The commonness of rarity: Global and future distribution of rarity across land plants

Brian J. Enquist[1,2] Xiao Feng[3], Brad Boyle[1], Brian Maitner[1], Erica A. Newman[1,3], Peter Møller Jørgensen[4], Patrick R. Roehrdanz[5], Barbara M. Theirs[6], Joseph R. Burger[3], Richard Corlett[7], John C. Donoghue[8], Wendy Foden[9], Jon C. Lovett[10], Pablo Marquet[11], Cory Merow[12], Guy Midgley[13], Naia Morueta-Holme[14], Nathan J.B. Kraft[15], Daniel S. Park[16], Robert K. Peet[17], Michiel Pillet[1], Josep M Serra-Diaz[18], Brody Sandel[19], Mark Schildhauer[20], Irena Šímová[21], Cyrille Violle[22], Susan Wiser[23], Lee Hannah[5], Jens-Christian Svenning[24], and Brian J. McGill[25]

[1]Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ 85721, USA; [2]The Santa Fe Institute, USA, 1399 Hyde Park Rd, Santa Fe, NM 87501, USA; [3]Institute of the Environment, University of Arizona, Tucson, AZ USA; [4]Missouri Botanical Garden, St. Louis, MO 63110, USA; [5]The Moore Center for Science, Conservation International, 2011 Crystal Dr., Arlington, VA 22202 [6]The New York Botanical Garden, 2900 Southern Blvd., Bronx, NY, 10348-5126 USA; [7]Centre for Integrative Conservation, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Menglun, Yunnan, China; [8]Center for Macroecology, Evolution, and Climate, Natural History Museum of Denmark, University of Copenhagen, Copenhagen 2100 Denmark; [9]Cape Research Centre, South African National Parks, Tokai, 7947, Cape Town, South Africa;[10]School of Geography, University of Leeds, Leeds, UK; Royal Botanic Gardens, Kew, Richmond, Surrey, UK; [11]The; [12]Department of Ecology and Evolutionary Biology, University of Connecticut, CT, USA; [13]Department of Botany and Zoology, Stellenbosch University, Stellenbosch, South Africa ; [14]The Center for Macroecology, Evolution and Climate; Natural History Museum of Denmark; University of Copenhagen, Universitetsparken 15, Building 3, DK-2100 Copenhagen Ø, Denmark. [15]Department of Ecology and Evolutionary Biology, University of California, Los Angeles, CA 90095, USA; [16]Department of Organismic and Evolutionary Biology, Harvard University, MA, USA; [17]The Department of Biology, University of North Carolina, NC, USA; [18]Université de Lorraine, AgroParisTech, INRA, Silva, 54000 Nancy, France; [19]The Department of Biology, Santa Clara University, Santa Clara, CA 95053; [20]National Center for Ecological Analysis and Synthesis, Santa Barbara, CA, USA; [21]Center for Theoretical Study, Charles University, Czech Republic; [22]Univ Montpellier, CNRS, EPHE, IRD, Univ Paul Valéry Montpellier 3, Montpellier, France; [23]Manaaki Whenua -- Landcare Research, Lincoln, New Zealand; [24]Center for Biodiversity Dynamics in a Changing World (BIOCHANGE), Department of Bioscience, Aarhus University, Ny Munkegade 114, DK-8000 Aarhus C, Denmark; Section for

Ecoinformatics and Biodiversity, Department of Bioscience, Aarhus University, Ny Munkegade 114, DK-8000 Aarhus C, Denmark; [25]School of Biology and Ecology, and Senator George J. Mitchell Center of Sustainability Solutions, University of Maine, Orono, ME

*Corresponding author. Email: benquist@email.arizona.edu

Abstract

A key feature of life's diversity is that some species are common but many more are rare. Nonetheless, at global scales, we do not know what fraction of biodiversity consists of rare species. Here, we present the largest compilation of global plant species observation data in order to quantify the fraction of Earth's extant land plant biodiversity that is common versus rare. Tests of different hypotheses for the origin of species commonness and rarity indicates that sampling biases and prominent models such as niche theory and neutral theory cannot account for the observed prevalence of rare species. Instead, the distribution of commonness is best approximated by heavy-tailed distributions like the Pareto or Poisson-lognormal distributions. As a result, a large fraction, ~36.5% of an estimated ~435k total plant species, are exceedingly rare. We also show that rare species tend to cluster in a small number of 'hotspots' mainly characterized by being in tropical and subtropical mountains and areas that have experienced greater climate stability. Our results indicate that (i) non-neutral processes, likely associated with reduced risk of extinction, have maintained a large fraction of Earth's plant species but that (ii) climate change and human impact appear to now and will disproportionately impact rare species. Together, these results point to a large fraction of Earth's plant species are faced with increased chances of extinction. Our results indicate that global species abundance distributions have important implications for conservation planning in this era of rapid global change.

Introduction

Why some species are common and others are rare has puzzled ecologists (*1*, *2*) at least since Darwin (*3*). Rare species are orders of magnitude more likely to go extinct (*4*, *5*), making it puzzling how so many rare species can be maintained (*6*). Understanding rarity and the maintenance of rare species are also central to conservation biology (e.g. (*7*)) and for understanding current and future changes in biodiversity due to global change (*8*). Despite this importance, at a global scale, it is surprising how little we know about the causes of commonness and rarity, and its maintenance (*9*, *10*).

Traditional quantification of species abundance uses abundance from local communities because measures or estimates of global taxon abundance have been difficult to obtain. However, there are two major limitations to focusing solely on local abundance. First, most species tend to be simultaneously common in a few parts of their ranges and rare in most of their ranges (*11*, *12*) making measures of local abundance a very noisy metric as well as a poor measure of how truly rare a species is globally. Second, at a global scale, a measure of rarity results from a combination of the average local abundance and the number of sites occupied throughout the species geographic range. Measures of local species abundance and species occupancy across the geographic range tend to be correlated (*12–14*). As a result, locally rare species tend to also show up in only a few local communities. This makes it likely that measures of global abundance will be more skewed to the rare, but this has rarely been tested (*15*). Thus, a more global measure of rarity can minimize potential issues associated with characterizing if a species is rare. With the rapid development of biodiversity databases and networks in the past decade, it is becoming increasingly possible to quantify continental and global patterns of biodiversity and test competing models for the origin and maintenance of diversity gradients, and assessing patterns of rarity, at a global scale (*16*).

Here, we utilized a global botanical database of unprecedented coverage to: (i) assess global patterns of plant rarity; (ii) test several proposed hypotheses underlying the generation and persistence of rare species; (iii) identify the regions that harbor 'hotspots' of rare species and explored the drivers such spatial patterns; and (iv) assess how current patterns of human impact and future climate change scenarios may impact plant diversity via impacts on rare species. Quantification of global patterns of abundance and rarity, however, has been hampered by several limitations and prevalent errors in global biodiversity data. These issues make their use in comprehensive biodiversity analyses difficult (*17*, *18*). In this paper we take a novel approach that overcomes many of these limitations. For all known land plants (Embryophyta), we compiled a global database of standardized botanical observation records - the integrated Botanical Information and Ecology Network (Figure 1, BIEN v4.1; http://bien.nceas.ucsb.edu/; see Supplementary

Materials, (*19*)). The BIEN data mainly comprise of herbarium collections, ecological plots and surveys, and trait observations. Together these data constitute over 200 million observations of primary botanical diversity data. Assembling such data involves overcoming numerous challenges of taxonomy, data quality, data exchange, provenance, interoperability, and scaling (Figure 1, (*20*)). After correcting misspelled or synonymous taxon names and removing records with invalid or suspect geocoordinates, incomplete or unresolvable taxon names, and observations of non-native species and cultivated plants, the final dataset consists of 34,902,348 observation records of 434,934 land plant species (See Figure 1 and Supplementary Materials for details of data cleaning and validation).

We quantified the distribution of global abundance for all land plant species (hereafter plant species) on Earth by using a metric of *global relative abundance* - the total number of unique observations of a species ever recorded in global databases. The distribution of the total number of global observations per species (the global species abundance distribution or *g*SAD) is an estimate of global abundance, and is still a sample as a count of all individuals on the planet would be impossible. Nonetheless, quantifying *g*SADs have a substantial practical advantage over other measures of abundance. First, we can combine data from different data sets including plots and surveys, herbarium specimens, and trait observations to increase sampling coverage. These datasets all share the common attribute of observing an individual of a given species in a given location and time. Second, comparing and integrating estimates of *g*SADs from different datasets (plots vs. herbarium specimens etc.) provide a way to assess potential biases in estimating species global abundance. For example, *g*SADs can be measured by compiling just plot or ecological survey data. In plot data, a global estimate of species abundance is quantified directly as each individual of that species within is summed within and across plots. As we discuss, our approach is less biased than data that uses only local plot-based abundance data that samples only a tiny fraction of the Earth's surface.

Traditionally, measures of rarity have been based on a multidimensional concept. For example, Rabinowitz (*21*) identified three major axes on which a species can be common or rare: local abundance, extent of the geographic range, and habitat specificity. Although conceptually these three dimensions are independent, they are often strongly positively correlated (*22*). Four of the five criteria the IUCN uses to evaluate extinction risk for their Red List (*23*) directly involve measurement of rarity via absolute levels of, or declines in abundance and geographic distribution, while the fifth involves computer simulations which are likely to incorporate population size and range size as well. Importantly, these criteria all point to the importance of measuring rarity at global scale (in contrast to local rarity). A species may be globally rare because it has few individuals at many local sites or many individuals at few sites.

Results

We generated two *g*SAD distributions based on summing species individual observations across all ecological plots and by summing all observations across all other botanical observation records. Our analyses reveal that a large fraction of the plants on earth are rare (Figs. 2 & 3). Analyzing the distribution of the number of observations per species reveals that the global-scale distribution is highly skewed and lacking of central tendency (i.e. the mode of the global species abundance distribution *g*SAD is at *N*=1; Fig. 2). Although we estimate that the total number of land plant species on the planet is ~435k, a large fraction of these species, 36.5% or 158,535 are quite rare (species that are represented by only 5 observations or fewer) and 28.3% or 123,149 and have 3 observations or less. The large number of rare species is consistent with past claims that when biodiversity observations are compiled at increasingly larger spatial (*15*) and temporal scales (*24*) rare species should be increasingly comprise the majority of species.

Global species abundance distribution (gSADs) - We tested several long-standing hypotheses concerning the processes creating and maintaining large-scale patterns of commonness and rarity. Specifically, we assessed whether the number of observed rare species follows predictions from biodiversity theory by comparing several proposed statistical distributions for the *g*SAD. First, we assessed two contrasting sets of predictions for the distribution of commonness and rarity of species (Fig. 2). Specifically, at increasingly larger geographic scales both the Unified Neutral Theory of Biogeography (UNTB) (*25*) and ecological niche theory (ENT) (*26*) predict that the global SAD (*g*SAD) will converge on Fisher's log-series distribution (*27*)

$$\hat{f} = \frac{\alpha x^n}{n} \tag{1}$$

where $\hat{f}$ is the expected number of species, n is the total number of observations per species, α is the diversity parameter, and x is a nuisance parameter that is defined by α and total number of individuals, N, sampled, $x = N/(N - \alpha)$. The UNTB further makes two predictions: (i) at increasingly large spatial scales (such as continental and global scales) the Fisher's log-series distribution will also increasingly converge to approximate a 'power law' (or a Pareto distribution) over most of the range of the distribution (*28*) (see Fig. 2A), where

$$\hat{f} = \frac{\beta - 1}{n_0}\left(\frac{n}{n_0}\right)^{\beta} \tag{2}$$

where (ii) the value of β , the scaling exponent or slope on a log-log plot, will equal -1.0.

For the continuous Pareto or power-law distribution, n_0 is the minimum scale of the distribution, and β is the scaling exponent (*29*). For the BIEN data, the minimum number of observations for a species is 1, so n_0 was set at 1.

The UNTB predicts that the *g*SAD (called the regional pool in neutral theory) will follow a log-series distribution. Pueyo (*28*) notes that the log-series distribution consists of two parts multiplied together: a Pareto distribution with exponent β=1 that is the result of neutral dynamics and an exponential "bend" that takes effect at very high abundances, which is due to the finite size assumption. Pueyo (*28*) also suggests a generalization of the Pareto and log-series that incorporates a Pareto where the exponent β is allowed to vary combined with an exponential finite size term which we call here "Pareto with exponential finite adjustment". Thus, testing whether the *g*SAD is best fit by a log-series (where β=1), or a Pareto distribution (where β is allowed to vary), or a 'Pareto with exponential finite adjustment' (where β is also allowed to vary) provides a test of neutral dynamics. In sum, both the UNTB and ETN predict that the log-series distribution will best fit gSADs but, at large geographic scales, this distribution will also converge to a Pareto distribution. Thus, fitting the Pareto or the "Pareto with exponential finite adjustment" provides a simultaneous test of whether neutral or niche dynamics underlie the data (*28*). A poor fit or a value of β not equal to one rejects neutral theory. A poor fit of the Pareto regardless of the value of beta further rejects niche theory. (*28*). Additionally, the value of β is then a useful ecological and evolutionary indicator, of whether the Earth has more rare species ($\beta < -1$, the slope of the function is steeper) or fewer ($\beta > -1$, the slope of the function is flatter) rare species than expected under zero-sum neutral evolutionary dynamics (*28*, *30*).

In contrast to the predictions from the UNTB and ENT, ecological sampling theory predicts that *g*SADs will be characterized by a lognormal distribution. If a species abundance is the result of several processes acting together (*31*), lognormal distributions are expected. Because of the central limit theorem (CLT), a lognormal distribution is expected any time many variables interact multiplicatively to influence abundance, such as many differing biotic and abiotic factors (see Refs in (*32*)). Indeed, common processes in ecology and evolution combine are known to interact multiplicatively to influence species abundance (*32*) (see Supplemental Document) .

Next, we fit several additional models and statistical distributions that have been proposed to describe the distribution of commonness and rarity (see (*33*, *34*) and Supplementary Materials). Using maximum-likelihood estimations, we fit each

distribution to the empirical *g*SAD both for all of the species observation records within the BIEN database as well as for just species recorded across all ecological plots. The best model varied with which measure of goodness of fit was used as well as the data set used. However, in general, the truncated Pareto (i.e. a modified Pareto distribution that adds an additional parameter to allow the right tail to drop down due to finite sample size (*28*)) and the Poisson-lognormal (*35*) both fit well. These models possess strong skew on a log scale, indicative of many rare species. Indeed all three models (at the estimated parameter values) show the mode at species with one individual. The log-series, while also showing a mode at one individual in a log plot, drastically underestimates the number of extremely rare species, and the remaining models fit the distribution even worse (and have an interior mode, incorrectly predicting the most common abundances will be intermediate).

The UNTB predicts that log-series distribution will be approximated by the fit of the Pareto power distribution, with an exponent, $\beta = -1.0$ (*28*). However, our fit of the log-series distribution shows that it was not the best fit and the fitted scaling exponent is significantly steeper than -1.0 (β MLE = -1.41 for all of the BIEN observations and β MLE = -1.43, for the observations from ecological plot data, Fig. 2). Thus, a Pareto power distribution needs an exponent significantly less than -1 to generate the number of rare species actually observed. Together, these results underscore that, at continental to global scales, only a few species abundance distribution models are capable of producing sufficient numbers of rare species to match the observed data, and that neutral dynamics under the UNTB is not one of them. Interestingly, the observed value of β for Embryophytes is similar to what has been reported for an extensive dataset for other taxa including animals and marine phytoplankton (*28*), suggesting that the shape of the SAD at increasingly larger spatial scales converges to a similar distribution across disparate taxa. In sum, the Poisson lognormal is best fit to both all data and the plots only and the Pareto exponent is markedly steeper than -1.0 (and the Pareto distribution is the 2nd best fit on 2 of the 3 metrics).

Assessment of sampling or taxonomic bias - Next, an open question is whether the observed number of rare species is the result of sampling or taxonomic bias. Data from herbarium records are known to exhibit biases in collection and sampling (*17*, *18*). However, do these biases influence our identification of whether a species is rare or not?

We conducted two tests:

First, we compared the distribution of global abundance in the total BIEN database (including plot surveys and herbarium and museum records) with the *g*SAD for just the plot datasets (Fig. 2). Ecological plots and surveys, in contrast to herbarium data, contain minimal sampling biases as a robust effort is made to ensure all individuals within the sampling design are surveyed within a given area. In many cases repeated visits ensure accurate identification to species. Thus, assessing if the *g*SAD from plot data is different from the *g*SAD from all botanical observations enables us to assess potential bias and sampling effectiveness. As discussed below, both empirical global species abundance distributions are described by similar statistical distributions (e.g. the shape of Fig. 2B is similar to Fig. 2A).

Next to further assess if rare species are truly rare or artifactual, we randomly sampled 300 rare species with 3 observations or fewer from the Americas. The Americas were chosen because our taxonomic expertise was more focused between these two continents. For each species selected, we consulted taxonomic experts at the Missouri Botanical Garden and the New York Botanical Garden to sort each species into several classifications (Fig. 3; see Supplementary Materials). Taxonomic experts largely confirm that the rare species identified by BIEN are indeed rare taxa, while only 7.3% were clearly erroneous and recognized as abundant or large-ranged species. We concluded that our results are not driven by taxonomic and sampling biases associated with herbarium data.

Our results from Figure 2 allow us to estimate the total number of native land plant species across the globe with estimates for taxonomic uncertainty. After correcting and standardizing data, we estimate that the total number of extant Embryophyte (land plant) species on earth is between ~358k to ~435k. The lower limit stems from subtracting 17.6% from the total (10.3% from the remaining presence of naturalized non-native species + 7.5% due to the over inflation of names due to 'old names'(basionyms) not yet corrected for by taxonomic data cleaning see Fig. 3). Our estimates are consistent with previous estimates of the total number of Embryophytes in the world of approximately 391,000 (see (*36*) and Supplementary Materials). However, now we can quantify that ~36% of these species are strikingly rare with very little information. In sum, our results from Figure 2 show that rarity is commonplace across the land plants. Little botanical information exists across the world's herbaria and ecological collections for between 13.6% to 11.2% (species with 1 observation) or between 30.0% to 36.5% (species with fewer than or equal to 5 total observations) of all vascular plant species.

'Hotspots' of rare species **-** To identify the regions that harbor 'hotspots' of rare species, we mapped the location of rare species across the world (Fig. 4). We controlled for variation in sampling effort by calculating both the Menhinick and Margalef indices (see

Methods). Plotting the sampling corrected number of rare species, reveals several patterns. Rare species cluster: *in the Americas* in (i) mountainous regions (particularly along the thin strip along the western flank of the Andean Mountains, Central America, and the southern Sierra Madre of Mexico); (*ii*) the Guiana shield in northern South America; (*iii*) relatively small climatic regions that are strongly distinct from surrounding areas (the Atlantic Forest, or Mata Atlântica in Brazil, the southern region of the California Floristic Province, and the Caribbean); *in Africa in* (iv) the Fynbos of South Africa; (*v*) mountainous regions of Madagascar; (*vi*) the coastal mountains of Cameroon, (*vii*) the Ethiopian highlands and the Somali peninsula; *in Asia*: (*viii*) south eastern China and the border regions of Myanmar, Laos, and Thailand; (*ix*) Malaysia; (*x*) New Guinea; (*xi*) the mountainous strip from Iran through Turkey and Georgia. In Europe there are several regions of notably high diversity of rarity in and around (*xii*) the Mediteranean including the Pyrenees.

There is a relative dearth of rare species throughout the Amazon basin confirming past claims that the Amazon consists of widespread and relatively abundant species (*37*). The areas identified by our methods show some overlap with areas independently identified as biodiversity 'hotspots'(*14*) (e.g. Mesoamerican highlands, Andes, Southeast Asia, New Guinea; but differ in other areas).

Drivers of the spatial distribution of rarity - To assess the drivers of the spatial distribution of rarity, we conducted ordinary least squares linear regression (OLS) and simultaneously autoregressive models (SAR) to analyze the relationship between rarity index and environmental variables, including present climate, historical climatic velocity or stability of climate, and topology. Our OLS models showed that all the variables (annual mean temperature, annual precipitation, temperature seasonality, precipitation seasonality, temperature velocity, precipitation velocity, elevation, and heterogeneity of elevation) have significant relationships with the Menhinick rarity index, with the largest effect from temperature velocity and heterogeneity of elevation (Table S1-3 and Fig. S1). In comparing the group models (present climate [annual mean temperature, annual precipitation, temperature seasonality, precipitation seasonality], stability of climate [temperature and precipitation velocity], and topology [elevation, and heterogeneity of elevation]), the model with stability of climate outperformed models with current climate and topography, while the full model showed the lowest AIC (Table S1-3). The exhaustively selected model did not include elevation as a predictor, though it had minor differences in model performance compared with the full model (Table S1-3).
A Moran's *I* test showed high spatial autocorrelation in the residuals of the OLS models, while we found no significant spatial autocorrelation in the residuals of the SAR models (Table S1-3). The coefficients of the SAR models were generally similar with those from

OLS models, with the exceptions that signs of annual mean temperature, precipitation seasonality, and precipitation velocity switched from positive to negative in the SAR model. Temperature velocity remains the largest negative effect and heterogeneity of elevation remains the largest positive effects in the SAR models. Models with stability of climate and topography also outperformed the model with current climate, while the full model remains the best performed SAR model. The modeling results based on Menhinick and Margalef rarity index showed comparable results (Tables S1-6 and Figs. S1-2).

To summarize, areas that contain a higher proportion of rare species have had a more stable climate. The best predictor of high numbers of rare species is the historical temperature velocity(see Tables S1-6). Climate velocity describes climate instability with ecologically relevant units (distance/time; see refs and discussion in Supplemental Document). In addition, mountainous areas as measured by the standard deviation of elevation, is also a significant predictor with positive effect (Tables S1-6 and Fig. S1). Adding short-term annual variation (annual seasonality) in temperature and precipitation, and mountainous conditions in addition to climate velocity does improve the explanation of the current spatial distribution of rarity (e.g. R^2 increased to 0.193 for the OLS model and to 0.518 for the SAR model of Menhinick rarity index; but less so for Margalef rarity index, 0.176 for the OLS model and 0.263 for the SAR model; Table S1-6). Together, these results are consistent with previous results (see Refs (*38*, *39*) and references therein), indicating that increased rates of climate change velocity negatively impacts the retention of endemic and rare species.

The overlaps between future climate velocity and human footprint and rarity indices- Our environment is facing rapid human changes at the global scale, thus we quantified the intensity of human impact on the area with rare species (*40*). Regions with rare species are currently characterized by higher human impact and will experience faster rates of future climate change (Fig. 5). Areas with rare species have human footprint values of 8.5 ± 5.8 which is ~1.6 times higher (Wilcoxon test, $p < 0.001$) than that of the globe on average (5.2 ± 5.8). Further, on average, areas with rare species are predicted to experience ~ 200 (± 58) times greater rates of temperature velocity than those same areas experienced historically and will experience ~1.2 times greater (Wilcoxon test, $p < 0.001$) rates of temperature velocity than the globe will experience on average (170 ± 77).

Predicted changes of rarity indices - With the previously calibrated OLS and SAR full models, we made predictions of rarity indices under future projected climate. The predictions showed worldwide decreases in rarity indices (Fig. 6), where the southern Andes and Southeast Asia are predicted to experience the highest decreases in rarity indices. These decreases were likely due to the accelerated future climate velocities to the

coming century under RCP8.5, which are two orders of magnitude higher than those experienced from Last Glacial Maximum (~21 Ka; LGM) to present day.

Discussion

Primary biodiversity data can be subject to various forms of collection bias, and large amounts of these data remain immobilized (*17*, *18*). Thus it is possible that that the patterns we observe may change with additional data. However, our dataset represents the most comprehensive assembly of plant diversity data, comprising both plots and museum specimen data globally from far more sources than ever before. Comparison between plot and all observation *g*SADs (Fig 2; Table 1) indicates that rare species Further, we made significant efforts in data cleaning and curation to assure that our analyses represent by far the best window yet into the notions of global commonness and rarity in plants.

Our results indicate that 'hotspots' of plant biodiversity largely reflect the accumulation of very rare species. Assessing the predictions of the Unified Theory of Neutral Biogeography (UTNB) to the distribution of commonness and rarity across species enables us to reveal likely drivers of rarity. The UTNB assumes that species overlap in their niches and are equivalent in their rates of speciation, extinction and dispersal (*25*). It implies that biodiversity arises at random, as each species follows a random walk so that the distribution of abundances across species is given by a dynamic equilibrium of speciation and extinction. This equilibrium is predicted to result in a skewed global species abundance distribution (*g*SAD) best fit by a log-series distribution and approximated with a slope $\beta = -1.0$ (Fig. 2A). However, if, at global scales, rates of speciation and extinction are not at equilibrium then $\beta \neq 1$. For example, if rarity is a result of poor adaptation to local conditions or if rare species are maintained by lower rates of extinction then we would expect $\beta > -1$ and $\beta < -1$, respectively (see Ref (*28*). Our results show that $\beta \approx -1.4$ indicating that the proportion of plant species that are rare is more than expected from neutral processes. Given that rare species are orders of magnitude more likely to go extinct (*4*, *5*) than more abundant species begs the question: why do we observe a larger proportion of observed rare species than expected from neutral theory?

Together, our analyses (Tables S1-6; Fig S1-S3) suggest that current 'hotspots' of rare species (Fig. 4) reflect in part a lower risk of historical extinction because they are often found in geographic localities that have had less variable and more stable climates that has likely lowered the probability of extinction (see (*4*, *5*)).

Although the power or Pareto distributions fit well (especially with an exponential correction for finite samples), the Poisson-lognormal fits better supporting the predictions of the CLT. In the end, it is not surprising that the abundance of plant species across the globe is not determined abundance by the predictions of ENT where niche partitioning drives abundance (they are too far apart spatially to compete). Similarly, at large scales environmental gradients become so strong it is hard to imagine that neutrality and the predictions of UNTB applying across such scales. The Poisson-lognormal models integer (such as count of observation) samples from a lognormal distribution. The lognormal distribution derives from the same central limit theorem (CLT) argument that explains the normal distribution except it applies to random factors that are multiplied together or variables that are on a logarithmic scale (such as abundance; see the Supplemental Document). Biologically this can be interpreted as abundance being a consequence of multiple (~5+) different processes (*34*). This statistical, complexity-oriented explanation often fits SADs better than either niche or neutral processes, especially in large scale data (*35*) . Abundance is such a central feature of ecology it is not surprising that many factors simultaneously work to determine relative abundance.

The distribution of the observed species abundance distribution (or SAD) for all botanical observations as well as for species only observed in ecological plots, is heavily right-skewed on both an arithmetic and a logarithmic scale and is best described by a truncated power-law and/or Poisson-lognormal. While past macroecological research has indicated that many species are likely rare (*14*) our results quantify the proportions of commonness vs. rarity. Furthermore, after controlling for number of species, rare species are spatially clumped in ways that support the mechanisms generating and maintaining rare species articulated by Simpson and Janzen. They speculated on the role of mountains and climate stability in influencing both rates of speciation as well as limiting dispersal. In 1964, G.G. Simpson (*41*) extended evolutionary theory to hypothesize that "*Small population ranges and numerous barriers against the spread and sympatry of related populations would therefore tend to increase density of species*." Janzen's 1967 (*42*) 'Why mountain passes are higher in the tropics' extended Simpson's hypothesis to predict that mountainous regions in the tropics will harbor proportionally more rare species than temperate mountains or even topographically uniform tropical regions because of less variability in climate. Our findings of disproportionate numbers of rare species in mainly tropical mountains and small climatic regions support these ideas.

Our results present important implications for conservation in the light of climate change and human impact. If ~36% of species are indeed rare, then our ability to forecast and anticipate how the majority of plant species will respond to climate change is severely limited. Methods to forecast species responses to global change, such as the use of

ecological niche models or species distribution models often require sample sizes on the order of 20 or more observations to generate accurate models (*43*). Currently, we estimate that only a quarter of all land plant species (26% or 112,699 species) have sufficient sample sizes to assess their geographic distribution. Thus, our knowledge of plant species ecology, natural history, and conservation knowledge disproportionately stems from just a quarter of all plant species (*44*). Though not all primary biodiversity data have been digitized, it is still remarkable that ~36% of all plant species known are only documented a small number of times. In addition, our analyses show that rapid rates of current human impact and projected future climate change appear to disproportionately impact regions that harbour most of these rare species (Fig. 5), whereas the rare species likely have been in relatively stable climate through their evolutionary history.

Current human impact and climate change are currently (and will) disproportionately impact rare plant species (Fig. 5). Indeed, our results show a marked contrast between the historical stable climate characteristic of current rarity 'hotspots' (Fig. 4; Tables S1-6; Fig S1-S3) and the predicted two orders of magnitude increase in climate velocity in the areas enriched with rare species (Fig. 5). As shown in maps of rarity indices and regression models, areas with rare species tend to occur in regions that have had comparatively stable climates (i.e. low decadal or century-scale velocity of climate change), which has been noted by others (38, 40). The contrast indicates that future rates of climate change is likely to disproportionately impact diversity by impacting species most prone to extinction— the rare species. Furthermore, this predicted greater exposure to climate change is much greater (2 orders of magnitude) that what those rare species have experienced through their recent evolutionary history.

Materials and Methods

Because the sampling intensity for plants across the Americas is not uniform, we assessed the rarified species diversity. For each 1° grid cell, we calculated the total number of observations or samples, *N* as well as the total number of observed rare species, *S;* rare species were defined as having 3 observation records or fewer. by calculating two separate rarified diversity measures for each 1° grid cell:

(*i*) Margalef diversity ($S_{Margalef}$) – which assumes that species richness increases with sampling intensity *N*, and in particular, increases non-linearly and approximately logarithmically with *N*.

$$S_{Margalef} = (S - 1)/\ln N \qquad (3)$$

(*ii*) Menhinick diversity ($S_{Menhinick}$) – In a similar vein, the Menhinick diversity measure assumes that species richness also increases nonlinearly with sampling intensity, *N*, but according to a square root function

$$S_{Menhinick} = S/\sqrt{N} \quad (4)$$

Comparing both measures of $S_{Margalef}$ and $S_{Menhinick}$ reveals similar spatial maps indicating that both measures result in identical conclusions.

***Competing different hypothesized gSADs*—** As we describe in Supplementary Materials, we fit several additional hypothesized univariate distributions to the global species abundance distribution (*g*SAD) using the following proposed biological and statistical distributions. Most theories produce SADs that are so similar to each other it is difficult to distinguish them given the noisy data and the fact that the differences are most pronounced in the tails which are by definition infrequently observed (*34*). In Table 1 we provide several different goodness of fit measures. Each emphasizes different aspects of fit (chi-square on log-binned data emphasizes the fit of each statistical distribution to rare species, calculating an r^2 on the predicted vs. empirical cumulative distribution function), CDF (describes the probability that a random variable, X, drawn from $f(x)$ is$\leq x$) emphasizes the abundances with the most species—usually intermediate abundances, while likelihood emphasizes avoidance of extreme outliers. As previously noted, it is common for different measures of fit to select different SAD theories as providing the best fit to a single data set (*32*). As a result, any claim of a superior fit must be robust by being superior on multiple measures.

***Methods for regression models*—** As described in the Supplementary Materials, we conducted ordinary least squares (OLS) linear regression models to analyze the relationship between environmental variables and rarity index. We included three groups of environmental variables that portray present climate (annual mean temperature, annual precipitation, temperature seasonality, precipitation seasonality), stability of climate (temperature velocity and precipitation velocity), and topology (elevation and heterogeneity of elevation).We also calculated the standardized deviation of elevations within each one by one degree window and considered this as the heterogeneity of elevation. We performed log-transformation of rarity index, temperature and precipitation velocity, elevation, and heterogeneity of elevation to get normally distributed residuals in the regression models. We standardized all variables to zero mean and one standard deviation to make the regression coefficients comparable. With 4,571 records, we conducted OLS linear regression models to explore the bivariate relationship between rarity index and each environmental variable.

We also constructed multiple regression models using each group of variables (present climate, stability of climate, and topology), as well as using all variables (full model). We conducted multiple regression models through exhaustive model selection based on AIC values using all environmental predictors. Lasty, to account for spatial autocorrelation in climate data, we performed Moran's *I* test and performed simultaneous autoregressive models for all the OLS models mentioned above.

Climate change and future predicted changes in rarity indices – With the previously calibrated full models (OLS and SAR models), we made predictions of rarity indices under future projected climate. We used the full models as they outperformed individual models or sub-group models, and had comparable performances with the exhaustively selected model. We obtained future climatic variables from WorldClim (https://www.worldclim.org/paleo-climate1) (*45*). We used the future climate in 2070 constructed by Community Climate System Model (CCSM4) under RCP 8.5 scenario, which has comparatively high greenhouse gas emissions (*46*). To match the resolution of the rarity map, we sampled the environmental variables (annual temperature, annual precipitation, temperature seasonality, and precipitation seasonality) to one degree cells. We further calculated the temperature and precipitation velocity between present and future following (*38*). The two topological variables (elevation and heterogeneity of elevation)were kept the same as present. After making the predictions, we compared the differences between predicted rarity indices under present and future climate.

Rarity & climate velocity – Using data sources and methods described above in regression model methods, we derived velocity of temperature change and velocity of precipitation change over the periods: Last Glacial Maximum (LGM) to baseline climate (~21 Ka to 1960-1990) and baseline climate to late century (1960-1990 to 2060-2080) under RCP8.5. Velocity was calculated using the neighborhood statistic approach originally described in Loarie *et al*. 2009 (*47*). We compared velocity values at locations where 1) there are rare species observations; 2) there are no rare-species observations; and to 3) background sampled locations. This comparison was conducted for both historical change since LGM and projected future change.

Rarity & the human footprint – Our environment is facing rapid human changes at the global scale, thus we quantified the intensity of human impact on the area with rare species. We downloaded global human footprint data (*40*) and resampled to the resolution of rarity map. We extracted the values of human footprint where rare species exist (i.e. one degree by one degree spatial windows where one or more rare species are observed), and compared the mean of those values with that of the global human footprint map using Wilcoxon test.

Supplementary Materials and Figures (see here)

Supplementary Tables (see here)

Acknowledgements
This work was conducted as a part of the Botanical Information and Ecology Network (BIEN) Working Group, 2008-2012. PIs B.J.E., R.C., B.B., S.D., R.K.P. were supported by the National Center for Ecological Analysis and Synthesis, a center funded by NSF (Grant #EF-0553768), the University of California, Santa Barbara, and the State of California. Additional support was also provided for J.C.D., the NCEAS Research Associate in the Group. The BIEN Working Group was also supported by The iPlant Collaborative (NSF #DBI-0735191). We thank all the data contributors, numerous herbaria who have contributed their data to various data compiling organizations (see Supplementary Materials) for the invaluable data and support provided to BIEN. We especially thank the New York Botanical Garden, Missouri Botanical Garden, Utrecht Herbarium, the UNC Herbarium as well as GBIF, REMIB, and SpeciesLink. The staff at iPlant and the Texas Advanced Computing Center at the University of Texas provided critical computational assistance. We thank the more than 50 scientists have participated in our various BIEN working group and sub-group meetings since 2008 including B. Blonder, K. Engemann, E. Fegraus, J. Cavender-Bares, B. Dobrin, K. Gendler, R. Jorgensen, G. Lopez-Gonzalez, L. Zhenyuan, S. McKay, O. Phillips, J. Pickering, N. Swenson, C. Vriesendorp, K. Woods, who participated in a working group meeting and D. Ackerly, E. Garnier, R. Guralnick, W. Jetz, J. Macklin, N. Matasci, S. Ramteke, and A. Zanne participated in sub-group meetings. We thank John J. Wiens and Michael Sanderson for prior comments and encouragement. We also acknowledge the critical support of iPlant from R. Jorgensen, S. Goff, N. Matasci, and R. Walls. Further, the long-term vision, encouragement, and support of F. Davis, S. Hampton, M. Jones, and the ever-helpful staff at NCEAS were critical for the completion of this first stage of the BIEN working group. N.M.-H. was supported by the European Union's Horizon 2020 research and innovation program under the Marie Sklodowska-Curie grant agreement No. 746334, and acknowledges the Danish National Research Foundation for support to the Center for Macroecology, Evolution and Climate (grant no. DNRF96). C.V. was supported by a Marie Curie International Outgoing Fellowship within the 7th European Community Framework Program (DiversiTraits project, no. 221060) by the European Research Council (ERC) Starting Grant Project (Grant ERC-StG-2014-639706-CONSTRAINTS). J.C.S. and B.J.E. acknowledge support from the Center for Informatics Research on Complexity in Ecology (CIRCE), funded by the Aarhus University Research Foundation under the AU Ideas program. J.C.S. also

considers this work a contribution to his VILLUM Investigator project "Biodiversity Dynamics in a Changing World" funded by VILLUM FONDEN (grant 16549). X.F. and E.A.N. were supported by the University of Arizona Bridging Biodiversity and Conservation Science program. C.M. acknowledges funding from NSF Grant DBI-1913673. We acknowledge the herbaria that contributed data to this work: HA, FCO, DUKE, MFU, UNEX, VDB, ASDM, AMD, BPI, BRI, BRM, CLF, CNPO, L, LPB, AD, A, TAES, FEN, FHO, ANSM, ASU, B, BCMEX, RAS, RB, TRH, AAH, ACOR, UI, AK, CAS, ALCB, AKPM, EA, AAU, ALTA, ALU, AMES, AMNH, AMO, CHAPA, GH, ANGU, ANSP, ARAN, AS, CICY, BAI, CIMI, AUT, BA, BAA, BAB, CMMEX, BACP, BAF, BAJ, BAL, COCA, CODAGEM, BARC, BAS, BBS, BC, BCN, BCRU, BEREA, BG, BH, BIO, BISH, SEV, BLA, BM, BOCH, MJG, BOL, CVRD, BOLV, BONN, DAV, BOUM, BR, DES, BREM, BRLU, BSB, BUT, C, DS, CALI, CAN, CANB, CAY, EBUM, CBM, CEN, CEPEC, CESJ, CHR, ENCB, CIIDIR, CINC, CLEMS, F, COA, COAH, FCME, COFC, CP, COL, COLO, CONC, CORD, CPAP, CPUN, CR, CRAI, FURB, CU, G, CRP, CS, CSU, CTES, CTESN, CUZ, DAO, HB, DBN, DLF, DNA, DR, DUSS, E, HUA, EAC, EIF, EIU, GES, GI, GLM, GMNHJ, K, GOET, GUA, EMMA, HUAZ, ERA, ESA, FAA, FAU, FB, UVIC, FI, GZU, H, FLAS, FLOR, HCIB, FR, FTG, FUEL, GB, HNT, GDA, HPL, GENT, HUAA, HUJ, CGE, HAL, HAM, IAC, HAMAB, HAO, HAS, IB, HASU, HBG, IBUG, HBR, HEID, IEB, HIP, IBGE, ICEL, ICN, ILL, SF, HO, HRCB, HRP, HSS, HU, HUAL, HUEFS, HUEM, HUFU, HUSA, HUT, IAA, HXBH, HYO, IAN, ILLS, HAC, IPRN, IMSSM, FCQ, ABH, INEGI, INIF, BAFC, BBB, INPA, IPA, NAS, INB, INM, MW, EAN, IZTA, ISKW, ISC, ISL, GAT, JEPS, IBSC, UCSB, ISTC, ISU, IZAC, JACA, JBAG, JE, SD, JUA, JYV, KIEL, ECON, KSC, TOYA, MPN, USF, TALL, RELC, CATA, AQP, KMN, KMNH, KOELN, KOR, FRU, KPM, KSTC, LAGU, TRTE, KSU, UESC, GRA, IBK, KTU, ACAD, MISSA, KU, PSU, KYO, LA, LOMA, LW, SUU, UNITEC, TASH, NAC, UBC, IEA, GMDRC, LD, M, LE, LEB, LIL, LINN, AV, HUCP, QFA, LISE, MBML, NM, MT, FAUC, MACF, CATIE, LTB, LISI, LISU, MEXU, LL, LOJA, LP, LPAG, MGC, LPD, LPS, IRVC, MICH, JOTR, LSU, LBG, WOLL, LTR, MNHN, CDBI, LYJB, MOL, DBG, AWH, NH, HSC, LMS, MELU, NZFRI, MA, UU, MU, CSUSB, MAF, MAK, MB, KUN, MARY, MASS, MBK, MBM, UCSC, UCS, JBGP, DSM, OBI, BESA, LSUM, FULD, MCNS, ICESI, MEL, MEN, TUB, MERL, CGMS, MFA, FSU, MG, HIB, MIL, DPU, TRT, BABY, ETH, YAMA, SCFS, SACT, ER, JCT, JROH, SBBG, SAV, PDD, MIN, SJSU, MMMN, PAMP, MNHM, OS, SDSU, BOTU, OXF, P, MOR, POM, MPU, MPUC, MSB, MSC, CANU, SFV, RSA, CNS, WIN, MSUN, CIB, MUR, MTMG, VIT, MUB, MVFA, SLPM, MVFQ, PGM, MVJB, MVM, MY, PASA, N, UCMM, HGM, TAM, BOON, UFS, MARS, CMM, NA, NU, UADY, UAMIZ, UC, NE, NHM, NHMC, NHT, UFMA, NLH, UFRJ, UFRN, ULS, UMO, UNL, UNM, US, NMB, NMNL, USP, NMR, NMSU, WIS, NSPM, XAL, NSW, NT, ZMT, BRIT, MO,

NCU, NY, TEX, U, UNCC, NUM, O, CHSC, LINC, CHAS, ODU, CDA, OSA, OSC, OSH, OULU, OWU, PACA, PAR, UPS, PE, PEL, SGO, PEUFR, PFC, PH, PKDC, SI, PLAT, PMA, PORT, PR, QM, PRC, TRA, PRE, PY, QCA, TROM, QCNE, QRS, UH, QUE, R, SAM, RBR, REG, RFA, RIOC, RM, RNG, RYU, S, SALA, SANT, SAPS, SASK, SBT, SEL, SIU, SJRP, SMDB, SMF, SNM, SOM, SP, SRFA, SPF, SPSF, SQF, STL, STU, SVG, TAI, TAIF, TAMU, TAN, TEF, TENN, TEPB, TFC, TI, TKPM, TNS, TO, TU, UAM, UB, UCR, UEC, UFG, UFMT, UFP, UGDA, UJAT, ULM, UME, UNA, UNB, UNR, UNSL, UPCB, UPEI, UPNA, USAS, USJ, USM, USNC, USZ, UT, UTC, UTEP, UWO, V, VAL, VALD, VEN, VMSL, VT, W, WAG, WAT, WII, WELT, WFU, WMNH, WS, WTU, WU, Z, ZSS, ZT, CUVC, LZ, AAS, AFS, BHCB, CHAM, FM, PERTH, SAN.

Table 1. Three different measures of goodness of fit (r^2 or percentage of variance explained in the cumulative distribution function, χ^2 on $\log_2$ bins and AIC) are shown for six different species abundance models (see (*34*)). Distributions were fitted for the number of observations per species across all species found (*i*)within ecological plots only, and (*ii*) across all datasets within the BIEN database. Sampling species found only in plots standardizes for sampling influences as all individuals within ecological plots are sampled and identified to species. Thus, the species abundance distribution from ecological plots is expected to more accurately describe the species abundance distribution. As predicted by the central limit theorem (CLT) the Poisson lognormal distribution provides the best fit to both *g*SADs. Nonetheless, Pareto, and Truncated Pareto also all fit well. The log-series distribution, predicted by ecological niche theory and neutral theory, falls behind these distribution across the different goodness of fit measures.

Model	**Plot Data Only**				**All Data**			
	CDF r^2	**Chi^2 log$_2$**	**AIC**	**ΔAIC**	**CDF r^2**	**Chi^2 log$_2$**	**AIC**	**ΔAIC**
Zipf Mand	0.929	54,188	139,822	25,848	0.447	738,849,47	7,402,206	3309517
Weibull	0.999	1.6E+10	127,111	13,137	0.999	3.01E+10	4,269,287	176,598
Log-series	0.991	1.57E+13	120,082	6,109	0.999	5.08E+13	4,119,057	26,368
Truncated Pareto	0.999	5.69E+13	115,244	1,270	0.999	1.46E+13	4,110,900	18,211
Poisson lognormal	0.999	**490**	**113,974**	**0**	0.999	**2,966**	**4,092,689**	**0**
Pareto with finite sample exponential adjustment	0.999	563	114,096	122	0.998	100,558	4,203,550	110,861

Table 2. Parameter fits for each of the fitted statistical distributions. The estimated slope values, β, of the *g*SAD are given in bold by fits of the Pareto and Truncated Pareto distributions. Note the estimated slope values differ from -1.0 expected from the unified neutral theory of biodiversity. Instead, the observed fitted slope, β, is steeper than expected from neutral theory (with fitted exponents more negative than -1.0). The steeper exponents indicate that of all of the observed plant species on Earth, proportionally more of them are rare species and that there are more rare species than expected by demographic and evolutionary neutral processes. Thus, the processes creating and maintaining rare species on earth generate proportionally more rare species.

Model	**Plot Data**	**All Data**
Zipf-Mand, *b*	13.3	1186.7
Zipf-Mand, *c*	1.4	1.2
Logseries, *c*	0.9	0.9
Pareto fitted exponent, β	**-1.4**	**-1.3**
Weibull scale	18.1	40.6
Weibull shape	0.4	0.5
Poisson Lognormal, *m*	4.07E-08	1.7
Poisson Lognormal, *s*	2.9	2.6
Pareto with exponential finite adjustment(*28*) **fitted Pareto exponent, β**	**-1.3**	**-1.1**
Pareto with exponential finite tial adjustment: exponential parameter, Omega	0.1	0.1

Figure 1 Computational workflow for creating global species abundance distributions (*g*SADs).

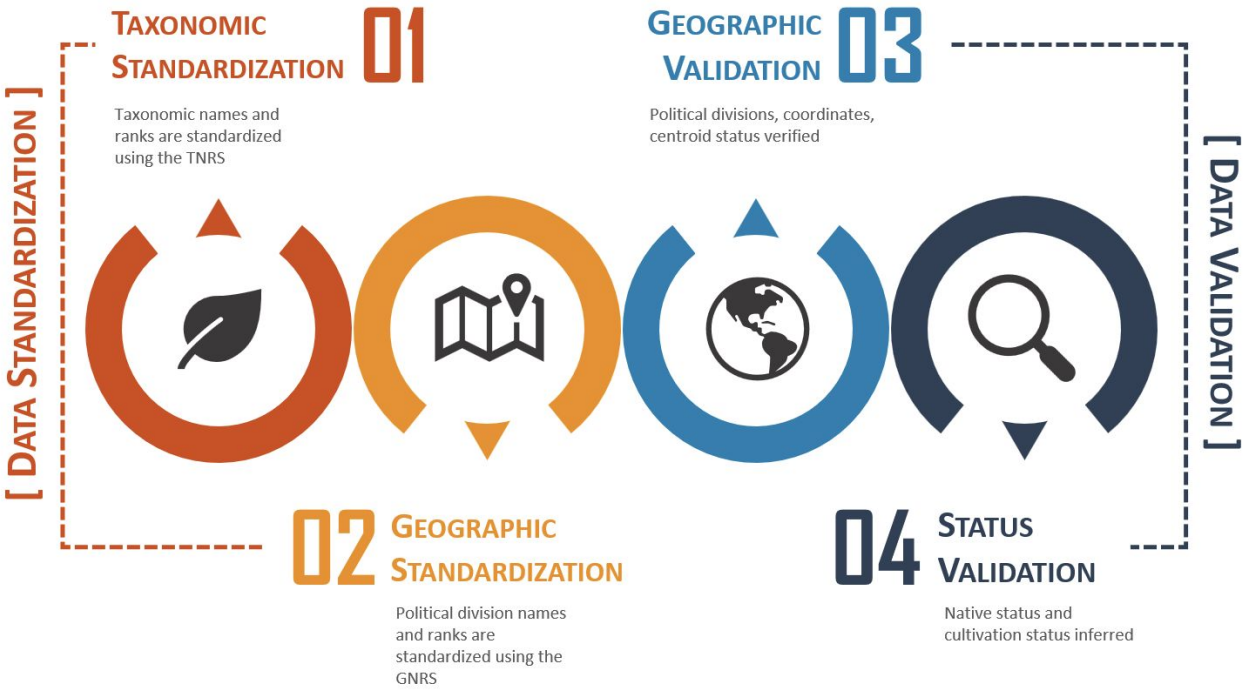


Figure 2. The global species abundance distribution, *g*SAD for all plant species. A. The expected distribution of the number of observations recorded for species, the global species abundance distribution or *g*SAD, based on expectations from theory (*28*) (see main text). In the inset, we list several differing predictions for the shape of the *g*SAD. **B.** Two estimates of the *g*SAD for all land plant species. The first distribution (black) is the observed number observations per species for all species found in ecological plots and community surveys. Each data point represents the total number of individuals observed for a given species. The second distribution (red) is all botanical observations combined (including herbarium specimens) per species. Each distribution is strongly modal at the lowest abundance, showing that most species have only been observed a very small number of times and only a few species are common. The distributions are shown on $\log_{10}$ transformed axes. Comparing the shape of the distributions the competing the fits of differing proposed *g*SAD distributions allows us to test differing hypotheses for the origin of the *g*SAD.

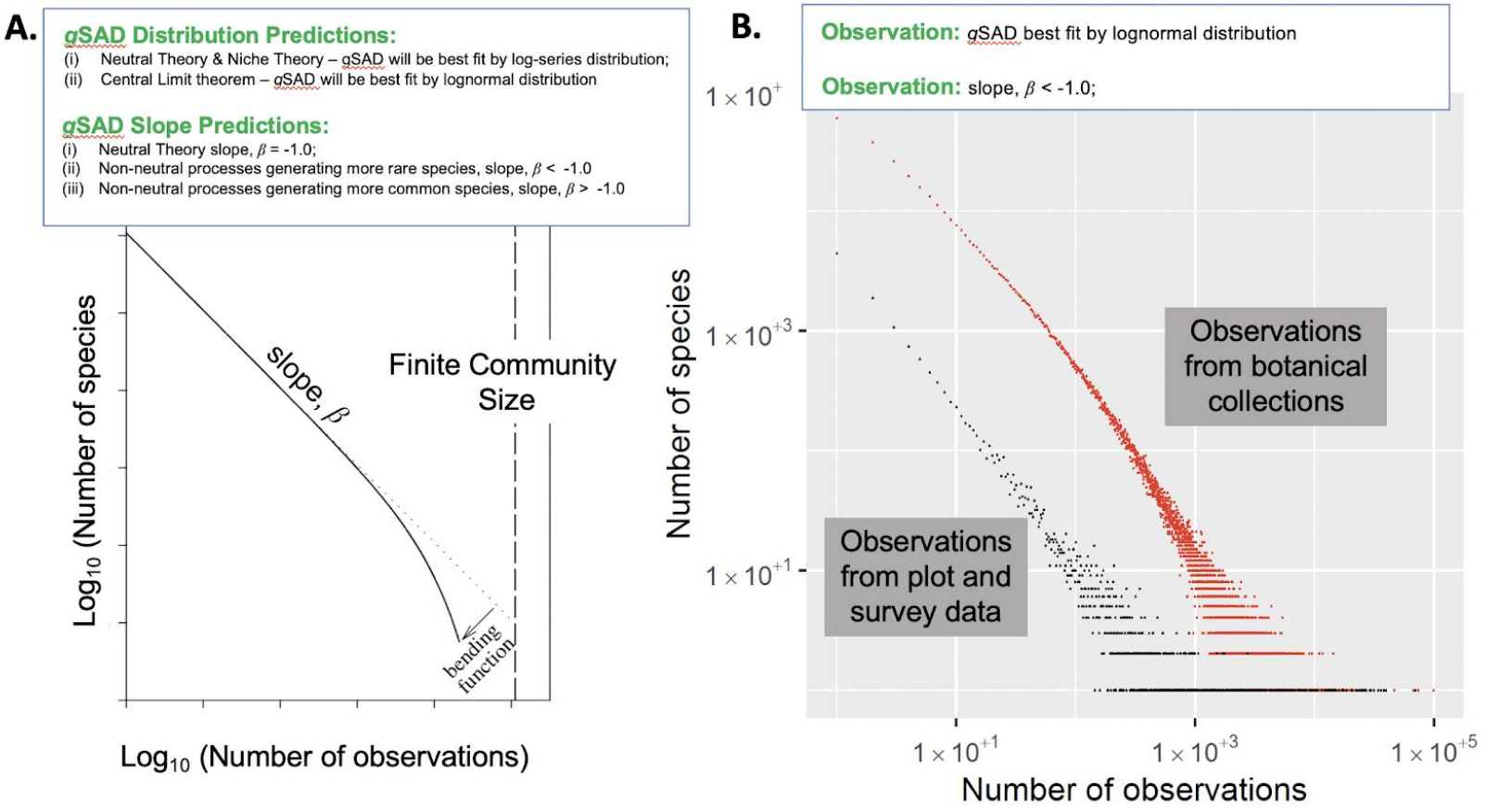


Figure 3. Does using the number of observations in botanical datasets provide a reliable measure of rarity? Assessments of rarity by taxonomic specialists at the Missouri Botanical Garden and the New York Botanical Garden for a random sample of 300 species with 3 observations or fewer in the BIEN database. Most species (72.7%) identified as 'rare' based on the number of unique occurrences within the BIEN database are also recognized as rare by experts. Approximately 7.3% of these species appear to be incorrectly characterized as rare, as they are recognized by experts as abundant or having large ranges. The apparent scarcity of approximately 7.5% of these taxa may reflect recent taxonomic splits or old names no longer used. Moreover, 10.3% are non-native species (which may or may not be rare). In sum, we estimate that between 72 to 90% of plant taxa (recognized as rare + recent name + unresolved + old name) identified by BIEN as being rare would be recognized as rare by other measures.

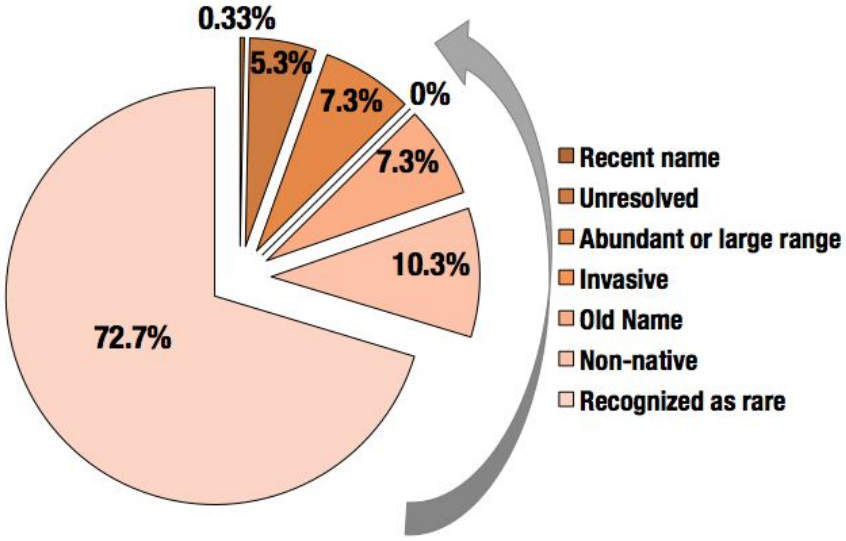


Figure 4. Where are rare species distributed geographically? Plotting the geographic coordinates for all the observations for species with 5 observations or fewer at a coarse, one-degree resolution reveals several patterns. Shown are: the sampling background (gray cells indicate regions with botanical observation records, white cells are areas with no georeferenced botanical sampling records). Colored cells correspond to areas with rare species (species with 5 observations or fewer) rarified to the sampling intensity using the Margalef index (see Supplementary Materials). Areas with a proportionally high number of rare species are red ('hotspots of rarity') while areas with relatively low numbers of rare species are green to blue. Areas with a high number of rare species tend to be clustered in a small number of locations including mountainous tropical and subtropical regions including New Guinea, Indonesia, SE China, Madagascar, The Andes (in Ecuador, Columbia and Peru), Central America (Costa Rica and Panama) and southern Mexico. Also, several notable temperate zone locations including the Fynbos in South Africa and south west Australia, Northern Iran/Georgia/Turkey, and the Iberian peninsula

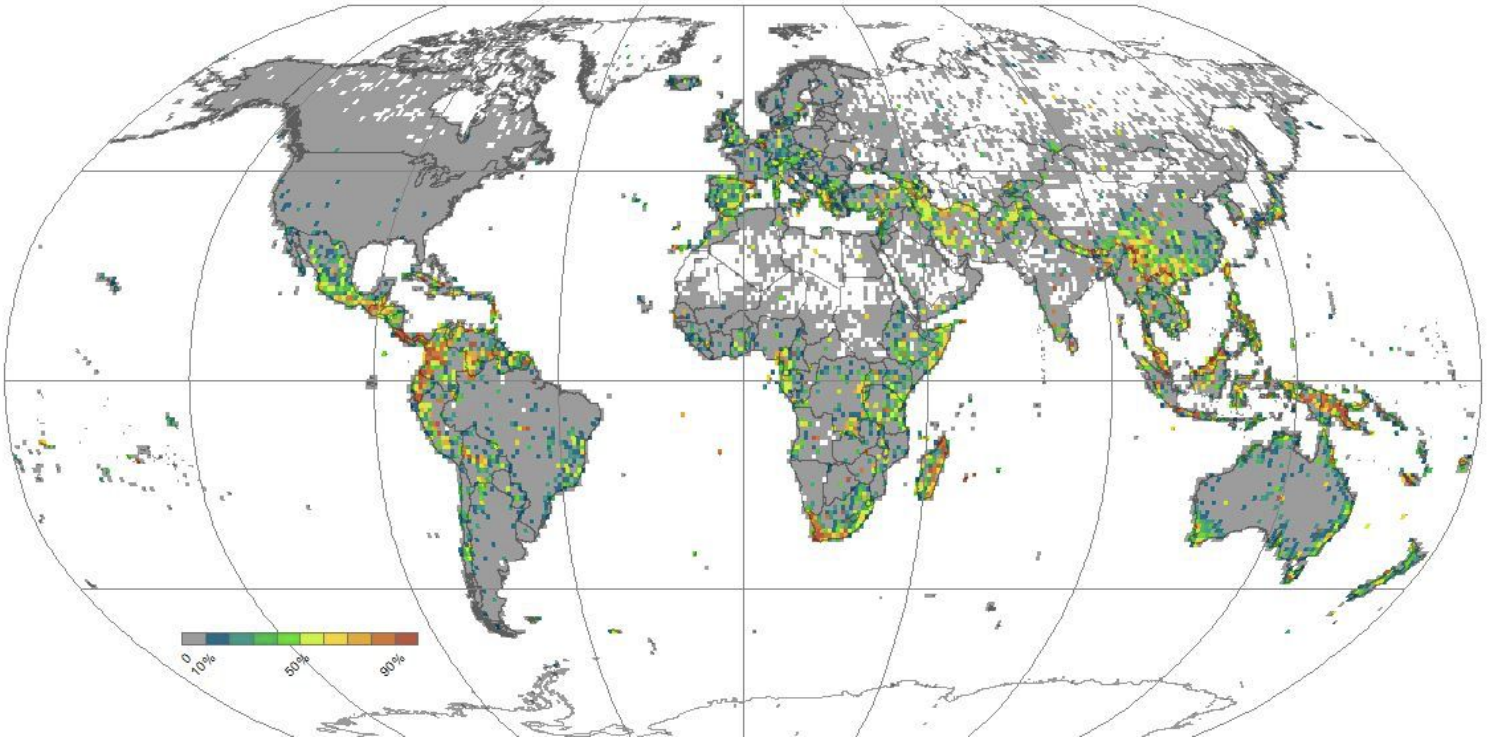

0
10%
50%
90%

Figure 5. Regions that currently have high numbers of rare species are also characterized by higher human impact and will experience faster rates of future climate change. A. Density plot of human footprint index in areas with rare species (light gray) and the global map (dark gray). Areas with rare species have on average human footprint values of 8.5 ± 5.8 which is ~1.6 times higher (Wilcoxon test, $p < 0.001$) the human impact than the globe on average (5.2 ± 5.8). **B**. Density plot of the ratio of future climatic (temperature) velocity vs. historical climatic velocity. On average, areas with rare species will experience ~ 200 (± 58) times greater rates of temperature velocity than those same areas experienced historically and will experience ~1.2 times greater (Wilcoxon test, $p < 0.001$) rates of temperature velocity change than the globe will experience on average (170 ± 77).

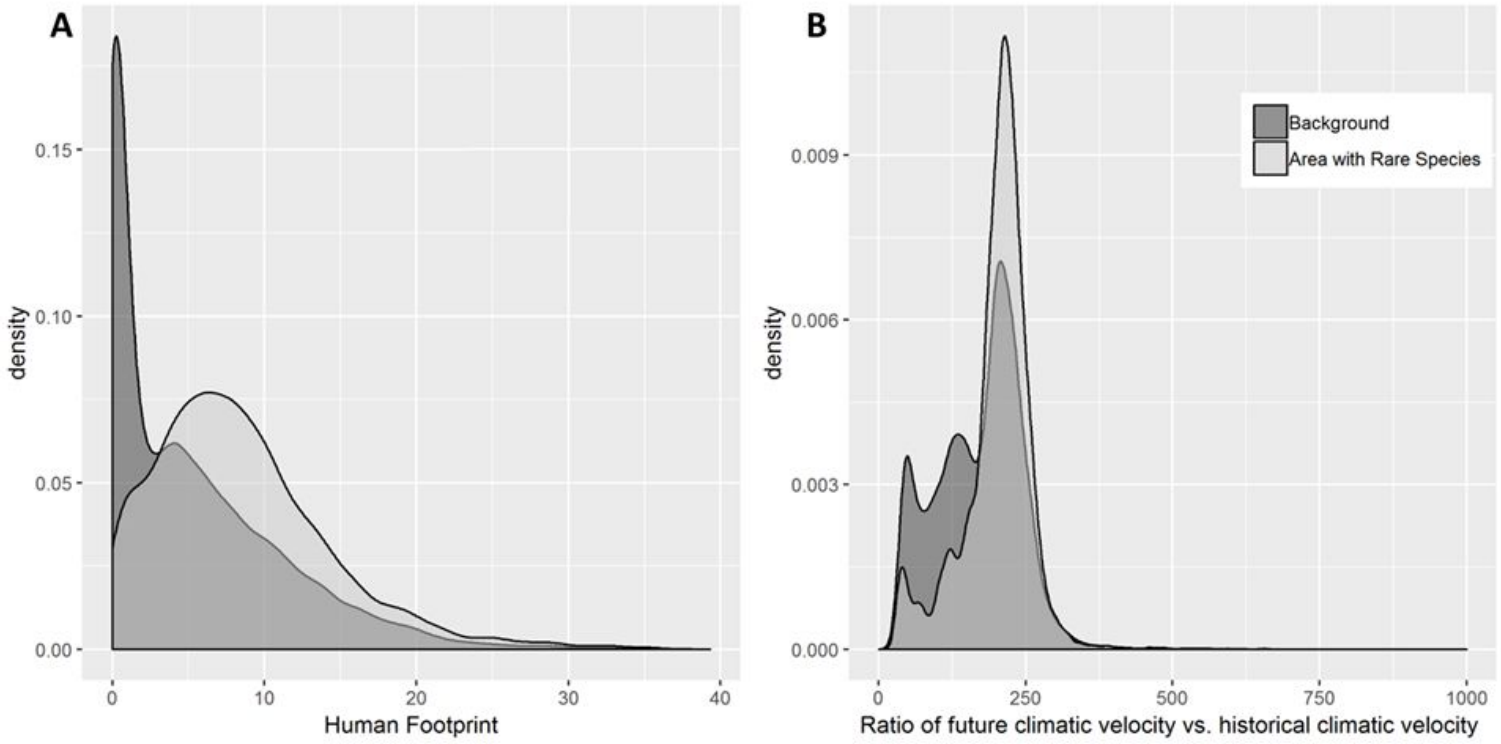
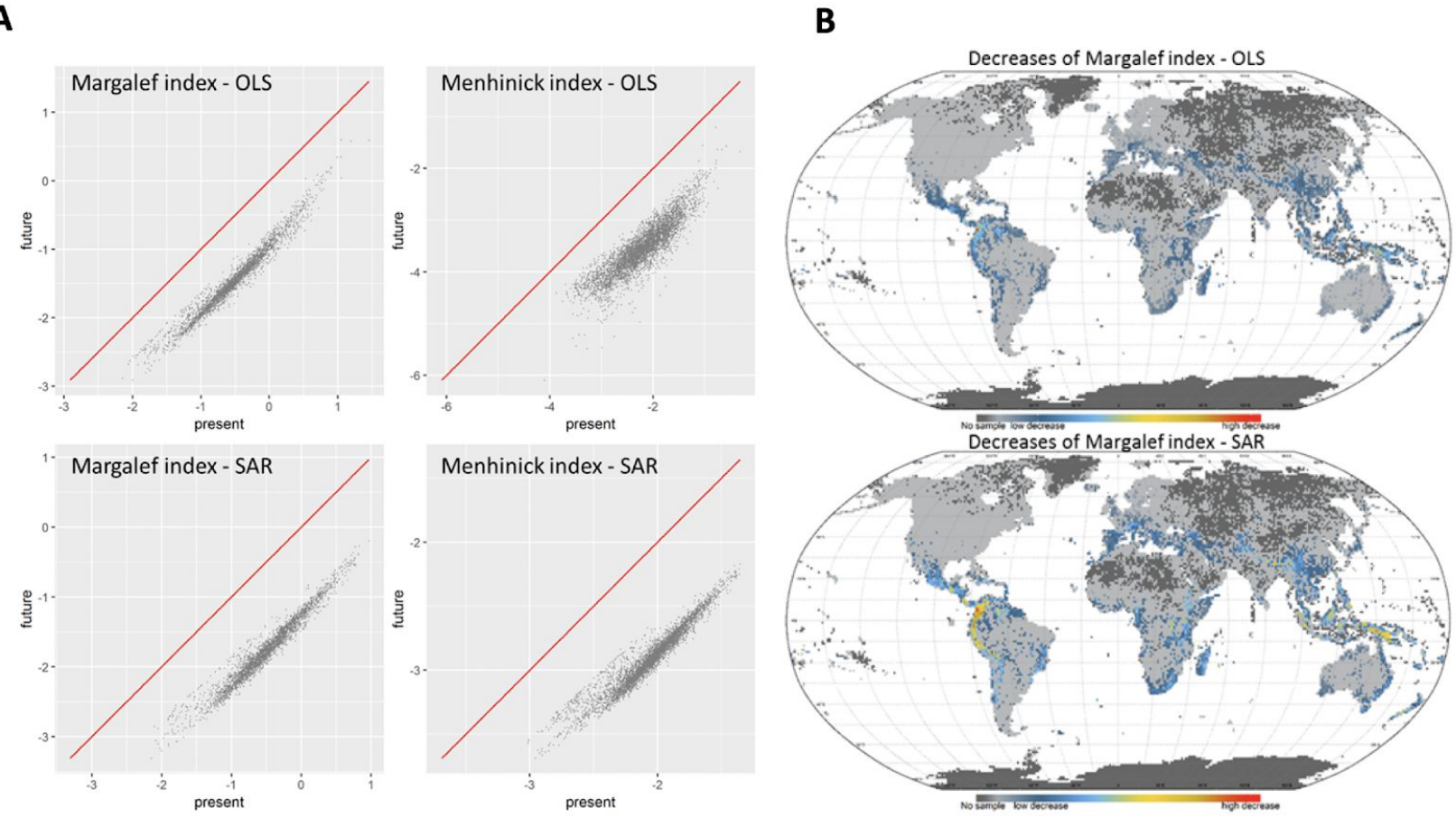


Figure 6. What will happen to rare species diversity with climate change? (A) The predicted Margalef or Menhinick rarity index for area (grid cells) where rare species occur in Fig. 4. Each panel shows the predicted rarity index using either ordinary least squares linear regression (OLS) or simultaneously autoregressive models (SAR) for both present and future climatic conditions. The rarity indices are log transformed. The diagonal 1:1 line (red) represents situations of no difference between the predicted current and future rarity index. All points in the scatter plot are below the diagonal line, indicating a reduction of rare species diversity across all the areas where they currently occur (**B**).



Bibliography

1. F. W. Preston, The canonical distribution of commonness and rarity: Part I. *Ecology*. **43**, 185–215 (1962).

2. I. Hanski, Dynamics of regional distribution: The core and satellite species hypothesis. *Oikos*. **38**, 210–221 (1982).

3. C. Darwin, *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life* (London: John Murray, 1859), *p. 502*.

4. J. M. Diamond, "Normal"extinctions of isolated populations. *Extinctions*, 191–246 (1984).

5. S. L. Pimm, H. L. Jones, J. Diamond, On the risk of extinction. *Am. Nat.* **132**, 757–785 (1988).

6. S. P. Hubbell, R. B. Foster, in *Conservation Biology: The Science of Scarcity and Diversity*, M. E. Soulé, Ed. (Sinauer, 1986), pp. 205–231.

7. M. E. Soulé, *Conservation Biology: The Science of Scarcity and Diversity* (Sinauer Associates, 1986).

8. H. Ter Steege *et al.*, Estimating the global conservation status of more than 15,000 Amazonian tree species. *Sci Adv*. **1**, e1500936 (2015).

9. A. Zizka, H. T. Steege, M. do C. R. Pessoa, A. Antonelli, Finding needles in the haystack: where to look for rare species in the American tropics. *Ecography* . **41**, 321–330 (2018).

10. H. Ter Steege *et al.*, The discovery of the Amazonian tree flora with an updated checklist of all known tree taxa. *Sci. Rep.* **6**, 29549 (2016).

11. J. H. Brown, D. W. Mehlman, G. C. Stevens, Spatial Variation in Abundance. *Ecology*. **76**, 2028–2043 (1995).

12. S. R. Abades, P. A. Marquet, Finite size scaling in the local abundances of geographic populations. *Biol. Res.* **44**, 107–112 (2011).

13. F. He, K. J. Gaston, Occupancy-abundance relationships and sampling scales. *Ecography* . **23**, 503–511 (2000).

14. J. H. Brown, *Macroecology* (University of Chicago Press, 1995).

15. B. J. McGill, Does Mother Nature really prefer rare species or are log-left-skewed SADs a sampling artefact? *Ecol. Lett.* **6**, 766–773 (2003).

16. C. N. Jenkins, S. L. Pimm, L. N. Joppa, Global patterns of terrestrial vertebrate diversity and conservation. *Proc. Natl. Acad. Sci. U. S. A.* **110**, E2602–10 (2013).

17. B. H. Daru *et al.*, Widespread sampling biases in herbaria revealed from large-scale digitization. *New Phytol.* **217**, 939–955 (2018).

18. C. Meyer, P. Weigelt, H. Kreft, Multidimensional biases, gaps and uncertainties in global plant occurrence information. *Ecol. Lett.* **19**, 992–1006 (2016).

19. B. J. Enquist, R. Condit, R. K. Peet, M. Schildhauer, B. M. Thiers, "Cyberinfrastructure for an integrated botanical information network to investigate the ecological impacts of global climate change on plant biodiversity" (PeerJ Preprints, 2016), (available at https://peerj.com/preprints/2615.pdf).

20. C. Thomas, Biodiversity. Biodiversity databases spread, prompting unification call. *Science*. **324**, 1632–1633 (2009).

21. Rabinowitz, in *The Biological Aspects of Rare Plant Conservation*, H. Synge, Ed. (Wiley, 1981), pp. 205–217.

22. J. Yu, F. S. Dobson, Seven forms of rarity in mammals. *J. Biogeogr.* **27**, 131–139 (2000).

23. IUCN, The IUCN Red List of Threatened Species. *The IUCN Red List of Threatened Species* (2019), (available at http://www. iucnredlist.org).

24. A. E. Magurran, P. A. Henderson, Explaining the excess of rare species in natural species abundance distributions. *Nature*. **422**, 714–716 (2003).

25. S. P. Hubbell, *The Unified Neutral Theory of Biodiversity and Biogeography* (2001).

26. R. A. Chisholm, S. W. Pacala, Niche and neutral models predict asymptotically equivalent species abundance distributions in high-diversity ecological communities. *Proceedings of the National Academy of Sciences*. **107**, 15821–15825 (2010).

27. R. A. Fisher, A. S. Corbet, C. B. Williams, The Relation Between the Number of Species and the Number of Individuals in a Random Sample of an Animal Population. *Journal of Animal Ecology*. **12**, 42–58 (1943).

28. S. Pueyo, Diversity: between neutrality and structure. *Oikos*. **112**, 392–405 (2006).

29. E. P. White, B. J. Enquist, J. L. Green, On estimating the exponent of power-law frequency

distributions. *Ecology* (2008) (available at https://onlinelibrary.wiley.com/doi/abs/10.1890/07-1288.1).

30. T. Zillio, R. Condit, The impact of neutrality, niche differentiation and species input on diversity and abundance distributions. *Methods*, 1–10 (2007).

31. J. Aitchison, J. A. C. Brown, The lognormal distribution with special reference to its uses in economics (1957) (available at http://cds.cern.ch/record/109563).

32. B. McGill, Strong and weak tests of macroecological theory. *Oikos*. **102**, 679–685 (2003).

33. F. P. Schoenberg, R. Peng, J. Woods, On the distribution of wildfire sizes. *Environmetrics*. **14**, 583–592 (2003).

34. B. J. McGill *et al.*, Species abundance distributions: moving beyond single prediction theories to integration within an ecological framework. *Ecol. Lett.* **10**, 995–1015 (2007).

35. S. R. Connolly, T. P. Hughes, D. R. Bellwood, R. H. Karlson, Community structure of corals and reef fishes at multiple scales. *Science*. **309**, 1363–1365 (2005).

36. S. L. Pimm, P. H. Raven, The Fate of the World's Plants. *Trends Ecol. Evol.* **32**, 317–320 (2017).

37. N. C. A. Pitman *et al.*, Dominance and Distribution of Tree Species in Upper Amazonian Terra Firme Forests. *Ecology*. **82**, 2101–2117 (2001).

38. B. Sandel *et al.*, The Influence of Late Quaternary Climate-Change Velocity on Species Endemism. *Science*. **334**, 660–664 (2011).

39. B. Sandel, A.-C. Monnet, R. Govaerts, M. Vorontsova, Late Quaternary climate stability and the origins and future of global grass endemism. *Ann. Bot.* **119**, 279–288 (2017).

40. O. Venter *et al.*, Sixteen years of change in the global terrestrial human footprint and implications for biodiversity conservation. *Nat. Commun.* **7**, 12558 (2016).

41. G. G. Simpson, Species density of North American recent mammals. *Syst. Zool.* **13**, 57–73 (1964).

42. D. H. Janzen, Why mountain passes are higher in the tropics. *Am. Nat.* **101**, 233–249 (1967).

43. A. S. J. van Proosdij, M. S. M. Sosef, J. J. Wieringa, N. Raes, Minimum required number of specimen records to develop accurate species distribution models. *Ecography* . **39**, 542–552 (2016).

44. W. Jetz, C. Rahbek, Geographic range size and determinants of avian species richness. *Science*. **297**, 1548–1551 (2002).

45. R. J. Hijmans, S. E. Cameron, J. L. Parra, P. G. Jones, A. Jarvis, Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* **25**, 1965–1978 (2005).

46. K. Riahi *et al.*, RCP 8.5—A scenario of comparatively high greenhouse gas emissions. *Clim. Change*. **109**, 33 (2011).

47. S. R. Loarie *et al.*, The velocity of climate change. *Nature*. **462**, 1052–1055 (2009).

Supplemental document

BIEN Data Workflow

The BIEN database (http://bien.nceas.ucsb.edu/bien/about/) is generated via a linked workflow that imports and integrates heterogeneous data structures (including Darwin Core (*1*) , plus a variety of project-specific schemas and exchange formats), and then performs multiple corrections and validations (Fig. 1). The BIEN workflow is described at http://bien.nceas.ucsb.edu/bien/tools/ and in the following references (*2*–*6*). In addition to correcting erroneous original content and standardizing variant spellings to a single canonical form, corrections also remove or flag erroneous content when the correct meaning cannot be determined). Validations delete erroneous records and add annotations that can be used to filter low-quality data and data useful for some analyses but not others (e.g., observations of introduced or cultivated species).

The two major classes of corrections are taxonomic name resolution and geographic name resolution. Taxonomy is standardized using the Taxonomic Name Resolution Service (TNRS; *2*), which corrects spelling errors in scientific names, standardizes variant spelling and updates synonyms to accepted names. Additional code detects cross-code homonyms (e.g. plant and animal species with identical names) and flags non-plant observations for removal. The names of political divisions (e.g. country, state/ province, county/parish) are standardized using the Geographic Name Resolution Service (GNRS; https://github.com/ojalaquellueva/gnrs) which corrects spelling errors and matches codes (e.g., ISO, FIPS), abbreviations, variant spellings and alternative names in multiple languages to standard political divisions in the GeoNames gazetteer (https://www.geonames.org).

The two major classes of validations are geographic validation and species status validation. Checks performed by geographic validations include (1) coordinate values outside coordinate system (e.g., longitude >180° or <-180°), (2) likely erroneous coordinate values (latitude is exactly 0 or 90 or longitude is exactly 0 or 180, (3) coordinates in the ocean, (4) coordinate matches a centroid (centroid detection) and (5) coordinates outside lowest declared political division (political division validation). Centroid detection and political division validation used administrative boundaries from the Database of Global Administrative Areas (GADM; http://www.gadm.org), with political division names standardized by the GNRS (see above). Species status validations checked for (1) species falling outside their native ranges and (2) observations of human-cultivated plants. Observations species outside of their native range were identified using the Native Species Resolver (NSR; https://github.com/ojalaquellueva/nsr), which uses published country and state checklists to determine if the observed species is native to the lowest declare political divisions. List of endemic taxa are also used to detect non-native occurrences outside the region of endemism. and endemism data. Observations were flagged as cultivated based on (1) keywords in the specimen locality data suggesting provenance from a farm or garden, or (2) geographic proximity (≤3 km) to a botanical garden or herbarium, or (3) original observation metadata indicating a cultivated origin.

For these analyses we excluded records if (1) they lacked a scientific name resolved to at least the species level; (2) they did not come from a land plant (Embryophyta); (3) they failed one or more geographic validations; (4) the species was flagged as potentially non-native to the region of observation; (5) the plant was flagged as potentially cultivated; 5) the observation did not originate from either plot or specimen data.

BIEN Data Sources –

The BIEN data mainly comprise herbarium collections, ecological plots and surveys (*7–13*), and trait observations. For details of specimen data sources see Table S2 of Maitner et al. 2017 (*3*). A full listing of the herbaria data used are given in the Acknowledgements section. The observations in the BIEN database are the product of contributions by 1076 different data contributors, including numerous individual herbaria, data indexers of herbarium or plot data. Of the herbaria, 550+ are listed in Index Herbariorum[1] . Additionally, BIEN 4.1 includes data from RAINBIO[2], TEAM[3], The Royal Botanical Garden of Sydney, Australia[4], and NeoTropTree[5]. Plot data within BIEN are from the CVS, NVS, SALVIAS, VEGBANK, CTFS, FIA, MADIDI, and TEAM data networks and datasets (see http://bien.nceas.ucsb.edu/bien/data-contributors/all/).

A summary of all of the botanical data in the (BIEN 4.1) database include: Total observations: 206,241,288 which stem from 63,498,238 observations from specimen dat, 17,430,379 observations from plot and survey observations. Plot observations originate from 364,477 plots. Multiple observations of the same species from a single plot were counted as a single 'plot' observation. For details of plot data sources see Table S1 of Maitner et al. 2017 (*3*). The final number of the total number of specimens observations used in our analyses after passing through our pipeline was 9,345,197. For each species we counted the total number of occurrences that were recorded in each dataset.

Estimates of the total number of land plant species.

There have been several estimates of the total number of Embryophyte species on Earth. Our estimates come from estimates from the following sources (*14–18*) as well as the most recent estimate from Kew(*19*).

Assessment of the accuracy of our rarity measure –

Potential confounding issues associated with characterizing rarity status - There are several potential issues with using the number of absolute observations as a measure of rarity. For example, small number of observations may reflect collection bias for under sampled species. Indeed, three types of errors could drive the pattern in Fig. 2a. First, sampling biases with botanical data may under-sample rare species so that the number of observations is a poor measure of rarity. Second, the high number of species names with small numbers of observations

[1] http://sweetgum.nybg.org/science/ih/
[2] http://rainbio.cesab.org/
[3] https://www.wildlifeinsights.org/team-network
[4] https://www.rbgsyd.nsw.gov.au/
[5] http://www.neotroptree.info/

may reflect taxonomic biases such as old names no longer in use or more recent taxonomic splits. In which case, the high numbers of rare species may reflect the predominance of old taxonomic names that are no longer in use. Third, using the number of observations as a measure of rarity may impose bias, if low sampled species are indeed common elsewhere and/or have large geographic distributions.

Methods - To assess the accuracy of our classification of rarity we randomly selected 350 taxa that we identified as 'rare' (having 3 unique observations or less). These names were then divvied up between the appropriate taxonomic experts at the Missouri Botanical Garden and The New York Botanical Garden. Experts were asked to classify each name as falling into one of seven classifications: (i) Recognized by taxonomists as rare; (ii) a non-native taxa; (iii) an 'old' taxonomic name that is no longer used or is synonymous with another taxa; (iv) a taxa known to be invasive; (v) as taxa not actually rare but instead abundant or having a large geographic range; (vi) a taxa having a name that is 'unresolved' or with a status that is unclear; (vii) a taxa having a 'recent name' - meaning that it was either recently discovered or recently taxonomically split.

Most species, 72.7%, identified by BIEN as being 'rare', (having 3 unique observations or less), are indeed taxa that are recognized by experts as rare. Only 7.3% of the remaining subsampled taxa appear to be incorrectly characterized as rare but recognized by experts as actually abundant or having large geographic ranges. The large number of rare taxa does not appear to be due to recent taxonomic splits or old names no longer apply, as ~7.5% of the remaining taxa were due to recent taxonomic splits. In total, 10.3% of the remaining species were identified as non-native species, which may indeed be rare in their naturalized range. Thus, we estimate that between 72% and 90% of plant taxa (the later value being equal to the 'recognized as rare' + 'Recent name' + 'Unresolved' + 'Old Name') identified by BIEN as being rare would be recognized as indeed rare species by other metrics.

Model fitting to Species Abundance Distribution -

We tested how well different proposed distributions fit the observed data by fitting several hypothesized distributions and statistical distributions(*20–22*) to the BIEN data. Each univariate distribution was fit using the Palamedes toolbox (*21*) developed in MATLAB which uses maximum-likelihood estimations for each distribution. All univariate distributions were fit to the continuous BIEN species observation data by maximizing the log-likelihood unless otherwise indicated.

We first visualized the fit of several candidate distributions including:

Fisher's log-series (*23*)
$$\hat{f} = \frac{\alpha x^n}{n} \tag{S1}$$

Pareto distribution (*24*)
$$\hat{f} = \frac{\theta - 1}{n_0}\left(\frac{n}{n_0}\right)^{-\theta} \tag{S2}$$

Weibull distribution(*25*) $$\hat{f} = \frac{k}{\lambda}\left(\frac{n-\delta}{\lambda}\right)^{k-1} \exp\left(-\left(\frac{n-\delta}{\lambda}\right)^{k}\right) \quad (S3)$$

where $\hat{f}$ is the expected number of species. For the log-series, n is the total number of observations per species, α is the diversity parameter, and x is a nuisance parameter and is defined by α and total number of individuals, N, sampled, $x = N/(N - \alpha)$. For the Pareto or power-law distribution where n_0 is the minimum scale of the distribution, and θ is the scaling exponent. For the BIEN data, the minimum number of observations for a species is 1, so n_0 was set at 1. For the Weibull distribution, k is the shape parameter, δ is the location parameter of the distribution, λ is the scale parameter.

Next, we fit several additional hypothesized univariate distributions to the species abundance distribution using the following proposed biological and statistical distributions. The following distributions were fit by plotting the species logarithmic abundances in decreasing order, or against ranks of species:

Broken Stick model of MacArthur (*26*) $$\hat{a}_r = \frac{N}{S}\sum_{k=r}^{S}\frac{1}{k} \quad (S4)$$

Niche-preemption or geometric series (*27*) $$\hat{a}_r = N\alpha(1-\alpha)^{r-1} \quad (S5)$$

The Log-normal or Preston distribution (*28*, *29*)

$$\hat{a}_r = \exp\left[\log(\mu) + \log(\sigma)\Phi\right] \quad (S6)$$

The Zipf-Mandelbrot distribution (*30*) $$\hat{a}_r = N\hat{p}_1 r^{\gamma} \quad (S7)$$

Here, $\hat{a}_r$ is the expected abundance of species at rank r, S is the number of species, N is the number of observations, Φ is a standard normal function, $\hat{p}_1$ is the estimated proportion of the most abundant species, and α, μ, σ, and γ are the estimated parameters in each model.

We compete each of these distributions by fitting each distribution to the gSADs. For these distributions, we fit each to the global species abundance distribution fits we followed the methodology of McGill 2011 (*20*).

Additional niche based models - The Broken Stick model was proposed by MacArthur in 1957 (*26*). The model assumes that a given resource in a community is then 'randomly' divided into species niches. Species niches are broken at random and the successive niches are chosen with a probability proportional to their size. This model can lead to a more even distribution, where larger niches are more likely to be broken, facilitating co-existence between species in equivalent sized niches. In contrast, the geometric series model, originally proposed in 1932 by Montomura

(*27*) has been proposed to equate with a model of niche preemption, in which species sequentially colonize a region and the first species to arrive receives the majority of resources (*8*).

The Lognormal Distribution and statistical limit theorems
The central limit theorem (CLT) of statistics predicts that that we should expect to find normal distributions when many variables interact additively. However, within biology many biological processes are multiplicative instead of additive (*31*). Many biological phenomena (e.g. fitness, growth, reproduction, metabolism, sensation) are fundamentally the result of multiplicative processes and likely conform more closely to a geometric error model (*32*, *33*). When many variables interact multiplicatively, we should find a lognormal distribution (*34–36*). We note that lognormal distributions are expected any time many variables interact multiplicatively to influence abundance, such as many differing biotic and abiotic factors (*34–36*), see also (*37*).

Competing model fits - Building on the arguments by McGill et al. (*9*), we next fit and then competed several different proposed candidate species abundance distributions:

- **Geometric distribution**. We utilized the methodology of He & Tang 2008 (*38*). Their regression method was used to fit this distribution to the distribution of commonness across species.
- **Logseries distribution**- We used the standard iterative method (*39*, *40*)**.**
- **Zipf-Mandelbrot distribution** – We used the methods per Izsak 2006 (*41*).
- **Power law or truncated pareto distribution** – We used the maximum likelihood estimate of the discrete, truncated power law following per White, E. P., B. J. Enquist, et al. (2008) (*24*).
- **Powbend distribution** – We used the iterative maximum likelihood methods as per Pueyo, S. (2006) (*42*).
- **Poisson Lognormal distribution** – We used an iterative maximum likelihood approach with integral approximations as per Bulmer, M. G. (1974) (*43*).
- **Lognormal distribution -** Was fit using standard maximum likelihood estimator formulas, e.g. Evans, M., N. Hastings, et al. (1993) (*44*)
- **Gamma distribution -** Was fit using standard maximum likelihood estimator formulas, e.g. Evans, M., N. Hastings, et al. (1993). (*44*)
- **Weibull distribution** – Was fit using standard maximum likelihood estimator formulas, e.g. Evans, M., N. Hastings, et al. (1993). (*44*)

We competed each model we followed the methodology of McGill (2003; (*45*)) by calculating several goodness of fit measures including: from the Cumulative Distribution Function (CDF) *r*2, the Chi-square log2 measure and AICc. For each distribution we calculated the Akaike's Information Criterion (AIC) value for the fitted function (*10*). We utilized the vegan package[6] in R to generate AIC values for the fitted Broken Stick, preemption, Zipf, Fisher's log-series, and log-normal distributions. Because the vegan package does not generate AIC values for the Fisher's log-series, we generated AIC values for the log-series by calculating the log-likelihood and then calculating the AIC values directly where AIC = 2 * No. of Parameters – 2 *

[6] http://vegan.r-forge.r-project.org/

log(Likelihood). To calculate AIC values for the Pareto distribution, we utilized the pareto.R function[7] and the log-likelihood value to calculate the AIC directly. The Wiebull distribution was fit using the fitdistr.R function[8]. Following Burnham and Anderson (*10*), the best-fit model is the one with the lowest AIC value.

Pareto or power-law distribution – We calculated the slope or the exponent of the power-law fit to the data, by calculating the maximum likelihood estimate (MLE) for the exponent (*2, 11*)

$$\hat{\theta} = 1 + n\left[\sum_{i=1}^{n} \ln \frac{n_i}{n_0}\right]^{-1} \quad \text{(S8)}$$

where n_i is the number of observations for a given species, *i*. We utilized the pareto.R function[2] to calculate the MLE and 95% confidence intervals for $\hat{\theta}$.

Sampling and Rarefaction –

For each 1° grid cell, we calculated the total number of samples, *N* as well as the total number of rare species, *S;* rare species were defined as having 3 observation records or less.

Because the sampling intensity for plants across the Americas is not uniform, we assessed the rarified species diversity by calculating two separate rarified diversity measures for each 1° grid cell:

(*i*) Margalef diversity ($S_{Margalef}$) – This measures stems from Margalef (*46*) who noted that species richness increases with *N*, and in particular, increases non-linearly and roughly logarithmically with *N*.

$$S_{Margalef} = (S - 1)/\ln N \quad \text{(S9)}$$

(*ii*) Menhinick diversity ($S_{Menhinick}$) – In a similar vein (*47*), Menhinick proposed adjusting species richness by the similarly shaped square root of *N*.

$$S_{Menhinick} = S/\sqrt{N} \quad \text{(S10)}$$

Comparing both measures of $S_{Margalef}$ and $S_{Menhinick}$ reveals similar spatial maps indicating that both measures result in identical conclusions.

Rarefaction produces subtle changes to the absolute rare species diversity map. For example, in Central America, Costa Rica and Panama each have large numbers of absolute rare species while also having a relatively large number of samples. Thus, rarefaction effectively 'demotes' the large number of rare species throughout Central America due to the heightened sampling intensity there.

[7] http://www.stat.cmu.edu/~cshalizi/uADA/12/pareto.R
[8] http://cran.r-project.org/package=fitdistrplus

Methods for regression models

We conducted ordinary least squares (OLS) linear regression models to analyze the relationship between environmental variables and rarity index. We included three groups of environmental variables that portray present climate (annual mean temperature, annual precipitation, temperature seasonality, precipitation seasonality), stability of climate (temperature velocity and precipitation velocity), and topology (elevation and heterogeneity of elevation), which are known to influence biodiversity patterns (*48–51*). We obtained the four variables for current climate and elevation from WorldClim (version 1.4) (*52*) at10 arc-minute resolution. We also calculated the temperature and precipitation velocity between present climate and Last Glacial Maximum climate [based on Community Climate System Model (CCSM4); https://www.worldclim.org/paleo-climate1] following (*53*). To match the resolution of rarity map, we aggregated the environmental variables and the derived velocity to one degree by one degree.

We performed Moran's *I* test and found the presence of spatial autocorrelation in the dataset analyzed here. Therefore, we performed simultaneous autoregressive models (SAR) (*54*) for all the OLS models mentioned above. We considered three different simultaneous autoregressive model types (lagged-response, lagged-mixed, and spatial error) and five different spatial neighborhood structures (lag distances between 200 and 1,000 km). Our preliminary analyses of Akaike's information criterion (AIC) and Moran's *I* values showed that SAR with spatial error and a lag distance of 600 km accounted best for the spatial structure in the analyzed data set. All analyses were conducted in R 3.5.1 (*55*), using raster (*56*), glmulti (*57*) and spdep (*58*) packages.

Bibliography

1. J. Wieczorek *et al.*, Darwin Core: an evolving community-developed biodiversity data standard. *PLoS One*. **7**, e29715 (2012).

2. B. Boyle *et al.*, The taxonomic name resolution service: an online tool for automated standardization of plant names. *BMC Bioinformatics*. **14**, 16 (2013).

3. B. S. Maitner *et al.*, The bien r package: A tool to access the Botanical Information and Ecology Network (BIEN) database. *Methods Ecol. Evol.*, doi:10.1111/2041-210X.12861.

4. G. R. Goldsmith *et al.*, Plant-O-Matic: a dynamic and mobile guide to all plants of the Americas. *Methods Ecol. Evol.* **7**, 960–965 (2016).

5. I. R. McFadden *et al.*, Temperature shapes opposing latitudinal gradients of plant taxonomic and phylogenetic β diversity. *Ecol. Lett.* (2019), doi:10.1111/ele.13269.

6. B. J. Enquist, R. Condit, R. K. Peet, M. Schildhauer, B. M. Thiers, "Cyberinfrastructure for an integrated botanical information network to investigate the ecological impacts of global climate change on plant biodiversity" (PeerJ Preprints, 2016), (available at https://peerj.com/preprints/2615.pdf).

7. E. Fegraus, Tropical Ecology Assessment and Monitoring Network (TEAM Network). *Biodivers. Ecol.* **4**, 287–287 (2012).

8. R. K. Peet, M. T. Lee, M. D. Jennings, D. Faber-Langendoen, VegBank: a permanent, open-access archive for vegetation plot data. *Biodivers. Ecol.* **4**, 233–241 (2012).

9. S. J. DeWalt, G. Bourdy, L. R. Chavez de Michel, C. Quenevo, Ethnobotany of the Tacana: Quantitative inventories of two permanent plots of Northwestern Bolivia. *Econ. Bot.* **53**, 237–260 (1999).

10. Forest Inventory and Analysis National Program (2013), (available at http://www.fia.fs.fed.us/).

11. S. K. Wiser, P. J. Bellingham, L. E. Burrows, Managing biodiversity information: development of New Zealand's National Vegetation Survey databank. *N. Z. J. Ecol.* **25**, 1–17 (2001).

12. K. J. Anderson-Teixeira *et al.*, CTFS-ForestGEO: a worldwide network monitoring forests in an era of global change. *Glob. Chang. Biol.* **21**, 528–549 (2015).

13. B. Enquist, B. Boyle, SALVIAS -- the SALVIAS vegetation inventory database. *Biodiversity & Ecology*. **4**, 288–288 (2012).

14. CRANDALL-STOTLER, B, Morphology and classification of the Marchantiophyta. *Bryophyte biology*, 1–54 (2008).

15. E. M. Gifford, A. S. Foster, *Morphology and Evolution of Vascular Plants* (W. H. Freeman, 1989;

https://play.google.com/store/books/details?id=Zc98QgAACAAJ).

16. B. Goffinet, Systematics of the Bryophyta (mosses): from molecules to a revised classification. *Molecular systematics of bryophytes. Monographs in Systematic Botany*. **98**, 205–239 (2004).

17. P. H. Raven, University Ray F Evert, University Susan E Eichhorn, R. F. Evert, S. E. Eichhorn, *Biology of Plants* (Macmillan, 2005; https://play.google.com/store/books/details?id=8tz2aB1-jb4C).

18. H. Bischler, R. M. Schuster, The Hepaticae and Anthocerotae of North America. *The Bryologist*. **96** (1993), p. 281.

19. S. L. Pimm, P. H. Raven, The Fate of the World's Plants. *Trends Ecol. Evol.* **32**, 317–320 (2017).

20. B. J. McGill, Species abundance distributions. *Biological Diversity: Frontiers In Measurement & Assessment*, 105–122 (2011).

21. B. J. McGill *et al.*, Species abundance distributions: moving beyond single prediction theories to integration within an ecological framework. *Ecol. Lett.* **10**, 995–1015 (2007).

22. B. A. Maurer, B. J. McGill, Measurement of species diversity. *Biological diversity: frontiers in measurement and assessment*, 55–65 (2011).

23. R. A. Fisher, A. S. Corbet, C. B. Williams, The relation between the number of species and the number of individuals in a random sample of an animal population. *J. Anim. Ecol.*, 42–58 (1943).

24. E. P. White, B. J. Enquist, J. L. Green, On estimating the exponent of power-law frequency distributions. *Ecology*. **89**, 905–912 (2008).

25. M. Evans, N. Hastings, B. Peacock, Mathematical Distributions (1993).

26. R. H. MacArthur, On the relative abundance of bird species. *Proceedings of the National Academy of Sciences*. **43** (1957), pp. 293–295.

27. I. Motomura, *Statistical Treatment of Associations* (1932; https://play.google.com/store/books/details?id=XkOstAEACAAJ).

28. F. W. Preston, The canonical distribution of commonness and rarity: Part I. *Ecology*. **43**, 185–215 (1962).

29. F. W. Preston, The Canonical Distribution of Commonness and Rarity: Part II. *Ecology*. **43**, 410–432 (1962).

30. B. Mandelbrot, Information Theory and Psycholinguistics: A Theory of Word Frequencies, Readings in Mathematical Social Sciences (1966).

31. A. J. Kerkhoff, B. J. Enquist, Multiplicative by nature: why logarithmic transformation is necessary in allometry. *J. Theor. Biol.* (2009) (available at https://arizona.pure.elsevier.com/en/publications/multiplicative-by-nature-why-logarithmic-transformation-is-necess).

32. F. Galton, The geometric mean, in vital and social statistics. *Proc. R. Soc. Lond.* **29**, 365–367 (1879).

33. P. D. Gingerich, Arithmetic or geometric normality of biological variation: an empirical test of theory. *J. Theor. Biol.* **204**, 201–221 (2000).

34. D. McAlister, XIII. The law of the geometric mean. *Proc. R. Soc. Lond.* **29**, 367–376 (1879).

35. R. May, Patterns of species abundance and diversity. *Ecology and Evolution of Communities*, 81–120 (1975).

36. R. H. MacArthur, On the relative abundance of species. *Am. Nat.* **94**, 25–36 (1960).

37. S. R. P. Halloy, A theoretical framework for abundance distributions in complex systems. *Complex Systems' 98*, 118 (1998).

38. F. He, D. Tang, Estimating the niche preemption parameter of the geometric series. *Acta Oecol.* **33**, 105–107 (2008).

39. A. E. Magurran, *Ecological Diversity and Its Measurement* (Princeton University Press, 1988; https://play.google.com/store/books/details?id=CuU9DwAAQBAJ).

40. B. C. McCarthy, A. E. Magurran, Measuring Biological Diversity. *Journal of the Torrey Botanical Society*. **131** (2004), p. 277.

41. J. Izsák, Some practical aspects of fitting and testing the Zipf-Mandelbrot model: A short essay. *Scientometrics*. **67**, 107–120 (2006).

42. S. Pueyo, Diversity: between neutrality and structure. *Oikos*. **112**, 392–405 (2006).

43. M. G. Bulmer, On fitting the Poisson lognormal distribution to species-abundance data. *Biometrics*, 101–110 (1974).

44. E. R. Ziegel, M. Evans, N. Hastings, B. Peacock, Statistical Distributions. *Technometrics*. **36** (1994), p. 229.

45. B. J. McGill, A test of the unified neutral theory of biodiversity. *Nature*. **422**, 881–885 (2003).

46. R. Margaleff, Temporal succession and spatial heterogeneity in phytoplankton. *Perspectives in marine biology*, 323–349 (1958).

47. E. F. Menhinick, A comparison of some species-individuals diversity indices applied to samples of field insects. *Ecology*. **45**, 859–861 (1964).

48. G. Feng *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.* **21**, 1687 (2019).

49. G. Feng, L. Mao, B. Sandel, N. G. Swenson, J.-C. Svenning, High plant endemism in China is partially linked to reduced glacial-interglacial climate change. *Journal of Biogeography*. **43** (2016), pp. 145–154.

50. H. Kreft, W. Jetz, Global patterns and determinants of vascular plant diversity. *Proceedings of the National Academy of Sciences*. **104** (2007), pp. 5925–5930.

51. D. S. Park, O. H. Razafindratsima, Anthropogenic threats can have cascading homogenizing effects on the phylogenetic and functional diversity of tropical ecosystems. *Ecography* . **42**, 148–161 (2019).

52. R. J. Hijmans, S. E. Cameron, J. L. Parra, P. G. Jones, A. Jarvis, Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* **25**, 1965–1978 (2005).

53. B. Sandel *et al.*, The Influence of Late Quaternary Climate-Change Velocity on Species Endemism. *Science*. **334**, 660–664 (2011).

54. W. D. Kissling, G. Carl, Spatial autocorrelation and the selection of simultaneous autoregressive models. *Glob. Ecol. Biogeogr.* **0**, 070618060123007–??? (2007).

55. R Development Core Team, *R: a language and environment for statistical computing* (2019; https://www.r-project.org/).

56. R. J. Hijmans *et al.*, Raster package in R (2013), (available at https://www.rspatial.org/raster/RasterPackage.pdf).

57. V. Calcagno, C. de Mazancourt, Others, glmulti: an R package for easy automated model selection with (generalized) linear models. *J. Stat. Softw.* **34**, 1–29 (2010).

58. R. Bivand *et al.*, spdep: Spatial dependence: weighting schemes, statistics and models (2011), (available at http://ftp.auckland.ac.nz/software/CRAN/src/contrib/Descriptions/spdep.html).

59. S. R. Loarie *et al.*, The velocity of climate change. *Nature*. **462**, 1052–1055 (2009).

60. B. Sandel, A.-C. Monnet, R. Govaerts, M. Vorontsova, Late Quaternary climate stability and the origins and future of global grass endemism. *Ann. Bot.* **119**, 279–288 (2017).

61. R. Jansson, Global patterns in endemism explained by past climatic change. *Proc. Biol. Sci.* **270**, 583–590 (2003).

62. N. Morueta-Holme *et al.*, Habitat area and climate stability determine geographical variation in plant species range sizes. *Ecol. Lett.* **16**, 1446–1454 (2013).

Table S1. Summary statics of ordinary least squares linear regressions models (OLS) and simultaneously autoregressive models (SAR) for predicting the Menhinick rarity index.

	OLS linear regression model				Simultaneous autoregressive model			
Variables	Coefficient	R^2	AIC	Moran's I	Coefficient	R^2_a	AIC	Moran's I
Annual mean temperature	**0.171*****	0.009	14240.324	0.530	**-0.251*****	0.501	11104.069	-0.011
Temperature seasonality	**-0.171*****	0.011	14234.316	0.534	**-0.165***	0.500	11121.533	-0.010
Annual precipitation	**0.195*****	0.036	14115.325	0.517	**0.076*****	0.500	11115.225	-0.010
Precipitation seasonality	**0.187*****	0.024	14173.469	0.519	-0.048	0.499	11124.369	-0.010
Temperature velocity	**-0.493*****	0.121	13696.249	0.477	**-0.275*****	0.513	10998.582	-0.011
Precipitation velocity	**0.068*****	0.002	14272.818	0.532	**-0.087*****	0.501	11105.332	-0.011
Elevation	**0.193*****	0.029	14147.199	0.535	**0.178*****	0.511	11017.718	-0.010
Heterogeneity of elevation	**0.340*****	0.077	13918.380	0.501	**0.179*****	0.512	11004.857	-0.011
Current climate	b	0.088	13869.881	0.470	b	0.504	11082.582	-0.009
Stability of climate	b	0.146	13563.481	0.436	b	0.513	10996.343	-0.011
Topography	b	0.078	13915.694	0.497	b	0.514	10990.681	-0.011
Full	b	0.193	13312.139	0.396	b	0.518	10966.549	-0.010
Exhaustive selection	b	0.194	13310.149	0.396	b	0.516	10980.676	-0.010

a: Nagelkerke pseudo R^2

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

b: coefficients of multiple regressions are shown in Tables S3 and S4

8

Table S2. Summary statics of ordinary least squares linear regressions models (OLS) for predicting the Menhinick rarity index.

	indv1	indv2	indv3	indv4	indv5	indv6	indv7	indv8	Current climate	Stability of climate	Topography	Full	Exhaustive selection
(Intercept)	**-2.29761*****	**-2.31324*****	**-2.29446*****	**-2.22792*****	**-2.52946*****	**-2.19005*****	**-2.26115*****	**-2.35542*****	**-2.26873*****	**-2.46921*****	**-2.35988*****	**-2.45199*****	**-2.45211*****
	(0.02058)	(0.02145)	(0.01767)	(0.01687)	(0.02020)	(0.01938)	(0.01716)	(0.01779)	(0.02088)	(0.02057)	(0.01790)	(0.02220)	(0.02217)
Annual mean temperature	**0.17127*****								**-0.13250****			**0.10838***	**0.10995****
	(0.02590)								(0.04142)			(0.04555)	(0.04261)
Temperature seasonality		**-0.17051*****							0.04285			**0.22031*****	**0.22097*****
		(0.02416)							(0.03990)			(0.03988)	(0.03931)
Annual precipitation			**0.19503*****						**0.31432*****			**0.24683*****	**0.24715*****
			(0.01489)						(0.01896)			(0.01846)	(0.01817)
Precipitation seasonality				**0.18725*****					**0.30987*****			**0.16460*****	**0.16407*****
				(0.01772)					(0.01940)			(0.01996)	(0.01922)
Temperature velocity					**-0.49288*****					**-0.56365*****		**-0.35213*****	**-0.35157*****
					(0.01969)					(0.02032)		(0.03240)	(0.03188)
Precipitation velocity						**0.06806*****				**0.23197*****		**0.20265*****	**0.20267*****
						(0.02047)				(0.01984)		(0.01982)	(0.01982)
Elevation							**0.19252*****				**-0.04821***	-0.00237	
							(0.01635)				(0.02227)	(0.02420)	
Heterogeneity of elevation								**0.33982*****			**0.37662*****	**0.18623*****	**0.18506*****
								(0.01742)			(0.02434)	(0.02922)	(0.02666)
Num. obs.	4571	4571	4571	4571	4571	4571	4571	4571	4571	4571	4571	4571	4571
Moran's I	0.529759	0.529759	0.529759	0.529759	0.529759	0.529759	0.529759	0.529759	0.529759	0.529759	0.529759	0.529759	0.529759
R^2	0.00948	0.01078	0.03620	0.02386	0.12063	0.00241	0.02946	0.07684	0.08779	0.14618	0.07779	0.19398	0.19398
AIC	14240.32398	14234.31597	14115.32505	14173.46878	13696.24909	14272.81792	14147.19866	13918.38030	13869.88099	13563.48143	13915.69445	13312.13906	13310.14865

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

9
10
11

12

Table S3. Summary statics of simultaneously autoregressive models (SAR) for predicting the Menhinick rarity index.

	indv1	indv2	indv3	indv4	indv5	indv6	indv7	indv8	Current climate	Stability of climate	Topography	Full	Exhaustive selection
(Intercept)	**-2.02150*****	**-2.22011*****	**-2.15563*****	**-2.13199*****	**-2.27854*****	**-2.14664*****	**-2.09769*****	**-2.16611*****	**-2.14515*****	**-2.27961*****	**-2.13755*****	**-2.28012*****	**-2.24749*****
	(0.08949)	(0.09100)	(0.08286)	(0.08469)	(0.07881)	(0.08456)	(0.08318)	(0.08053)	(0.09371)	(0.07959)	(0.08162)	(0.08861)	(0.08833)
Annual mean temperature	**-0.25057*****								**-0.31994*****			0.04633	-0.10215
	(0.05128)								(0.05376)			(0.06902)	(0.05713)
Temperature seasonality		**-0.16549***							**-0.22966****			-0.07875	-0.07171
		(0.06816)							(0.07415)			(0.07543)	(0.07552)
Annual precipitation			**0.07633*****						**0.07224****			**0.05058***	**0.04577***
			(0.02182)						(0.02338)			(0.02324)	(0.02325)
Precipitation seasonality				-0.04759					-0.01627			-0.02312	-0.01618
				(0.02708)					(0.02784)			(0.02744)	(0.02743)
Temperature velocity					**-0.27502*****					**-0.26268*****		**-0.13433*****	**-0.15011*****
					(0.02398)					(0.02468)		(0.03560)	(0.03541)
Precipitation velocity						**-0.08703*****				**-0.03882***		-0.02970	-0.03085
						(0.01848)				(0.01880)		(0.01893)	(0.01895)
Elevation							**0.17845*****				**0.09309*****	**0.11097*****	
							(0.01693)				(0.02310)	(0.02904)	
Heterogeneity of elevation								**0.17867*****			**0.11832*****	0.03697	**0.08640*****
								(0.01602)			(0.02189)	(0.02680)	(0.02351)
Num. obs.	4571	4571	4571	4571	4571	4571	4571	4571	4571	4571	4571	4571	4571
Moran's I	-0.01054	-0.01054	-0.01054	-0.01054	-0.01054	-0.01054	-0.01054	-0.01054	-0.01054	-0.01054	-0.01054	-0.01054	-0.01054
Nagelkerke pseudo R^2	0.501466	0.501466	0.501466	0.501466	0.501466	0.501466	0.501466	0.501466	0.501466	0.501466	0.501466	0.501466	0.501466
AIC (Spatial model)	11104.06869	11121.53292	11115.22491	11124.36879	10998.58222	11105.33218	11017.71828	11004.85706	11082.58216	10996.34255	10990.68112	10968.09917	10980.67615

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

13
14
15
16
17
18
19

Table S4. Summary statics of ordinary least squares linear regressions models (OLS) and simultaneously autoregressive models (SAR) for predicting the Margalef rarity index.

	OLS linear regression model				Simultaneous autoregressive model			
Variables	Coefficient	R^2	AIC	Moran's I	Coefficient	R^2_a	AIC	Moran's I
Annual mean temperature	**0.200*****	0.011	9105.447	0.211	**-0.356*****	0.175	8571.756	-0.008
Temperature seasonality	**-0.317*****	0.030	9047.746	0.211	**-0.374*****	0.176	8569.524	-0.006
Annual precipitation	**0.233*****	0.056	8965.718	0.186	**0.167*****	0.179	8559.861	-0.004
Precipitation seasonality	-0.012	0.000	9136.231	0.231	**-0.133*****	0.173	8581.107	-0.006
Temperature velocity	**-0.506*****	0.115	8778.627	0.195	**-0.538*****	0.235	8349.668	-0.005
Precipitation velocity	**-0.098*****	0.005	9122.341	0.236	**-0.192*****	0.182	8548.623	-0.006
Elevation	**0.204*****	0.033	9039.071	0.242	**0.331*****	0.214	8430.225	-0.006
Heterogeneity of elevation	**0.388*****	0.090	8860.037	0.226	**0.425*****	0.242	8325.336	-0.005
Current climate	b	0.068	8936.720	0.184	b	0.200	8488.695	-0.006
Stability of climate	b	0.115	8779.622	0.193	b	0.238	8341.996	-0.006
Topography	b	0.091	8856.571	0.222	b	0.243	8322.144	-0.005
Full	b	0.176	8581.069	0.151	b	0.263	8253.847	-0.004
Exhaustive selection	b	0.176	8578.428	0.153	b	0.260	8262.023	-0.003

a: Nagelkerke pseudo R^2

***p < 0.001, **p < 0.01, *p < 0.05

b: coefficients of multiple regressions are shown in Tables S5 and S6

27

Table S5. Summary statics of ordinary least squares linear regressions models (OLS) for predicting the Margalef rarity index.

	indv1	indv2	indv3	indv4	indv5	indv6	indv7	indv8	Current climate	Stability of climate	Topography	Full	Exhaustive selection
(Intercept)	**-0.62222*****	**-0.72481*****	**-0.64103*****	**-0.52260*****	**-0.92820*****	**-0.57586*****	**-0.58309*****	**-0.74361*****	**-0.67410*****	**-0.91999*****	**-0.75418*****	**-1.02524*****	**-1.03314*****
	(0.02748)	(0.02972)	(0.02234)	(0.02111)	(0.02873)	(0.02528)	(0.02160)	(0.02393)	(0.02955)	(0.02987)	(0.02434)	(0.03373)	(0.03251)
Annual mean temperature	**0.19953*****								**-0.24280*****			**0.18939****	**0.18100*****
	(0.03571)								(0.05627)			(0.06345)	(0.05108)
Temperature seasonality		**-0.31702*****							**-0.22142*****			0.00388	
		(0.03340)							(0.05404)			(0.05444)	
Annual precipitation			**0.23280*****						**0.24959*****			**0.20441*****	**0.20641*****
			(0.01756)						(0.02345)			(0.02261)	(0.02136)
Precipitation seasonality				-0.01232					**0.11038*****			-0.04983	-0.04709
				(0.02286)					(0.02533)			(0.02620)	(0.02586)
Temperature velocity					**-0.50559*****					**-0.51243*****		**-0.18810*****	**-0.18192*****
					(0.02593)					(0.02681)		(0.04426)	(0.04298)
Precipitation velocity						**-0.09837*****				0.02551		0.02867	
						(0.02610)				(0.02546)		(0.02502)	
Elevation							**0.20416*****				**-0.06516***	**0.11574*****	**0.11642*****
							(0.02052)				(0.02787)	(0.03253)	(0.03200)
Heterogeneity of elevation								**0.38751*****			**0.43973*****	**0.24519*****	**0.24141*****
								(0.02277)			(0.03188)	(0.04009)	(0.03916)
Num. obs.	2940	2940	2940	2940	2940	2940	2940	2940	2940	2940	2940	2940	2940
Moran's I	0.210599	0.210599	0.210599	0.210599	0.210599	0.210599	0.210599	0.210599	0.210599	0.210599	0.210599	0.210599	0.210599
R^2	0.01051	0.02974	0.05644	0.00010	0.11462	0.00481	0.03260	0.08976	0.06761	0.11492	0.09145	0.17609	0.17571
AIC	9105.44673	9047.74614	8965.71758	9136.23088	8778.62704	9122.34102	9039.07139	8860.03720	8936.71989	8779.62229	8856.57095	8581.06921	8578.42755

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

28
29
30
31
32
33
34

Table S6. Summary statics of simultaneously autoregressive models (SAR) for predicting the Margalef rarity index.

	indv1	indv2	indv3	indv4	indv5	indv6	indv7	indv8	Current climate	Stability of climate	Topography	Full	Exhaustive selection
(Intercept)	**-0.40532*****	**-0.79384*****	**-0.64092*****	**-0.56641*****	**-0.95994*****	**-0.66699*****	**-0.61932*****	**-0.76853*****	**-0.70418*****	**-0.98656*****	**-0.75496*****	**-1.05468*****	**-0.98129*****
	(0.07613)	(0.07235)	(0.05449)	(0.06041)	(0.05704)	(0.06169)	(0.06211)	(0.05737)	(0.07771)	(0.05868)	(0.05888)	(0.06920)	(0.06320)
Annual mean temperature	**-0.35574*****								**-0.63121*****			-0.01922	0.11957
	(0.06950)								(0.07772)			(0.09100)	(0.07879)
Temperature seasonality		**-0.37425*****							**-0.62355*****			**-0.25575****	
		(0.07522)							(0.09573)			(0.09131)	
Annual precipitation			**0.16726*****						**0.13354*****			**0.10484*****	**0.12950*****
			(0.02784)						(0.03245)			(0.03070)	(0.02923)
Precipitation seasonality				**-0.13261*****					-0.04807			**-0.09598****	**-0.10082****
				(0.03717)					(0.03888)			(0.03661)	(0.03627)
Temperature velocity					**-0.53821*****					**-0.51263*****		**-0.18387*****	**-0.23255*****
					(0.03362)					(0.03469)		(0.05207)	(0.04976)
Precipitation velocity						**-0.19155*****				**-0.08810****		**-0.05863***	
						(0.02840)				(0.02814)		(0.02791)	
Elevation							**0.33075*****				**0.07852***	**0.11786****	**0.13223*****
							(0.02541)				(0.03411)	(0.04065)	(0.04014)
Heterogeneity of elevation								**0.42538*****			**0.37072*****	**0.22060*****	**0.21085*****
								(0.02538)			(0.03489)	(0.04334)	(0.04236)
Num. obs.	2940	2940	2940	2940	2940	2940	2940	2940	2940	2940	2940	2940	2940
Moran's I	-0.00773	-0.00773	-0.00773	-0.00773	-0.00773	-0.00773	-0.00773	-0.00773	-0.00773	-0.00773	-0.00773	-0.00773	-0.00773
Nagelkerke pseudo R^2	0.175334	0.175334	0.175334	0.175334	0.175334	0.175334	0.175334	0.175334	0.175334	0.175334	0.175334	0.175334	0.175334
AIC	8571.75610	8569.52444	8559.86115	8581.10732	8349.66815	8548.62336	8430.22536	8325.33619	8488.69545	8341.99638	8322.14435	8253.84651	8262.02337

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

35

36

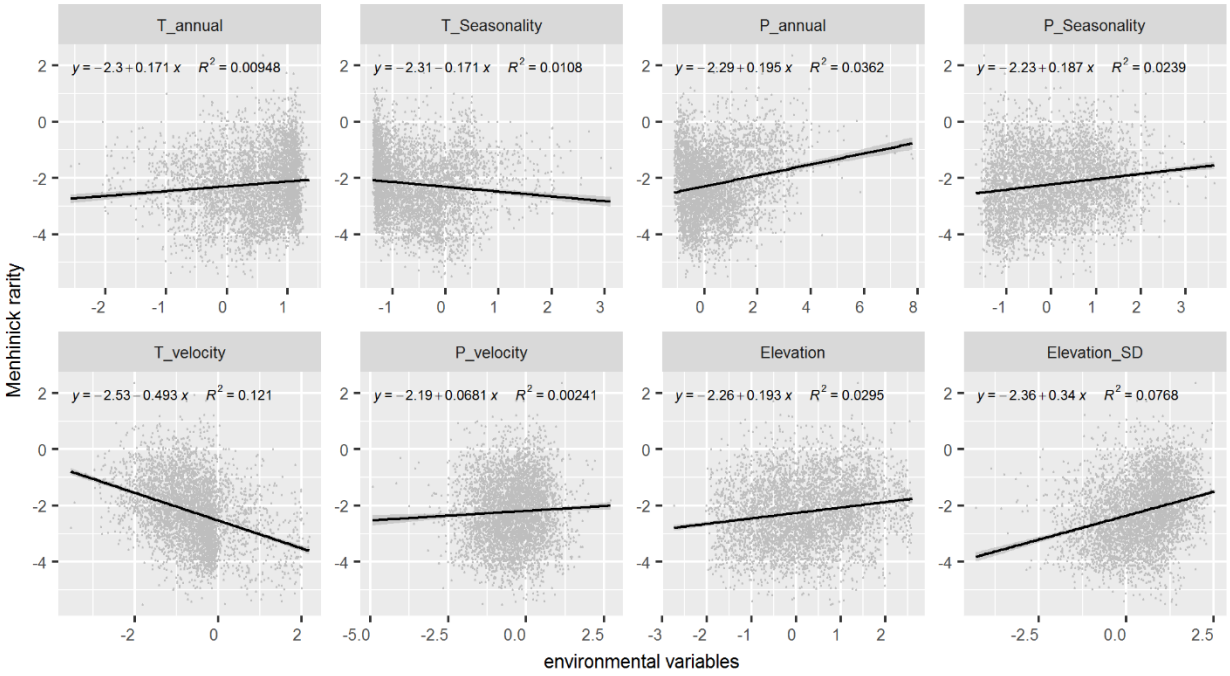


Figure S1. Scatter plots showing the relationships between bivariate relationship between Menhinick rarity index and environmental variables.

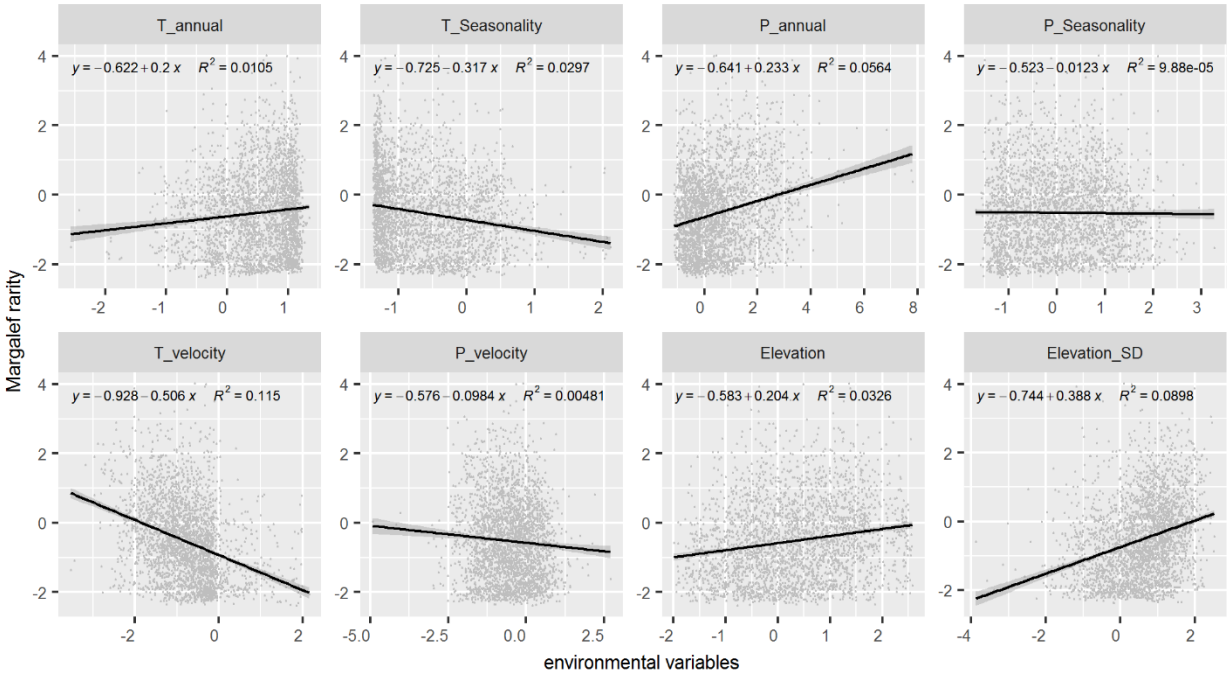


Figure S2. Scatter plots showing the relationships between bivariate relationship between Margalef rarity index and environmental variables.

Figure S3. Predicted changes of Margalef rarity index. Warm color (red) indicates higher decreases, and cold color (blue) indicates lower decreases.