nature ecology & evolution

ARTICLES

https://doi.org/10.1038/s41559-018-0709-x

A global test of ecoregions

Jeffrey R. Smith[1,2*], Andrew D. Letten[1,3], Po-Ju Ke[1], Christopher B. Anderson[1,2], J. Nicholas Hendershot[1,2], Manpreet K. Dhami[1,4], Glade A. Dlott[1], Tess N. Grainger[5], Meghan E. Howard[1], Beth M. L. Morrison[1], Devin Routh[6], Priscilla A. San Juan[1], Harold A. Mooney[1], Erin A. Mordecai[1], Thomas W. Crowther[6] and Gretchen C. Daily[1,2,7,8]

A foundational paradigm in biological and Earth sciences is that our planet is divided into distinct ecoregions and biomes demarking unique assemblages of species. This notion has profoundly influenced scientific research and environmental policy. Given recent advances in technology and data availability, however, we are now poised to ask whether ecoregions meaningfully delimit biological communities. Using over 200 million observations of plants, animals and fungi we show compelling evidence that ecoregions delineate terrestrial biodiversity patterns. We achieve this by testing two competing hypotheses: the sharp-transition hypothesis, positing that ecoregion borders divide differentiated biotic communities; and the gradual-transition hypothesis, proposing instead that species turnover is continuous and largely independent of ecoregion borders. We find strong support for the sharp-transition hypothesis across all taxa, although adherence to ecoregion boundaries varies across taxa. Although plant and vertebrate species are tightly linked to sharp ecoregion boundaries, arthropods and fungi show weaker affiliations to this set of ecoregion borders. Our results highlight the essential value of ecological data for setting conservation priorities and reinforce the importance of protecting habitats across as many ecoregions as possible. Specifically, we conclude that ecoregion-based conservation planning can guide investments that simultaneously protect species-, community- and ecosystem-level biodiversity, key for securing Earth's life support systems into the future.

The spatial distribution of species, communities and ecosystems is one of the most fundamental and well-investigated topics in ecology[1–7]. Early ecologists and biogeographers focused on mapping and predicting delineations between ecosystems from one another at regional (ecoregions) and global (biomes) spatial scales[4,8–10]. These efforts relied principally on topographic features (that is, mountains, rivers and other water bodies), gradients of temperature and rainfall, as well as broad vegetation patterns[8,9,11], although more recent efforts have incorporated soil conditions, geologic history and recent human land-management regimes[12,13]. These geographical groupings quickly emerged as significant forces in the structuring of modern ecological research and modelling[14–21].

Early ecoregion maps were intended to delineate ecosystems principally from a biophysical standpoint, under the assumption that similar environmental conditions would lead to similar ecosystem function, regardless of species identity[8,9]. The resulting concept, terminology and maps, however, were quickly adopted and modified by community and landscape ecologists who compiled data from many disparate data sources to adapt maps of ecoregions to better reflect species- and community-level biodiversity patterns[22,23]. As the field of biodiversity science has evolved, it has become canonical that these modified ecoregion borders represent meaningful and sharp boundaries between biological communities of all taxa[24,25].

Despite the widespread use of ecoregion maps for guiding decision making throughout environmental sciences and management[23,26–29], current research of biodiversity patterns might now suggest that sharp boundaries between ecoregions may not exist[30,31]. Remote sensing of the Earth's surface has revealed that environmental gradients rarely show the sharp changes across narrow geographical bands that would suggest functionally distinct ecoregions[32–34]. Moreover, species distributions under both present and historical conditions suggest that, in general, individual species rarely associate with a single community over a broad enough spatial (regional to continental) or temporal (decades to centuries) extent[35–37]. Therefore, despite the recognized value of categorically defined ecoregions in the development of modern biodiversity science, it is possible that they are obsolete and may even misinform research and conservation efforts that increasingly utilize continuous predictor and response variables.

Here, we ask: do ecoregion borders represent meaningful delineations between distinct biotic communities? We address this question by testing two competing hypotheses: the sharp-transition hypothesis and the gradual-transition hypothesis. The sharp-transition hypothesis predicts a rapid accumulation of new species on crossing into a new ecoregion (Fig. 1b) visible as a spike in the number of new species discovered (Fig. 1c), while the gradual-transition hypothesis predicts species should be discovered gradually (Fig. 1d) with no significant change in the rate of novel species discovery at ecoregion borders (Fig. 1e). We tested these hypotheses using over 200 million point occurrences (including plants, arthropods, birds, mammals, reptiles, amphibians and fungi) from the Global Biodiversity Information Facility (GBIF; Supplementary Table 1)[38] and the most commonly used conservation-relevant map of global ecoregions[22,23]. We generated pseudo-random sampling transects by tracing a path through 10-km grid cells within ecoregions and across boundaries, simulating a naturalist moving across the landscape at random, and recording the location and identity of species encountered (Fig. 1a)[39].

---

[1]Department of Biology, Stanford University, Stanford, CA, USA. [2]Center for Conservation Biology, Stanford University, Stanford, CA, USA. [3]School of Biological Sciences, University of Canterbury, Christchurch, New Zealand. [4]Landcare Research, Lincoln, New Zealand. [5]Department of Ecology and Evolutionary Biology, University of Toronto, Toronto, Canada. [6]Institute of Integrative Biology, ETH Zurich , Zurich, Switzerland. [7]Woods Institute for the Environment, Stanford University, Stanford, CA, USA. [8]The Natural Capital Project, Stanford University, Stanford, CA, USA. *e-mail: jrsmith7@stanford.edu

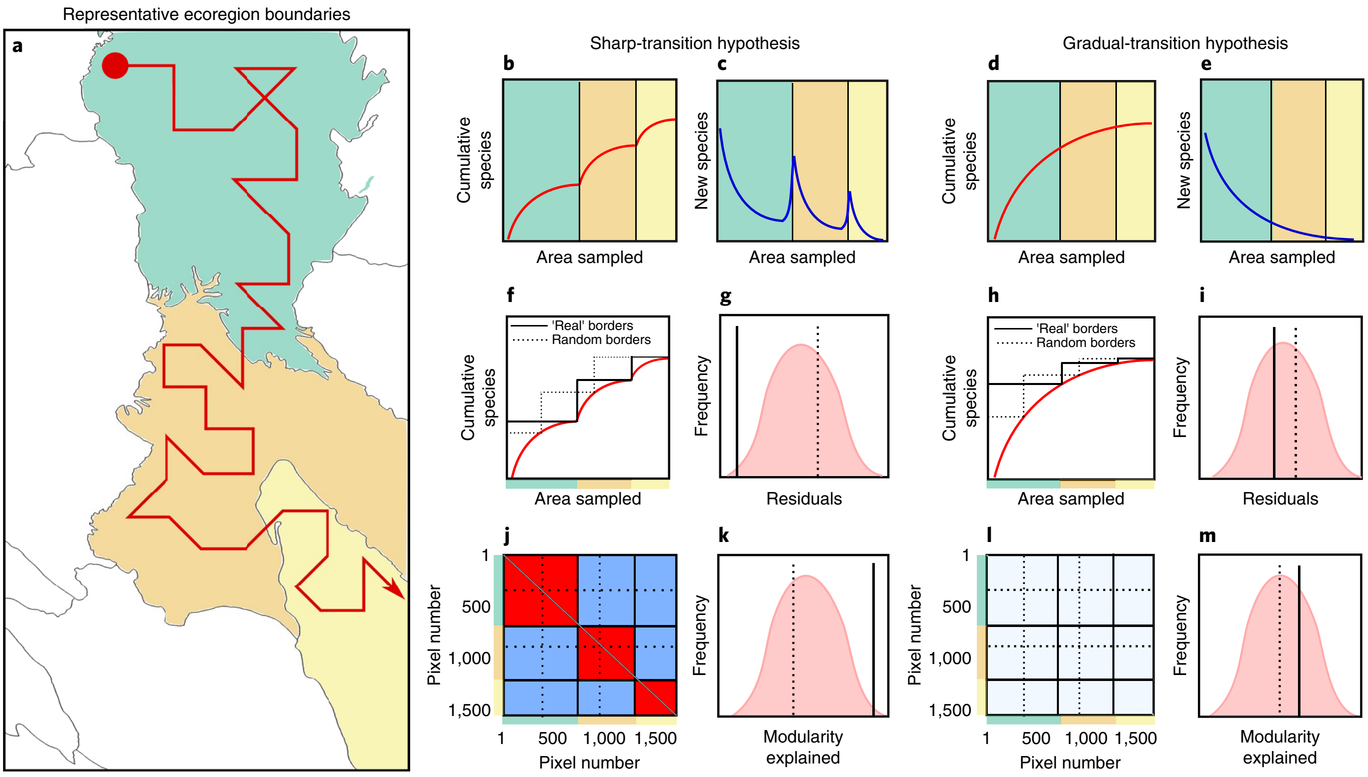

Fig. 1 | Our approach to testing the sharp-transition and gradual-transition hypotheses. a–m, Here we lay out our approach to moving across ecoregion boundaries (**a**) and testing our two hypotheses: the sharp-transition hypothesis, where we expect to see sharp spikes in species discovery at ecoregion boundaries (**b**,**c**) and the gradual-transition hypothesis, where we expect to see a gradual accumulation of species with increased sampling area, regardless of ecoregion boundary locations (**d**,**e**). We fit stepwise regressions (**f**,**h**) to actual ecoregion borders (solid) and sample randomly permuted borders (dotted), which show either an improved fit of the actual borders over permuted borders as expected in the sharp-transition hypothesis (**f**) or only a random change in quality of fit between actual and permuted borders in the gradual-transition hypothesis (**h**). Residuals from the real borders will fall to the left of the histogram of randomly generated residuals for transects supporting the sharp-transition hypothesis (**g**), while residuals from transects adhering to the gradual-transition hypothesis will fall in the middle of the distribution (**i**). We also generate full pairwise pixel to pixel matrices (**j**,**l**) displaying the difference between the observed community similarity (**j**) and the hypothetical community similarity (**l**) generated using geographical distance alone. In this figure, red regions are more similar than distance would predict while blue regions are less similar than distance would predict overlaid with the 'real' ecoregion borders (solid) and an 'average', but random, set of permuted borders (dotted). We then quantify how much similarity in community composition can be explained by knowing where the 'real' ecoregion borders fall. Transects supporting the sharp-transition hypothesis will have more modularity explained when using the 'real' borders and therefore fall to the extreme right of the histogram of possible results (**k**), while those adhering to the gradual-transition hypothesis will lack distinct clusters and therefore should fall in the middle of the distribution (**m**).

Results

To test the relative fit of the sharp and gradual-transition hypotheses, we assessed whether a stepwise function describing species accumulation along ecoregion boundaries outperformed stepwise functions fit with randomly generated boundaries. At an individual transect level (Figs. 1f–h and 2a,b), we evaluated significance as the proportion of randomly generated boundaries that perform better than the known ecoregion boundaries (Figs. 1g–i and 2c,d). We found that accumulation along ecoregion borders performed better than random, distance-based accumulation for amphibians (33.42% (percentage of the transects that met a significance threshold of $P<0.05$)), mammals (29.07%), reptiles (27.97%), plants (26.85%), birds (25.07%), arthropods (17.45%) and fungi (9.82%).

Our findings are more powerful, however, once aggregated across all transects conducted for each taxon (Fig. 3). The probability distributions generated for all seven taxa vary significantly from the uniform distribution that would be expected under the gradual-transition hypothesis ($F\geq 0.2121$, $P<0.001$), providing a strong basis for rejecting the gradual-transition hypothesis. This suggests that ecoregion boundaries better predict biogeographic patterns of species accumulation than random species accumulation by distance. We then asked if there were certain taxonomic groups which were more or less strongly associated with ecoregion boundaries (that is, we were better able to describe species-accumulation patterns for taxa more strongly tied to ecoregion boundaries). This analysis suggests that ecoregion borders best describe species accumulation in amphibians and mammals, followed by reptiles, birds, plants and arthropods, and finally fungi, which show the least adherence to ecoregion boundaries.

To verify that our findings were not an artefact of our methods, we complemented this approach using a community similarity-based test that asked whether ecoregions contain biotic communities that are more similar to one another than spatial distance alone would predict. In this framing, the sharp-transition hypothesis posits that two communities within the same ecoregion will be relatively similar even if they are separated by a significant geographical distance, while communities from different ecoregions will be relatively dissimilar even if they are geographically close to one another. By contrast, the gradual-transition hypothesis posits that similarity between communities should vary strictly as a function of the geographical distance between them, irrespective of whether they are in the same or different ecoregions.

To test these predictions, we generated two matrices for each transect, one that calculated the Euclidean distance between each

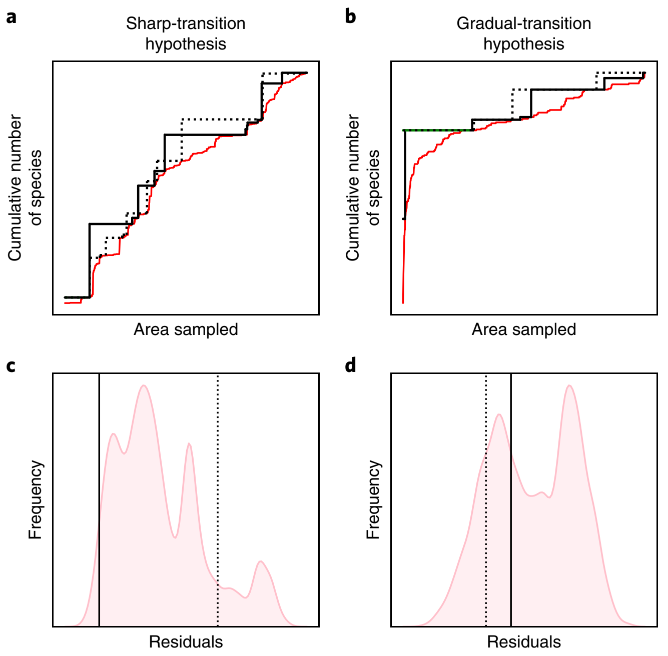

Fig. 2 | Species-accumulation curves from transects representing our two hypotheses. a–d, Here we show the species-accumulation curves from two transects: one representative of the sharp-transition hypothesis (**a**) and one representative of the gradual-transition hypothesis (**b**) fit with stepwise functions derived from either the 'real' ecoregion borders (solid) or a randomly permuted set of borders (dotted). We show the residuals generated from 5,000 possible stepwise functions derived from random ecoregion borders in the pink histograms (**c**,**d**), with the solid line showing the residual from the real ecoregion borders and the dotted line showing the residual from the random permutation shown above. We calculated the *P* value at an individual transect level as the proportion of randomly generated boundaries that perform better than the known ecoregion boundaries. This means that transects supporting the sharp-transition hypothesis fall to the extreme left of the histogram (**c**) and those reflective of the gradual-transition hypothesis falling in the middle of the histogram (**d**).

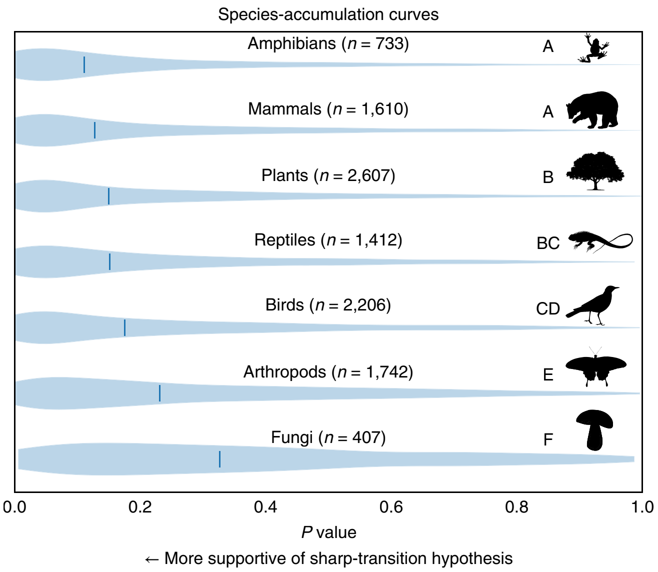

Fig. 3 | Summary of results from species-accumulation curve tests. Across all taxa sampled, we find that ecoregions improve our capacity to describe the rate of species discovery as we move across a landscape, underlying their enduring importance for both the ecological sciences and conservation community. We show the probability distribution resulting from aggregating the *P* values across all transects within a taxon. A skew towards 0 indicates support for the sharp-transition hypothesis while the gradual-transition hypothesis would predict a uniform distribution centred around 0.5. Distributions marked with the same letters are not statistically different from one another, while those not sharing any letters in common are significantly different distributions according to a two-tailed Kolmogorov–Smirnov test ($P < 0.05$).

pair of grid cells (distance matrix) and one that calculated the Jaccard coefficient of similarity between each pair of grid cells (similarity matrix). We compressed these two matrices into a distance-similarity matrix such that positive values indicate that grid cells have more similar species assemblages than distance alone would predict, while negative values indicate that grid cells are more dissimilar than expected by distance[40]. We then applied a modified modularity analysis[41], where we fixed the number and size of the modules based on known ecoregion boundaries and asked how much similarity or dissimilarity was contained in the grid cell comparisons deemed to be in the same ecoregion based on either real or randomly permuted boundaries. At an individual transect level (Figs. 1j–l and 4a,b), we evaluated significance as the proportion of randomly generated boundaries that better explain community similarity (above what distance alone predicts) than the known ecoregion boundaries (Figs. 1k–m and 4c,d).

We find that community similarity, after accounting for distance, is better explained by known ecoregion borders for mammals (28.26% (percentage of transects significant at $P < 0.05$)), birds (22.12%), plants (23.09%), reptiles (18.80%), amphibians (17.70%), arthropods (13.62%) and fungi (9.61%). After aggregating all transects for each taxon we found that the probability distributions generated for all taxa vary significantly from the uniform distribution that would be expected under the gradual-transition hypothesis (Fig. 5, $F \geq 0.120$, $P < 0.001$). As with our tests of species-accumulation rates, we again tested which taxa show the highest amount of community similarity explained by ecoregion boundaries. The results from this approach suggests that ecoregion borders best describe communities of mammals, birds and plants, followed by amphibians and reptiles, then arthropods and finally fungi, differing slightly from the relative ranking of ecoregion fidelity across taxa in the species-accumulation analysis.

Despite the value of presence-only observations for detecting the signals of biodiversity patterns, these approaches have several known limitations: they were not collected systematically, either taxonomically or geographically, and they do not record absences[42]. We address this in two ways. First, we conducted the same analysis described above (species-accumulation tests and distance-similarity matrix tests) on a subset of the GBIF data that includes only point observations coming from contributing datasets spanning five or more ecoregions. This was done to ensure that the results we showed were not biased by certain studies or projects specifically aimed at sampling certain taxa in certain ecoregions. Although this reduces the total number of data points in the analysis (Supplementary Table 1) our results remain essentially unchanged (Supplementary Figs. 1 and 2) and we are therefore confident our results were not biased in this manner.

To further assess the effect of these limitations on our analysis, we also compiled high-quality data from approximately 300,000 plots of tree community composition from the United States Forest Service Forest Inventory and Analysis (USFS FIA) Program (https://www.fia.fs.fed.us/). Although neither global in their distribution nor available across taxa, the abundance values and known absences of the data allows for testing with more comprehensive and rigorous data. Consistent with the global results across taxa using presence-only data, we found that tree communities within the

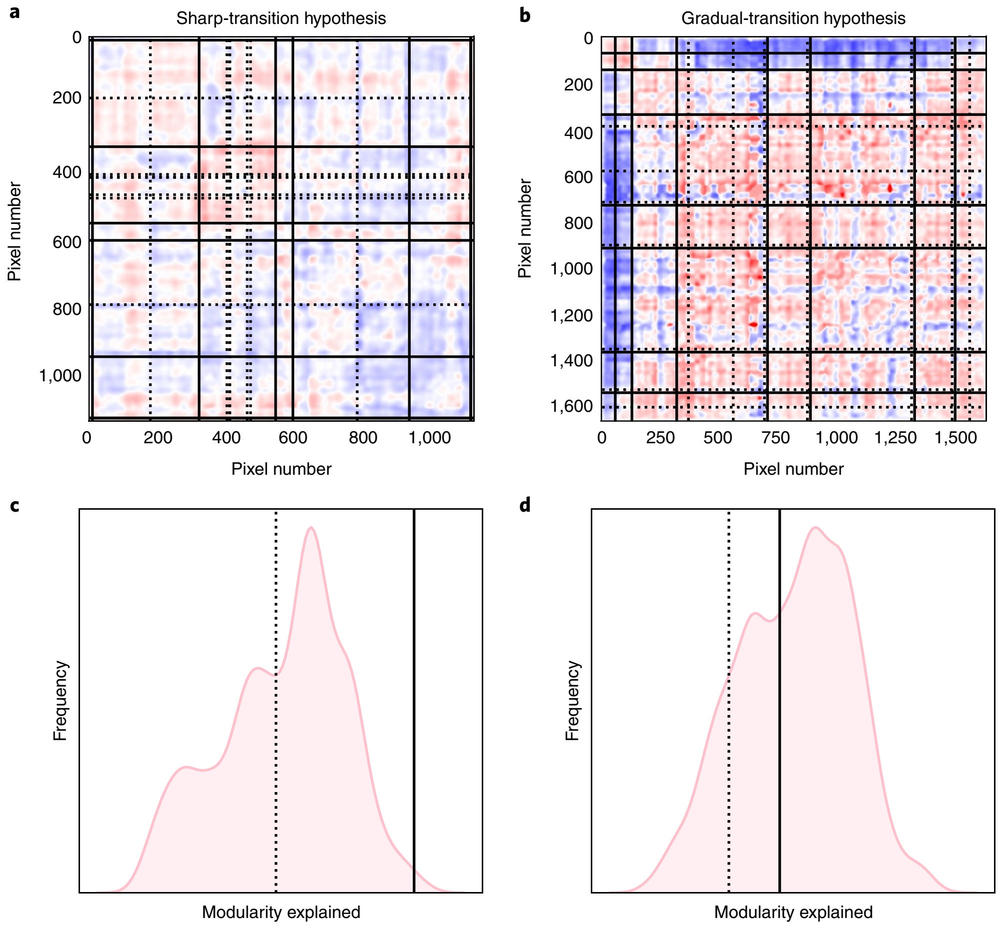

Fig. 4 | Distance-similarity matrices from transects representing our two hypotheses. a – d, Here we show the distance-similarity matrices from two transects: one representative of the sharp-transition hypothesis (**a**) and one representative of the gradual-transition hypothesis (**b**). Here, red areas indicate more similar community composition than expected by distance alone and blue areas are less similar in community composition than expected by distance. The magnitude of the colour corresponds to the size of the effect. We then overlay the ecoregion groupings drawn with the 'real' ecoregion borders (solid) or a random set of permuted boundaries (dotted). For ease of visualization we show the heatmaps after applying a Gaussian smoothing function that influences only the visualization, but not the underlying statistics. The amount of community similarity that randomly placed modules capture from 5,000 possible sets of random ecoregion borders is shown in the pink histograms (**c**,**d**), with the solid line showing the residual from the real ecoregion borders and the dotted line showing the residual from the random permutation shown above. The *P* value at an individual transect level is calculated as the proportion of randomly generated boundaries that perform better than the known ecoregion boundaries. This means that transects reflective of the sharp-transition hypothesis fall to the extreme right of the histogram (**c**) and those reflective of the gradual-transition hypothesis fall in the middle of the histogram (**d**).

same ecoregion are more similar to one another in species composition than they are to plots in other ecoregions, even after controlling for geographical distance (Fig. 6). That is, plots that are separated by ecoregion boundaries were often more dissimilar than others within the same ecoregion that are located hundreds of kilometres away.

Discussion

In our analyses we found strong support for the notion that there are relatively sharp boundaries that can be used to describe both patterns in species accumulation and community similarity. Our results give credence to the persisting importance of ecoregions in understanding ecological principles[18,19] and setting conservation priorities across all taxonomic groups considered[23,43–46]. Although the relative importance of ecoregions to different taxa varies by test, in both of our main analyses we discovered that fungi are the least strongly linked to ecoregion borders. This finding adds to evidence that although fungi do follow biogeographic patterns established based on plants and animals[47,48], such relationships may be weaker than they are for above-ground organisms[49,50]. It may, however, also reflect known biases due to the relatively poor sampling and incomplete, or inaccurate, taxonomic identification of fungi compared to plants and animals[51,52]. Furthermore, separation into functional guilds (for example, saprotrophic, parthenogenic, mycorrhizal) may also elucidate distinct patterns[53,54]. We similarly find that arthropods are relatively poorly described by this set of ecoregion boundaries, possibly reflecting their small body size and dispersal limitations that would emphasize local-scale patterns over macroecological patterns[55]. As with fungi, we caution that this may also be a function of significant biases in the taxa of arthropods represented in the datasets as well as a general under-representation of their total taxonomic diversity[51,52,56] because detailed field studies of insects have found moderate to strong structuring by ecoregion[57–60]. Based on the results here, we cannot make any strong claims for differences among plants, reptiles, amphibians, birds and mammals, but we were surprised to find that birds, despite their wide-ranging nature, show strong fidelity to ecoregions boundaries[61,62].

Since our analyses here focus principally on high-level biogeographic patterns, elucidating the mechanisms underlying observed

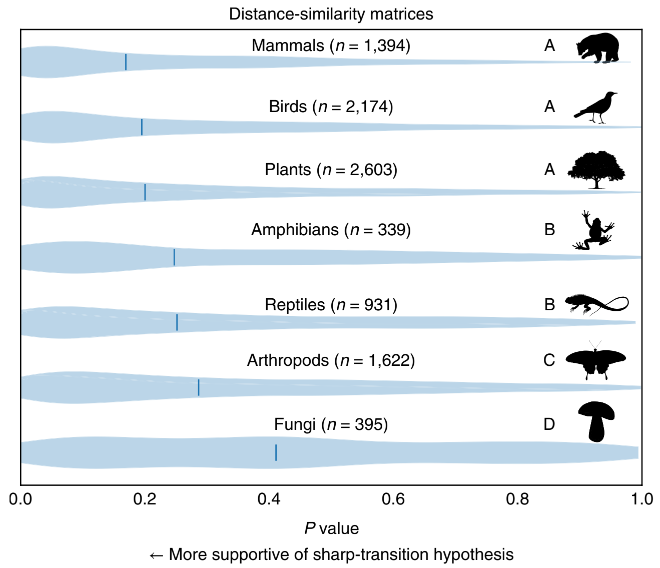

Fig. 5 | Summary of results from distance-similarity matrices tests. Across all taxa sampled, we find that ecoregions improve our capacity to describe community composition as we move across a landscape, underlying the enduring importance for both the ecological sciences and conservation community. Here, we show the summary of the *P* values across all taxa and transects. A skew towards 0 indicates support for the sharp-transition hypothesis, while the gradual-transition hypothesis would predict a uniform distribution centred around 0.5. Distributions marked with the same letters are not statistically different from one another, while those not sharing any letters in common are significantly different distributions according to a two-tailed Kolmogorov–Smirnov test ($P < 0.05$).

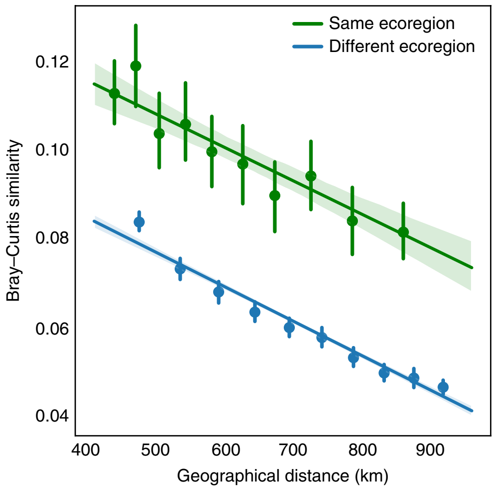

Fig. 6 | Relationship between geographical distance and community similarity in USFS FIA tree plots. Our results showing ecoregions as a foundational biogeographic principle is reinforced here by showing that tree communities (including abundances and known absences) within the same ecoregion are more similar to one another than they are to communities in a different ecoregion that are hundreds of kilometres closer. We show the relationship between the geographical distance of 1 million randomly drawn pairs of USFS FIA plots and the corresponding Bray–Curtis similarity metric (between 0 (completely dissimilar) and 1 (completely similar)). The lines are shown with 99% confidence intervals and are truncated to the range where there are sufficient plot–plot comparisons within the same ecoregion and across different ecoregions to make meaningful conclusions. Rather than showing the underlying scatterplot (which is completely saturated with points), we show binned summary statistics (mean and standard deviation). These are simply graphical and do not influence the regression lines.

ecoregion borders remains a vital research frontier. In particular, understanding why certain taxonomic groups respond more strongly to ecoregion borders than others remains a pressing challenge. Future work should concentrate on using finer grains of phylogenetic resolution or functional traits that may reveal mechanisms that delineate ecoregion boundaries. Another major gap is quantifying the combinations of abiotic and biotic gradients that are responsible for shaping ecoregion boundaries and, crucially, how they might respond to future environmental change[63]. It is also possible that for some taxa there are sharp boundaries between ecoregions, but they do not align with the maps of ecoregions we analyse in this paper[22,23]. This point will become particularly important for conservation plans that are constrained to certain taxa or certain regions because we found high levels of variability across taxa and ecoregions in how well they explain species accumulations rates (Supplementary Tables 2 and 3).

By providing strong and consistent support for the ecoregion concept, these results further our understanding of foundational biogeographic principles[14,15,64] and improve their potential application to conservation efforts[29,43]. Although continuous gradients in environmental conditions certainly shape biological communities, and probably underpin the emergent properties that distinguish ecoregions, our results demonstrate that ecoregion borders represent regions of rapid species turnover. Our findings show how maps of ecoregion boundaries should continue to play a vital role in designing effective large-scale conservation plans[44–46] in two ways. First, we show that not only can ecoregions continue to indicate ecosystem-level diversity, but that they can also serve as proxies for community- and species-level biodiversity in decision contexts where high-quality data is lacking. Second, conservation efforts that aim to maximize the number of species protected should distribute the areas they protect across a diverse array of ecoregion types. These ideas are gaining momentum within governments (for example, United States Environmental Protection Agency[65], European Environment Agency[66]) and international conservation groups (for example, The Nature Conservancy[67], World Wildlife Fund[68]), but our analysis provides compelling empirical evidence that these strategies together may help secure holistic protection of different levels of biodiversity. Indeed, a foundational understanding of the patterns and processes governing the distribution of biodiversity is fundamental to securing the vital elements of nature into the future.

Methods

Acquisition and cleaning of GBIF data. We obtained species occurrence data for this study from the GBIF (www.gbif.org)[38]. GBIF is the world's largest online repository of biodiversity information, with data being principally stored as occurrence data, which includes the geographical coordinates where an individual of a given species was observed. Using GBIF data in global models has immense value because of its scope and coverage, but has several limitations. In particular, the data lacks known absences (that is, there are no records of a place where a species is known not to occur), has significant spatial bias in the density of observations and is prone to misidentification of species. However, one can address these problems with proper cleaning of the data (see later) and statistical considerations[42,69].

We downloaded point occurrence data on 7 May 2018 for seven groups, corresponding to the seven taxonomic groups presented in the main text using the following searches: (1) Plants–Kingdom = 'Plantae'; (2) Arthropods–Phylum = 'Arthropoda'; (3) Reptiles–Class = 'Reptilia'; (4) Amphibians–Class = 'Amphibia'; (5) Mammals–Class = 'Mammalia'; (6) Birds–Class = 'Aves' and (7) Fungi–Kingdom = 'Fungi' (Supplementary Table 1). We proceeded to reproject all point occurrence data from WGS 1984 (world geodetic system) to EPSG:3410 NSIDC EASE-Grid Global (National Snow and Ice Data Center equal-area scaleable Earth)[70]. We chose this projection system because it provided equal area coverage across the entirety of the study area (global except for polar biomes, which were excluded from this study), resulting in a rectangular, uninterrupted grid, crucial for employing our transect-based approach.

We applied the following cutoffs for a point occurrence to be included in subsequent analysis. We first eliminated all records without georeferenced latitude and longitude points. We then removed all points where either the latitude or longitude was exactly 0, indicating a missing value[71]. We also removed any point where the margin of error was greater than 10 km (our selected pixel size). The margin of error is calculated as "the horizontal distance (in metres) from the given decimalLatitude and decimalLongitude describing the smallest circle containing the whole of the Location"[72]. We also restricted observations to those characterized as: 'observation', 'human observation', 'machine observation', 'specimen', 'preserved specimen', 'living specimen' or 'material sample'. The two most significant groups that were excluded for this reason were fossil records and observations with an unknown basis. We finally excluded records collected prior to 1950. It has been suggested that one of the main sources of location and georeferencing errors in GBIF is associated with old collections[73]. Our aim here was to balance losing records from setting a cutoff that is too restrictive (there were approximately 1 million records per year collected in the 1950s) with adding highly inaccurate data. We felt 1950 represented a good compromise between these two competing aims.

We used all data points remaining after this first set of cutoffs (Supplementary Table 1); however, we recognize that there may be certain datasets within the GBIF database that were collected to sample a specific taxon within one or across a small number of ecoregions. Therefore, we also conducted a follow-up set of data cleaning, in which we included point occurrences coming only from datasets that spanned at least five ecoregions. We chose this number to exclude datasets that spanned only a few ecoregions, potentially indicating studies that focused on a specific taxonomic group within one ecoregion or across a small number of ecoregions. This reduced the number of points used in the analysis (Supplementary Table 1); however, the results were qualitatively unchanged. We therefore present the results that includes all data to increase the robustness of our findings, but we also provide the results stemming from the subset of data in the Supplementary Information (Supplementary Figs. 1 and 2).

Preparation of ecoregion maps. We obtained the ecoregion shapefile map first published by Olson et al. (2001 map)[22] from The Nature Conservancy's website (http://maps.tnc.org/gis_data.html). However, a new version of the global ecoregions map was recently published by Dinerstein et al. (2017 map)[23], which we obtained from Resolve (https://ecoregions2017.appspot.com/). Similar to the point data, we reprojected these maps from WGS 1984 to EPSG:3410. We then rasterized the maps with square $10 \times 10\,\mathrm{km}^2$ grid cells. The value of these new raster maps corresponds to the unique identifier included for each ecoregion in the initial shapefile. We then used these maps to plot the transect paths .

We present results based on the updated 2017 map in the main text because it is a direct update of the 2001 map. However, given the importance of the 2001 map in shaping scientific research and conservation policy, we also conducted all analyses with the 2001 map and provide the results in the Supplementary Information (Supplementary Figs. 3 and 4). The results are qualitatively similar.

Transect selection. Transects began at a random point on the globe, pixels were not sampled more than once, a given ecoregion could only be entered and exited once, water was not entered and transects were terminated on turning in on themselves (when all adjacent pixels had either been sampled or were water pixels and therefore ineligable for sampling). Moves for each interval were chosen randomly from moving either one pixel away in any of the eight cardinal or diagonal directions, or two grid cells away in any of the four cardinal directions with equal probability of each possible move being selected. We did this to prevent transects from terminating by turning in on themselves too quickly.

A transect had to contain at least 750 grid cells (75,000 km^2 sampled) to be included in subsequent analyses. Furthermore, we performed sensitivity analyses to address variability in the spatial distribution of point occurrence data. To carry out these analyses we varied the minimum number of pixels along the transect containing species occurrence records, the minimum number of ecoregions traversed, the minimum number of species and the minimum number of total points in the transect (see section on Sensitivity analysis in Supplementary methods).

Calculating residuals. The species-accumulation curve tests determine the similarity of the observed data to a stepwise function fit to either 'real' ecoregion boundaires or randomly drawn ecoregion boundaries. After generating the species-accumulation curves, we calculated the position along the observed data curve to dictate where the stepwise function would intersect the curve. The points where the stepwise function intersected the observed data were selected by first taking the index for the last pixel sampled in each ecoregion before crossing into the next ecoregion (x value) and then calculating the cumulative number of species up to and including that pixel (y value). In the first case, these (x, y) values are based on known ecoregion boundaries and in subsequent permutations are varied randomly (see section on Permutation analysis in Supplementary Methods). In all cases, the stepwise function serves as a ceiling to the observed data.

To judge the fit of all the stepwise functions generated from known and random bounds, we calculated the residual by taking the difference of our modelled stepwise curve and the observed data and then squaring the result. This is comparable to a standard least-squares regression. However, rather than fitting the best curve, we assessed whether stepwise functions drawn from known ecoregion borders better fit the observed data than stepwise functions drawn from random ecoregion borders.

Making distance-similarity matrices. To calculate the distance-similarity matrices, we first generated a pairwise matrix of geographical distance between grid cells using Euclidean distance. We also created a pairwise matrix of community similarity distance between all grid cells using Jaccard's index of dissimilarity (which gives entirely similar grid cells a score of 0 and entirely dissimilar grid cells a score of 1). For ease of interpreting results, we converted this to a similarity coefficient by subtracting all values from 1.

As the point occurrence data can only represent the presence and not the absence, we eliminated all comparisons between two grid cells that did not share any species. We felt it was unfair to say two grid cells where entirely dissimilar when we did not have full community data, particularly because we lacked known absences. We therefore performed a presence-only Jaccard test between cells, which subset the data to only include comparisons that have at least one overlapping species.

We log-transformed the community similarity matrix to improve normality, but did not convert the geographical distance matrix. This enabled testing an exponential decay function fit between the two curves. We varied the decay relationship (quadratic, linear) in subsequent tests investigating the robustness of our conclusions to the shape of the decay rate between community similarity and geographical distance (see section on Varying assumptions of distance-similarity matrices in Supplementary Methods). We then converted both matrices to z-scores using the mean and standard deviation. For the similarity matrix, positive z-scores indicated higher than average community similarity and negative z-scores indicated lower than average community similarity. For the distance matrix, positive z-scores indicated a larger geographical distance and negative z-scores indicated a smaller geographical distance. We inverted the z-scores for geographical distance so that positive z-scores represented comparisons that were either similar in community composition (for the community similarity matrix) or nearby (for the geophonic distance matrix), while negative z-scores represented comparisons that were either dissimilar in community makeup (for the similarity matrix) or far away (for the geographical distance matrix).

We linearly regressed community similarity against geographical distance. Based on the above transformations we expected a positive relationship between the two z-scores. If there was a perfect relationship between the two variables with no environmental noise, we would expect a regression with a positive slope and an intercept near 0. We took the geographical distance matrix and applied the equation derived from the regression to generate a predicted similarity score for each comparison based on distance alone. We then subtracted the actual similarity scores from the scores predicted solely by distance. In this method, positive values indicated pairwise comparisons that are more similar than distance alone predicts, while negative values indicated pairwise comparisons less similar than distance predicts. We based our permutation analyses on this matrix, that is, the difference between expected and observed community similarity, which, by definition, is 0 across all grid cells. To calculate the amount of residual community similarity that can be explained by any given set of ecoregion boxes, we added the values contained in ecoregion modules located along the main diagonal. We then compared the value from known ecoregion boundaries to random boundaries, with a higher value indicating that the set of ecoregion boundaries is better at describing community similarity.

Permutation analysis. Both model-based tests (species-accumulation curves and distance-similarity matrices) relied on permutation analyses to determine whether the community along an individual transect adheres to the established ecoregion boundaries rather than any boundaries that could be drawn. Since the transects were sampled in an order-dependent fashion (that is, pixel 2 always comes after pixel 1) and the 'correct' ecoregion boundaries can be slotted directly in between any two grid cells, we were able to ask what would happen if we maintained the order of the grid cells, but changed the location in the transect order at which the ecoregion breaks were inserted. This enabled us to compare the results of analyses conducted with the correct borders versus random iterations of where the borders might be inserted into the otherwise properly ordered transect.

Importantly, we fixed the distribution of grid cells across each ecoregion. In other words, although the borders could be shuffled, the number of grid cells contained within each ecoregion was conserved. To illustrate with a simple example, imagine there are 100 grid cells, divided into four ecoregions, with 25 in the first ecoregion, 40 in the second ecoregion, 15 in the third ecoregion and 20 in the final ecoregion. These boundaries could be reordered so that the number of grid cells falls across the four ecoregions in any of the following ratios: 40–25–20–15, 15–20–25–40, 25–20–15–40 and so on, but never 30–30–20–20. This is crucial because it ensures that any changes to calculations due to permuting ecoregion borders derived not from changing the widths of bin, but rather from changing the difference between the expected results of the model within that bin from the actual data.

We used the species-accumulation curve and distance-similarity matrix tests, presented in the main text, and the 5,000 randomly permuted ecoregion boundaries, to generate the histograms against which we compared our result from the 'real' borders. In the case of the species-accumulation curve, the histogram reflected the residuals of all stepwise functions generated using random boundaries. In the case of the distance-similarity matrix, it represented how much of the modularity of the matrix can be explained by a given set of random ecoregion boundaries. In this type of analysis, the P value is equal to one minus the proportion of randomly permuted boundaries that the known ecoregion borders outperforms. In the species-accumulation curve, outperforming means that the known boundaries have a lower residual, while in the distance-similarity matrix outperforming indicates that known ecoregion boundaries better explain the modularity of the distance-similarity matrix. This allowed us to consider a discretized significant/non-significant threshold for each transect (that is, whether known boundaries for an individual transect outperformed 95% of ecoregion borders) and also a continuous distribution of P values from many transects (that is, the total probability distribution of all transects). We compared the aggregated probability distributions for each taxon against a normal distribution and against one other using a one-sided and two-sided Kolmogorov–Smirnov test, respectively.

Preparation of USFS FIA data. We obtained data for the USFS FIA plots through their online web portal (https://www.fia.fs.fed.us/). We cleaned the data to select for trees greater than 12.7 cm (5 inches) in diameter at breast height (DBH), the definition used by the FIA Program to define a tree. We then thinned the data to select only the most recent sample from a plot (if the plot was surveyed more than once). Finally, we arranged the data into a community composition matrix, with plot identity along one axis and species identity along the other and values in the middle representing the number of individuals of a given species in a given plot.

Since these plots are not arranged in a similar transect fashion to the above analysis, we are only able to compare them using a distance-similarity approach. We randomly sampled 1 million pairs of plots, calculating both geographical distance and Bray–Curtis similarity, to explore whether species assemblages differed more between or across ecoregion borders when controlling for distance. Since plots in the same ecoregion will generally be much closer than those in different ecoregions, we constrained the figure to focus on the region of overlap. We did this by trimming the dataset to only include comparisons that were further away than 95% of the comparisons between ecoregions and closer together than 95% of the comparisons within ecoregions. This allowed us to focus on comparisons of community compositions within distances where it was realistic to have many comparisons between plots in the same ecoregion and plots in different ecoregions. We chose to present this test in the main text because it allows us to include information on abundance; however, we also provide analysis using the Jaccard coefficient of similarity (Supplementary Fig. 12) and the 2001 map of ecoregions published by Olson et al. (Supplementary Figs. 13 and 14). In all cases, the results are qualitatively similar.

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

Code availability. All code used in data analysis is available from https://github.com/jeffreysmith-jrs/testingEcoregions

Data availability

All data used in this study are publicly available from either www.gbif.org or www.fia.fs.fed.us

Received: 15 February 2018; Accepted: 3 October 2018;

References

1. Ebach, M. C. *Origins of Biogeography* (Springer, Dordrecht, 2015).
2. Gleason, H. A. On the relation between species and area. *Ecology* **3**, 158–162 (1922).
3. Clements, F. E. *Plant Succession. An Analysis of the Development of Vegetation* (Carnegie Institution of Washington, Washington DC, 1916).
4. von Humboldt, A. & Bonpland, A. *Essay on the Geography of Plants* (University of Chicago Press, Chicago, 2013).
5. Tjørve, E., Calf Tjørve, K. M., Šizlingová, E. & Šizling, A. L. Great theories of species diversity in space and why they were forgotten: the beginnings of a spatial ecology and the Nordic early 20th-century botanists. *J. Biogeogr.* **45**, 530–540 (2018).
6. Wallace, A. R. What are zoological regions? *Nature* **49**, 610–613 (1894).
7. Wallace, A. R. *The Geographical Distribution of* Animals*: With a Study of the Relations of Living and Extinct Faunas as Elucidating the Past Changes of the Earth's Surface* (Harper and Brothers, New York, 1876).
8. Holdridge, L. R. Determination of world plant formations from simple climatic data. *Science* **105**, 367–368 (1947).
9. Whittaker, R. H. Classification of natural communities. *Bot. Rev.* **28**, 1–239 (1962).
10. Carpenter, J. R. The biome. *Am. Midl. Nat.* **21**, 75–91 (1939).
11. Hutchins, L. W. The bases for temperature zonation in geographical distribution. *Ecol. Monogr.* **17**, 325–335 (1947).
12. Bailey, R. G. Identifying ecoregion boundaries. *Environ. Manage.* **34**, S14–S26 (2004).
13. Ellis, E. C. & Ramankutty, N. Putting people in the map: anthropogenic biomes of the world. *Front. Ecol. Environ.* **6**, 439–447 (2008).
14. Crowther, T. W. et al. Mapping tree density at a global scale. *Nature* **525**, 201–205 (2015).
15. Snyder, P. K., Delire, C. & Foley, J. A. Evaluating the influence of different vegetation biomes on the global climate. *Clim. Dynam.* **23**, 279–302 (2004).
16. Naidoo, R. et al. Global mapping of ecosystem services and conservation priorities. *Proc. Natl Acad. Sci. USA* **105**, 9495–9500 (2008).
17. Sala, O. E. et al. Global biodiversity scenarios for the year 2100. *Science* **287**, 1770–1774 (2000).
18. Holt, B. G. et al. An update of Wallace's zoogeographic regions of the world. *Science* **339**, 74–78 (2013).
19. Rueda, M., Rodríguez, M. Á. & Hawkins, B. A. Identifying global zoogeographical regions: lessons from Wallace. *J. Biogeogr.* **40**, 2215–2225 (2013).
20. Di Marco, M., Watson, J. E. M., Currie, D. J., Possingham, H. P. & Venter, O. The extent and predictability of the biodiversity–carbon correlation. *Ecol. Lett.* **21**, 365–375 (2018).
21. Kreft, H. & Jetz, W. Comment on 'An update of Wallace's zoogeographic regions of the world'. *Science* **341**, 343 (2013).
22. Olson, D. M. et al. Terrestrial ecoregions of the world: a new map of life on Earth. *Bioscience* **51**, 933–938 (2001).
23. Dinerstein, E. et al. An ecoregion-based approach to protecting half the terrestrial realm. *Bioscience* **67**, 534–545 (2017).
24. Bailey, R. G. *Ecoregions* (Springer, New York, 2014).
25. Lamoreux, J. F. et al. Global tests of biodiversity concordance and the importance of endemism. *Nature* **440**, 212–214 (2006).
26. Ricketts, T. H. et al. *Terrestrial Ecoregions of North America: A Conservation Assessment* (Island Press, Washington DC, 1999).
27. Groves, C., Valutis, L. & The Nature Conservancy (US) *Guidelines for Representing Ecological Communities in Ecoregional Conservation Plans* (The Nature Conservancy, Arlington, 1999).
28. Ricketts, T. & Imhoff, M. Biodiversity, urban areas, and agriculture: locating priority ecoregions for conservation. conservation. *Conserv. Ecol.* **8**, 1 (2003).
29. Olson, D. M. & Dinerstein, E. The global 200: priority ecoregions for global conservation. *Ann. Mo. Bot. Gard.* **89**, 199–224 (2002).
30. Ricklefs, R. E. Disintegration of the ecological community. *Am. Nat.* **172**, 741–750 (2008).
31. McDonald, R. et al. Species compositional similarity and ecoregions: do ecoregion boundaries represent zones of high species turnover? *Biol. Conserv.* **126**, 24–40 (2005).
32. Kerr, J. T. & Ostrovsky, M. From space to species: ecological applications for remote sensing. *Trends Ecol. Evol.* **18**, 299–305 (2003).
33. Higgins, S. I., Buitenwerf, R. & Moncrieff, G. R. Defining functional biomes and monitoring their change globally. *Glob. Change Biol.* **22**, 3583–3593 (2016).
34. Anderson, C. B. Biodiversity monitoring, Earth observations and the ecology of scale. *Ecol. Lett.* **21**, 1572–1585 (2018).
35. Jetz, W., McPherson, J. M. & Guralnick, R. P. Integrating biodiversity distribution knowledge: toward a global map of life. *Trends Ecol. Evol.* **27**, 151–159 (2012).
36. Hubbell, S. P. *The Unified Neutral Theory of Biodiversity and Biogeography* (Princeton Univ. Press, Princeton, 2001).
37. Davis, M. B. in *Forest Succession* (eds West, D. C., Shugart, H. H. & Botkin, D. B.) 132–153 (Springer, New York, 1981).
38. *What is GBIF?* (Global Biodiversity Information Facility, 2018); https://www.gbif.org/what-is-gbif
39. Codling, E. A., Plank, M. J. & Benhamou, S. Random walk models in biology. *J. R. Soc. Interface* **5**, 813–834 (2008).
40. Soininen, J., McDonald, R. & Hillebrand, H. The distance decay of similarity in ecological communities. *Ecography* **30**, 3–12 (2007).
41. Fortuna, M. A. et al. Nestedness versus modularity in ecological networks: two sides of the same coin? *J. Anim. Ecol.* **79**, 811–817 (2010).
42. *Report of the Task Group on GBIF Data Fitness for Use in Distribution Modelling* (Global Biodiversity Information Facility, 2017); https://www.gbif.org/document/82612/report-of-the-task-group-on-gbif-data-fitnessfor-use-in-distribution-modelling.
43. Hoekstra, J. M., Boucher, T. M., Ricketts, T. H. & Roberts, C. Confronting a biome crisis: global disparities of habitat loss and protection. *Ecol. Lett.* **8**, 23–29 (2005).
44. Watson, J. E. M. & Venter, O. Ecology: a global plan for nature conservation. *Nature* **550**, 48–49 (2017).
45. Wilson, E. O. *Half-Earth: Our Plant's Fight for Life* (Liveright, New York, 2017).
46. Noss, R. F. et al. Bolder thinking for conservation. *Conserv. Biol.* **26**, 1–4 (2012).
47. Peay, K. G., Bidartondo, M. I. & Arnold, A. E. Not every fungus is everywhere: scaling to the biogeography of fungal–plant interactions across roots, shoots and ecosystems. *New Phytol.* **185**, 878–882 (2010).
48. Peay, K. G., Kennedy, P. G. & Talbot, J. M. Dimensions of biodiversity in the Earth mycobiome. *Nat. Rev. Microbiol.* **14**, 434–447 (2016).

49. Hendershot, J. N., Read, Q. D., Henning, J. A., Sanders, N. J. & Classen, A. T. Consistently inconsistent drivers of microbial diversity and abundance at macroecological scales. *Ecology* **98**, 1757–1763 (2017).
50. Meyer, K. M. et al. Why do microbes exhibit weak biogeographic patterns? *ISME J.* **12**, 1404–1413 (2018).
51. Troia, M. J. & McManamay, R. A. Filling in the GAPS: evaluating completeness and coverage of open-access biodiversity databases in the United States. *Ecol. Evol.* **6**, 4654–4669 (2016).
52. Troudet, J., Grandcolas, P., Blin, A., Vignes-Lebbe, R. & Legendre, F. Taxonomic bias in biodiversity data and societal preferences. *Sci. Rep.* **7**, 9132 (2017).
53. Sato, H., Tsujino, R., Kurita, K., Yokoyama, K. & Agata, K. Modelling the global distribution of fungal species: new insights into microbial cosmopolitanism. *Mol. Ecol.* **21**, 5599–5612 (2012).
54. Tedersoo, L. et al. Global diversity and geography of soil fungi. *Science* **346**, 1256688 (2014).
55. Hortal, J., Roura-Pascual, N., Sanders, N. J. & Rahbek, C. Understanding (insect) species distributions across spatial scales. *Ecography* **33**, 51–53 (2010).
56. Lobo, J. M. The use of occurrence data to predict the effects of climate change on insects. *Curr. Opin. Insect Sci.* **17**, 62–68 (2016).
57. Lightfoot, D. C., Brantley, S. L. & Allen, C. D. Geographic patterns of ground-dwelling arthropods across an ecoegional transition in the North American southwest. *West. N. Am. Naturalist* **68**, 83–102 (2008).
58. González-Reyes, A. X., Corronca, J. A. & Rodriguez-Artigas, S. M. Changes of arthropod diversity across an altitudinal ecoregional zonation in northwestern Argentina. *PeerJ* **5**, e4117 (2017).
59. González-Reyes, A. X., Corronca, J. A. & Arroyo, N. C. Differences in alpha and beta diversities of epigeous arthropod assemblages in two ecoregions of northwestern Argentina. *Zool. Stud.* **51,** 1367–1379 (2012).
60. Anderson, D. J. & Vondracek, B. Insects as indicators of land use in three ecoregions in the Prairie Pothole Region. *Wetlands* **19**, 648–664 (1999).
61. Melo, A. S., Rangel, T. F. L. V. B. & Diniz-Filho, J. A. F. Environmental drivers of beta-diversity patterns in New-World birds and mammals. *Ecography* **32**, 226–236 (2009).
62. Van Rensburg, B. J., Koleff, P., Gaston, K. J. & Chown, S. L. Spatial congruence of ecological transition at the regional scale in South Africa. *J. Biogeogr.* **31**, 843–854 (2004).
63. Beaumont, L. J. et al. Impacts of climate change on the world's most exceptional ecoregions. *Proc. Natl Acad. Sci. USA* **108**, 2306–2311 (2011).
64. Karp, D. S. et al. Intensive agriculture erodes β-diversity at large scales. *Ecol. Lett.* **15**, 963–970 (2012).
65. *Ecoregions* (US Environmental Protection Agency, 2015); https://www.epa.gov/eco-research/ecoregions
66. *Digital Map of European Ecological Regions* (European Environment Agency, 2003); https://www.eea.europa.eu/data-and-maps/data/digital-map-of-european-ecological-regions
67. *The Nature Conservancy Ecoregional Priorities* (LandScope America, 2018); http://www.landscope.org/focus/understand/tnc_portfolio/
68. Ecoregions (World Wildlife Fund, 2018); https://www.worldwildlife.org/biomes
69. Beck, J., Böller, M., Erhardt, A. & Schwanghart, W. Spatial bias in the GBIF database and its effect on modeling species' geographic distributions. *Ecol. Inform.* **19**, 10–15 (2014).
70. *EPSG:3410 NSIDC EASE-Grid Global* (National Snow and Ice Data Center, 2018); https://epsg.io/3410
71. Yesson, C. et al. How global is the Global Biodiversity Information Facility? *PLoS ONE* **2**, e1124 (2007).
72. GBIF replacing coordinateAccuracy w/ coordinateUncertaintyInMeters & coordinatePrecision, Issue #206 (GitHub, 2018); https://github.com/ropensci/rgbif/issues/206
73. Maldonado, C. et al. Estimating species diversity and distribution in the era of Big Data: to what extent can we trust public databases. *Glob. Ecol. Biogeogr* **24,** 973–984 (2015).

Acknowledgements

This work was supported financially by the National Science Foundation's Graduate Research Fellowship Program Division of Graduate Education No. 1656518, the Stanford Department of Biology and the Ward Wilson Woods Jr Environmental Studies Fund (to J.R.S.). Some of the computing for this project was performed on the Sherlock cluster. We would like to thank Stanford University and the Stanford Research Computing Center for providing computational resources and support that contributed to these research results. We thank K. Peay, T. Fukami, B. Brosi and B. Bryant for discussions that increased the quality of the manuscript.

Author contributions

J.R.S., A.D.L. and P.J.K developed the original concept. J.R.S., A.D.L., P.J.K., C.B.A., J.N.H., M.K.D., G.A.D., T.N.G., M.E.H., B.M.L.M. and P.A.S.J. developed the model. J.R.S., C.B.A., D.R. and T.W.C. carried out the spatial analysis. J.R.S. and T.W.C. gathered and analysed supplementary data from the USFS FIA. J.R.S. wrote and edited the manuscript with input from all authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41559-018-0709-x

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

Correspondence and requests for materials should be addressed to J.R.S.

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

nature research

Corresponding author(s): Jeffrey R. Smith

Reporting Summary

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

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. 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
☐	☒	An indication of 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 statistics 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
☐	☒	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	Data is available online from www.gbif.org or https://www.fia.fs.fed.us/
Data analysis	A combination of custom Python scripts were used for statistical analyses and are available upon request from jrsmith7@stanford.edu.

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

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

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

Data is available online from www.gbif.org or https://www.fia.fs.fed.us/

Field-specific reporting

Please select 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/authors/policies/ReportingSummary-flat.pdf

Ecological, evolutionary & environmental sciences study design

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

Study description	Our study analyzed patterns of biodiversity across ecoregion boundaries and asked if these borders meaningfully separate distinct biotic communities or if community composition more so gradually changes continuously across the landscape.
Research sample	A research sample is considered to be a single transect tracing GBIF data points somewhere around the globe, collecting biodiversity information of 7 focal taxa (GBIF data). In the USFS NFI data section a sample is considered to be 2 paired forest service plots.
Sampling strategy	The location of each sample was chosen randomly around the globe (GBIF transect data) or within the United States (USFS NFI data) Sample sizes were chosen arbitrarily, but are well in excess of thresholds needed to use standard tests of significance. If there are issues raised regarding sample size in any component of the paper it is easily addressable by increasing the number of simulations.
Data collection	Data were obtained from other data provider who detailed their data collection. Please see data collection protocols from GBIF (https://www.gbif.org/publishing-data) and USFS NFI (https://www.srs.fs.usda.gov/pubs/gtr/gtr_wo091.pdf).
Timing and spatial scale	Spatial scale - Data were obtained for global (GBIF) or national (USFS NFI) scales. Temporal scale - Data were restricted to include only biodiversity observations from 1950 onwards.
Data exclusions	Some of the initial biodiversity data from GBIF was excluded from analysis. This is common in the field as this data is not collected following standardized methods. A complete explanation of which points were excluded, and why, can be found in the methods section.
Reproducibility	The results come from publicly available data using computer codes available upon request for reproduction of our results.
Randomization	All randomization was done using the random package in Python.
Blinding	Not applicable to this study

Did the study involve field work? ☐ Yes ☒ No

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study	
☒	☐	Unique biological materials
☒	☐	Antibodies
☒	☐	Eukaryotic cell lines
☒	☐	Palaeontology
☒	☐	Animals and other organisms
☒	☐	Human research participants

Methods

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