the term co-limitation emphasizes the reduced effect of bioclimate on tree species richness at low latitudes, although bioclimate is the globally predominant driver of species richness, recognizing that several local subordinate factors synergistically contribute to increased tree species richness in this latitudinal range. We thus argue that the inclusion of co-limitation could substantially improve the explanatory power of biodiversity models in estimating alpha diversity by considering multiple subordinate factors where single-factor dominance is lacking, especially in the tropics. At high latitudes, bioclimatic conditions, particularly environmental temperature, are the major limiting factors and thus the dominant drivers of tree species diversity. As the latitude declines, the influence of bioclimatic conditions dwindles and the maintenance of tree species richness is moderated by many interacting drivers without a clear dominance, which is especially well expressed between 5° N and 15° S (Fig. 5). This prevalence of co-limiting factors is thus not a mere coincidence as to why the observed LDG at low latitudes is almost double that predicted by MTB (Fig. 2). While each of the existing hypotheses underpinning LDG addresses a certain process[10,12] (for example, selection, drift, dispersal or speciation), the evidence of co-limitation highlights synergistic interactions of local processes across the latitudinal gradient.

Concluding remarks. More research is needed to fully elucidate patterns of LDG driven by climatic and other influences, especially those outlined in competing hypotheses. First, our analyses lack explicit consideration of some evolutionary, ecological and historical factors. These include mid-domain stochastic effects[37], the legacies of the poleward expansion of tree species after the Last Glacial Maximum[38,39] and recent human land use/management. Alternative hypotheses, such as niche conservatism or climatic history, are more difficult to test due to data limitations. In addition, long-term effects at geological and millennial time scales also play a role, but it is difficult to disentangle these effects due to collinearity[40]. A major source of uncertainty in our results (Fig. 4b) came from an uneven sample coverage between developed and developing countries (Extended Data Fig. 1). To address this gap, we argue that there needs to be a shared responsibility among forestry agencies at various levels of government, scientists, indigenous communities and other biodiversity monitoring groups to improve sample coverage of forest inventories in developing countries. Innovative biodiversity funding mechanisms, for example, forest inventories funded by carbon

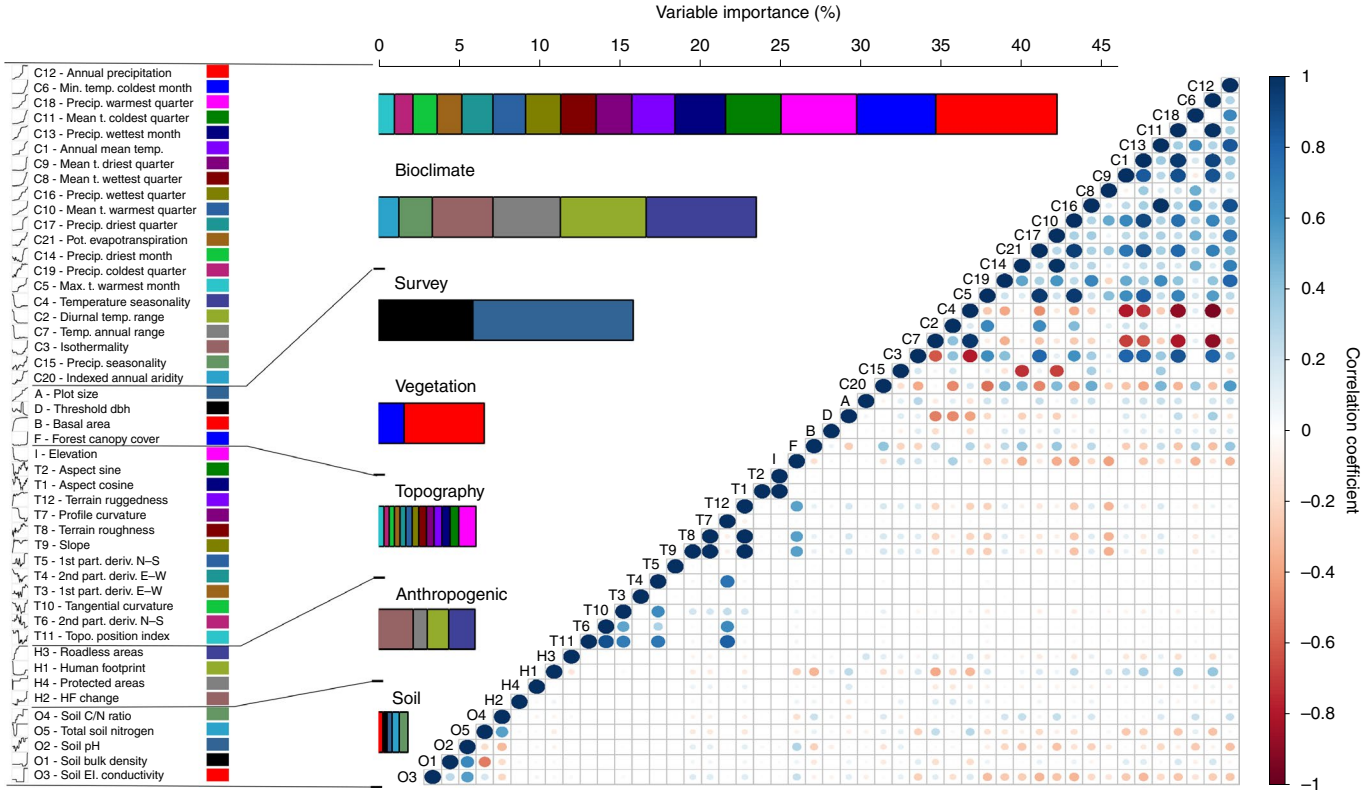


Fig. 3 | A total of 47 explanatory variables in five categories were used in random forests models to predict local tree species richness and quantify LDG. According to standardized variable importance values (horizontal bar plots to the left), bioclimatic variables contributed the most to LDG, followed by vegetation and survey, topographic, anthropogenic and soil variables. The correlogram to the right illustrates correlations between any two variables by the colour and size of a disk. The partial dependence plots to the left (see Extended Data Fig. 3) show the effect of each predictor variable on the species richness, while all the other predictors remained constant at their sample mean. The metadata of the training dataframe provide a detailed description of the explanatory variables. Min., minimum; Temp., temperature; Precip., precipitation; t., temperature; Pot., potential; Max., maximum; dbh, diameter at breast height; part. deriv., partial derivative; Topo., topographic; HF, human footprint; El., electrical.

initiatives such as REDD+, should be incorporated into a comprehensive global forest biodiversity database. Meanwhile, the severe shortage of experts and database management infrastructures, especially in developing countries, poses another major challenge to address this gap[41]. The education and training of new generations of forest scientists, taxonomists and foresters can bring tangible benefits to biodiversity monitoring while improving local economies as well.

Considering co-limitation in addition to MTB enables a refined description of the biogeographic distribution of biodiversity and mechanisms underlying LDG. Our analysis has resulted in the production of a high-resolution map of tree species richness across the global forest range, along with visuals of those factors responsible for the moderation of local tree species richness. Such tools are necessary for conservation management which requires assessments of factors responsible for biodiversity patterns at multiple scales that matter—from local and regional to global scales. Patterns of local tree species richness and associated drivers may provide insights into how and why the diversity of other forest flora, fauna and microbes[42,43] vary across space and time. Furthermore, the high-resolution map of local tree species richness presented here provides a benchmark for evaluating the impact of biodiversity loss on the productivity and functioning of forest ecosystems[31,44]. Finally, aligned with current international calls for spatially explicit monitoring of ecosystem attributes[45], this study delivers detailed biogeographic information to support international endeavours[46] focused on valuing natural capital and advancing global conservation.

Methods

As illustrated in Fig. 1, we conducted data analyses and modelling in three stages.

Stage 1. For this study, we compiled individual in situ tree data from all the regional and national GFBi forest inventory datasets (Supplementary Table 2) into a standardized GFBi dataframe, that is, the GFBi tree list. In this standardized GFBi dataframe, each row represents an individual tree, and columns represent nine key tree- and plot-level attributes. These attributes are tree ID (FID), a unique number assigned to each individual tree; plot ID (PLT), a unique string assigned to each plot; plot coordinates (LAT and LON); tree species name (SPCD); DBH or above buttress; year of measurement; and dataset name (DSN), a unique number assigned to each forest inventory dataset (Supplementary Table 2). With a total of 56 million trees surveyed, GFBi individual-based dataframe represents 1/50,000 of the approximately 2.7 trillion trees[47] worldwide. Because all trees in each sample plot were identified and measured, GFBi data make it possible to quantify forest community structure, composition and species distribution.

To ensure consistency and maximum accuracy in species names, we standardized observations from different forest inventory datasets with the following protocol. First, all multi-stem trees were divided so that each stem represents an individual tree. The scientific names were extracted from original datasets, keeping only the genus and species (authority names were removed). Next, all the species names were compiled into five general species lists, one for each continent. We verified individual species names against 23 online taxonomic databases or web application programming interfaces (API) using the gnr_resolve() function from the 'taxize' package[48] of R[49]. We then manually verified and corrected all the names that did not match with the majority of the online taxonomic databases, that is, the names with a matching score lower than 0.9. For individuals denoted by morphospecies, we assigned each a unique name comprising the genus name and a unique species code. The unique species code consisted of the string 'spp', plus the dataset name followed by a unique number denoting if two individuals belong to the same species. For example, '*Aidia* sppCDi1' and '*Aidia* sppCDi2' represented two different species under the genus '*Aidia*', and both species have been observed in a forest inventory of the Democratic Republic of

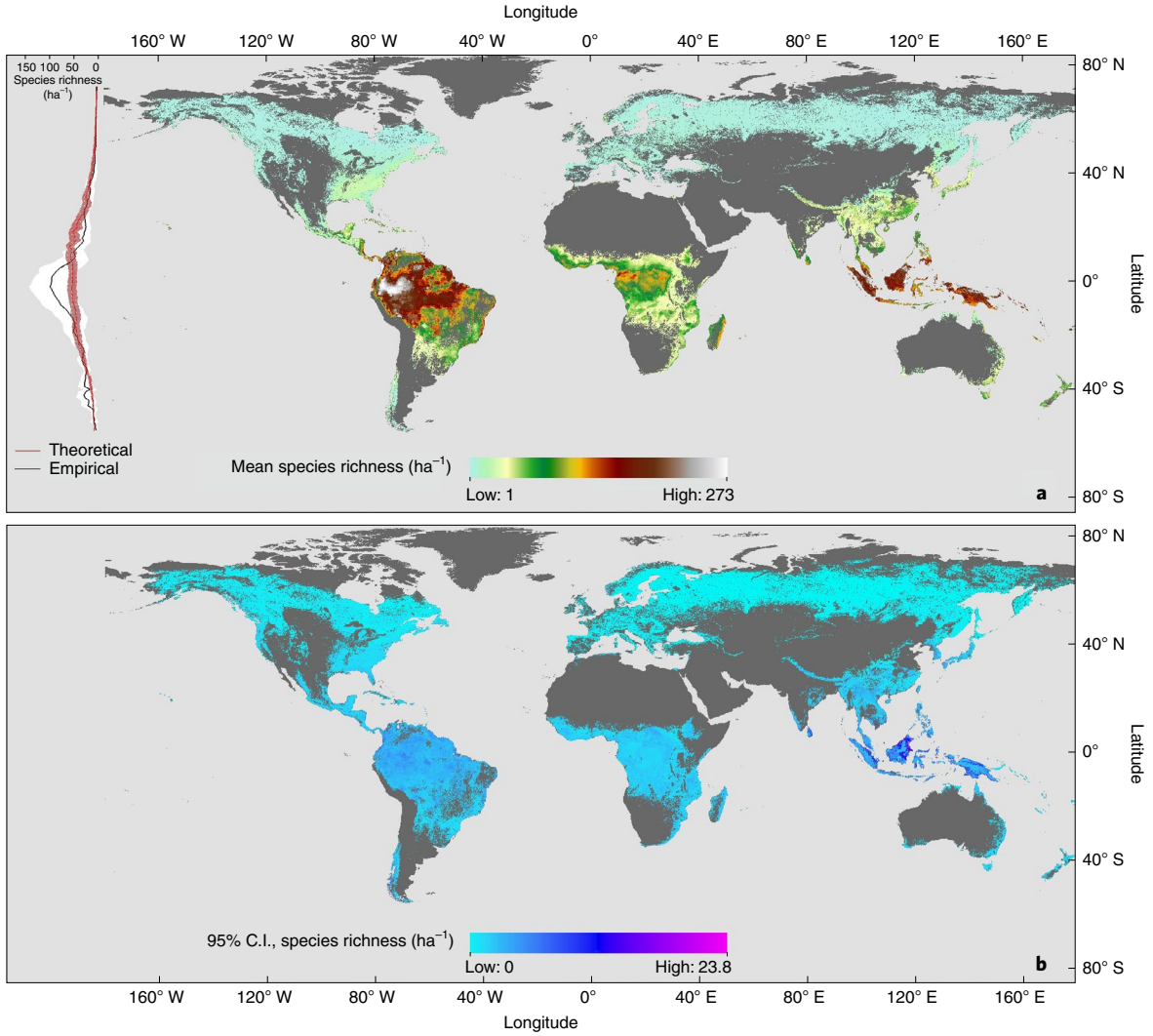


Fig. 4 | Estimated tree species richness per hectare in forested areas worldwide. a, Tree species richness per hectare was first derived for the ~1.3 million GFBi plots across the world and then imputed to the global forest extent. Top left, curves (solid lines representing the mean and shaded areas representing the standard deviation of the mean) represent the observed LDG (black) of tree species diversity in comparison with LDG (red) predicted by the MTB based on local mean annual temperatures (Fig. 2a). **b**, Width of the 95% confidence interval (C.I.) for the estimated tree species richness per hectare. All map layers are displayed at a 0.025° × 0.025° resolution with an equirectangular projection (Plate-Carrée) for better illustration of the latitudinal gradients.

the Congo named 'CDi'. To maximize our species coverage, a tree was defined in this study as a perennial plant with an elongated woody stem that supports branches and leaves, including woody angiosperms, gymnosperms and taller palms (Arecaceae). Tree ferns (Cyatheales) and bamboos (Bambusoideae) were excluded from our analysis.

From the GFBi individual tree-level dataframe, we derived a global species abundance matrix. The global species abundance matrix consisted of the number of individuals by species (column vectors) within individual sample plots (row vectors). The global species abundance matrix consisted of two complementary datasets: Phase I dataset contained 1,255,444 sample plots and Phase II dataset contained 22,131 sample plots, most of which are located in unsampled and under-sampled regions of Phase I dataset. Phase I sample plots cover 394 ecoregions across the world, and Phase II sample plots cover an additional 30 ecoregions in Africa, South America, Southeast Asia, Mexico, India and Japan. Together, our ground-based forest sample plots cover 424 of 435 (97.5%) forested ecoregions across the world. The global species abundance matrix contains ~1.3 million rows (plots) by 32,608 columns (species). Key plot-level information was added to the matrix, including PLT, DSN, plot coordinates, basal area (B), the total cross-sectional areas (m^2) of living trees per ha calculated from DBH and TPH (expansion factor) and the year of measurement. TPH denotes the number of trees per hectare represented by each sampled individual. It ranged from 1 to 5,244 across the GFBi data, with a mean of 48 trees per ha.

We quantified for each sample plot tree species richness (S), which is the total number of tree species in a community. Due to the difference in plot size (standard deviation = 0.09 ha) and threshold DBH values (standard deviation = 2.52 cm) across GFBi sample plots, we developed machine learning models to standardize tree species richness for a common basis of 1 ha in area and 5 cm in threshold DBH. The models incorporated both plot area (A) and threshold DBH (D) as predictors to account for the underlying species-area relationship[50–52] and species-individual size distribution[53] in a rarefaction-based approach[54]. This standardization approach justifies compiling direct tree species diversity estimates from GFBi in situ data of different sources and sampling protocols[55–57], an issue highlighted in earlier large-scale—although less extensive—forest biodiversity studies[57,58]. To evaluate the accuracy of this standardization approach, we tested the machine learning models using cross-sample validation and compared our global maps of estimated tree species diversity against other standardization approaches based on sample completeness (Model Evaluation below).

The machine learning models employed 47 environmental covariates to predict tree species richness. These covariates, derived from satellite-based remote sensing and ground-based survey data, can be summarized into five general categories: *bioclimatic* (for example, annual mean temperature, total annual precipitation, potential evapotranspiration and indexed annual aridity); *soil* (bulk density, pH, electrical conductivity, C/N ratio and total nitrogen); *topographic*, including elevation, slope, aspect and terrain features; *vegetation and survey* attributes

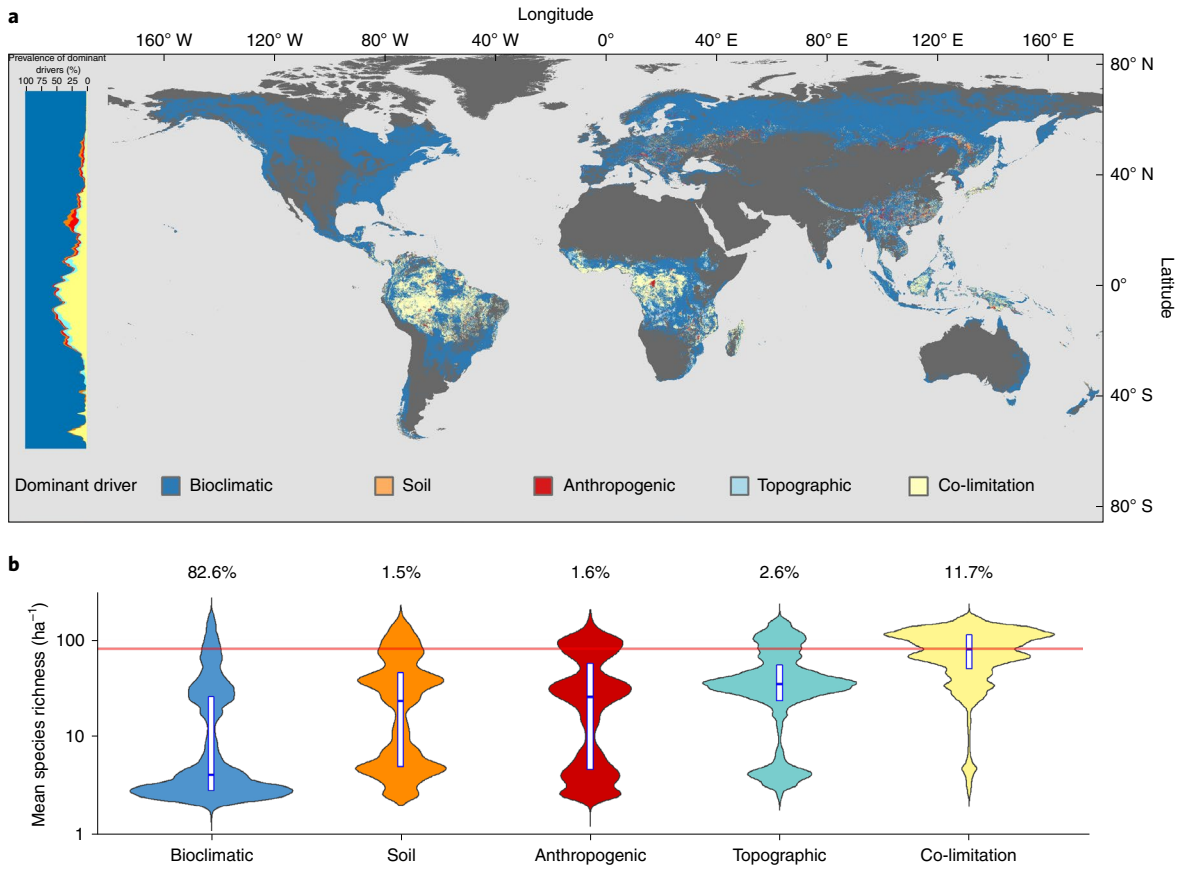


Fig. 5 | Dominant drivers of tree species richness in forested areas worldwide. a, Driver dominance was derived for each pixel from four driver categories (that is, bioclimatic, topographic, anthropogenic and soil) and co-limitation, which represents a lack of clear dominance among the four foregoing categories. The pixel-level drivers were then aggregated by 0.5° latitudinal bins to show the percentage prevalence of dominant drivers by latitude (top left). **b**, The violin charts show the kernel probability density of tree species richness per ha for different drivers. Inside boxes indicate the median (line in the centre) and interquartile range (bounds of boxes). The numbers on top of the violin charts indicate the percentage of forested pixels globally that corresponds to each driver category. The red line represents the mean and 95% confidence interval of tree species richness per ha (81.1 ± 0.1) for all the 0.025° × 0.025° pixels of co-limitation. The vertical axis is on a logarithmic scale for better illustration.

(plot size, basal area, threshold diameter and percent forest canopy cover); and *anthropogenic* variables (human footprint, roadless areas and size of protected areas) (Supplementary Table 1). We extracted all geospatial covariate values from raster datasets to point locations of GFBi plots using ArcMap 10.3 (ref. [59]) and R 3.4.1 (ref. [49]), to build a standardized plot-level dataframe.

Stage 2. We trained random forests (RF)[60], an ensemble learning method that detects general trends present in the data using a multitude of decision trees, to estimate standardized community-level tree species diversity. The RF algorithm applies the general technique of bootstrap aggregating (bagging) with a modified tree learning algorithm that selects at each candidate split in the learning process a random subset of the features (that is, feature bagging). Because a random subset of variables is chosen for each tree, the RF algorithm based on bagged tree ensembles avoids overfitting[60] and mitigates the multicollinearity issue[61] posed by high correlations between some of the predictors variables (Fig. 3). Using subsamples of GFBi data as the training set (that is, training dataframe) with response S, bagging repeatedly for B times selects a random sample with replacement of the training set and trains a regression tree f_b. After training, RF can predict for unseen samples X' with the response variable S being tree species richness per ha:

$$S = \frac{1}{B}\sum_{b=1}^{B} f_b(X'). \quad (1)$$

For rigorous model evaluation, we employed three very different cross validation approaches: randomized cross validation (RCV), spatial cross validation (SCV) and post-sample validation (PSV). In RCV, a model was trained for each continent with a random subsample that accounted for 90% of the training data from that continent, and the remaining 10% of the training data were used as the testing set. This process was repeated 20 times with sample replacement to examine the accuracy of estimated tree species diversity values. In SCV, all sample data from an ecoregion[62] were reserved for testing the model that was trained with the remaining samples from the larger continent within which the ecoregion is situated. We decided to use ecoregions as spatial blocks because (1) unlike political units such as countries and provinces, ecoregions are delineated based on ecological and bioclimatic conditions; and (2) with a total of ~700 terrestrial ecoregions across the world, each ecoregion encompasses 1,800 sample plots on average, which is a large enough sample size for training RF models. This process was repeated until all the forested ecoregions across the world had been tested. SCV was more rigorous than RCV because samples from an entire ecoregion rather than random samples were withheld for validation. PSV was the most rigorous among the three validation processes. For PSV, we have collated an independent sample dataset from 22,131 forest sample plots, which we named Phase II sample plots to highlight their independence from the original GFBi dataset (that is, Phase I sample plots). In PSV, we used Phase II data as the testing set to evaluate the accuracy of the predictive models that were trained for each continent with the Phase I data.

Using these three cross validation processes, we also evaluated the performance of the RF model against two other predictive models, including multiple regression with ordinary least squares (OLS) and Extreme Gradient Boosting (XGBoost). For each model, we derived predicted values of tree species richness of the testing sets and compared these predicted values against observed data using mean absolute error, root-mean-squared error (RMSE), and coefficient of determination (R^2) (ref. [63]). The process was repeated 20 times to select the best model for each continent.

The OLS model estimated values of standardized point diversity for non-sampled point location s, based on spatially explicit values of covariates:

$$Y(\mathbf{s}) = \boldsymbol{\alpha} \cdot X(\mathbf{s}) + \mathbf{e}(\mathbf{s}), \quad (2)$$

where $Y(s)$ is tree species richness at location s; X a design matrix for the predictor variables at location s; $\boldsymbol{\alpha}$ is a vector of coefficients; and $\mathbf{e}$ is a random vector

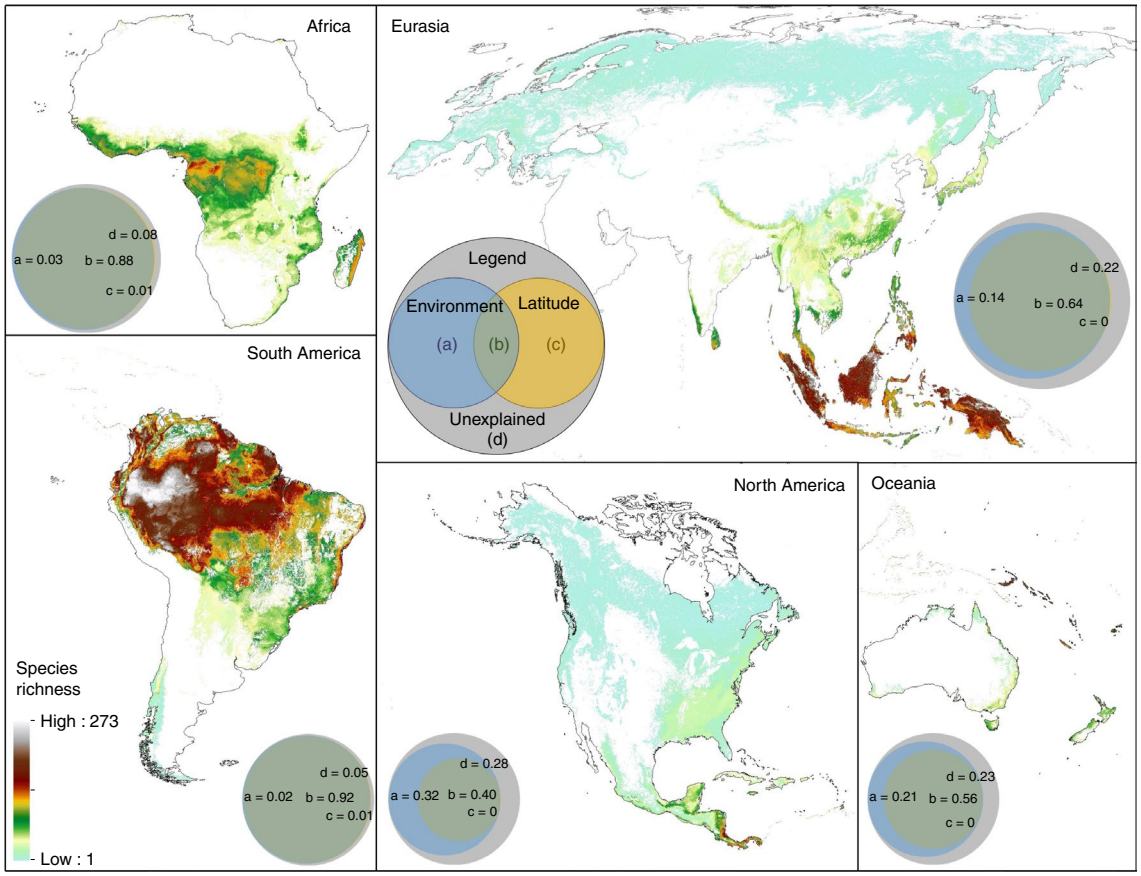


Fig. 6 | Patterns and variance of local tree species richness per ha by continent. The collage of maps shows the zoomed-in view of the distribution of predicted local tree species richness per ha (Fig. 4a) by continent. Circular Venn diagrams (with the legend in the centre) show, for each continent, the spatial variance in observed tree species richness partitioned as follows: **a** (mean = 14.3%) represents the fraction of variance uniquely explained by environmental factors (that is, bioclimatic, topographic, anthropogenic and soil variables) after latitudinal effects had been accounted for. **b** (mean = 68.2%) stands for the fraction of variance jointly explained by environmental factors and latitudinal effects. **c** (mean = 0.3%) represents the fraction of variance explained by latitudinal effects after removing environmental effects. **d** (mean = 17.2%) represents the fraction of unexplained variance in tree species richness. The fractions were based on contrasting the amount of local richness variations in sample data from ~1.3 million plots explained by the R^2 statistics from the continental-scale random forest models with the full set of factors versus those with targeted factors removed.

following a Gaussian probability density function, with an expected value of zero and variance of σ^2. Spatial autocorrelation[64] was not accounted for here due to computational limitations. GFBi data collected from sample plots of various sizes were harmonized to represent local forest community populations per ha using the expansion factor[65], and we used the standardized species richness per ha values for the response variables. We fit a model (2) for each continent. To mitigate the multicollinearity issue[66], we selected for the OLS model the best subset of predictor variables for each continent from the predictor variables used in the RF models, using step-wise regression and Akaike information criterion[67].

XGBoost is a scalable machine learning system[68] that implements the gradient boosting decision tree algorithm[69]. With this ensemble technique, an initial model was trained, with new models added sequentially to correct for errors made by each existing model until no further improvements could be made. Then, new and initial models were merged to make a final prediction that minimized errors. With its algorithm engineered for efficiency in computing time and memory resources, XGBoost is widely used by data scientists to achieve state-of-the-art results on a number of machine learning challenges[68]. In this study, the XGBoost model estimated tree species diversity values in three steps. First, an initial model F_0 was defined to predict the target variable Y. This model was associated with a residual ($Y - F_0$). Second, a new model h_1 was fit to the residuals from the previous step, and F_0 and h_1 were combined to form the boosted model F_1:

$$F_1(x) = F_0(x) + h_1(x), \quad (3)$$

of which the mean squared error was lower than that from F_0. Finally, to improve the performance of F_1, we modelled after the residuals of F_1 to create a new model, F_2, and repeated it for m iterations until the mean squared error converged:

$$F_m(x) = F_{m-1}(x) + h_m(x). \quad (4)$$

Before training RF and XGBoost models, we fine tuned four key hyper-parameters, two for each model. Using 20 bootstrapping iterations on random training sets consisting of 90% of the samples, we first evaluated the sensitivity of RMSE of the testing sets (consisting of the remaining 10% of the samples) to the number of trees to grow and the number of variables randomly sampled as candidates at each split for the RF model and selected the optimal hyper-parameter values (Extended Data Fig. 5). Similarly, we selected the optimal values of the maximum number of boosting iterations (that is, number of rounds) and the maximum depth of a tree for the XGBoost model (Extended Data Fig. 6). As a result, we obtained a preliminary RF model.

Because the RF model emerged as the most accurate model from all three cross validation processes (Extended Data Fig. 2), we selected the RF as the final model, and re-calibrated the final RF model using all the sample data (Phase I and Phase II data).

Stage 3. *Global map of local tree species richness.* To map community-level tree species richness over the global forest range, we first derived the global forest range map from version 1.3 of the Global Forest Change database[70] (years 2000–2015). To ensure consistency with the definition of forest[71] by the Food and Agriculture Organization of the United Nations (FAO), the global forest range in this study was defined as forested areas with ≥10% tree crown coverage per unit area. The tiled 'treecover2000', 'loss' and 'gain' datasets were integrated to obtain current forest cover estimates for the year 2015. To minimize processing artefacts, the ~1 arcsecond spatial resolution tiles were spatially aggregated to an even multiple of their native resolution that approximated the resolution of our covariates. The datasets were then converted to vector point files before being reconverted to raster format with the exact resolution and origin of our covariates. After mosaicking each set of tiles, we computed 'treecover' (scaled) – 'loss' + 'gain' to obtain the 2015 global forest cover, represented as

percent forest cover per ~30 arcsecond pixel. Artefacts in the original data led to 0.08% of all terrestrial pixels having forest cover estimates greater than 100% and 1.9% of terrestrial pixels having estimates less than 0%. These values were truncated to 100% and inflated to 0%, respectively. Finally, the global forest range consisted of those pixels with a percent forest cover ≥10% in 2015. In total, each map consisted of 9,944,908 pixels of 0.025° × 0.025° (hereafter, the pixel) of forested areas. This range is rather conservative and potentially underestimates many remnant forests in drylands and grasslands[72].

We then estimated tree species richness at a 1 ha scale for all pixels within a continent based on the final RF model trained for that continent, using both Phase I and Phase II data. Spatially explicit local environmental covariate data across the global forest range were used for the imputation, except that plot size and threshold DBH were set as 1 ha and 5 cm, respectively. For ecoregions with extremely low sample coverage, we further fine tuned the RF model using samples of similar environment characteristics from other continents. More specifically, we first identified two ecoregions of extremely low sample coverage, that is, the temperate forests in South America and the tropical forests in Oceania, as there were fewer than 1,000 sample plots for the entire biome on those continents. We then trained a new RF model for each ecoregion, using all the sample data from the same biome across the world and fine tuned the mapping data for that ecoregion using the biome-specific RF model.

We computed and mapped the width of the *95%* confidence interval for our local estimates of tree species richness per ha across the global forest range. To this end, we employed a rigorous spatial-block approach, analogous to the spatial cross validation, to derive the 95% confidence interval. More specifically, we computed the width of the 95% confidence interval for each 0.025° × 0.025° mapping pixel by ecoregion. For a pixel p in ecoregion e, we trained 20 RF models using random subsamples that accounted for 90% of the training data from the same continent, which included all samples except those from ecoregion e. We then derived the standard error and the width of the 95% confidence interval for this pixel p in ecoregion e from the predictions of the 20 RF models trained for this ecoregion. This process was repeated until all the forested ecoregions across the world had been assessed and mapped.

Uncertainty in our global diversity estimates was caused by two types of error. The first was measurement error from in situ forest inventories. We mitigated this type of error by implementing stringent species name check and data standardization protocols (Stage 1 Data Standardization). The second arose from the imputation process to map tree species diversity. We minimized this type of error using the three cross validation approaches introduced in Stage 2.

Metabolic theory of biodiversity. Using the global standardized tree species richness values predicted from the final RF models, we quantified the global LDG of tree species richness and tested the effect of environmental temperature based on the MTB[19]:

$$\ln(S) = \alpha \frac{1,000}{T_{\mathrm{env}}} + \beta, \tag{5}$$

where S represents species richness and T_{env} here represents absolute environmental temperature (mean annual temperature +273.15 K); α and β represent coefficients to be estimated by ordinary least squares. According to both the original and extended MTB[19,20], the slope α is expected to range between −7.5 K and −9.0 K, under the assumption that tree community abundance per area does not vary with latitude.

Variance partitioning. We used variance partitioning[73] based on the sample data from ~1.3 million plots to quantify the unique and joint fractions of spatial variance in tree species richness explained by environmental factors and latitude. Due to the correlation between species and environment and between the spatially explicit environmental factors, the variance partitioning approach mitigates type I error inflated by spatial autocorrelation[74]. With variance partitioning, we tested the significance of environmental effects on tree species richness in a series of nested RF models. (A) *The full model* (Extended Data Fig. 4a) consisted of latitude and 47 environmental variables (including 21 bioclimatic ones). (B) *The reduced model I* (Extended Data Fig. 4b) consisted of all but the 21 bioclimatic variables. (C) *The reduced model II* (Extended Data Fig. 4c) consisted of only a zero constant. The overall significance of all environmental factors plus latitude was tested in an one-tailed F-test by comparing the residual sum of squares of error (RSS) of model (A) and model (C):

$$F = \frac{\frac{\mathrm{RSS_C} - \mathrm{RSS_A}}{n_{\mathrm{C}} - n_{\mathrm{A}}}}{\frac{\mathrm{RSS_A}}{n - n_{\mathrm{A}}}}, \tag{6}$$

where $n - n_{\mathrm{A}}$ and $n_{\mathrm{C}} - n_{\mathrm{A}}$ stand for the degree of freedom for the full model and the difference in the degrees of freedom between the full model and the reduced model II, respectively.

The significance of bioclimatic factors, with the effect of latitude being controlled, was tested in an one-tailed F-test by comparing RSS of model (A) and model (B):

$$F = \frac{\frac{\mathrm{RSS_B} - \mathrm{RSS_A}}{n_{\mathrm{B}} - n_{\mathrm{A}}}}{\frac{\mathrm{RSS_A}}{n - n_{\mathrm{A}}}}, \tag{7}$$

where $n_{\mathrm{B}} - n_{\mathrm{A}} = 21$ stands for the difference in the degrees of freedom between the full model and the reduced model I.

We partitioned the spatial variance in observed species richness into four components: a represents the fraction of variance uniquely explained by environmental factors (that is, bioclimatic, topographic, anthropogenic and soil variables), after latitudinal effects have been taken into account; b represents the fraction of variance jointly explained by environmental factors and latitudinal effects; c represents the fraction of variance explained by latitudinal effects after removing environmental effects; and d represents the fraction of variance not explained by the full RF model. Then, the total fraction of variance explained by both environmental factors and latitude was $a + b + c$, the fraction of variance explained by environmental factors was $a + b$, and the fraction of variance explained by latitude was $b + c$. Components $a + b + c$, $a + b$ and $b + c$ were estimated by the R^2 statistics from the RF models trained for each continent using all factors, environmental factors and latitude, respectively (Stage 2 Model Training and Evaluation). Components a, b and c were computed from the previous components using arithmetic relationships that ensure that $a + b + c + d = 100\%$.

Model sensitivity. Based on the final RF models and sample data from ~1.3 million plots, we mapped the dominant drivers of tree species richness with a 0.025° × 0.025° resolution (that is, global map of co-limitation), following a standard procedure for model sensitivity analysis[75]:

Step 1: using the full RF model, and the values of environmental factors $X(s)$ specific to a 0.025°-pixel s, we had already estimated local tree species richness $S_{\mathrm{full}}(s)$:

$$S_{\mathrm{full}}(s) = f(X(s)), \tag{8}$$

where $f()$ represents the RF model, and $X(s)$ environmental factors in four categories, namely E1: bioclimatic, E2: topographic, E3: anthropogenic and E4: soil.

Step 2: for the above-mentioned pixel, we estimated a new local tree species richness value $S_{-\mathrm{E1}}(s)$, using a reduced RF model in which all E1 (bioclimatic) variables were removed:

$$S_{-\mathrm{E1}}(s) = f_{-\mathrm{E1}}((X - \mathrm{E1})(s)), \tag{9}$$

where $f_{-\mathrm{E1}}()$ represents the RF model trained with all but 21 bioclimatic variables, and $(X - \mathrm{E1})(s)$ encompassed environmental factors in three categories, namely E2: topographic, E3: anthropogenic and E4: soil.

Step 3: for a given pixel, we calculated the relative sensitivity of predicted species richness to E1:

$$R(\mathrm{E1}) = |S_{\mathrm{full}}(s) - S_{-\mathrm{E1}}(s)| / S_{\mathrm{full}}(s). \tag{10}$$

Step 4: we repeated steps 2 and 3 to calculate, for a given pixel, the relative sensitivity of each of the following categories (that is, E2: topographic, E3: anthropogenic and E4: soil), respectively. The dominant driver (that is, limiting factor) for this pixel was then the category with the highest relative sensitivity, provided that this relative sensitivity was greater than or equal to 0.2.

Step 5: if the relative sensitivities were less than 0.2 for all categories, we considered that this was a scenario of joint effects of multiple categories of factors (that is, co-limitation), rather than dominance of a single category. Where clear dominance of a single category was lacking, we denoted the dominant driver of this pixel as 'E5: co-limitation.'

Step 6: we repeated the steps above to calculate, for all the remaining pixels of the global grid, the relative sensitivity of each of the five categories of environmental factors, namely E1: bioclimatic, E2: topographic, E3: anthropogenic, E4: soil and E5: co-limitation. On the basis of these values, we created a wall-to-wall map of dominant drivers of tree species richness across the global forest range by labelling the category with the highest relative sensitivity for each pixel (Fig. 5a).

Step 7: on the basis of the relative sensitivity obtained from the steps 1–6, we computed percent prevalence (0–100%) of bioclimatic, topographic, anthropogenic and soil factors and a lack of dominance (co-limitation) in all the forested pixels along each latitudinal band.

Inclusion and ethics statement. The international research collaboration leading to this research paper was conducted via Science-i.org, a transparent and FAIR (Findable, Accessible, Interoperable and Reusable) web platform for international research collaboration. Through this platform and our partner initiatives including the Global Forest Biodiversity Initiative (GFBI), we pursue excellence and high standards of performance, professionalism and ethical conduct. Science-i strictly prohibits any form of discrimination against individuals on the basis of gender, race, age, religion, sexual orientation, veteran status or disability status. Science-i continuously seeks and encourages underrepresented and underprivileged people and groups and the unique voices in global scientific research collaboration.

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

Data availability

The global map of tree species richness is available under license CC BY 4.0, with the identifier: 10.6084/m9.figshare.17232491. This map can be downloaded in two formats. One is a geoTIFF file (S_mean_raster.tif) containing the fully geo-referenced map of tree species richness worldwide at a 0.025°×0.025° resolution. The other is a comma-separated file (S_mean_grid.csv) with the following attributes: *S* is local average tree species richness per ha and *x, y* are centroid coordinates of all 0.025°×0.025° pixels. The global map of co-limitation is available under license CC BY 4.0, with the identifier: 10.6084/m9.figshare.17234339. The metadata of the entire training dataframe—including the characteristics and references of all the in situ Phase I and Phase II datasets and the definitions, units and summary statistics of the environmental covariates—is available under license CC BY 4.0 with the identifier: 10.6084/m9.figshare.19733449.v1. The public version of the training dataframe, including the plot-level species richness and all the covariates, which is needed to reproduce the models and results presented here, is available at: https://doi.org/10.6084/m9.figshare.20055488. The maps and dataframe are also available on the international web research platform: Science-i (https://science-i.org/). Raw forest inventory data are commonly subject to a wide array of confidentiality clauses in regard to open access policies. Despite recent efforts to make some of these data fully open[76,77], some governments and private data owners, especially those from the developing countries generally have decided to keep their data confidential. This decision is based on well-founded arguments to protect certain trees or forests (because of their large size or protected taxonomic status) from illegal logging or trespassing and to protect landowners' privacy against the misuse of plot information such as the geographic coordinates. The sensitive information in the training dataframe, including the plot coordinates and tree-level information, will be available from the corresponding author (albeca.liang@gmail.com) upon a request via Science-i (https://science-i.org/) or GFBI (https://www.gfbinitiative.org/) and an approval from data contributors.

Code availability

All the models in this study were constructed using command line applications written in the R programming language, which processed and restructured the input data, trained the model and performed cross validation. Due to the massive amount of data, we used Purdue University's Brown supercomputing cluster to accelerate the training process. The development of the GFBi database, tabular data cleaning, creation of species abundance matrices, evaluation of diversity determinants and geostatistical imputation were conducted in R[49] (v.3.4.2) through the use of several Linux-based high-performance computing resources at Purdue University and a custom HPC interface developed using Amazon Web Services, each designed for batch processing, scalable resource distribution, embarrassingly parallel computations and/or large RAM jobs. Compute nodes with up to 1 TB of RAM and clusters of up to 64 nodes were employed in this study. Portions of the covariate preparation, mapping and quality control assessment were conducted on Windows-based operating systems with up to 128 GB of RAM. Final continental-level RF models and the R codes we developed to train the models are available under license MIT with the identifier: 10.6084/m9.figshare.17234729.

Received: 29 June 2021; Accepted: 15 June 2022;

References

1. Sutherland, W. J. et al. Identification of 100 fundamental ecological questions. *J. Ecol.* **101**, 58–67 (2013).
2. Gaston, K. J. Global patterns in biodiversity. *Nature* **405**, 220–227 (2000).
3. Pickering, C. *The Geographical Distribution of Animals and Plants* (Little, Brown & Co., 1854).
4. Humboldt, A. V. & Bonpland, A. *Essai sur la Géographie des Plantes Accompagné d'un Tableau Physique des Régions Équinoxiales* (Levrault, Schoell & Co., 1805).
5. Hillebrand, H. On the generality of the latitudinal diversity gradient. *Am. Nat.* **163**, 192–211 (2004).
6. Crame, J. A. Taxonomic diversity gradients through geological time. *Diversity Distrib.* **7**, 175–189 (2001).
7. Gough, L. & Field, R. Latitudinal diversity gradients. *eLS* https://doi.org/10.1002/9780470015902.a0003233.pub2 (2007).
8. Willig, M. R., Kaufman, D. M. & Stevens, R. D. Latitudinal gradients of biodiversity: pattern, process, scale, and synthesis. *Annu. Rev. Ecol., Evolution, Syst.* **34**, 273–309 (2003).
9. Pontarp, M. et al. The latitudinal diversity gradient: novel understanding through mechanistic eco-evolutionary models. *Trends Ecol. Evol.* **34**, 211–223 (2019).
10. Vellend, M. *The Theory of Ecological Communities (MPB-57)* Vol. 75 (Princeton Univ. Press, 2016).
11. Brown, J. H. Why are there so many species in the tropics? *J. Biogeogr.* **41**, 8–22 (2014).
12. 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).
13. Baldeck, C. A. et al. Soil resources and topography shape local tree community structure in tropical forests. *Proc. R. Soc. B* **280**, 20122532 (2013).
14. Qian, H. & Ricklefs, R. E. Large-scale processes and the Asian bias in species diversity of temperate plants. *Nature* **407**, 180–182 (2000).
15. 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).
16. Sponsel, L. E. in *Encyclopedia of Biodiversity* (ed. Scheiner, S.M.) 137–152 (Academic Press, 2013).
17. Sullivan, M. J. et al. Diversity and carbon storage across the tropical forest biome. *Sci. Rep.* **7**, 39102 (2017).
18. Cazzolla Gatti, R. et al. The number of tree species on Earth. *Proc. Natl Acad. Sci. USA* **119**, e2115329119 (2022).
19. Allen, A. P., Brown, J. H. & Gillooly, J. F. Global biodiversity, biochemical kinetics, and the energetic-equivalence rule. *Science* **297**, 1545–1548 (2002).
20. Stegen, J. C., Enquist, B. J. & Ferriere, R. Advancing the metabolic theory of biodiversity. *Ecol. Lett.* **12**, 1001–1015 (2009).
21. Wright, D. H. Species-energy theory: an extension of species-area theory. *Oikos* **41**, 496–506 (1983).
22. Hawkins, B. A. et al. Energy, water, and broad-scale geographic patterns of species richness. *Ecology* **84**, 3105–3117 (2003).
23. Staal, A., Dekker, S. C., Xu, C. & van Nes, E. H. Bistability, spatial interaction, and the distribution of tropical forests and savannas. *Ecosystems* **19**, 1080–1091 (2016).
24. de L. Dantas, V., Batalha, M. A. & Pausas, J. G. Fire drives functional thresholds on the savanna–forest transition. *Ecology* **94**, 2454–2463 (2013).
25. Bodart, C. et al. Continental estimates of forest cover and forest cover changes in the dry ecosystems of Africa between 1990 and 2000. *J. Biogeogr.* **40**, 1036–1047 (2013).
26. Hubau, W. et al. The persistence of carbon in the African forest understory. *Nat. Plants* **5**, 133–140 (2019).
27. Šímová, I. & Storch, D. The enigma of terrestrial primary productivity: measurements, models, scales and the diversity–productivity relationship. *Ecography* **40**, 239–252 (2017).
28. Kreft, H. & Jetz, W. Global patterns and determinants of vascular plant diversity. *Proc. Natl Acad. Sci. USA* **104**, 5925–5930 (2007).
29. Clark, D. A. et al. Net primary production in tropical forests: an evaluation and synthesis of existing field data. *Ecol. Appl.* **11**, 371–384 (2001).
30. Pärtel, M., Laanisto, L. & Zobel, M. Contrasting plant productivity–diversity relationships across latitude: the role of evolutionary history. *Ecology* **88**, 1091–1097 (2007).
31. Liang, J. et al. Positive biodiversity–productivity relationship predominant in global forests. *Science* **354**, aaf8957 (2016).
32. Hammond, W. M. et al. Global field observations of tree die-off reveal hotter-drought fingerprint for Earth's forests. *Nat. Commun.* **13**, 1761 (2022).
33. Martens, C. et al. Large uncertainties in future biome changes in Africa call for flexible climate adaptation strategies. *Glob. Change Biol.* **27**, 340–358 (2021).
34. Chapin, F. S., Bloom, A. J., Field, C. B. & Waring, R. H. Plant responses to multiple environmental factors. *BioScience* **37**, 49–57 (1987).
35. Tilman, D. *Resource Competition and Community Structure (MPB-17)* Vol. 17 (Princeton Univ. Press, 1982).
36. Harpole, W. S. et al. Nutrient co-limitation of primary producer communities. *Ecol. Lett.* **14**, 852–862 (2011).
37. Colwell, R. K., Rahbek, C. & Gotelli, N. J. The mid-domain effect and species richness patterns: what have we learned so far? *Am. Nat.***163**, E1–E23 (2004).
38. Feng, G. et al. Species and phylogenetic endemism in angiosperm trees across the Northern Hemisphere are jointly shaped by modern climate and glacial–interglacial climate change. *Glob. Ecol. Biogeogr.* **28**, 1393–1402 (2019).
39. Fraser, C. I., Nikula, R., Ruzzante, D. E. & Waters, J. M. Poleward bound: biological impacts of Southern Hemisphere glaciation. *Trends Ecol. Evol.* **27**, 462–471 (2012).
40. Algar, A. C., Kerr, J. T. & Currie, D. J. A test of metabolic theory as the mechanism underlying broad-scale species-richness gradients. *Glob. Ecol. Biogeogr.* **16**, 170–178 (2007).
41. Wilson, E. O. A global biodiversity map. *Science* **289**, 2279–2279 (2000).
42. Castagneyrol, B. & Jactel, H. Unraveling plant–animal diversity relationships: a meta-regression analysis. *Ecology* **93**, 2115–2124 (2012).
43. Steidinger, B. S. et al. Climatic controls of decomposition drive the global biogeography of forest–tree symbioses. *Nature* **569**, 404–408 (2019).
44. Bongers, F. J. et al. Functional diversity effects on productivity increase with age in a forest biodiversity experiment. *Nat. Ecol. Evol.* **5**, 1594–1603 (2021).
45. UN Statistical Commission. *System of Environmental–Economic Accounting—Ecosystem Accounting: Final Draft* (United Nations, 2021).
46. *Green Bond Impact Report* (The World Bank Treasury, 2019).
47. Crowther, T. et al. Mapping tree density at a global scale. *Nature* **525**, 201–205 (2015).

48. Chamberlain, J. L., Prisley, S. & McGuffin, M. Understanding the relationships between American ginseng harvest and hardwood forests inventory and timber harvest to improve co-management of the forests of eastern United States. *J. Sustainable For.* **32**, 605–624 (2013).
49. R Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2020).
50. Borda-de-Água, L., Hubbell, S. P. & McAllister, M. Species-area curves, diversity indices, and species abundance distributions: a multifractal analysis. *Am. Nat.* **159**, 138–155 (2002).
51. Connor, E. F. & McCoy, E. D. The statistics and biology of the species-area relationship. *Am. Nat.* **113**, 791–833 (1979).
52. He, F. & Legendre, P. On species-area relations. *Am. Nat.* **148**, 719–737 (1996).
53. White, E. P., Ernest, S. K. M., Kerkhoff, A. J. & Enquist, B. J. Relationships between body size and abundance in ecology. *Trends Ecol. Evol.* **22**, 323–330 (2007).
54. Gotelli, N. J. & Colwell, R. K. Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecol. Lett.* **4**, 379–391 (2001).
55. Chirici, G., Winter, S. & McRoberts, R. E. *National Forest Inventories: Contributions to Forest Biodiversity Assessments* Vol. 20 (Springer Science & Business Media, 2011).
56. Tomppo, E. et al. *National Forest Inventories: Pathways for Common Reporting* (Springer, 2010).
57. Winter, S., Böck, A. & McRoberts, R. E. Estimating tree species diversity across geographic scales. *Eur. J. For. Res.* **131**, 441–451 (2012).
58. McRoberts, R. E., Winter, S., Chirici, G. & LaPoint, E. Assessing forest naturalness. *For. Sci.* **58**, 294–309 (2012).
59. Release 10.3 of Desktop, ESRI ArcGIS (Environmental Systems Research Institute, 2014).
60. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32 (2001).
61. Strobl, C., Boulesteix, A.-L., Kneib, T., Augustin, T. & Zeileis, A. Conditional variable importance for random forests. *BMC Bioinf.* **9**, 307 (2008).
62. Olson, D. M. & Dinerstein, E. The global 200: priority ecoregions for global conservation. *Ann. Mo. Bot. Gard.* **89**, 199–224 (2002).
63. Hyndman, R. J. & Koehler, A. B. Another look at measures of forecast accuracy. *Int. J. Forecast.* **22**, 679–688 (2006).
64. Legendre, P. Spatial autocorrelation: trouble or new paradigm? *Ecology* **74**, 1659–1673 (1993).
65. Husch, B., Beers, T. W. & Kershaw Jr, J. A. *Forest Mensuration* 4th edn (John Wiley & Sons, 2003).
66. Farrar, D. E. & Glauber, R. R. Multicollinearity in regression analysis: the problem revisited. *Rev. Econ. Stat.* **49**, 92–107 (1967).
67. Venables, W. N. & Ripley, B. D. *Modern Applied Statistics with S-PLUS* (Springer Science & Business Media, 2013).
68. Chen, T. & Guestrin, C. XGBoost: a scalable tree boosting system. In *Proc. 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 785–794 (Association for Computing Machinery, 2016).
69. Friedman, J. H. Greedy function approximation: a gradient boosting machine. *Ann. Stat.* **29**, 1189–1232 (2001).
70. Hansen, M. C. et al. High-resolution global maps of 21st-century forest cover change. *Science* **342**, 850–853 (2013).
71. *Global Forest Resources Assessment 2015—How are the World's Forests Changing?* (Food and Agriculture Organization of the United Nations, 2015).
72. Bastin, J.-F. et al. The extent of forest in dryland biomes. *Science* **356**, 635–638 (2017).
73. Peres-Neto, P. R., Legendre, P., Dray, S. & Borcard, D. Variation partitioning of species data matrices: estimation and comparison of fractions. *Ecology* **87**, 2614–2625 (2006).
74. Peres-Neto, P. R. & Legendre, P. Estimating and controlling for spatial structure in the study of ecological communities. *Glob. Ecol. Biogeogr.* **19**, 174–184 (2010).
75. Saltelli, A. et al. *Global Sensitivity Analysis: the Primer* (John Wiley & Sons, 2008).
76. Liang, J. & Gamarra, J. G. P. The importance of sharing global forest data in a world of crises. *Sci. Data* **7**, 424 (2020).
77. *Towards Open and Transparent Forest Data for Climate Action—Experiences and Lessons Learned* (United Nations Food and Agriculture Organization, 2022).

Acknowledgements

The team collaboration and manuscript development are supported by the web-based team science platform: science-i.org, with the project number 202205GFB2. We thank the following initiatives, agencies, teams and individuals for data collection and other technical support: the Global Forest Biodiversity Initiative (GFBI) for establishing the data standards and collaborative framework; United States Department of Agriculture, Forest Service, Forest Inventory and Analysis (FIA) Program; University of Alaska Fairbanks; the SODEFOR, Ivory Coast; University Félix Houphouët-Boigny (UFHB, Ivory Coast); the Queensland Herbarium and past Queensland Government Forestry and Natural Resource Management departments and staff for data collection for over seven decades; and the National Forestry Commission of Mexico (CONAFOR). We thank M. Baker (Carbon Tanzania), together with a team of field assistants (Valentine and Lawrence); all persons who made the Third Spanish Forest Inventory possible, especially the main coordinator, J. A. Villanueva (IFN3); the French National Forest Inventory (NFI campaigns (raw data 2005 and following annual surveys, were downloaded by GFBI at https://inventaire-forestier.ign.fr/spip.php?rubrique159; site accessed on 1 January 2015)); the Italian Forest Inventory (NFI campaigns raw data 2005 and following surveys were downloaded by GFBI at https://inventarioforestale.org/; site accessed on 27 April 2019); Swiss National Forest Inventory, Swiss Federal Institute for Forest, Snow and Landscape Research WSL and Federal Office for the Environment FOEN, Switzerland; the Swedish NFI, Department of Forest Resource Management, Swedish University of Agricultural Sciences SLU; the National Research Foundation (NRF) of South Africa (89967 and 109244) and the South African Research Chair Initiative; the Danish National Forestry, Department of Geosciences and Natural Resource Management, UCPH; Coordination for the Improvement of Higher Education Personnel of Brazil (CAPES, grant number 88881.064976/2014-01); R. Ávila and S. van Tuylen, Instituto Nacional de Bosques (INAB), Guatemala, for facilitating Guatemalan data; the National Focal Center for Forest condition monitoring of Serbia (NFC), Institute of Forestry, Belgrade, Serbia; the Thünen Institute of Forest Ecosystems (Germany) for providing National Forest Inventory data; the FAO and the United Nations High Commissioner for Refugees (UNHCR) for undertaking the SAFE (Safe Access to Fuel and Energy) and CBIT-Forest projects; and the Amazon Forest Inventory Network (RAINFOR), the African Tropical Rainforest Observation Network (AfriTRON) and the ForestPlots.net initiative for their contributions from Amazonian and African forests. The Natural Forest plot data collected between January 2009 and March 2014 by the LUCAS programme for the New Zealand Ministry for the Environment are provided by the New Zealand National Vegetation Survey Databank https://nvs.landcareresearch.co.nz/. We thank the International Boreal Forest Research Association (IBFRA); the Forestry Corporation of New South Wales, Australia; the National Forest Directory of the Ministry of Environment and Sustainable Development of the Argentine Republic (MAyDS) for the plot data of the Second National Forest Inventory (INBN2); the National Forestry Authority and Ministry of Water and Environment of Uganda for their National Biomass Survey (NBS) dataset; and the Sabah Biodiversity Council and the staff from Sabah Forest Research Centre. All TEAM data are provided by the Tropical Ecology Assessment and Monitoring (TEAM) Network, a collaboration between Conservation International, the Missouri Botanical Garden, the Smithsonian Institution and the Wildlife Conservation Society, and partially funded by these institutions, the Gordon and Betty Moore Foundation and other donors, with thanks to all current and previous TEAM site manager and other collaborators that helped collect data. We thank the people of the Redidoti, Pierrekondre and Cassipora village who were instrumental in assisting with the collection of data and sharing local knowledge of their forest and the dedicated members of the field crew of Kabo 2012 census. We are also thankful to FAPESC, SFB, FAO and IMA/SC for supporting the IFFSC. This research was supported in part through computational resources provided by Information Technology at Purdue, West Lafayette, Indiana.This work is supported in part by the NASA grant number 12000401 'Multi-sensor biodiversity framework developed from bioacoustic and space based sensor platforms' (J. Liang, B.P.); the USDA National Institute of Food and Agriculture McIntire Stennis projects 1017711 (J. Liang) and 1016676 (M.Z.); the US National Science Foundation Biological Integration Institutes grant NSF-DBI-2021898 (P.B.R.); the funding by H2020 VERIFY (contract 776810) and H2020 Resonate (contract 101000574) (G.-J.N.); the TEAM project in Uganda supported by the Moore foundation and Buffett Foundation through Conservation International (CI) and Wildlife Conservation Society (WCS); the Danish Council for Independent Research | Natural Sciences (TREECHANGE, grant 6108-00078B) and VILLUM FONDEN grant number 16549 (J.-C.S.); the Natural Environment Research Council of the UK (NERC) project NE/T011084/1 awarded to J.A.-G. and NE/S011811/1; ERC Advanced Grant 291585 ('T-FORCES') and a Royal Society-Wolfson Research Merit Award (O.L.P.); RAINFOR plots supported by the Gordon and Betty Moore Foundation and the UK Natural Environment Research Council, notably NERC Consortium Grants 'AMAZONICA' (NE/F005806/1), 'TROBIT' (NE/D005590/1) and 'BIO-RED' (NE/N012542/1); CIFOR's Global Comparative Study on REDD+ funded by the Norwegian Agency for Development Cooperation, the Australian Department of Foreign Affairs and Trade, the European Union, the International Climate Initiative (IKI) of the German Federal Ministry for the Environment, Nature Conservation, Building and Nuclear Safety and the CGIAR Research Program on Forests, Trees and Agroforestry (CRP-FTA) and donors to the CGIAR Fund; AfriTRON network plots funded by the local communities and NERC, ERC, European Union, Royal Society and Leverhume Trust; a grant from the Royal Society and the Natural Environment Research Council, UK (S.L.L.); National Science Foundation CIF21 DIBBs: EI: number 1724728 (A.C.C.); National Natural Science Foundation of China (31800374) and Shandong Provincial Natural Science Foundation (ZR2019BC083) (H.L.). UK NERC Independent Research Fellowship (grant code: NE/S01537X/1) (T.J.); a Serra-Húnter Fellowship provided by the Government of Catalonia (Spain) (S.d.-M.); the Brazilian National Council for Scientific and Technological Development (CNPq, grant 442640/2018-8, CNPq/Prevfogo-Ibama number 33/2018) (C.A.S.); a grant from the Franklinia Foundation (D.A.C.); Russian Science Foundation project number 19-77-300-12 (R.V.); the Takenaka Scholarship Foundation (A.O.A.); the German Research Foundation (DFG), grant number Am 149/16-4 (C.A.); the Romania National Council for Higher Education Funding, CNFIS, project number CNFIS-FDI-2022-0259 (O.B.); Natural Sciences and Engineering

Research Council of Canada (RGPIN-2019-05109 and STPGP506284) and the Canadian Foundation for Innovation (36014) (H.Y.H.C.); the project SustES—Adaptation strategies for sustainable ecosystem services and food security under adverse environmental conditions (CZ.02.1.01/0.0/0.0/16_019/0000797) (E.C.); Consejo de Ciencia y Tecnología del estado de Durango (2019-01-155) (J.J.C.-R.); Science and Engineering Research Board (SERB), New Delhi, Government of India (file number PDF/2015/000447)—'*Assessing the carbon sequestration potential of different forest types in Central India in response to climate change*' (J.A.D.); Investissement d'avenir grant of the ANR (CEBA: ANR-10-LABEX-0025) (G.D.); National Foundation for Science & Technology Development of Vietnam, 106-NN.06-2013.01 (T.V.D.); Queensland government, Department of Environment and Science (T.J.E.); a Czech Science Foundation Standard grant (19-14620S) (T.M.F.); European Union Seventh Framework Program (FP7/2007–2013) under grant agreement number 265171 (L. Finer, M. Pollastrini, F. Selvi); grants from the Swedish National Forest Inventory, Swedish University of Agricultural Sciences (J.F.); CNPq productivity grant number 311303/2020-0 (A.L.d.G.); DFG grant HE 2719/11-1,2,3; HE 2719/14-1 (A. Hemp); European Union's Horizon Europe research project OpenEarthMonitor grant number 101059548, CGIAR Fund INIT-32-MItigation and Transformation Initiative for GHG reductions of Agrifood systems RelaTed Emissions (MITIGATE+) (M.H.); General Directorate of the State Forests, Poland (1/07; OR-2717/3/11; OR.271.3.3.2017) and the National Centre for Research and Development, Poland (BIOSTRATEG1/267755/4/NCBR/2015) (A.M.J.); Czech Science Foundation 18-10781 S (S.J.); Danish of Ministry of Environment, the Danish Environmental Protection Agency, Integrated Forest Monitoring Program—NFI (V.K.J.); State of São Paulo Research Foundation/FAPESP as part of the BIOTA/FAPESP Program Project Functional Gradient-PELD/BIOTA-ECOFOR 2003/12595-7 & 2012/51872-5 (C.A.J.); Danish Council for Independent Research—social sciences—grant DFF 6109–00296 (G.A.K.); Russian Science Foundation project 21-46-07002 for the plot data collected in the Krasnoyarsk region (V.K.); BOLFOR (D.K.K.); Department of Biotechnology, New Delhi, Government of India (grant number BT/PR7928/NDB/52/9/2006, dated 29 September 2006) (M.L.K.); grant from Kenya Coastal Development Project (KCDP), which was funded by World Bank (J.N.K.); Korea Forest Service (2018113A00-1820-BB01, 2013069A00-1819-AA03, and 2020185D10-2022-AA02) and Seoul National University Big Data Institute through the Data Science Research Project 2016 (H.S.K.); the Brazilian National Council for Scientific and Technological Development (CNPq, grant 442640/2018-8, CNPq/Prevfogo-Ibama number 33/2018) (C.K.); CSIR, New Delhi, government of India (grant number 38(1318)12/EMR-II, dated: 3 April 2012) (S.K.); Department of Biotechnology, New Delhi, government of India (grant number BT/ PR12899/ NDB/39/506/2015 dated 20 June 2017) (A.K.); Coordination for the Improvement of Higher Education Personnel (CAPES) #88887.463733/2019-00 (R.V.L.); National Natural Science Foundation of China (31800374) (H.L.); project of CEPF RAS 'Methodological approaches to assessing the structural organization and functioning of forest ecosystems' (AAAA-A18-118052590019-7) funded by the Ministry of Science and Higher Education of Russia (N.V.L.); Leverhulme Trust grant to Andrew Balmford, Simon Lewis and Jon Lovett (A.R.M.); Russian Science Foundation, project 19-77-30015 for European Russia data processing (O.M.); grant from Kenya Coastal Development Project (KCDP), which was funded by World Bank (M.T.E.M.); the National Centre for Research and Development, Poland (BIOSTRATEG1/267755/4/NCBR/2015) (S.M.); the Secretariat for Universities and of the Ministry of Business and Knowledge of the Government of Catalonia and the European Social Fund (A. Morera); Queensland government, Department of Environment and Science (V.J.N.); Pinnacle Group Cameroon PLC (L.N.N.); Queensland government, Department of Environment and Science (M.R.N.); the Natural Sciences and Engineering Research Council of Canada (RGPIN-2018-05201) (A.P.); the Russian Foundation for Basic Research, project number 20-05-00540 (E.I.P.); European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement number 778322 (H.P.); Science and Engineering Research Board, New Delhi, government of India (grant number YSS/2015/000479, dated 12 January 2016) (P.S.); the Chilean Government research grants Fondecyt number 1191816 and FONDEF number ID19 10421 (C.S.-E.); the Deutsche Forschungsgemeinschaft (DFG) Priority Program 1374 Biodiversity Exploratories (P.S.); European Space Agency projects IFBN (4000114425/15/NL/FF/gp) and CCI Biomass (4000123662/18/I-NB) (D. Schepaschenko); FunDivEUROPE, European Union Seventh Framework Programme (FP7/2007–2013) under grant agreement number 265171 (M.S.-L.); APVV 20-0168 from the Slovak Research and Development Agency (V.S.); Manchester Metropolitan University's Environmental Science Research Centre (G.S.); the project 'LIFE+ ForBioSensing PL Comprehensive monitoring of stand dynamics in Białowieża Forest supported with remote sensing techniques' which is co-funded by the EU Life Plus programme (contract number LIFE13 ENV/PL/000048) and the National Fund for Environmental Protection and Water Management in Poland (contract number 485/2014/WN10/OP-NM-LF/D) (K.J.S.); Global Challenges Research Fund (QR allocation, MMU) (M.J.P.S.); Czech Science Foundation project 21-27454S (M.S.); the Russian Foundation for Basic Research, project number 20-05-00540 (N. Tchebakova); Botanical Research Fund, Coalbourn Trust, Bentham Moxon Trust, Emily Holmes scholarship (L.A.T.); the programmes of the current scientific research of the Botanical Garden of the Ural Branch of Russian Academy of Sciences (V.A.U.); FCT—Portuguese Foundation for Science and Technology—Project UIDB/04033/2020. Inventário Florestal Nacional—ICNF (H. Viana); Grant from Kenya Coastal Development Project (KCDP), which was funded by World Bank (C.W.); grants from the Swedish National Forest Inventory, Swedish University of Agricultural Sciences (B.W.); ATTO project (grant number MCTI-FINEP 1759/10 and BMBF 01LB1001A, 01LK1602F) (F.W.); ReVaTene/PReSeD-CI 2 is funded by the Education and Research Ministry of Côte d'Ivoire, as part of the Debt Reduction-Development Contracts (C2Ds) managed by IRD (I.C.Z.-B.); the National Research Foundation of South Africa (NRF, grant 89967) (C.H.). The Tropical Plant Exploration Group 70 1 ha plots in Continental Cameroon Mountains are supported by Rufford Small Grant Foundation, UK and 4 ha in Sierra Leone are supported by the Global Challenge Research Fund through Manchester Metropolitan University, UK; the National Geographic Explorer Grant, NGS-53344R-18 (A.C.-S.); University of KwaZulu-Natal Research Office grant (M.J.L.); Universidad Nacional Autónoma de México, Dirección General de Asuntos de Personal Académico, Grant PAPIIT IN-217620 (J.A.M.). Czech Science Foundation project 21-24186M (R.T., S. Delabye). Czech Science Foundation project 20-05840Y, the Czech Ministry of Education, Youth and Sports (LTAUSA19137) and the long-term research development project of the Czech Academy of Sciences no. RVO 67985939 (J.A.). The American Society of Primatologists, the Duke University Graduate School, the L.S.B. Leakey Foundation, the National Science Foundation (grant number 0452995) and the Wenner-Gren Foundation for Anthropological Research (grant number 7330) (M.B.). Research grants from Conselho Nacional de Desenvolvimento Científico e Tecnologico (CNPq, Brazil) (309764/2019; 311303/2020) (A.C.V., A.L.G.). The Project of Sanya Yazhou Bay Science and Technology City (grant number CKJ-JYRC-2022-83) (H.-F.W.). The Ugandan NBS was supported with funds from the Forest Carbon Partnership Facility (FCPF), the Austrian Development Agency (ADC) and FAO. FAO's UN-REDD Program, together with the project on 'Native Forests and Community' Loan BIRF number 8493-AR UNDP ARG/15/004 and the National Program for the Protection of Native Forests under UNDP funded Argentina's INBN2.

Author contributions

Conceptualization: J. Liang and C.H. Methodology: J. Liang, C.H., J.G.P.G. and N. Picard. Data coordination: J. Liang, M.Z., S.d.-M., T.W.C., G.-J.N., P.B.R., F. Slik, K.v.G., J.G.P.G. and N. Picard. Writing, revision and editing: all.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41559-022-01831-x.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41559-022-01831-x.

Correspondence and requests for materials should be addressed to Jingjing Liang or Cang Hui.

Peer review information *Nature Ecology & Evolution* thanks Jennifer Baltzer and Chengjin Chu for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permissions information is available at www.nature.com/reprints.

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 2022

Jingjing Liang[1]✉, Javier G. P. Gamarra[2], Nicolas Picard[3], Mo Zhou[1], Bryan Pijanowski[4], Douglass F. Jacobs[4], Peter B. Reich[5,6,7], Thomas W. Crowther[8], Gert-Jan Nabuurs[9,10], Sergio de-Miguel[11,12], Jingyun Fang[13], Christopher W. Woodall[14], Jens-Christian Svenning[15,16], Tommaso Jucker[17], Jean-Francois Bastin[18], Susan K. Wiser[19], Ferry Slik[20], Bruno Hérault[21,22], Giorgio Alberti[23,24,25], Gunnar Keppel[26], Geerten M. Hengeveld[27,28], Pierre L. Ibisch[29], Carlos A. Silva[30], Hans ter Steege[31], Pablo L. Peri[32], David A. Coomes[33], Eric B. Searle[34],

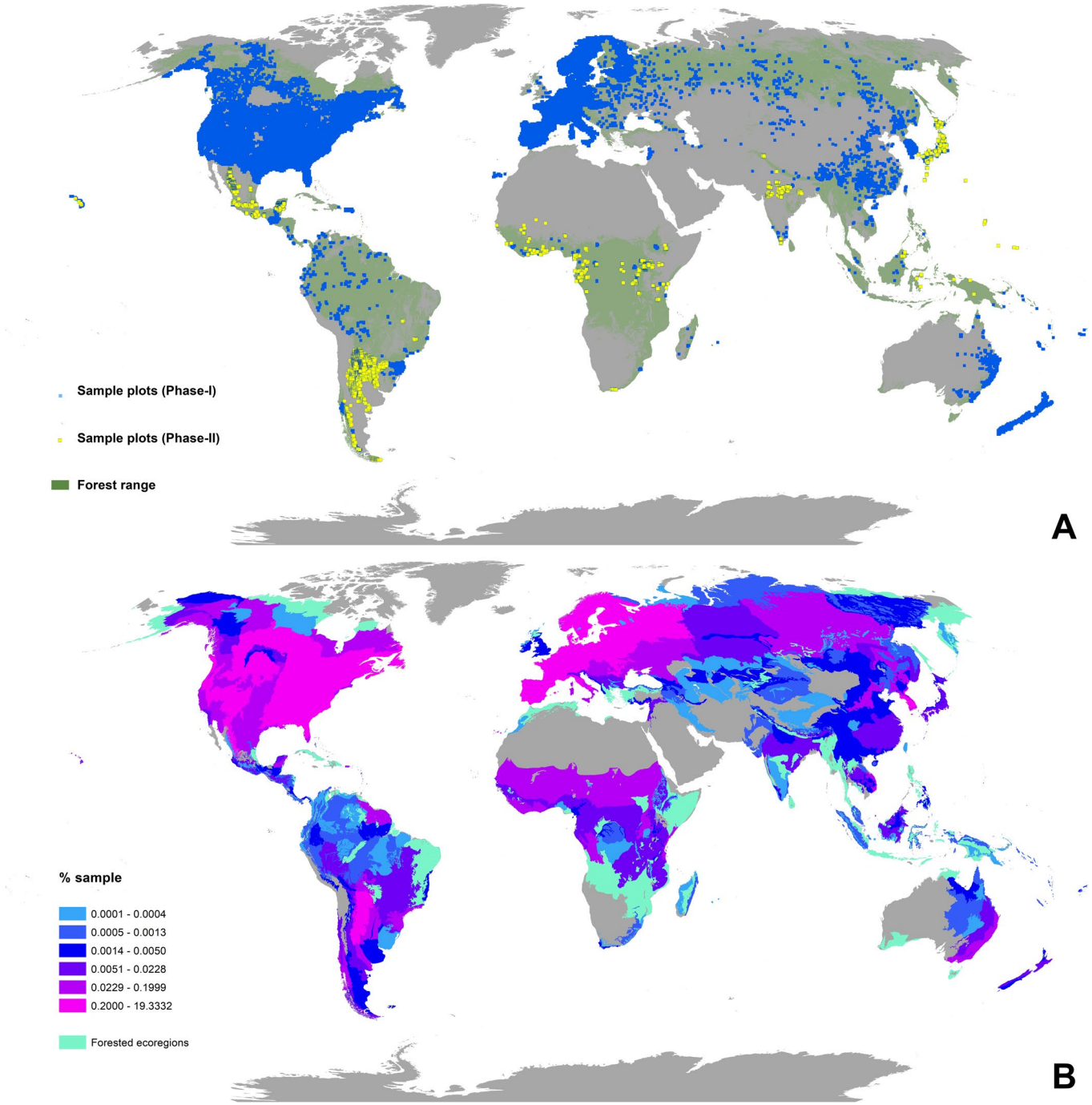
Klaus von Gadow[35,36,37], Bogdan Jaroszewicz[38], Akane O. Abbasi[1], Meinrad Abegg[39], Yves C. Adou Yao[40], Jesús Aguirre-Gutiérrez[41,42], Angelica M. Almeyda Zambrano[43], Jan Altman[44,45], Esteban Alvarez-Dávila[46], Juan Gabriel Álvarez-González[47], Luciana F. Alves[48], Bienvenu H. K. Amani[49], Christian A. Amani[50], Christian Ammer[51], Bhely Angoboy Ilondea[52], Clara Antón-Fernández[53], Valerio Avitabile[54], Gerardo A. Aymard[55], Akomian F. Azihou[56], Johan A. Baard[57], Timothy R. Baker[58], Radomir Balazy[59], Meredith L. Bastian[60,61], Rodrigue Batumike[62], Marijn Bauters[63,64], Hans Beeckman[65], Nithanel Mikael Hendrik Benu[66], Robert Bitariho[67], Pascal Boeckx[64], Jan Bogaert[68], Frans Bongers[10], Olivier Bouriaud[69], Pedro H. S. Brancalion[70], Susanne Brandl[71], Francis Q. Brearley[72], Jaime Briseno-Reyes[73], Eben N. Broadbent[30], Helge Bruelheide[74,75], Erwin Bulte[76], Ann Christine Catlin[77], Roberto Cazzolla Gatti[78], Ricardo G. César[70], Han Y. H. Chen[34], Chelsea Chisholm[79], Emil Cienciala[80,81], Gabriel D. Colletta[82], José Javier Corral-Rivas[73], Anibal Cuchietti[83], Aida Cuni-Sanchez[84,85], Javid A. Dar[86,87,88], Selvadurai Dayanandan[89], Thales de Haulleville[65,68], Mathieu Decuyper[10], Sylvain Delabye[90,91], Géraldine Derroire[92], Ben DeVries[93], John Diisi[94], Tran Van Do[95], Jiri Dolezal[44,96], Aurélie Dourdain[92], Graham P. Durrheim[57], Nestor Laurier Engone Obiang[97], Corneille E. N. Ewango[98], Teresa J. Eyre[99], Tom M. Fayle[91,100], Lethicia Flavine N. Feunang[101], Leena Finér[102], Markus Fischer[103], Jonas Fridman[104], Lorenzo Frizzera[105], André L. de Gasper[106], Damiano Gianelle[105], Henry B. Glick[107], Maria Socorro Gonzalez-Elizondo[108], Lev Gorenstein[77], Richard Habonayo[109], Olivier J. Hardy[110], David J. Harris[111], Andrew Hector[112], Andreas Hemp[113], Martin Herold[114], Annika Hillers[115,116], Wannes Hubau[65,117], Thomas Ibanez[118], Nobuo Imai[119], Gerard Imani[120], Andrzej M. Jagodzinski[121,122], Stepan Janecek[90], Vivian Kvist Johannsen[123], Carlos A. Joly[124], Blaise Jumbam[125,126], Banoho L. P. R. Kabelong[101], Goytom Abraha Kahsay[127], Viktor Karminov[128], Kuswata Kartawinata[129], Justin N. Kassi[130], Elizabeth Kearsley[131], Deborah K. Kennard[132], Sebastian Kepfer-Rojas[123], Mohammed Latif Khan[133], John N. Kigomo[134], Hyun Seok Kim[135,136,137,138], Carine Klauberg[30], Yannick Klomberg[90], Henn Korjus[139], Subashree Kothandaraman[87,88], Florian Kraxner[140], Amit Kumar[141], Relawan Kuswandi[66], Mait Lang[139,142], Michael J. Lawes[143], Rodrigo V. Leite[144], Geoffrey Lentner[77], Simon L. Lewis[58,145], Moses B. Libalah[101,146], Janvier Lisingo[147], Pablito Marcelo López-Serrano[148], Huicui Lu[149], Natalia V. Lukina[150], Anne Mette Lykke[151], Vincent Maicher[90,91,152], Brian S. Maitner[153], Eric Marcon[92,154], Andrew R. Marshall[155,156,157], Emanuel H. Martin[158], Olga Martynenko[128], Faustin M. Mbayu[98], Musingo T. E. Mbuvi[159], Jorge A. Meave[160], Cory Merow[161], Stanislaw Miscicki[162], Vanessa S. Moreno[70], Albert Morera[12], Sharif A. Mukul[163], Jörg C. Müller[164,165], Agustinus Murdjoko[166], Maria Guadalupe Nava-Miranda[148], Litonga Elias Ndive[167], Victor J. Neldner[99], Radovan V. Nevenic[168], Louis N. Nforbelie[101], Michael L. Ngoh[169,170], Anny E. N'Guessan[40], Michael R. Ngugi[99], Alain S. K. Ngute[163,171], Emile Narcisse N. Njila[101], Melanie C. Nyako[101], Thomas O. Ochuodho[172], Jacek Oleksyn[121], Alain Paquette[173], Elena I. Parfenova[174], Minjee Park[4], Marc Parren[10], Narayanaswamy Parthasarathy[88], Sebastian Pfautsch[175], Oliver L. Phillips[58], Maria T. F. Piedade[176], Daniel Piotto[177], Martina Pollastrini[178], Lourens Poorter[10], John R. Poulsen[152], Axel Dalberg Poulsen[111], Hans Pretzsch[179], Mirco Rodeghiero[105,180], Samir G. Rolim[177], Francesco Rovero[181,182], Ervan Rutishauser[183], Khosro Sagheb-Talebi[184], Purabi Saikia[185], Moses Nsanyi Sainge[169,186], Christian Salas-Eljatib[187,188,189], Antonello Salis[2], Peter Schall[51], Dmitry Schepaschenko[140,174,190], Michael Scherer-Lorenzen[191], Bernhard Schmid[192],

Jochen Schöngart[176], Vladimír Šebeň[193], Giacomo Sellan[72,194], Federico Selvi[178], Josep M. Serra-Diaz[195], Douglas Sheil[10,196], Anatoly Z. Shvidenko[140], Plinio Sist[197], Alexandre F. Souza[198], Krzysztof J. Stereńczak[59], Martin J. P. Sullivan[72], Somaiah Sundarapandian[88], Miroslav Svoboda[45], Mike D. Swaine[199], Natalia Targhetta[176], Nadja Tchebakova[174], Liam A. Trethowan[200], Robert Tropek[90,91], John Tshibamba Mukendi[201], Peter Mbanda Umunay[202], Vladimir A. Usoltsev[203], Gaia Vaglio Laurin[204], Riccardo Valentini[204], Fernando Valladares[205], Fons van der Plas[206], Daniel José Vega-Nieva[73], Hans Verbeeck[131], Helder Viana[207,208], Alexander C. Vibrans[209], Simone A. Vieira[210], Jason Vleminckx[211], Catherine E. Waite[212], Hua-Feng Wang[213], Eric Katembo Wasingya[98], Chemuku Wekesa[214], Bertil Westerlund[104], Florian Wittmann[215], Verginia Wortel[216], Tomasz Zawiła-Niedźwiecki[217], Chunyu Zhang[218], Xiuhai Zhao[218], Jun Zhu[219], Xiao Zhu[77], Zhi-Xin Zhu[213], Irie C. Zo-Bi[220] and Cang Hui[221,222] ✉

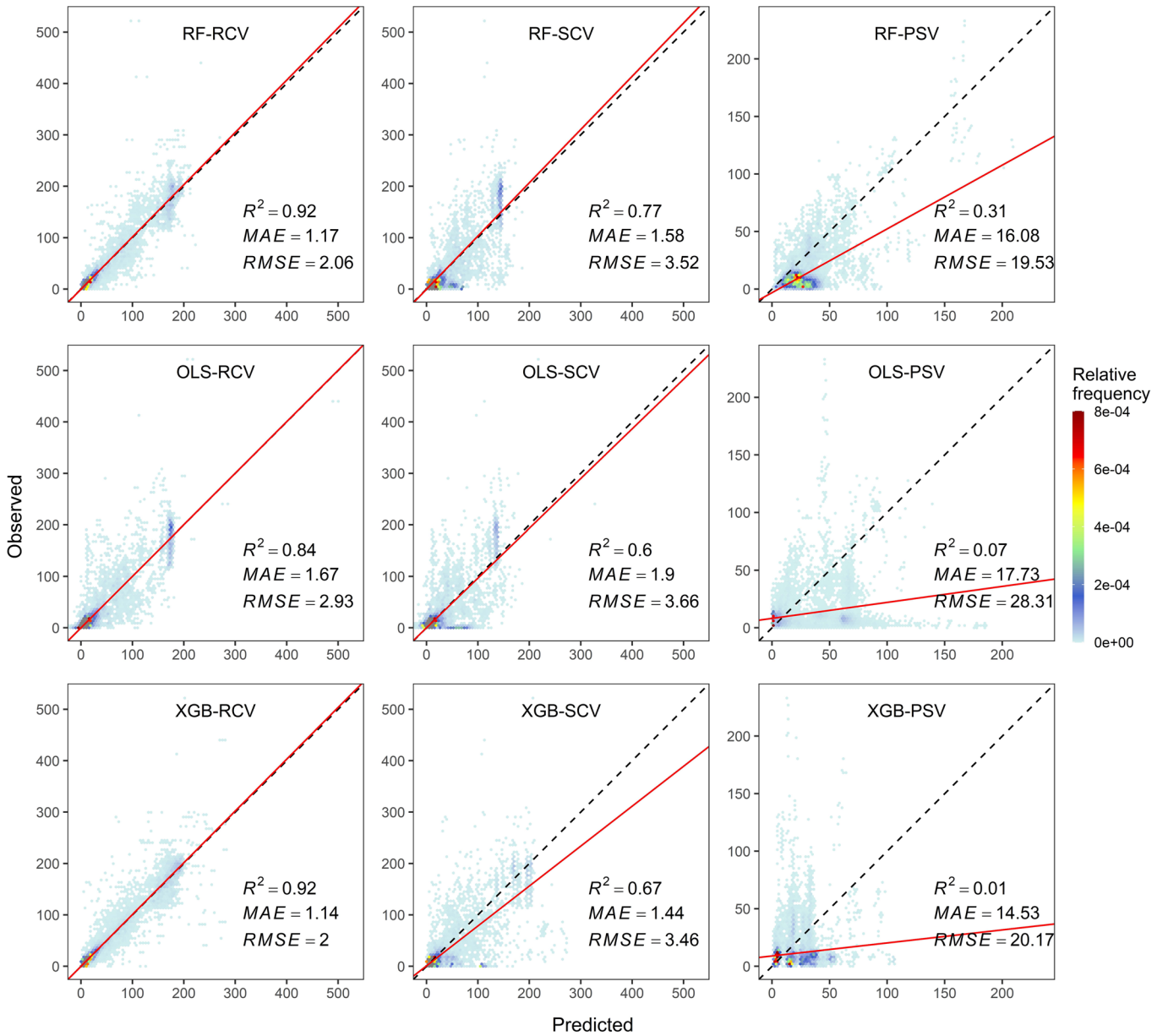
[1]Forest Advanced Computing and Artificial Intelligence Laboratory (FACAI), Department of Forestry and Natural Resources, Purdue University, West Lafayette, IN, USA. [2]Forestry Division, Food and Agriculture Organization of the United Nations, Rome, Italy. [3]GIP ECOFOR, Paris, France. [4]Department of Forestry and Natural Resources, Purdue University, West Lafayette, IN, USA. [5]Institute for Global Change Biology, School for Environment and Sustainability, University of Michigan, Ann Arbor, MI, USA. [6]Department of Forest Resources, University of Minnesota, St. Paul, MN, USA. [7]Hawkesbury Institute for the Environment, Western Sydney University, Penrith, New South Wales, Australia. [8]Crowther Lab, Department of Environmental Systems Science, Institute of Integrative Biology, ETH Zürich, Zürich, Switzerland. [9]Wageningen Environmental Research, Wageningen University and Research, Wageningen, Netherlands. [10]Forest Ecology and Forest Management Group, Wageningen University and Research, Wageningen, Netherlands. [11]Department of Crop and Forest Sciences, University of Lleida, Lleida, Spain. [12]Joint Research Unit CTFC—Agrotecnio—CERCA, Solsona, Spain. [13]Institute of Ecology and Key Laboratory for Earth Surface Processes of the Ministry of Education, College of Urban and Evironmental Sciences, Peking University, Beijing, China. [14]Northern Research Station, USDA Forest Service, Durham, NH, USA. [15]Center for Biodiversity Dynamics in a Changing World (BIOCHANGE), Department of Biology, Aarhus University, Aarhus C, Denmark. [16]Section for Ecoinformatics and Biodiversity, Department of Biology, Aarhus University, Aarhus C, Denmark. [17]School of Biological Sciences, University of Bristol, Bristol, UK. [18]TERRA Teaching and Research Centre, Gembloux Agro Bio-Tech, University of Liege, Gembloux, Belgium. [19]Manaaki Whenua Landcare Research, Lincoln, New Zealand. [20]Environmental and Life Sciences, Faculty of Science, Universiti Brunei Darussalam, Gadong, Brunei Darussalam. [21]Centre de Coopération Internationale en Recherche Agronomique pour le Développement, Montpellier, France. [22]INP-HB (Institut National Polytechnique Félix Houphouet-Boigny), University of Montpellier, Yamoussoukro, Ivory Coast. [23]Department of Agricultural, Food, Environmental and Animal Sciences, University of Udine, Udine, Italy. [24]Faculty of Science and Technology, Free University of Bolzano, Bolzano, Italy. [25]Institute of Bioeconomy, CNR, Sesto, Italy. [26]Natural and Built Environments Research Centre, School of Natural and Built Environments, University of South Australia, Adelaide, South Australia, Australia. [27]Biometris, Wageningen University and Research, Wageningen, Netherlands. [28]Wageningen University & Research, Forest and Nature Conservation Policy Group, Wageningen, Netherlands. [29]Centre for Econics and Ecosystem Management, Eberswalde University for Sustainable Development, Eberswalde, Germany. [30]School of Forest, Fisheries, and Geomatics Sciences, Institute of Food & Agricultural Sciences, University of Florida, Gainesville, FL, USA. [31]Naturalis Biodiversity Center, Leiden, Netherlands. [32]Instituto Nacional de Tecnología Agropecuaria (INTA), Santa Cruz, Argentina. [33]Department of Plant Sciences, University of Cambridge, Cambridge, UK. [34]Faculty of Natural Resources Management, Lakehead University, Thunder Bay, Ontario, Canada. [35]University of Göttingen, Göttingen, Germany. [36]Beijing Forestry University, Beijing, China. [37]University of Stellenbosch, Stellenbosch, South Africa. [38]Białowieża Geobotanical Station, Faculty of Biology, University of Warsaw, Białowieża, Poland. [39]Swiss National Forest Inventory/Swiss Federal Institute for Forest, Snow and Landscape Research WSL, Birmensdorf, Switzerland. [40]UFR Biosciences, University Félix Houphouët-Boigny, Abidjan, Ivory Coast. [41]Environmental Change Institute, School of Geography and the Environment, University of Oxford, Oxford, UK. [42]Biodiversity Dynamics, Naturalis Biodiversity Center, Leiden, Netherlands. [43]Center for Latin American Studies, University of Florida, Gainesville, FL, USA. [44]Institute of Botany, Academy of Sciences of the Czech Republic, Trebon, Czech Republic. [45]Faculty of Forestry and Wood Sciences, Czech University of Life Sciences in Prague, Praha-Suchdol, Czech Republic. [46]Escuela ECAPMA, National Open University and Distance (Colombia) | UNAD, Bogotá, Colombia. [47]Departamento de Ingeniería Agroforestal, Universidad de Santiago de Compostela, Lugo, Spain. [48]Center for Tropical Research, Institute of the Environment and Sustainability, University of California, Los Angeles, CA, USA. [49]Université Jean Lorougnon Guédé, Daloa, Ivory Coast. [50]Université Officielle de Bukavu, Bukavu, Democratic Republic of Congo. [51]Silviculture and Forest Ecology of the Temperate Zones, University of Göttingen, Goettingen, Germany. [52]Institut National pour l'Etude et la Recherche Agronomiques, Kinshasa, Democratic Republic of Congo. [53]Norwegian Institute of Bioeconomy Research (NIBIO), Division of Forestry and Forest Resources, Ås, Norway. [54]European Commission, Joint Research Centre, Ispra, Italy. [55]Compensation International Progress S.A., Bogotá, Colombia. [56]Laboratory of Applied Ecology, University of Abomey-Calavi, Cotonou, Benin. [57]Scientific Services, South African National Parks, Knysna, South Africa. [58]School of Geography, University of Leeds, Leeds, UK. [59]Department of Geomatics, Forest Research Institute, Sekocin Stary, Raszyn, Poland. [60]Proceedings of the National Academy of Sciences, Washington, DC, USA. [61]Department of Evolutionary Anthropology, Duke University, Durham, NC, USA. [62]Department of Environment, Universtité du Cinquantenaire de Lwiro, Bukavu, Democratic Republic of Congo. [63]Department of Environment, Ghent University, Ghent, Belgium. [64]Department of Green Chemistry and Technology, Ghent University, Ghent, Belgium. [65]Service of Wood Biology, Royal Museum for Central Africa, Tervuren, Belgium. [66]Balai Penelitian dan Pengembangan Lingkungan Hidup dan Kehutanan, Manokwari, Indonesia. [67]Institute of Tropical Forest Conservation, Mbarara University of Science and Technology, Mbarara, Uganda. [68]Université de Liège, Gembloux Agro-Bio Tech, Gembloux, Belgium. [69]Integrated Center for Research, Development and Innovation in Advanced Materials, Nanotechnologies, and Distributed Systems for Fabrication and Control (MANSiD), University Stefan cel Mare of Suceava, Suceava, Romania. [70]Department of Forestry Sciences, 'Luiz de Queiroz' College of Agriculture, University of São Paulo, Piracicaba, Brazil. [71]Bavarian State Institute of Forestry, Freising, Germany. [72]Department of Natural Sciences, Manchester Metropolitan University, Manchester, UK. [73]Facultad de Ciencias Forestales, Universidad Juárez del Estado de Durango, Durango, Mexico. [74]Institute of Biology and Botanical Garden, Martin Luther University Halle-Wittenberg, Halle (Saale), Germany. [75]German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Leipzig, Germany. [76]Development Economics Group, Wageningen University, Wageningen, Netherlands. [77]Rosen Center for Advanced Computing (RCAC), Purdue University,

West Lafayette, IN, USA. [78]Department of Biological, Geological and Environmental Sciences (BiGeA), University of Bologna, Bologna, Italy. [79]Institute of Integrative Biology, ETH Zürich, Zürich, Switzerland. [80]IFER – Institute of Forest Ecosystem Research, Jilove u Prahy, Czech Republic. [81]Global Change Research Institute of the CAS, Brno, Czech Republic. [82]Programa de Pós-graduação em Biologia Vegetal, Instituto de Biologia, Universidade Estadual de Campinas, Campinas CEP, Biologia, Brazil. [83]Dirección Nacional de Bosques (DNB), Ministerio de Ambiente y Desarrollo Sostenible (MAyDS), Ciudad Autónoma de Buenos Aires, Buenos Aires, Argentina. [84]Department of International Environment and Development Studies (Noragric), Faculty of Landscape and Society, Norwegian University of Life Sciences (NMBU), Ås, Norway. [85]Department of Environment and Geography, University of York, York, UK. [86]Department of Environmental Science, School of Engineering and Sciences, SRM University-AP, Guntur, India. [87]Department of Botany, Dr. Harisingh Gour Vishwavidyalaya (A Central University), Madhya Pradesh, India. [88]Department of Ecology and Environmental Sciences, Pondicherry University, Puducherry, India. [89]Centre for Structural and Functional Genomics & Quebec Centre for Biodiversity Science, Biology Department, Concordia University, Montreal, Quebec, Canada. [90]Department of Ecology, Faculty of Science, Charles University, Prague, Czech Republic. [91]Biology Centre of the Czech Academy of Sciences, Institute of Entomology, Ceske Budejovice, Czech Republic. [92]Cirad, UMR EcoFoG (AgroParistech, CNRS, Inrae, Université des Antilles, Université de la Guyane), Campus Agronomique, Kourou, French Guiana. [93]Department of Geography, Environment and Geomatics, University of Guelph, Guelph, Ontario, Canada. [94]National Forest Authority, Kampala, Uganda. [95]Department of Silviculture Foundation, Silviculture Research Institute, Vietnamese Academy of Forest Sciences, Hanoi, Vietnam. [96]Department of Botany, Faculty of Science, University of South Bohemia, Bohemia, Czech Republic. [97]IPHAMETRA, IRET, CENAREST, Libreville, Gabon. [98]Faculté de Gestion de Ressources Naturelles Renouvelables, Université de Kisangani, Kisangani, Democratic Republic of Congo. [99]Queensland Herbarium, Department of Environment and Science, Toowong, Queensland, Australia. [100]School of Biological and Behavioural Sciences, Queen Mary University of London, London, UK. [101]Department of Plant Biology, Faculty of Science, University of Yaoundé I, Yaoundé, Cameroon. [102]Natural Resources Institute Finland, Joensuu, Finland. [103]Institute of Plant Sciences, University of Bern, Bern, Switzerland. [104]Department of Forest Resource Management, Swedish University of Agricultural Sciences, Umea, Sweden. [105]Research and Innovation Centre, Fondazione Edmund Mach, San Michele all'Adige, Italy. [106]Herbário Dr. Roberto Miguel Klein, Universidade Regional de Blumenau, Blumenau, Brazil. [107]Glick Designs, LLC, Hadley, MA, USA. [108]CIIDIR Durango, Instituto Politécnico Nacional, Durango, Mexico. [109]Département des Sciences et Technologies de l'Environnement, Université du Burundi, Bujumbura, Burundi. [110]Faculté des Sciences, Evolutionary Biology and Ecology Unit, Université Libre de Bruxelles, Brussels, Belgium. [111]Royal Botanic Garden Edinburgh, Edinburgh, UK. [112]Department of Plant Sciences, University of Oxford, Oxford, UK. [113]Department of Plant Systematics, Bayreuth University, Bayreuth, Germany. [114]Helmholtz GFZ German Research Centre for Geosciences, Section 1.4 Remote Sensing and Geoinformatics, Potsdam, Germany. [115]Wild Chimpanzee Foundation, Liberia Representation, Monrovia, Liberia. [116]Centre for Conservation Science, The Royal Society for the Protection of Birds, Sandy, UK. [117]Department of Environment, Laboratory for Wood Technology (UGent-Woodlab), Ghent University, Ghent, Belgium. [118]AMAP, University of Montpellier, CIRAD, CNRS, INRAE, IRD, Montpellier, France. [119]Department of Forest Science, Tokyo University of Agriculture, Tokyo, Japan. [120]Biology Department, Université Officielle de Bukavu, Bukavu, Democratic Republic of Congo. [121]Institute of Dendrology, Polish Academy of Sciences, Kórnik, Poland. [122]Poznan University of Life Sciences, Faculty of Forestry and Wood Technology, Department of Game Management and Forest Protection, Poznan, Poland. [123]Department of Geosciences and Natural Resource Management, University of Copenhagen, Copenhagen, Denmark. [124]Plant Biology Department, Biology Institute, University of Campinas (UNICAMP), Campinas, Brazil. [125]Department of Botany and Plant Pathology, Purdue University, West Lafayette, IN, USA. [126]Institute of Agricultural Research for Development (IRAD), Nkolbisson, Ministry of Scientific Research and Innovation, Yaounde, Cameroon. [127]Department of Food and Resource Economics, University of Copenhagen, Copenhagen, Denmark. [128]Forestry Faculty, Bauman Moscow State Technical University, Mytischi, Russia. [129]Integrative Research Center, The Field Museum, Chicago, IL, USA. [130]Labo Botanique, Université Félix Houphouët-Boigny, Abidjan, Ivory Coast. [131]Computational and Applied Vegetation Ecology Lab, Ghent University, Ghent, Belgium. [132]Department of Physical and Environmental Sciences, Colorado Mesa University, Grand Junction, CO, USA. [133]Department of Botany, Dr. Harisingh Gour Vishwavidalaya (A Central University), Sagar, India. [134]Kenya Forestry Research Institute, Department of Forest Resource Assessment, Nairobi, Kenya. [135]Department of Forest Sciences, Seoul National University, Seoul, Republic of Korea. [136]Interdisciplinary Program in Agricultural and Forest Meteorology, Seoul National University, Seoul, Republic of Korea. [137]National Center for Agro Meteorology, Seoul, Republic of Korea. [138]Research Institute for Agriculture and Life Sciences, Seoul National University, Seoul, Republic of Korea. [139]Institute of Forestry and Engineering, Estonian University of Life Sciences, Tartu, Estonia. [140]International Institute for Applied Systems Analysis, Laxenburg, Austria. [141]Department of Geoinformatics, Central University of Jharkhand, Ranchi, India. [142]Tartu Observatory, University of Tartu, Tõravere, Estonia. [143]School of Life Sciences, University of KwaZulu-Natal, Pietermaritzburg, South Africa. [144]Department of Forest Engineering, Federal University of Viçosa (UFV), Viçosa, Brazil. [145]Department of Geography, University College London, London, UK. [146]Plant Systematics and Ecology Laboratory (LaBosystE), Higher Teacher's Training College, University of Yaoundé I, Yaoundé, Cameroon. [147]Laboratoire d'Écologie et Aménagement Forestier, Département d'Ecologie et de Gestion des Ressources Végétales, Université de Kisangani, Kisangani, Democratic Republic of Congo. [148]Instituto de Silvicultura e Industria de la Madera, Universidad Juarez del Estado de Durango, Durango, Mexico. [149]Faculty of Forestry, Qingdao Agricultural University, Qingdao, China. [150]Center for Forest Ecology and Productivity RAS (CEPF RAS), Moscow, Russia. [151]Department of Ecoscience, Aarhus University, Silkeborg, Denmark. [152]Nicholas School of the Environment, Duke University, Durham, NC, USA. [153]Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ, USA. [154]AgroParisTech, UMR AMAP, University of Montpellier, CIRAD, CNRS, INRAE, IRD, Montpellier, France. [155]University of the Sunshine Coast, Sippy Downs, Queensland, Australia. [156]University of York, York, UK. [157]Flamingo Land Ltd., North Yorkshire, UK. [158]Department of Wildlife Management, College of African Wildlife Management, Mweka, Tanzania. [159]Kenya Forestry Research Institute, Headquarters, Nairobi, Kenya. [160]Departamento de Ecología y Recursos Naturales, Facultad de Ciencias, Universidad Nacional Autónoma de México, Mexico City, Mexico. [161]Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT, USA. [162]Department of Forest Management and Forest Economics, Warsaw University of Life Sciences, Warsaw, Poland. [163]Tropical Forests and People Research Centre, University of the Sunshine Coast, Maroochydore DC, Queensland, Australia. [164]Fieldstation Fabrikschleichach, Julius-Maximilians University Würzburg, Würzburg, Germany. [165]Bavarian Forest Nationalpark, Grafenau, Germany. [166]Fakultas Kehutanan, Universitas Papua, Jalan Gunung Salju Amban, Manokwari Papua Barat, Indonesia. [167]Limbe Botanic Garden, Limbe, Cameroon. [168]Institute of Forestry, Belgrade, Serbia. [169]Tropical Plant Exploration Group (TroPEG), Buea, Cameroon. [170]Department of Ecology and Evolutionary Biology, University of Tennessee, Knoxville, TN, USA. [171]Applied Biology and Ecology Research Unit, University of Dschang, Dschang, Cameroon. [172]Department of Forestry and Natural Resources, University of Kentucky, Lexington, KY, USA. [173]UQAM, Centre for Forest Research, Montreal, Quebec, Canada. [174]V.N. Sukachev Forest Institute of FRC KSC SB RAS, Krasnoyarsk, Russia. [175]Urban Management and Planning, School of Social Sciences, Western Sydney University, Penrith, New South Wales, Australia. [176]Instituto Nacional de Pesquisas da Amazônia—INPA, Grupo Ecologia. Monitoramento e Uso Sustentável de Áreas Úmidas MAUA, Manaus, Brazil. [177]Centro de Formação em Ciências Agroflorestais, Universidade Federal do Sul da Bahia, Ilhéus, Brazil. [178]Department of Agriculture, Food, Environment and Forestry, University of Firenze, Firenze, Italy. [179]Technical University of Munich, School of Life Sciences Weihenstephan, Chair of Forest Growth and Yield Science, Munich, Germany. [180]Centro Agricoltura, Alimenti, Ambiente, University of Trento, San Michele all'Adige, Italy. [181]Department of Biology, University of Florence, Sesto Fiorentino, Italy. [182]MUSE—Museo delle Scienze, Trento, Italy. [183]Infoflora c/o Botanical Garden of Geneva, Geneva, Switzerland. [184]Agricultural Research, Education and Extension Organization (AREEO), Research Institute of Forests and Rangelands (RIFR), Tehran, Iran. [185]Department of Environmental Sciences, Central University of Jharkhand, Ranchi, India. [186]Institute of International Education Scholar Rescue Fund (IIE-SRF), One World Trade Center, New York, NY, USA. [187]Centro de Modelación y Monitoreo de Ecosistemas, Facultad de Ciencias, Universidad Mayor, Santiago, Chile. [188]Vicerrectoría de

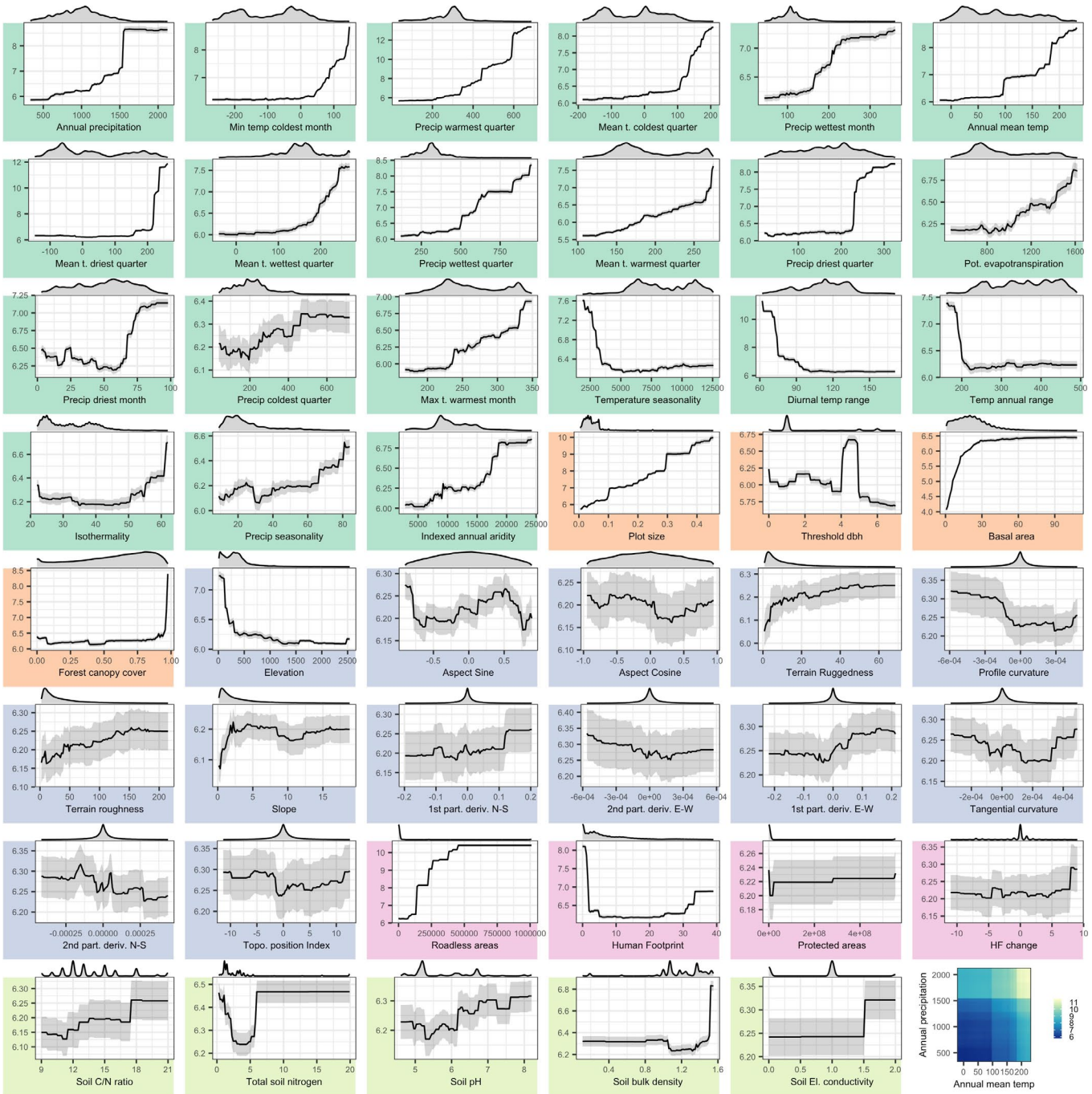
Investigación y Postgrado, Universidad de La Frontera, Temuco, Chile. [189]Departamento de Silvicultura y Conservación de la Naturaleza, Universidad de Chile, Santiago, Chile. [190]Peoples Friendship University of Russia (RUDN University), Moscow, Russia. [191]University of Freiburg, Faculty of Biology, Freiburg, Germany. [192]Institution with City, Department of Geography, University of Zurich, Zurich, Switzerland. [193]National Forest Centre, Zvolen, Slovak Republic. [194]CNRS-UMR LEEISA, Campus Agronomique, Kourou, French Guiana. [195]Universite de Lorraine, AgroParisTech, INRA, Nancy, France. [196]Center for International Forestry Research (CIFOR), Situ Gede, Bogor Barat, Indonesia. [197]Cirad, University of Montpellier, Montpellier, France. [198]Universidade Federal do Rio Grande do Norte, Departamento de Ecologia, Natal, Brazil. [199]School of Biological Sciences, University of Aberdeen, Aberdeen, UK. [200]Herbarium Kew, Royal Botanic Gardens Kew, London, UK. [201]Faculté des Sciences Appliquées, Université de Mbujimayi, Mbujimayi, Democratic Republic of Congo. [202]Yale School of Forestry and Environmental Studies, New Haven, CT, USA. [203]Ural State Forest Engineering University, Botanical Garden, Ural Branch of the Russian Academy of Sciences, Yekaterinburg, Russia. [204]DIBAF Department, Tuscia University, Viterbo, Italy. [205]LINCGlobal, MNCN, CSIC, Madrid, Spain. [206]Plant Ecology and Nature Conservation Group, Wageningen University, AA Wageningen, Netherlands. [207]Agricultural High School, ESAV, Polytechnic Institute of Viseu, IPV, Viseu, Portugal. [208]Centre for the Research and Technology of Agro-Environmental and Biological Sciences, CITAB, UTAD, Quinta de Prados, Vila Real, Portugal. [209]Department of Forest Engineering, Universidade Regional de Blumenau, Blumenau, Brazil. [210]Nucleo de Estudos e Pesquisas Ambientais, Universidade Estadual de Campinas, Campinas (UNICAMP), SP, Campinas, Brazil. [211]International Center for Tropical Botany, Department of Biological Sciences, Florida International University, Miami, FL, USA. [212]Forest Research Institute, University of the Sunshine Coast, Sippy Downs, Queensland, Australia. [213]Sanya Nanfan Research Institute, Hainan Yazhou Bay Seed Laboratory, Hainan University, Sanya, China. [214]Kenya Forestry Research Institute, Taita Taveta Research Centre, Wundanyi, Kenya. [215]Department of Wetland Ecology, Institute for Geography and Geoecology, Karlsruhe Institute for Technology, Rastatt, Germany. [216]Department of Forest Management, Centre for Agricultural Research in Suriname, Paramaribo, Suriname. [217]Polish State Forests-Coordination Centre for Environmental Projects, Warsaw, Poland. [218]Research Center of Forest Management Engineering of State Forestry and Grassland Administration, Beijing Forestry University, Beijing, China. [219]Department of Statistics, University of Wisconsin–Madison, Madison, WI, USA. [220]Institut National Polytechnique Félix Houphouët-Boigny, DFR Eaux, Forêts et Environnement, BP, Yamoussoukro, Ivory Coast. [221]Centre for Invasion Biology, Department of Mathematical Sciences, Stellenbosch University, Matieland, South Africa. [222]African Institute for Mathematical Sciences, Muizenberg, South Africa. ✉e-mail: albeca.liang@gmail.com; chui@sun.ac.za



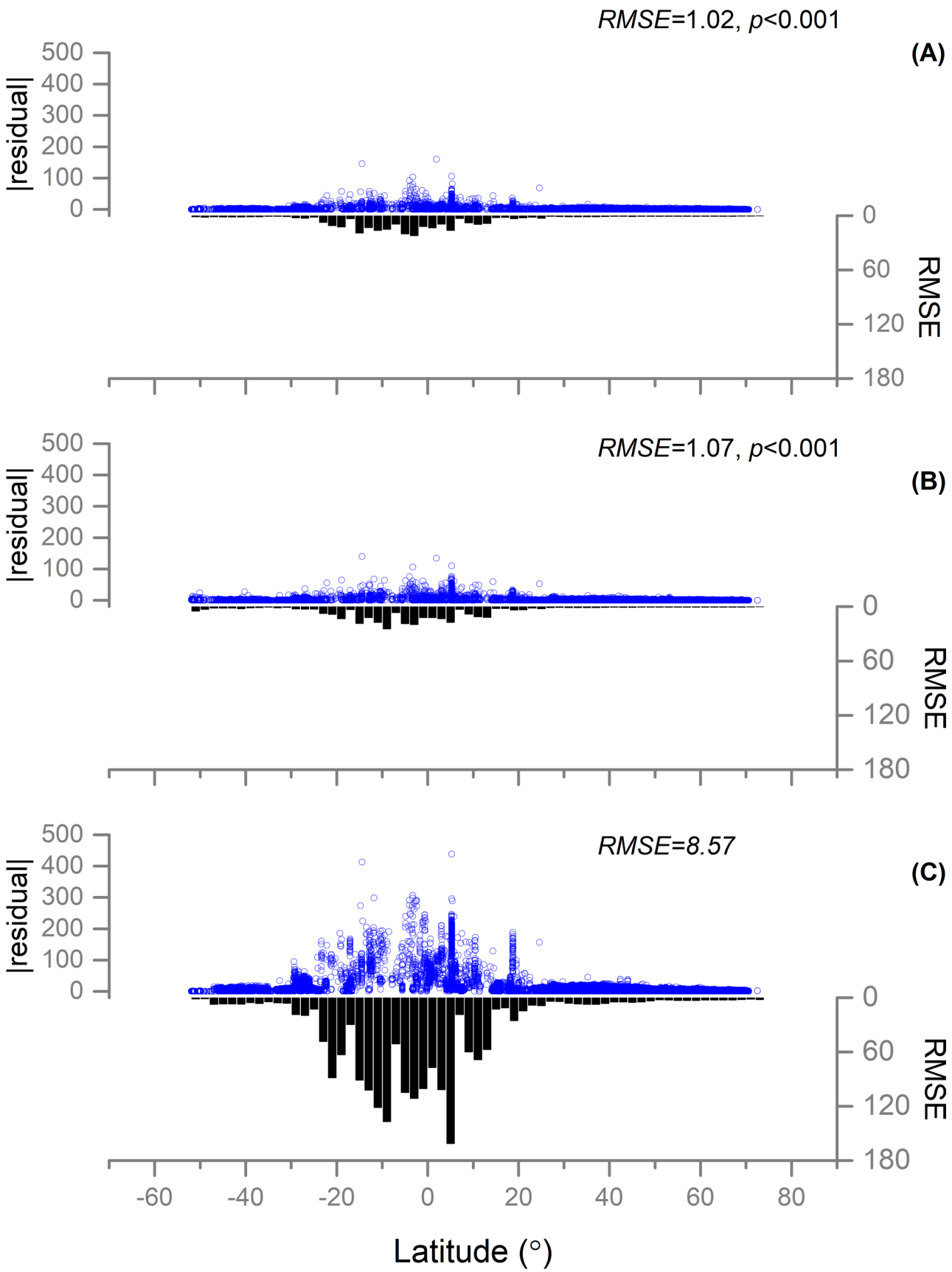
Extended Data Fig. 1 | Distribution of sample plots in the global forest biodiversity individual-based database (GFBi). (**A**) GFBi data consists of *in situ* measurements of approximately 1,255,444 forest sample plots (Phase I sample plots, blue). Together with an independent and complementary dataset of 22,131 forest sample plots (Phase II sample plots, yellow). (**B**) GFBi sample plots cover 424 ecoregions across the world, with the proportion of sample plots by ecoregion ranging from 0.001% to 19.333%. Phase I sample plots cover 394 ecoregions across the world, and Phase II sample plots cover an additional 30 ecoregions in Africa, South America, Southeast Asia, Mexico, India, and Japan. Together, our ground-based forest sample plots cover 424 of 435 forested ecoregions across the world.



Extended Data Fig. 2 | Examination of estimated tree species richness using three cross-validation approaches. (**Left column**) In randomized cross-validation (RCV), three imputation models—random forests (RF, top), multiple regression with ordinary least squares (OLS, centre), and XGBoost (XGB, bottom)—were trained for each continent with a random subsample comprising 90% of the training data from that continent; the remaining 10% of the training data were used as the testing set. This process was repeated 20 times with sample replacement to examine the accuracy of estimated tree species richness values for each sample plot. (**Centre column**) In spatial cross-validation (SCV), all sample data from an ecoregion were reserved for testing the three imputation models, trained with the remaining samples from the same continent. This process was repeated until all the forested ecoregions across the world had been tested. (**Right column**) For post-sample validation (PSV), we collated a new sample dataset from 22,131 forest sample plots (Phase II data), and used Phase II data as the testing set to evaluate the accuracy of the predictive models that were trained for each continent with the Phase I data. Scatter plots show observed (vertical axis) vs. predicted (horizontal axis) values of tree species richness per hectare, from which we calculated mean absolute error (MAE), root-mean-squared error (RMSE), and the coefficient of determination (R^2) of the trendlines (red) between the predicted and observed values.

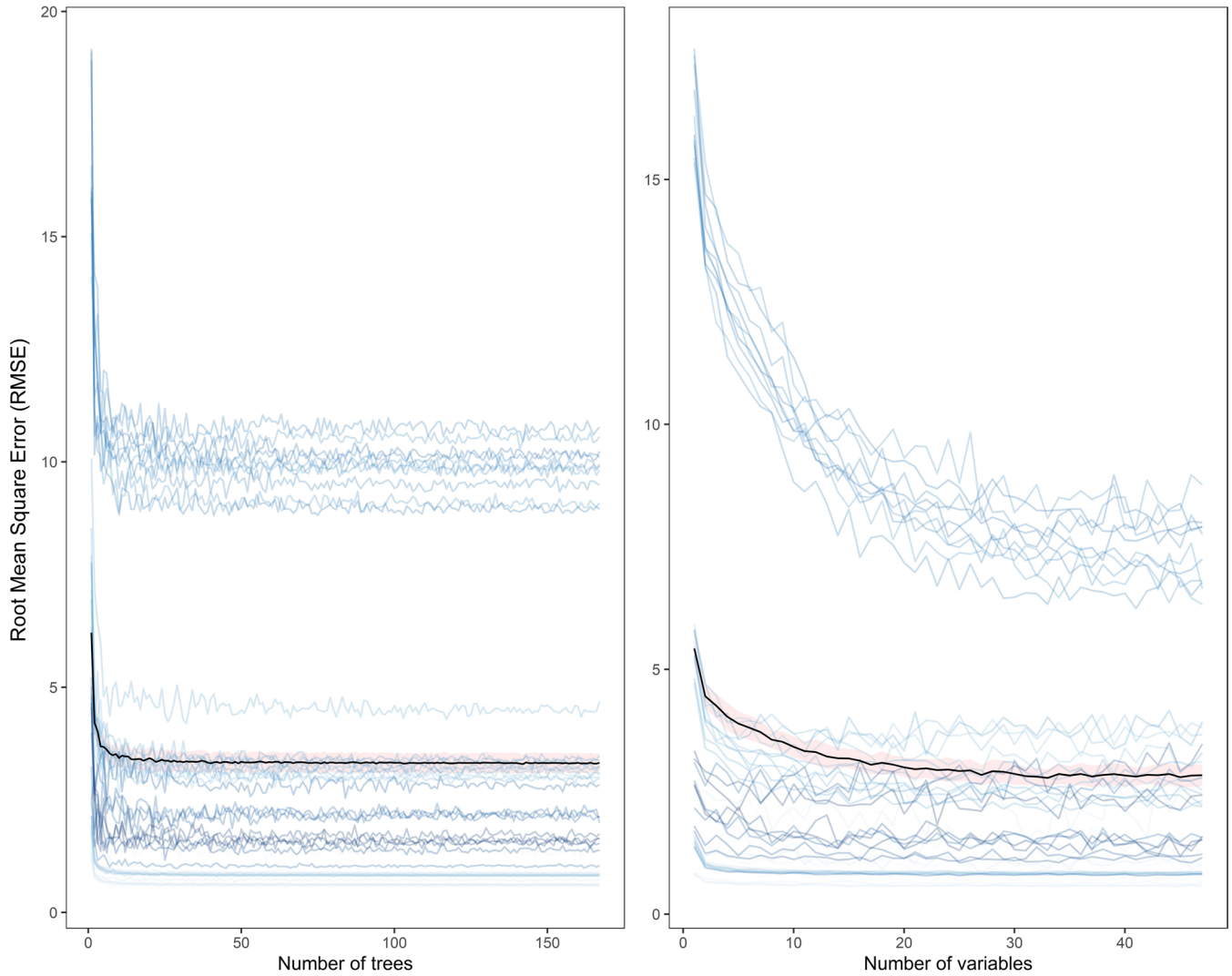


Extended Data Fig. 3 | Partial dependence of tree species richness on each predictor variable. We drew a partial dependence plot (with 95% confidential interval in shaded bands) for each of the 47 explanatory variables in five categories, that is, bioclimatic (green), vegetation and survey (orange), topographic (blue), anthropogenic (pink), and soil (lime). The final panel in the lower right corner is a 2D partial dependence plot showing how tree species richness changes by annual mean temperature and total annual precipitation, while all the other predictors remained constant at their sample mean. Curves on top of all partial dependence plots depict the density of sample data for each explanatory variable. See Supplementary Table 1 for the units and a detailed description of the explanatory variables.

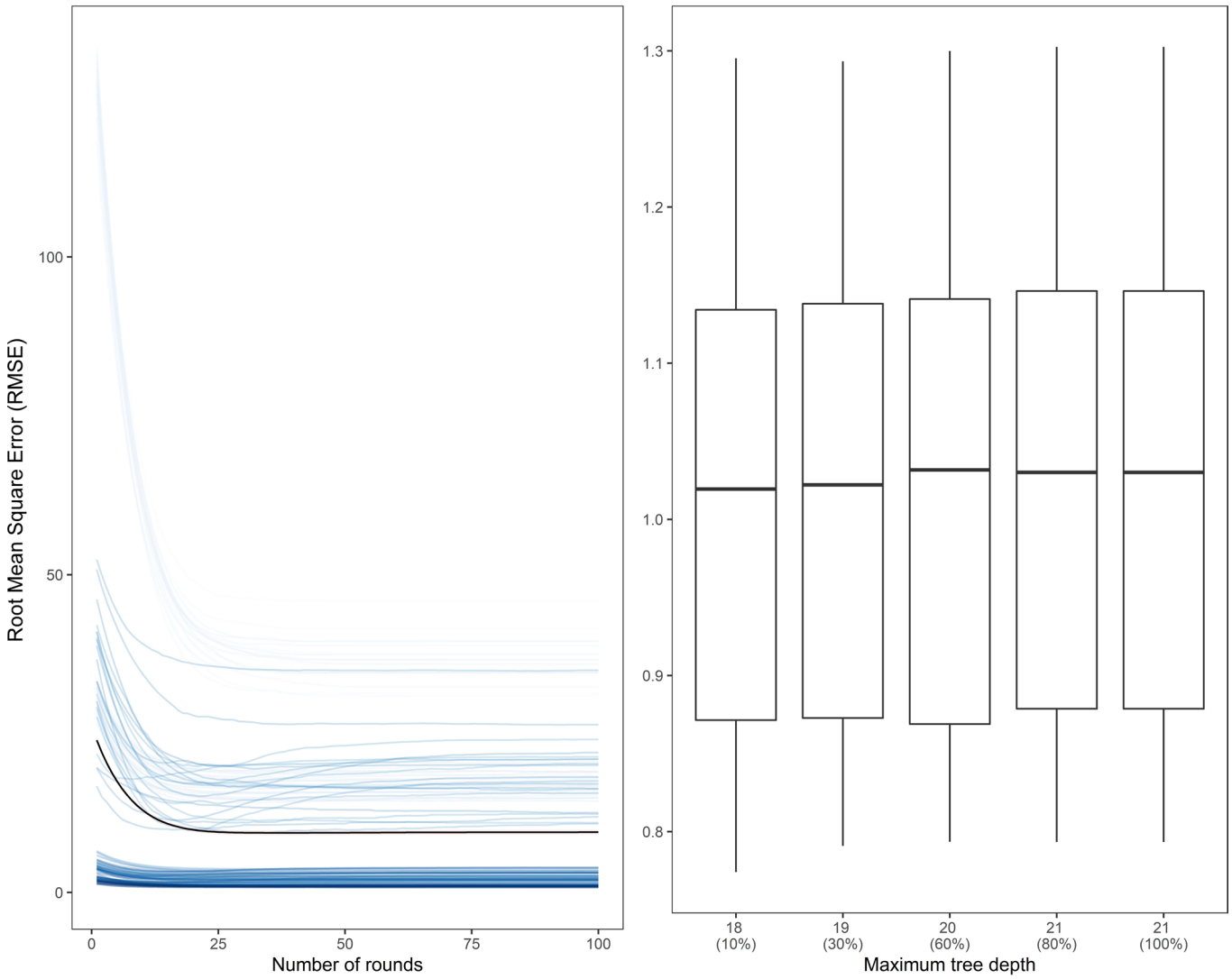


Extended Data Fig. 4 | See next page for caption.

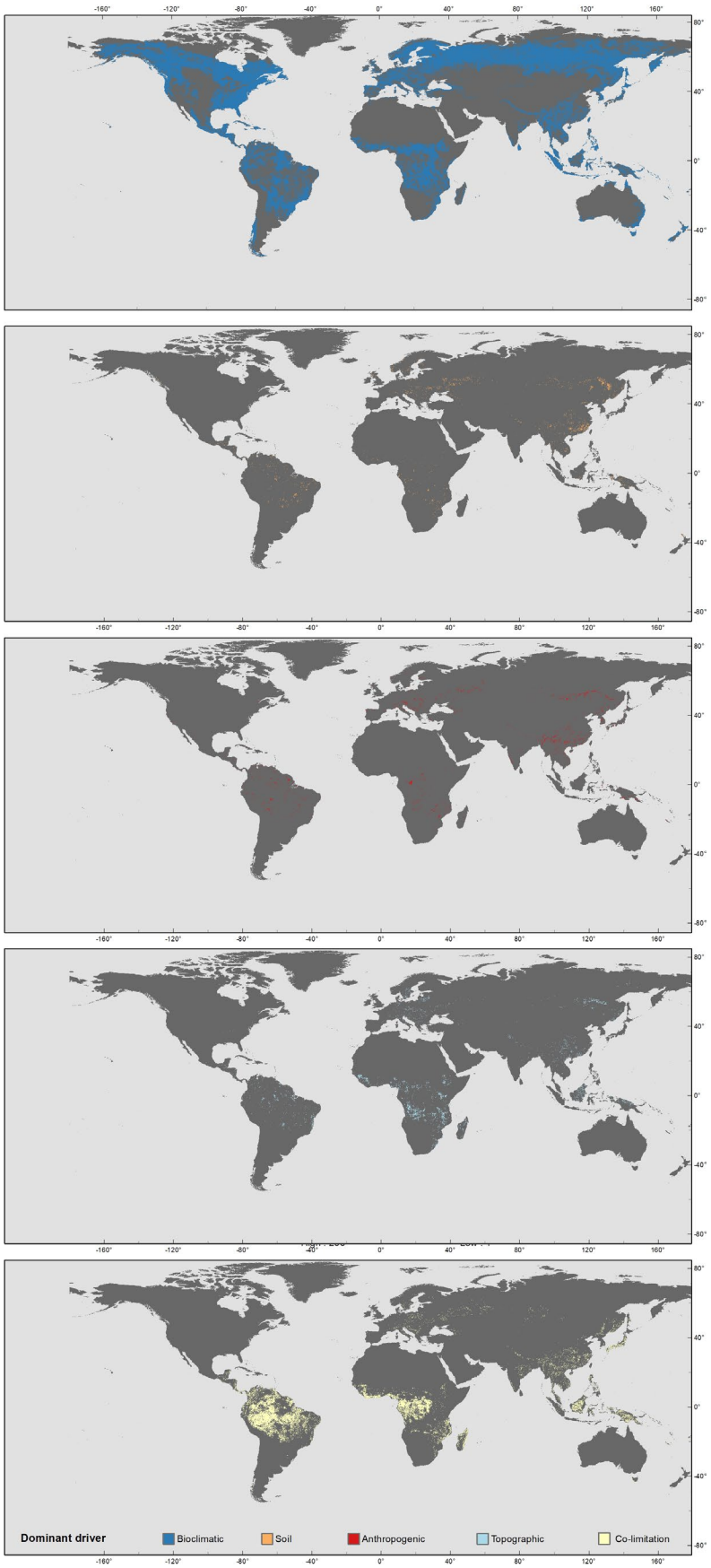
Extended Data Fig. 4 | Analyses of the latitudinal gradients in the residuals from three nested models. For each model, we show a scatter plot of absolute residual values (blue circles) and a column plot of root-mean-squared error (RMSE) calculated from the training dataset, with 2° bins and vertical axis in reverse order (black columns). Model (**A**) consisted of all 47 explanatory variables plus latitude. Model (**B**) contained all the explanatory variables in (**A**) except for the 21 bioclimatic variables. Model (**C**) as the base model only contained a zero constant. RMSE values in the upper-right corner of each panel represent the total RMSE, with *p*-values calculated from the one-sided F-test (see **Variance Partitioning** for details). *p*(A) stands for the p-value of all 47 explanatory variables plus latitude, and *p*(B) stands for the p-value of 21 bioclimatic variables.



Extended Data Fig. 5 | Hyper-parameter selection for the random forests model. Using 10 bootstrapping iterations on random training sets consisting of 1000 random sample for each continent, we calculated the sensitivity of the root-mean-squared error (RMSE) to (**left**) the number of trees to grow and (**right**) the number of variables randomly sampled as candidates at each split in the random forests model. As RMSE converged at 50 trees and 10 variables, we selected them as optimal hyper-parameters for the random forests model. Black lines represent the mean of all the iterations (with red bands showing the standard error of the mean), and blues lines represent each iteration.



Extended Data Fig. 6 | Hyper-parameter selection for the XGBoost model. Using 20 bootstrapping iterations on random training sets consisting of 90% of sample for each continent, we calculated the sensitivity of the root-mean-squared error (RMSE) of the testing sets (consisting of the remaining 10% of sample) to (**left**) the maximum number of boosting iterations (that is number of rounds), and (**right**) the maximum depth of a tree for the XGBoost model. As RMSE converged at 50 rounds and 20 depth, we selected them as optimal hyper-parameters for the XGBoost model. The box plot represents the 25th and 75th percentiles (bounds of box), median (centre line), and the maximum and minimum (upper and lower whiskers) of the RMSE values for each level of tree depth.



Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | Co-limiting environmental drivers of tree species richness per hectare in forested areas worldwide. Driver dominance was derived for each pixel from four driver categories (*that is* bioclimatic, topographic, anthropogenic, and soil), as well as 'co-limitation' which represents a lack of clear dominance among the four categories above. All map layers are displayed at a 0.025° pixel with an equirectangular projection (Plate–Carrée). See Fig. 5 for a map of all categories overlaid on top of each other.

nature portfolio

Corresponding author(s): Jingjing Liang
Cang Hui

Last updated by author(s): 05/16/2022

Reporting Summary

Nature Portfolio 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 Portfolio policies, see our Editorial Policies 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	No software was used
Data analysis	The development of the GFBi database, tabular data cleaning, creation of species abundance matrices, evaluation of diversity determinants, and geostatistical imputation were conducted in R49 (v.3.4.2) through the use of several Linux-based high-performance computing (HPC) resources at Purdue University, and a custom HPC interface developed using Amazon Web Services, each designed for batch processing, scalable resource distribution, embarrassingly parallel computations, and/or large RAM jobs. Compute nodes with up to 1TB of RAM and clusters of up to 64 nodes were employed in this study. Portions of the covariate preparation, mapping, and quality control assessment were conducted on Windows-based operating systems with up to 128 GB of RAM. Final continental-level RF models and the R codes we developed to train the models are available under license MIT, with the identifier: 10.6084/m9.figshare.17234729.

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 and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio 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 description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Data availability

● The global map of tree species richness is available under license CC BY 4.0, with the identifier: 10.6084/m9.figshare.17232491. This map can be downloaded in two formats. One is a geoTIFF file (S_mean_raster.tif) containing the fully geo-referenced map of tree species richness worldwide at a 0.025°×0.025° resolution. The other is a comma-separated file (S_mean_grid.csv) with the following attributes:
S is local average tree species richness per hectare
x, y are centroid coordinates of all 0.025°×0.025° pixels;
● The global map of co-limitation is available under license CC BY 4.0, with the identifier: 10.6084/m9.figshare.17234339.
● The metadata of the entire training dataframe – including the characteristics and references of all the in situ Phase-I and Phase-II datasets, as well as the definitions, units, and summary statistics of the environmental covariates – is available under license CC BY 4.0, with the identifier: 10.6084/m9.figshare.19733449.v1
● The public version of the training dataframe including the plot-level species richness and all the covariates, which is needed to reproduce the models and results presented here, will be deposited to a public data repository Figshare upon the final acceptance of this paper. Meanwhile, we will also publish this dataframe on two international web research platforms: science-i.org, and gfbinitiative.org.
● Raw forest inventory data are commonly subject to a wide array of confidentiality clauses in regards to open access policies. Despite recent efforts to make some of these data fully open76,77, some governments and private data owners, especially those from the developing countries generally have decided to keep their data confidential. This decision is based on well-founded arguments to protect certain trees or forests (because of their large size or protected taxonomic status) from illegal logging or trespassing, and to protect landowners' privacy, against the misuse of plot information such as the geographic coordinates. The sensitive information in the training dataframe, including the plot coordinates and tree-level information, will be available from the corresponding author (albeca.liang@gmail.com) upon a request via Science-I or GFBI, and an approval from data contributors.

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

Ecological, evolutionary & environmental sciences study design

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

Study description	We compiled a ground-sourced forest inventory database containing ca. 55 million trees in ~1.3 million forest sample plots, based on which we developed the first global maps of tree species diversity and global map of co-limilation
Research sample	In situ measurement of ~1.3 million forest sample plots spanning across 110 countries and territories
Sampling strategy	Sampling strategy is reported separately by the 96 national, regional, and local forest inventories underlying our database. See Table S2 for a complete list of all the forest inventories and references to the associated reports.
Data collection	Data collection is reported separately by the 96 national, regional, and local forest inventories underlying our database. See Table S2 for a complete list of all the forest inventories and references to the associated reports.
Timing and spatial scale	The overall GFBi database has a global coverage. The metadata of the entire training dataframe – including the characteristics and references of all the in situ Phase-I and Phase-II datasets, as well as the definitions, units, and summary statistics of the environmental covariates – is available under license CC BY 4.0, with the identifier: 10.6084/m9.figshare.19733449.v1
Data exclusions	Trees with incorrect or unrecognisable species names may have been excluded, according to the section Data collection and standardization.
Reproducibility	This study uses observational data from natural forest ecosystems.
Randomization	Random forests model was trained to estimate global forest tree species richness and co-limilation
Blinding	Blinding was not relevant to this study, because this study is based primarily on observational tree data from natural forest ecosystems.

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	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

View publication stats