Patterns of species rarity as a driving mechanism for species richness gradients

Mark Giles

Thesis submitted to the University of Ottawa
in partial Fulfillment of the requirements for the
Master's degree in the Ottawa-Carleton Institute of Biology

Department of Biology
Faculty of Science
University of Ottawa

Abstract

Broad scale geographic variation in species diversity correlates with environmental variables in most taxa, but a mechanistic understanding of this relationship has remained elusive. More than a half-century ago, F.W. Preston observed that the number of individuals per species in species assemblages is log-normally distributed (with two parameters: the total number of individuals, I, and the number of individuals of the rarest species, m). Here, we show that ϕ, a proxy for m, is correlated with environmental variables in several datasets of trees, birds, fish, and invertebrates. Moreover, variation in species richness is more strongly related to this measure of rarity than to environment. In all the datasets we examined, structural equation models are consistent with the hypothesis that environmental variables affect species richness principally by affecting rarity, which in turn affects richness. We propose that geographic variation in the ability of species to persist at low densities provides a possible unifying explanation for global gradients of species richness. Our findings may have important implications regarding Earth's biodiversity, highlighting the rarest species as those most at-risk but also important indicators for the ongoing consequences of climate change.

Résumé

Aux grandes échelles, la variation géographique en diversité des espèces est en corrélation avec les variables environnementales dans la plupart des taxons, mais une compréhension mécaniste de cette relation est restée difficile à atteindre. Il y a plus d'un demi-siècle, FW Preston a observé que le nombre d'individus par espèce dans les assemblages d'espèces est log-normalement distribué (avec deux paramètres: le nombre total d'individus, *I*, et le nombre d'individus de l'espèce les plus rares, *m*). Ici, nous montrons que ϕ, un proxy pour m, est corrélé avec des variables environnementales dans plusieurs ensembles de données sur les arbres, les oiseaux, les poissons et les invertébrés. De plus, la variation de la richesse en espèces est plus fortement liée à cette mesure de rareté qu'à l'environnement. Dans tous les ensembles de données que nous avons examinés, les modèles d'équations structurelles sont cohérents avec l'hypothèse que les variables environnementales affectent la richesse des espèces principalement en affectant la rareté, qui à son tour affecte la richesse. Nous proposons que la variation géographique de la capacité des espèces à persister à de faibles densités fournit une explication unificatrice possible des gradients mondiaux de la richesse des espèces. Nos résultats peuvent avoir des implications importantes en ce qui concerne la biodiversité de la Terre, mettant en évidence les espèces les plus rares comme les plus menacées, mais également des indicateurs importants des conséquences actuelles du changement climatique.

Thesis Acknowledgments

I would like to thank everyone who played a part in supporting me throughout my years as a graduate student- science, as well as life, is a collaborative effort and I have been very fortunate. Thank you to Dr. David Currie, my supervisor, for always providing insightful and inspiring feedback and suggestions, as well as for enjoyable conversation, your admirable scientific knowledge and critical thinking abilities, and especially for having the confidence and trust in me, and patience for my idiosyncrasies, to allow me to work my way. Thank you to Jessica Forrest and Tom Sherratt for serving as my committee members. Thank you to numerous fellow graduate students, along with countless scientists and researchers around the world, for your input, contributions, and time. I am thankful to my parents, the University, and NSERC for the financial support that has allowed me to complete this thesis. Finally, I thank my partner Sophie, for always bringing balance and joy to my life- in many ways you have been the glue that has held me, and my research, together.

Table of Contents

List of Tables

List of Figures

Figure S2. The bounds of the possible relationship between richness and singletons in a sample, where singletons cannot exceed richness, and richness cannot be lower than singletons. Especially for low to medium numbers of singletons, the bounds on the possible values are quite wide.

Figure S3. The relationship between a theoretically computed m, based on equation (1), and our proxy ϕ, in each of the five datasets. The strong correlation justifies our use of the proxy to estimate m. The correlation is negative because greater rarity is indicated by decreasing m, but by increasing ϕ.

General Introduction

It is impossible to observe biological communities at a global scale without beginning to wonder what underlies the variation and differences in these communities. The simplest difference, which has been a pattern and question of note in ecology since its inception, is that some places have *more* species than others (Sclater 1858, Wallace 1878). While countless other more complex characteristics of the ecology and biology of living communities differ across the globe, the variation in species richness (the number of different species) seems tantalizingly simple and solvable. One of the papers that forms the bedrock of modern macroecological theory, MacArthur's (1965) *Patterns of species diversity*, aptly begins by simply stating that "patterns of species diversity exist", highlighting the prominence and ease by which these patterns can be observed. However, the intervening decades of research have shown that while the patterns are easy to find, the answers to questions about these patterns are elusive.

Palmer (1994) notes that the problem of species richness gradients has not followed the 'normal' process of science, by which the discovery of an unexplained phenomenon is followed by the proposal of a small set of plausible hypotheses, which ideally are pared down to only one through investigation and experimentation. Instead, we have well over one hundred proposed hypotheses that address the patterns of species richness from different angles, and many with enough support that a large culling, resulting in a small and manageable explanatory mechanism, seems unlikely to occur. What can these seemingly endless hypotheses tell us about what drives gradients in species richness?

Studies of islands led to the identification of area as a determinant of species richness. Preston (1960, 1962) and MacArthur and Wilson (1963) contributed to the concept that greater area provides more space for a greater number of species, and indeed larger islands show exponentially greater species richness than do smaller ones. On all islands, there are new colonizations and local extinctions, but they seem to occur in equilibrium, and the number of species at which these processes equalize is area-dependant. Following Preston (1962), Srivastava and Lawton (1998) refined this into the 'more individuals hypothesis', noting that more space allows for more total individuals, which then will tend to be represented by more species. However, most species in the world do not live on islands, and simply considering area cannot address patterns of richness observed in oceans, or on continents. It is important to note, however, that by comparing differently sized patches of habitat on continents, essentially creating 'islands', the species-area curve is observed (Rosenzweig 1995).

Storch et al. (2018) state that the consensus among ecologists is that climatic variables are mostly responsible for the species richness gradients, and it is certain that climate and environment are important factors. In a meta-analysis, Field et al. (2009) surveyed almost 400 tests of the drivers of species richness gradients across many different taxa, categorizing the results to see which general mechanisms were most often supported. The best-supported category of hypothesis was found to be climate and productivity. There are several proposed explanations for why climate should drive richness. Hutchinson (1959) suggested that the total metabolic energy of a system is finite, and that species partition the energy among themselves. Warmer, wetter environments provide more energy, and allow for more species to get a 'piece of the pie' because more individuals can persist where productivity is greater. Alternatively, Klopfer (1959) and Pianka (1966) proposed that it is climatic *stability* that matters, as it allows for species to become more specialized and thus exploit different niches. Finally, more benign climates may accommodate a wider range of physiological tolerances, and therefore support more species, the so-called 'physiological tolerance' hypothesis (Kleidon & Mooney 2000), which was further supported by Spasojevic et al. (2014). Kleidon & Mooney (2000) found that higher numbers of simulated plant species (generated with random combinations of physiological parameters) could persist in climatically benign areas.

Climate is not the only distinguishing feature of a location. Physical, topographical, and geographic factors all contribute to the concept of 'habitat heterogeneity', which states that an increase in the physical diversity of a landscape creates more habitats, and therefore more dimensions for niche partitioning (MacArthur & MacArthur 1961). This process facilitates specialization, and inhibits competition, both of which contribute to an increase in the number of species. Support for the habitat heterogeneity hypothesis has been found in both aquatic and terrestrial environments, successional and disturbance areas, and along latitudinal, elevational, and humidity gradients (Cramer & Willig 2005). Kerr and Packer (1997) found that, in a case where the climatic rules do not seem to apply, the mammals of non-Arctic North America, it is instead habitat heterogeneity that becomes the strong predictor of richness along the gradient. This is an example that helps to illustrate why the drivers of species richness gradients are hard to pin onto a single theory- different locations and communities do seem to respond to different factors.

Another group of hypotheses could be categorized as 'historical', and their explanations for species richness gradients look away from the area, climate, or habitat heterogeneity of the *current* environment and instead look into the past, to see if processes or events that shaped the landscape are instead responsible for today's biological communities. Latham and Ricklefs (1993) noted that geographically separated, but climatically similar, angiosperm habitats such as mangroves and scrub can often show very different numbers of species, and therefore concluded that climate does not tell the whole story. Instead, they proposed that the evolutionary histories of the plants in these areas contribute to the current community. A historical event that has been linked to species richness gradients is the advance and subsequent recession of glaciers across much of the planet's temperate regions. Svenning and Skov (2007) found that, especially for species with restricted ranges, examining historical data of glacial cover reveals that species tend to occur in the places to which they were restricted by the ice. Finally, historical-type hypotheses need not extend into deep time and invoke glaciation or evolution- Jacquemyn et al. (2001) and other similar studies have found that ongoing processes that have occurred in the more recent past, such as dispersal, succession, and species extinctions and colonizations, influence local species richness in a significant fashion.

A final domain of hypotheses that address the species richness gradient question take a further step back, even beyond the historical hypotheses, and ask whether biological factors need be considered at all. Colwell and Hurtt (1993) acknowledged the various ideas discussed above, but noted that it is important to determine what patterns of species richness would look like *without* any biological factors. In other words, what should the default expectation, or null model, be, for the distribution of biological communities? They showed that many of the patterns, such as the latitudinal richness gradient, could develop 'on their own', and that effects of sampling could create some of the patterns as well. This proposed 'mid-domain effect' challenges other notions of what creates the gradients so commonly observed in species richness. However, Hawkins et al. (2005) and others (Currie and Kerr 2008, Fox 2011) have challenged the core assumptions of the mid-domain effect, and question whether it truly represents a null model, or whether existing environmental gradients in species richness are required even for its effects to properly model gradients observed in nature.

Any new research that seeks to ‘solve’ the species richness gradient problem should consider the breadth of theory discussed above, and the much larger pool of hypotheses that were not even mentioned, and attempt to do more than simply add one additional theory. Could there instead be patterns or processes that underlie several, or even most, of the existing hypotheses that could serve to connect them? It is the goal of the current study to highlight a previously undiscovered mechanism underlying the species richness gradient, and to add it as a new piece into a larger framework that models the complicated and interconnected process that controls why certain places have more species than others.

Introduction

An elusive problem in macroecology has been a mechanistic explanation for global-scale geographic variation in species richness. Put more simply, why do some places have few species, and other places so many? Commonly, species richness (the number of species observed in a particular area) decreases moving poleward from the equator, and from low to high elevation (Gaston 2000, McCain & Grytnes 2001, Hillebrand 2004). These are certainly not the only known gradients; species richness varies with depth in aquatic environments (Rex et al. 2000, Woolley et al. 2016), longitudinally across continents (Liu et al. 2008, Mee et al. 2018), and through time due to evolutionary diversification and historical events such as glaciation (Wiens & Donoghue 2004, Belmaker & Jetz 2015). Along latitudinal and elevational gradients, richness in most taxa correlates with temperature, precipitation, or an interaction between the two (Currie 1991, Francis & Currie 2003, Field et al. 2009). Species richness in specific taxa correlates with other environmental variables, such as NDVI (a measure of the density of vegetation cover) for birds (Hurlbert & Haskell 2002), and ambient salinity and nitrate concentrations for coral-reef fish (Mellin et al. 2010). Many of these environmental variables have been interpreted as measures of resource/energy availability (Currie 1991, Gaston 2000, Hawkins et al. 2003). Geographic factors such as habitat heterogeneity have also been found to be important (Rahbek & Graves 2001, Seibold et al. 2016). However, among seemingly endless correlations, no well-supported causal mechanism has emerged to broadly predict variation in species richness along these different environmental gradients. There has been no shortage of proposed mechanisms, and indeed the sheer volume of hypotheses could be said to have contributed to the confusion. According to Palmer (1994), at least 120 hypotheses relating to the question of species richness variation have been proposed, and this number has only increased in the intervening decades.

What, if anything, do all these patterns have in common? Is it possible that these variables are indirect correlates of some other variable that directly affects richness?

According to a classic study by Preston (1962), the distribution of individuals among species in any species assemblage follows a regular distribution, which Preston called "canonical log-normal". The log normal distribution has two free parameters and two coefficients (that Preston determined empirically). According to this model, there is a necessary relationship between the total number of individuals in the assemblage (*I*), the minimum number of individuals for a species to persist, which Preston denoted by *m*, and total species richness (N) (Figure S1). If two of these values are known, the third is fixed:

(1) $$N = 2.075 \times ({}^{I}/_{m})^{0.262}$$

(2) $$\log(N) = 0.31 + 0.262\ (\log I - \log m)$$

This suggests that any mechanism governing variation in species richness could be driven entirely by variation in these two variables, *I* and *m*. In other words, if environmental and geographic variables influence species richness, that influence must be exerted through effects on *I* and *m*. Preston discussed the fact that the well-known relationship between species richness and area (for example, on islands of differing areas, following MacArthur & Wilson (1967)) can be derived from the canonical log-normal model, as more area allows for more individuals (Preston 1962). Preston did not, however, extend his model to other patterns of species richness such as the latitudinal gradient.

Is it reasonable to pursue a lead from work more than a half-century old? The primary point that Preston wished to put forward was that species abundance distributions in many different taxa are log-normally distributed (Preston 1962): i.e., the variation in number of individuals per species approximately fits a log-normal curve. This was supported by empirical examples of his own, and later work has provided further evidence of the generality of the log-normal species abundance distribution (Ulrich et al. 2010). Other models to describe species-abundance distributions have been proposed (Hubbell et al. 2008, Chao et al. 2015), but they are close to log-normal, and the log-normal model often out-performs the alternatives (McGill 2003, McGill et al. 2006). Although Preston's (1962) analysis inspired the present study, the exact species-abundance distribution is not critical; it may well vary to some degree among taxa (McGill 2003). Rather, our hypothesis is that richness depends upon the proportion of rare species in an assemblage.

In contrast, the energy-richness hypothesis proposes that larger areas, as well as areas of higher primary productivity, are able to support more species because there is more 'energy' available to support more individuals (Hutchinson 1959, Wright 1983). A highly productive environment, which supports a large number of individuals, is therefore home to a large number of species, as the individuals are able to represent more species than would be possible with smaller populations. Crucially, however, do the warmer, wetter areas support more individuals? Currie et al. (2004) found that spatial variation in the total number of individuals (Preston's *I*) was insufficient to explain broad-scale variations in richness. Although this 'more individuals hypothesis' would be a straightforward explanation for the observed energy-richness patterns, studies and models based on it have not been well-supported (Storch et al. 2018). But note that, again following from Preston, variation in I should be only part of the explanation for variation in richness.

Therefore, if Preston's model does indeed provide a general description of the geographic variation of species-abundance distributions, then variation in rarity – *m* -- must be responsible for the variation in richness left unexplained by the total number of individuals. Places and populations with lots of species, without enough ‘extra’ individuals, must have species persisting at densities lower than the global average.

Following Preston (1962), here we hypothesize that observed patterns of variation in species richness are all to some extent mediated by variation in rarity, irrespective of the climatic or environmental gradient along which variation occurs. This hypothesis leads to several predictions.

1. Gradients of species richness should be correlated with some physical and/or biotic characteristic(s) of the environment (e.g. climate), depending on the taxon, as has been well-established previously.
2. m should show sufficient spatial variation (in a way that I does not) to explain these known patterns of variation in species richness.
3. When richness is strongly correlated with an environmental variable along some geographic gradient, m should be correlated with both the climatic, spatial, and/or environmental variables, and also with richness.
4. In the hypothetical causal pathway whereby environment influences rarity, which in turn influences richness, the proximal correlations should be stronger than distal climate-richness correlation, and this pathway should be better supported by path analysis than models that omit the environment → rarity → richness pathway.
5. Preston’s argument is agnostic to taxon or environmental gradient. It suggests that these predictions should hold for any taxonomic group (birds, trees, fish, etc.) along any richness gradient (latitudinal, altitudinal, depth, etc.).

Materials and Methods

To address our predictions, we sought datasets that met several important criteria. A dataset needed to provide survey-based, abundance data identified to the species level. Additionally, the datasets needed to be comprised of several sites with sufficient spatial coverage to capture important gradients of species richness. Finally, we needed datasets representing a variety of taxa, and different potentially driving environmental or climatic variables. We identified five datasets that met each of these criteria (Table 1).

The first dataset is composed of nearly 200 0.1 hectare forest plots located around the world (concentrated in the neotropics), sampled by Alwyn Gentry, on which all plants with stem diameters of at least 2.5 cm were counted and identified (Phillips and Miller 2002). The dataset included 149 sites for which we could estimate *m* (see below). Second, the Barva dataset consists of eleven 1-hectare forest plots sampled by Leiberman et al. (1996), at elevations of 100 m to 2600 m on the Barva volcano in central Costa Rica. Third, the North American Breeding Bird Survey (BBS) is a program established in 1966 and now represented by an extensive dataset of roughly 4000 sites across the United States and southern Canada, consisting of 40 km 'routes', along which experienced ornithologists do point counts every 1.0 km. We used routes censused during 2013 for which *m* could be estimated (see below). The sample includes 2660 sites. Fourth, the Reef Life Survey (RLS) is another extremely extensive dataset, representing 500 m^2 transects of reef habitat that have been sampled for reef fish by experienced scuba divers. The final sample used in our analysis, after eliminating sites for which *m* could not be estimated (see below), includes 1164 sites around the world, concentrated in the tropics but with some temperate reef sites. Fifth, the Ocean Basin Benthos dataset represents epibenthic sled collections of deep-see gastropods at different depths. Our analysis includes 40 sites.

All five datasets differ detectably from log-normality, based on Komolgorov-Smirnov tests, but none departs dramatically (Table S1). Failure to pass normality tests is very common in large datasets.

Climatic data were acquired from several sources. Temperature and precipitation at a resolution of 30 arc seconds were acquired from the ‘Climatologies at high resolution for the earth's land surface areas’ (CHELSA) dataset (Karger et al. 2017). The distribution of precipitation data was strongly positively skewed. We therefore square-root transformed precipitation to normalize its frequency distribution, thereby reducing the statistical weight of extreme values. NDVI (at a resolution of 250 metres) and ocean chlorophyll *a* (at a resolution of 4 km) were acquired from NASA’s Moderate Resolution Imaging Spectroradiometer (MODIS) datasets (Didan 2015, NASA 2015).

It is impossible in practice to directly measure m, the abundance of the rarest species in a species assemblage. An estimate of m must therefore be acquired by proxy from the survey-based data. In principle, one could hypothesize that the distribution of individuals among species conforms to a given distribution. One could then fit that distribution to the number of individuals per species and estimate the abundance of the rarest species using the fitted parameters of the distribution (Matthews et al. 2014). The drawback to this approach is that the estimate will depend upon the fit to the entire distribution. We found that this method is more sensitive to the abundances of the most common species, not the rarest ones. Instead, we chose to focus on the left tail (the rare side) of the distribution.

An indication of rarity in a sample is the number of singletons, species represented in the sample by a single individual (Colwell & Gotelli 2011). Singletons represent a unique signal of rarity in the data that is independent of N and I. Although the possible numbers of singletons in a sample are bounded (since it is impossible to have more singletons than species, or to have fewer species than singletons), these bounds are not so strict as to impose circularity (Figure S2). In other words, it is not mathematically necessary that samples with more singletons have more species.

However, the raw number of singletons is inappropriate as a measure of how rare species are because it is sensitive to sample size. The proportion of species that are represented by singletons is also inappropriate because it involves dividing by species richness, which is the variable we are trying to predict.

The proxy we arrived at, which we denote as ϕ, was the number of singletons in uniformly rarefied samples. Rarefaction involves resampling the data from individual sites to a given sample size. It has been used to make species assemblage data more comparable and useful (Colwell et al. 2012). Its advantage is that it eliminates sample size effects without 'contaminating' the proxy (which we want to use as an explanatory variable) with species richness (our response variable). We found that estimates of ϕ were relatively insensitive to the rarefaction threshold (the number of individuals to which each site is resampled) over a fairly wide range of thresholds. We chose a threshold for each dataset that included at least 80% of the sites, and that fell well within the range of rarefaction thresholds over which ϕ varied little.

We evaluated the proxy ϕ based on the following logic. For any given data set, *N* and *I* are known. If Preston's model correctly describes a general relationship between richness and rarity (H1), then one can generate a theoretical value for *m* by rearranging equation (1):

$$m_{theo} = 16.2 \times I / N^{3.817} \qquad (3)$$

If it is also true that ϕ is a good proxy for m (H2), then ϕ and m should be strongly correlated. Strong relationships between ϕ and m_{theo} would be consistent with the conjuction of propositions H1 and H2.

Note that I in Preston's model is the number of individuals in the regional species assemblage. I is also not directly observable. Here, we assume that regional variation in the density of individuals among samples is linearly proportional to the geographic variation in the number of individuals among regional assemblages.

To test the first and third predictions, that relationships exist between the environment, richness, and rarity in the data as predicted, regression models were created to predict species richness using Preston's two variables (I and m, via the proxy ϕ) and the environmental variables most strongly correlated with richness in the original studies. In all models, we transformed the environmental variables as necessary to meet the assumptions of linear regression. To test the second prediction, that m varies spatially, the ranges of N, I, and ϕ in each dataset were compared. To test the fourth prediction, path models were created to compare the pathway (environment → richness) with our hypothesized pathway (environment → rarity → richness). Finally, to test the fifth prediction, we plotted the richness/rarity relationships for all the species assemblages on the same graph, to visualize the degree of universality of the richness-rarity relationship.

All statistical analyses were performed in R version 3.60 (R Core Team 2013), except for the path models, which were performed in SPSS Amos v23.0 (Arbuckle 2014).

Results

The number of singletons in a rarefied sample, ϕ, is strongly related to the theoretical m derived from Preston's model (i.e., predicted from N and I), with ϕ and m having correlations (r) of -0.98, -0.94, -0.61, -0.75, and -0.86 for the Gentry, Barva, BBS, RLS, and Ocean Basin Benthos datasets, respectively (Figure S3). The correlation is negative because a rare species would have a lower m (indicating a lower density of individuals) but a higher ϕ (indicating more singletons in the sample).

Consistent with our first prediction, we found relationships between species richness and environmental variables in each dataset (Table 2). This is not a surprise, as earlier work has already established these correlations. In the Gentry dataset, tree species richness was related to temperature and precipitation (Currie 1991, Francis & Currie 2003, Hawkins et al. 2003, Field et al. 2009). In the Barva dataset, tree species richness was strongly negatively related to elevation (Lieberman 1996, Sharma et al. 2009). In the BBS dataset, bird species richness was related to NDVI, a measure of the density of vegetation cover (Hurlbert & Haskell 2002, Hurlbert & White 2005) as well as with precipitation (Coops et al. 2018). In the Reef Life Survey dataset, fish species richness was related to sea surface temperature (Parravicini et al. 2013) and chlorophyll *a* concentration (Mellin et al. 2010). In the Ocean Basin Benthos dataset, gastropod species richness was related to depth (Rex et al. 2000).

The second prediction was that ϕ, our proxy for the abundance of the rarest species, should show significant variability along richness gradients. We found that ϕ is more variable than is species richness (comparing coefficients of variation) in all five data sets (Table 3). Currie et al. (2004) argued that variation in I, the total number of individuals was insufficient to account for variation in richness. We found that the variability of I was less than the variability of richness in the two data sets that focus on trees (Barva and Gentry); however, the number of individuals was *more* variable than richness or rarity in the three other datasets (Table 3). Nevertheless, in the regression models, the total number of individuals I statistically accounted for relatively little variance in species richness in all datasets, and I was only weakly related to environmental and climatic variables, especially so for the Gentry, BBS, and RLS datasets. This consistent with the weaknesses of the 'more-individuals hypothesis' (Storch et al. 2018).

In all five datasets, rarity (ϕ) was related to the environmental and climatic variables for each taxon, consistent with the third prediction (Table 2). Furthermore, ϕ was related to richness, as predicted, and more strongly than richness was related to environment/climate in 4 of the 5 datasets (Table 2). In the fifth dataset, the Barva volcano transect, the richness-rarity relationship was essentially identical to the rarity-environment/climate relationship (Table 2).

Adding the effect of individuals as a predictor for richness to the stronger rarity predictor, and modelling species richness as a function of ϕ and I, increased the explained variance in all five datasets (Table 2). This is consistent with Preston's hypothesis that these two variables determine richness. Our hypothesis is supported by the fact that ϕ accounts for more variance in N than does I.

Path models, used to test whether hypothesized causal pathways are consistent with the data, were largely consistent with the third prediction. Path models that postulated a direct effect of rarity on richness, and no effect of environmental/climatic variable(s) on richness, performed better than models that postulate the inverse (Figure 1). The path model allowing both rarity and environment to influence richness was better than both restricted models in all five datasets (Figure 1). The number of individuals was included in all models, and contributed significantly, but less strongly, than rarity in all five datasets (Figure 1).

Richness varies as a function of rarity similarly for all taxa (Figure 2). Multiple regression models predicting richness from ϕ were compared. An ANCOVA model that allows for different slopes for each dataset ($R^2 = 0.695$), or an ANCOVA that allows for different intercepts but the same slope ($R^2 = 0.687$). account for only marginally more variance than a regression in which datasets are pooled ($R^2 = 0.680$). The ANCOVAs are significantly superior (p= <0.00001), given that statistical power is very high (n=4024). However, the most striking difference among the five datasets is the variance in richness for a given value of ϕ: variance is small for the gastropods, Barva trees, and Gentry's trees, but much larger for the Breeding Bird Survey and the Reef Life Survey. Differences in variance are primarily among groups. Within groups, variance is approximately homoscedastic with respect to ϕ.

Discussion

Our results are consistent with the hypothesis that environmental variables influence species richness through effects on rarity, irrespective of the taxon or the environmental gradient. Many correlations between these environmental variables and species richness have been documented in the literature. Here, we propose a mechanistic explanation underlying these observations: site-level species richness is high where species persist at low densities. Species persist at lower densities at the more favourable, perhaps more benign, ends of environmental gradients. This pattern held broadly, whether the benign end of the gradient represented conditions that were warmer, wetter, lower altitude, lesser depth, greater vegetation cover or other measures of resource availability. Furthermore, the pattern was found in a wide range of taxa. Therefore, in the spirit of Preston's work that laid the foundation for this model, and which he presented as 'canonical', we suggest that this mechanism is a universal feature of species assemblages, and that the rarity-richness relationship is also canonical.

As Preston's models predict, all of our models include significant effects of both the total number of individuals I and rarity ϕ. We found that I does vary widely in several of the datasets, in contrast with the findings and conclusions made by Currie et al. (2004). Our datasets that represented the same taxon (trees) as the Currie et al. (2004) study did replicate their results. However, other taxa showed plenty of variability in the numbers of individuals. Preston's model predicts that I and m (for which ϕ is a proxy) should have equal effects on richness (see eq. 2 above). In contrast, in our regressions, the coefficient on $\log(\phi)$ was roughly twice the coefficient on $\log(I)$ (Appendix). This may occur because the relationship between ϕ and Preston's m is not linear, or perhaps because the species-abundance distributions are not exactly log-normal, as Preston postulated (Magurran 1988).

The path models also revealed that while much of the explained variance in richness does 'flow' from the environmental variables through rarity, there is still a small contribution of the environmental variables *directly* to richness. This suggests that some aspects of the environment-richness relationship are mediated neither by rarity nor by number of individuals. While this technically contradicts Preston's model, which presents I and m as the only contributors to richness, it might reflect statistical artifacts such as our use of ϕ as a proxy for m, bias in census estimates of I (e.g., undetected species in point counts of bird abundance), or lack-of-fit in the log-linear relationship (perhaps because the species-abundance distribution is not exactly log-normal). Alternatively, it is also not clear that Preston intended his model to explain *all* variance in richness. For example, Preston did not address the factors that cause geographic gradients in richness. More study will be needed to identify the source of this statistical effect.

Is it possible that the causal effect is actually from richness to rarity (Environment$\rightarrow N \rightarrow \phi$), rather than our hypothesized pathway from rarity to richness

(Environment→ϕ→N)? In principle, one could create a new model by reversing the direction of the arrow in the $\phi \rightarrow N$ pathway in the saturated model (i.e., with all possible links) in the center column of Fig. 1, and then compare the modified model with the model from Fig. 1. Unfortunately, there are insufficient degrees of freedom to compare two saturated models. Comparing the sub-models consisting of our hypothesized pathway (Environment→ϕ→N) and a competing sub-model with the causation reversed (Environment→N→ϕ), we found that neither sub-model is consistent with the data: as Figure 1 showed, there is a direct link between environment and richness. Moreover, neither sub-model is clearly superior in all data sets (Table S2). Therefore, based on this analysis, we cannot unequivocally eliminate the competing hypothesis that environment affects richness, which then affects rarity.

Although richness may be higher in more benign environments, why should these environments allow for persistence of species at lower densities? In a sense, we have merely pushed back by one step the question of the mechanism(s) that control(s) species richness. A possible answer is found in the fluid nature of real-world species assemblages. While it is convenient to treat a natural community as static and unchanging, a sample represents only a snapshot of a continually changing and evolving network of species. One important factor that contributes to this continual remodelling of the assemblage is local species extinction. Any given species in an assemblage could become locally extinct, but what controls that probability? The likelihood of extinction is known to be inversely correlated with population size, as rarer species are, by definition, closer to becoming extinct than common ones (Pimm et al. 1988). An interesting possibility is that the slope of probability of local extinction as a function of abundance may vary among different environments and climates. While research has focused on the possible correlates of extinction risk, the possible connection to resource availability or the favourability of the environment has not often been considered (Purvis et al. 2000, O'Grady et al. 2004). Boucher-Lalonde et al. (2014) found that the year-to-year risk of local extinction was independent of climate in the Breeding Bird Survey data. They concluded that climate-dependent extinction is not the driver of richness-climate patterns in North American passerine birds. It is unknown whether this is also true for other taxa.

Colonization occurs alongside local extinction. If extinction is not climate dependent, perhaps instead the minimum number of individuals that are necessary to successfully colonize a new location is dependent on the favourability of the local environment. Schurr et al. (2007) found that a plant's colonization ability was a strong predictor of its ability to fill its potential range. It is possible that it is easier for new species to invade established species assemblages in benign environments because there are more different ways to survive physiologically than in harsher environments (Kleidon and Mooney 2000). This climate-dependent colonization hypothesis could be tested using the approach of Boucher-Lalonde et al. (2014).

An effect of rarity may not be the only possible interpretation of our data. Assume that species-abundance distributions are indeed log-normal, as Preston states, or a similar but still universally applicable distribution, as other authors have suggested (Hubbell et al. 2008, Chao et al. 2015). Following Preston (1962), we focused on the abundance of the rarest species, but the shape of the canonical log-normal is quite constrained. Consequently, m is strongly collinear with other parts of the distribution (e.g., the abundance of the most abundant species in the assemblage). We cannot eliminate the possibility that the observed relationships between environmental variables and our rarity proxy, and between richness and our rarity proxy, depend upon some other part of the distribution that is highly correlated with the left tail. Further research, perhaps designed to focus more directly on the rarest species, will be needed to test this possibility. It is also important to note that despite the current study having been inspired by Preston's log-normal, its conclusions are not dependent on it, as log-normality is not an assumption of any of the analyses.

We found that different taxa accumulate species at a given level of rarity approximately in the same way (Figure 2). Although the slopes and intercepts do differ between datasets, these differences add only a little explanatory power to a regression model of log(richness) as portrayed in Figure 2. Site-level richness varies by about a factor of ~100, whereas expected richness, given ϕ, varies by about a factor of only ~2 among taxa. Differences in the rarity-richness relationships among taxa are small. As far as we are aware, Preston never suggested that his model should explain differences in richness among different taxonomic groups, yet his model is presented as being universal. Nonetheless, it remains surprising to us that richness relates to rarity very similarly taxa in different Kingdoms and in different biomes. What differs most strikingly among the datasets is the amount of residual variation, with the tree datasets (Gentry and Barva) having very little variation, and the animal datasets, much more (Figure 2). This may reflect biological differences among the taxa, or methodological differences in the sampling: in other words, whether tree richness responds to rarity with less variation, or if we are simply able to capture and estimate the relationship better.

How do our findings fit into the larger picture of species richness gradients, among Palmer's (1994) list of over one hundred hypotheses? For example, Rahbek et al. (2006) found that climate-richness patterns in South American birds mainly reflect the distributions of the most common, widespread species. Small ranges do not overlap sufficiently to produce areas with high species richness. This appears to contradict our focus on the rarest species. If richness gradients result mainly from the distributions of the most common species, then rare species should have minimal influence. Note that the conclusions of Rahbek et al. (2006) were based on patterns derived from overlapping range maps, rather than site-level sampling or survey observations. Geographic variation in site-level richness is poorly related to range-map richness (Hurlbert and White 2005; De Camargo and Currie 2019) because species do not occupy all sites with their range. The datasets used in the present study characterize α-richness, or the actual richness (the species that are observed at a location) while range-map data characterize γ, the regional richness, which represents the pool of species that *could be* observed at a location. Range-filling can be quite low, quite variable, and potentially dependent upon historical factors (Svenning and Skov 2004) and species characteristics (Schurr et al. 2007).

If we are to generate testable predictions about the mechanism behind species richness gradients, we must acknowledge the inherent interaction between multiple effects. Therefore, we present a possible model that seeks to unify several proposed theories into a larger framework (Figure 3). Climate acts on rarity, and thereby on α richness, as proposed in the present study and in Preston (1962). Climate also acts on γ richness, via range size effects mediated by the physiological tolerances of the species concerned (Kleidon & Mooney 2000), and subsequent overlapping of those ranges (Rahbek et al. 2006). Meanwhile, the relationship between the two measures of richness is interactive, consisting of top-down processes such as range filling (Cornell & Lawton 1992) and also bottom-up processes such as the introduction of exotic species (Sax et al. 2002). The intention is not to propose this framework as a rigorously supported model, but rather to fit our puzzle piece of rarity driving species richness gradients into the larger body of literature.

The mechanism we have outlined has potential interest in the context of a changing climate and an increasingly impacted planet. If, as we propose, richness depends upon rarity, then any changes due to human activity that threaten the rarest species, even only slightly reducing their abundance or persistence, are likely to lead to overall species losses. Our data align with the existing body of literature that emphasizes a focus on the rarest species, both as those at greatest risk of extinction but also as indicators for the direction a natural community could be heading (Ohlemüller et al. 2008, Leitao et al. 2016). Furthermore, if the rarity-richness mechanism does indeed represent a piece of the process by which natural communities are assembled, and how they change through time, its inclusion in predictive models should improve our forecasts of species losses due to climate change.

In summary, our results provide a mechanism that may underlie many of the existing hypotheses pertaining to gradients of species richness in local species assemblages. As macroecology is already overburdened with hypotheses addressing the question of why some places are more species-rich than others, we are pleased to suggest a possible connection between them all. We do not establish *why* a specific climatic variable is related to richness, but rather *how* it exerts the observed influence. Importantly, we propose a mechanism that leads to testable predictions, the evaluation of which will provide benefits beyond the addition of one more hypothesis to the crowd. Such predictions include that many areas currently losing species should be those where the rarest species are being removed or put at risk, that effects of environment and climate on richness not considered in the present study should also be found to be mediated via rarity, and more broadly, that perhaps for any gradient of species richness in any taxon, that is found to be meaningfully correlated with any variable, a portion of the relationship will be mediated via rarity. It would be especially interesting to apply that third prediction to non-environmental gradients such as those throughout time, or across space. These tests will clarify how species accumulate in an assemblage, and perhaps more importantly, how they may be disappearing and could be expected to continue doing so. In a rapidly changing ecosystem, understanding why some places are more species-rich than others is essential if we hope to keep them that way.

Study Acknowledgements

This study was inspired by unpublished work carried out by Kevin Walker. The study benefitted from feedback from Jessica Forrest and Tom Sherratt. We acknowledge Alwyn H. Gentry, the Missouri Botanical Garden, and collectors who assisted Gentry or contributed data for specific sites. We acknowledge the thousands of U.S. and Canadian participants who annually perform and coordinate the BBS survey. We thank the many Reef Life Survey divers, researchers and managers who participated in data collection and provide ongoing expertise and commitment to the programme. We thank Michael A. Rex, Carol T. Stuart, Gary Hartshorn, Diana Lieberman, Milton Lieberman and Rodolfo Peralta for providing the gastropod and Barva tree data. The work was supported by funding from the Natural Sciences and Engineering Research Council of Canada.

References

Allen, A. P., Brown, J. H., & Gillooly, J. F. (2002). Global biodiversity, biochemical kinetics, and the energetic-equivalence rule. Science, 297, 1545-1548.

Arbuckle, J. L. (2014). Amos (Version 23.0) [Computer Program]. Chicago: IBM SPSS.

Belmaker, J., & Jetz, W. (2015). Relative roles of ecological and energetic constraints, diversification rates and region history on global species richness gradients. Ecology Letters, 18(6), 563-571.

Boucher-Lalonde, V., Theriault, F. L., & Currie, D. J. (2014). Can climate explain interannual local extinctions among bird species? Journal of Biogeography, 41(3), 443-451.

Chao, A., Hsieh, T. C., Chazdon, R. L., Colwell, R. K., & Gotelli, N. J. (2015). Unveiling the species-rank abundance distribution by generalizing the Good-Turing sample coverage theory. Ecology, 96(5), 1189-1201.

Colwell, R. K., & Hurtt, G. C. (1994). Nonbiological gradients in species richness and a spurious Rapoport effect. The American Naturalist, 144(4), 570-595.

Colwell, R. K., Chao, A., Gotelli, N. J., Lin, S. Y., Mao, C. X., Chazdon, R. L., & Longino, J. T. (2012). Models and estimators linking individual-based and sample-based rarefaction, extrapolation and comparison of assemblages. Journal of plant ecology, 5(1), 3-21.

Coops, N. C., Rickbeil, G. J., Bolton, D. K., Andrew, M. E., & Brouwers, N. C. (2018). Disentangling vegetation and climate as drivers of Australian vertebrate richness. Ecography, 41(7), 1147-1160.

Cornell, H. V., & Lawton, J. H. (1992). Species interactions, local and regional processes, and limits to the richness of ecological communities: a theoretical perspective. Journal of animal ecology, 1-12.

Cramer, M. J., & Willig, M. R. (2005). Habitat heterogeneity, species diversity and null models. Oikos, 108(2), 209-218.

Currie, D. J. (1991). Energy and large-scale patterns of animal-and plant-species richness. The American Naturalist, 137(1), 27-49.

Currie, D. J., Mittelbach, G. G., Cornell, H. V., Field, R., Guégan, J. F., Hawkins, B. A., ... & Turner, J. R. G. (2004). Predictions and tests of climate-based hypotheses of broad-scale variation in taxonomic richness. Ecology letters, 7(12), 1121-1134.

Currie, D. J. and J. T. Kerr (2008). "Tests of the mid-domain hypothesis: A review of the evidence." Ecological Monographs 78(1): 3-18.

De Camargo, R. X, & Currie, D. J. (2019). Is habitat conversion likely to impede the ability of bird species to track changing climate? *Frontiers of Biogeography*. http://dx.doi.org/10.21425/F5FBG41613

Didan, K. (2015). MOD13Q1 MODIS/Terra Vegetation Indices 16-Day L3 Global 250m SIN Grid V006 [Data set]. NASA EOSDIS Land Processes DAAC.

Edgar, G. J., & Stuart-Smith, R. D. (2014). Systematic global assessment of reef fish communities by the Reef Life Survey program. Scientific Data, 1, 140007.

Field, R., Hawkins, B. A., Cornell, H. V., Currie, D. J., Diniz-Filho, J. A. F., Guégan, J. F., ... & O'Brien, E. M. (2009). Spatial species-richness gradients across scales: a meta-analysis. Journal of biogeography, 36(1), 132-147.

Francis, A. P., & Currie, D. J. (2003). A globally consistent richness-climate relationship for angiosperms. The American Naturalist, 161(4), 523-536.

Fox, J.A. (2011) Zombie ideas in ecology. In: http://oikosjournal.wordpress.com/2011/06/17/zombie-ideas-in-ecology/

Gaston, K. J. (2000). Global patterns in biodiversity. Nature, 405(6783), 220.

Gotelli, N. J., & Colwell, R. K. (2011). Estimating species richness. Biological diversity: frontiers in measurement and assessment, 12, 39-54.

Hawkins, B. A., Field, R., Cornell, H. V., Currie, D. J., Guégan, J. F., Kaufman, D. M., ... & Porter, E. E. (2003). Energy, water, and broad-scale geographic patterns of species richness. Ecology, 84(12), 3105-3117.

Hawkins, B. A., Diniz-Filho, J. A. F., & Weis, A. E. (2005). The mid-domain effect and diversity gradients: is there anything to learn?. The American Naturalist, 166(5), E140-E143.

Hessler, R. R., & Sanders, H. L. (1967, February). Faunal diversity in the deep-sea. In Deep Sea Research and Oceanographic Abstracts (Vol. 14, No. 1, pp. 65-78). Elsevier.

Hillebrand, H. (2004). On the generality of the latitudinal diversity gradient. The American Naturalist, 163(2), 192-211.

Hubbell, S. P., He, F., Condit, R., Borda-de-Água, L., Kellner, J., & ter Steege, H. (2008). How many tree species are there in the Amazon and how many of them will go extinct?. Proceedings of the National Academy of Sciences, 105(Supplement 1), 11498-11504.

Hurlbert, A. H., & Haskell, J. P. (2002). The effect of energy and seasonality on avian species richness and community composition. The American Naturalist, 161(1), 83-97.

Hurlbert, A. H., & White, E. P. (2005). Disparity between range map-and survey-based analyses of species richness: patterns, processes and implications. Ecology Letters, 8(3), 319-327.

Hutchinson, G. E. (1959). Homage to Santa Rosalia or why are there so many kinds of animals?. The American Naturalist, 93(870), 145-159.

Jacquemyn, H., Butaye, J., & Hermy, M. (2001). Forest plant species richness in small, fragmented mixed deciduous forest patches: the role of area, time and dispersal limitation. Journal of Biogeography, 28(6), 801-812.

Karger, D. N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R. W., ... & Kessler, M. (2017). Climatologies at high resolution for the earth's land surface areas. Scientific data, 4, 170122.

Kerr, J. T., & Packer, L. (1997). Habitat heterogeneity as a determinant of mammal species richness in high-energy regions. Nature, 385(6613), 252.

Kleidon, A. and H. Mooney, A, (2000). "A global distribution of biodiversity inferred from climatic constraints: results from a process-based modelling study." Global Change Biology 6: 507-523.

Klopfer, P. H. (1959). Environmental determinants of faunal diversity. The American Naturalist, 93(873), 337-342.

Latham, R. E., & Ricklefs, R. E. (1993). Global patterns of tree species richness in moist forests: energy-diversity theory does not account for variation in species richness. Oikos, 325-333.

Leitao, R. P., Zuanon, J., Villéger, S., Williams, S. E., Baraloto, C., Fortunel, C., ... & Mouillot, D. (2016). Rare species contribute disproportionately to the functional structure of species assemblages. Proceedings of the Royal Society B: Biological Sciences, 283(1828), 20160084.

Lieberman, D., Lieberman, M., Peralta, R., & Hartshorn, G. S. (1996). Tropical forest structure and composition on a large-scale altitudinal gradient in Costa Rica. Journal of Ecology, 137-152.

Liu, Z., Liu, G., Fu, B., & Zheng, X. (2008). Relationship between plant species diversity and soil microbial functional diversity along a longitudinal gradient in temperate grasslands of Hulunbeir, Inner Mongolia, China. Ecological Research, 23(3), 511-518.

MacArthur, R. H., & MacArthur, J. W. (1961). On bird species diversity. Ecology, 42(3), 594-598.

MacArthur, R. H., & Wilson, E. O. (1963). An equilibrium theory of insular zoogeography. Evolution, 17(4), 373-387.

MacArthur, R. H. (1965). Patterns of species diversity. Biological reviews, 40(4), 510-533.

Magurran, A. E. (1988). Ecological Diversity and its Measurement. Princeton NJ: Princeton University Press.

Matthews, T. J., Borregaard, M. K., Ugland, K. I., Borges, P. A. V., Rigal, F., Cardoso, P., & Whittaker, R. J. (2014). The gambin model provides a superior fit to species abundance distributions with a single free parameter: evidence, implementation and interpretation. Ecography, 37(10), 1002-1011.

McCain, C. M., & Grytnes, J. A. (2001). Elevational gradients in species richness. e LS.

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

McGill, B. J., Maurer, B. A., & Weiser, M. D. (2006). Empirical evaluation of neutral theory. Ecology, 87(6), 1411-1423.

Mee, J. A., Robins, G. L., & Post, J. R. (2018). Patterns of fish species distributions replicated across three parallel rivers suggest biotic zonation in response to a longitudinal temperature gradient. Ecology of freshwater fish, 27(1), 44-61.

Mellin, C., Bradshaw, C. J. A., Meekan, M. G., & Caley, M. J. (2010). Environmental and spatial predictors of species richness and abundance in coral reef fishes. Global Ecology and Biogeography, 19(2), 212-222.

NASA Ocean Biology Processing Group. (2015). MODIS-Aqua Level 2 Ocean Color Data Version 2014 [Data set].

O'Grady, J. J., Reed, D. H., Brook, B. W., & Frankham, R. (2004). What are the best correlates of predicted extinction risk?. Biological Conservation, 118(4), 513-520.

Ohlemüller, R., Anderson, B. J., Araújo, M. B., Butchart, S. H., Kudrna, O., Ridgely, R. S., & Thomas, C. D. (2008). The coincidence of climatic and species rarity: high risk to small-range species from climate change. Biology letters, 4(5), 568-572.

Palmer, M. W. (1994). Variation in species richness: towards a unification of hypotheses. Folia Geobotanica et Phytotaxonomica, 29(4), 511.

Pardieck, K.L., D.J. Ziolkowski Jr., M. Lutmerding, V. Aponte and M-A.R. Hudson. 2019. North American Breeding Bird Survey Dataset 1966 - 2018, version 2018.0. U.S. Geological Survey, Patuxent Wildlife Research Center. https://doi.org/10.5066/P9HE8XYJ.

Parravicini, V., Kulbicki, M., Bellwood, D. R., Friedlander, A. M., Arias-Gonzalez, J. E., Chabanet, P., ... & Mouillot, D. (2013). Global patterns and predictors of tropical reef fish species richness. Ecography, 36(12), 1254-1262.

Pianka, E. R. (1966). Latitudinal gradients in species diversity: a review of concepts. The American Naturalist, 100(910), 33-46.

Preston, F. W. (1960). Time and space and the variation of species. Ecology, 41(4), 611-627.

Preston, F. W. (1962). The canonical distribution of commonness and rarity: Part I. Ecology, 43(2), 185-215.

Phillips, O., & Miller, J. S. (2002). Global patterns of plant diversity: Alwyn H. Gentry's forest transect data set.

Pimm, S. L., Jones, H. L., & Diamond, J. (1988). On the risk of extinction. The American Naturalist, 132(6), 757-785.

Purvis, A., Gittleman, J. L., Cowlishaw, G., & Mace, G. M. (2000). Predicting extinction risk in declining species. Proceedings of the royal society of London. Series B: Biological Sciences, 267(1456), 1947-1952.

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Rahbek, C., & Graves, G. R. (2001). Multiscale assessment of patterns of avian species richness. Proceedings of the National Academy of Sciences, 98(8), 4534-4539.

Rahbek, C., Gotelli, N. J., Colwell, R. K., Entsminger, G. L., Rangel, T. F. L., & Graves, G. R. (2006). Predicting continental-scale patterns of bird species richness with spatially explicit models. Proceedings of the Royal Society B: Biological Sciences, 274(1607), 165-174.

Rex, M. A., Stuart, C. T., & Coyne, G. (2000). Latitudinal gradients of species richness in the deep-sea benthos of the North Atlantic. Proceedings of the National Academy of Sciences, 97(8), 4082-4085.

Rosenzweig, M. L. (1995). Species Diversity in Time and Space. Cambridge UK, Cambridge University Press. Chapter 2.

Sax, D. F., Gaines, S. D., & Brown, J. H. (2002). Species invasions exceed extinctions on islands worldwide: a comparative study of plants and birds. The American Naturalist, 160(6), 766-783.

Schurr, F. M., Midgley, G. F., Rebelo, A. G., Reeves, G., Poschlod, P., & Higgins, S. I. (2007). Colonization and persistence ability explain the extent to which plant species fill their potential range. Global Ecology and Biogeography, 16(4), 449-459

Sclater, P. L. (1858). "On the general geographical distribution of the members of the class Aves." Journal of the Linneaen Society of London, Zoology 2: 118-133 (reprinted in Lomolino, M. V., D. F. Sax, and J. H. Brown. 2004. Foundations of Biogeography, University of Chicago Press pp. 2118-2133).

Seibold, S., Bässler, C., Brandl, R., Büche, B., Szallies, A., Thorn, S., ... & Müller, J. (2016). Microclimate and habitat heterogeneity as the major drivers of beetle diversity in dead wood. Journal of Applied Ecology, 53(3), 934-943.

Sharma, C. M., Suyal, S., Gairola, S., & Ghildiyal, S. K. (2009). Species richness and diversity along an altitudinal gradient in moist temperate forest of Garhwal Himalaya. Journal of American Science, 5(5), 119-128.

Srivastava, D. S., & Lawton, J. H. (1998). Why more productive sites have more species: an experimental test of theory using tree-hole communities. The American Naturalist, 152(4), 510-529.

Storch, D., Bohdalková, E., & Okie, J. (2018). The more-individuals hypothesis revisited: the role of community abundance in species richness regulation and the productivity–diversity relationship. Ecology letters, 21(6), 920-937.

Svenning, J. C., & Skov, F. (2007). Could the tree diversity pattern in Europe be generated by postglacial dispersal limitation?. Ecology letters, 10(6), 453-460.

Ulrich, W., Ollik, M., & Ugland, K. I. (2010). A meta-analysis of species–abundance distributions. Oikos, 119(7), 1149-1155.

Wallace, A. R. (1878). Tropical Nature and other Essays. New York, Macmillan.

Werner, E. E., Skelly, D. K., Relyea, R. A., & Yurewicz, K. L. (2007). Amphibian species richness across environmental gradients. Oikos, 116(10), 1697-1712.

Wiens, J. J., & Donoghue, M. J. (2004). Historical biogeography, ecology and species richness. Trends in ecology & evolution, 19(12), 639-644.

Woolley, S. N., Tittensor, D. P., Dunstan, P. K., Guillera-Arroita, G., Lahoz-Monfort, J. J., Wintle, B. A., ... & O'Hara, T. D. (2016). Deep-sea diversity patterns are shaped by energy availability. Nature, 533(7603), 393.

Wright, D. H. (1983). Species-energy theory: an extension of species-area theory. Oikos, 496-506.

Tables and Figures

Table 1. The five datasets used in the analysis, the taxon and location they represent, and the source of the data.

Dataset	Taxon	Gradient	Location	Source
Gentry	Trees	Climatic (temperature and precipitation)	Global, mostly in the neo-tropics	Phillips & Miller 2002
Barva	Trees	Elevational	Barva volcano, Costa Rica (elevational gradient)	Lieberman et al. 1996
Breeding Bird Survey	Birds	Vegetation cover	North America	Pardieck et al. 2019
Reef Life Survey	Reef fish	Temperature and resource availability (chlorophyll *a*)	Global, mostly in the tropics	Edgar & Stuart-Smith 2014
Ocean Basin Benthos	Gastropods	Depth	North Atlantic	Hessler & Sanders 1967

Table 2. Variance explained (adjusted R^2) by various regression models, for each of the five datasets. Column headers indicate the regression, based on the following abbreviations: N = species richness, E = environmental variable(s)[a], I = total number of individuals, ϕ = proxy for rarity. The symbol ~ denotes a general linear model. All models are significant except for where indicated by an asterisk (*).

Dataset	$N \sim E$	$N \sim I$	$N \sim \phi$	$N \sim \phi+I$	$\phi \sim E$	$I \sim E$
Barva	0.94	0.08*	0.92	0.91	0.93	0.20*
Gastropods	0.36	0.14	0.39	0.76	0.16	0.26
Gentry	0.44	0.12	0.94	0.96	0.40	-0.02
Breeding Bird Survey	0.23	0.28	0.52	0.63	0.12	0.10
Reef Life Survey	0.32	0.28	0.59	0.77	0.23	0.09

a) Environmental variables for each dataset are as follows:
Barva = elevation,
Gastropods = depth,
Gentry = temperature and precipitation,
BBS = NDVI, temperature, precipitation,
RLS = temperature, chlorophyll *a*.

Table 3. The variability in species richness, number of individuals (I), and ϕ (a proxy for rarity) in each of the five datasets used in this study. Both the coefficient of variation and orders of magnitude are measures of variability. The variable (I or ϕ) with the highest variability in each dataset is indicated in bold.

	Coefficient of variation (SD/mean) as a percentage		
Dataset	*Richness*	*Individuals*	ϕ
Barva	44.2%	12.5%	**59.7%**
Gastropods	47.0%	**245.5%**	53.0%
Gentry	47.1%	25.9%	**58.8%**
Breeding Bird Survey	24.5%	**54.4%**	35.7%
Reef Life Survey	75.5%	**176.0%**	72.4%

	Variation, Orders of magnitude		
Dataset	*Richness*	*Individuals*	ϕ
Barva	0.71	0.18	**0.98**
Gastropods	1.06	**2.08**	1.23
Gentry	1.24	0.60	**2.17**
Breeding Bird Survey	0.93	1.47	**1.48**
Reef Life Survey	1.93	**2.61**	1.46

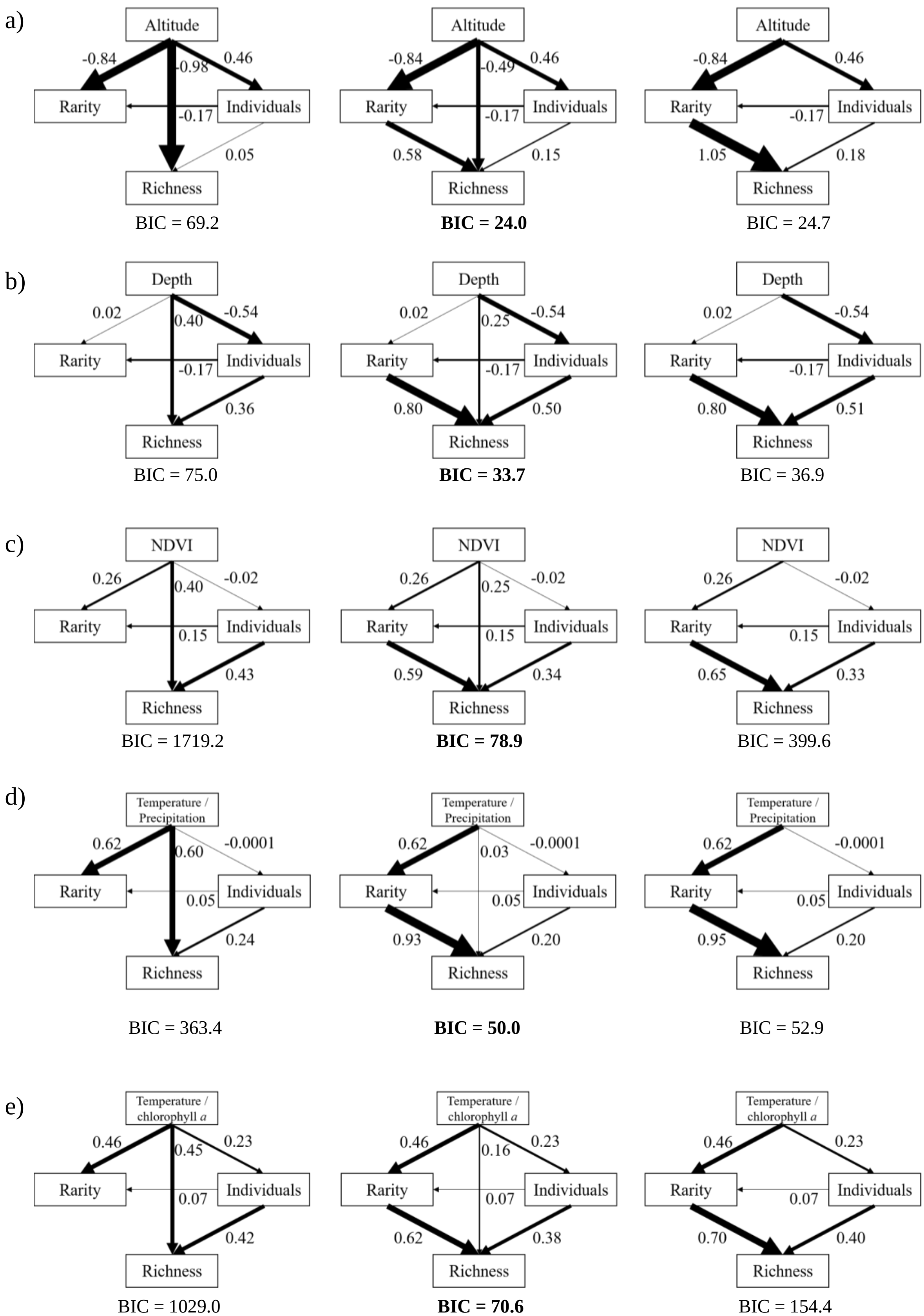
a)
Altitude
-0.84
-0.98
0.46
Rarity
-0.17
Individuals
0.05
Richness
BIC = 69.2
Altitude
-0.84
-0.49
0.46
Rarity
-0.17
Individuals
0.58
0.15
Richness
BIC = 24.0
Altitude
-0.84
0.46
Rarity
-0.17
Individuals
1.05
0.18
Richness
BIC = 24.7
b)
Depth
0.02
0.40
-0.54
Rarity
-0.17
Individuals
0.36
Richness
BIC = 75.0
Depth
0.02
0.25
-0.54
Rarity
-0.17
Individuals
0.80
0.50
Richness
BIC = 33.7
Depth
0.02
-0.54
Rarity
-0.17
Individuals
0.80
0.51
Richness
BIC = 36.9
c)
NDVI
0.26
0.40
-0.02
Rarity
0.15
Individuals
0.43
Richness
BIC = 1719.2
NDVI
0.26
0.25
-0.02
Rarity
0.15
Individuals
0.59
0.34
Richness
BIC = 78.9
NDVI
0.26
-0.02
Rarity
0.15
Individuals
0.65
0.33
Richness
BIC = 399.6
d)
Temperature / Precipitation
0.62
0.60
-0.0001
Rarity
0.05
Individuals
0.24
Richness
BIC = 363.4
Temperature / Precipitation
0.62
0.03
-0.0001
Rarity
0.05
Individuals
0.93
0.20
Richness
BIC = 50.0
Temperature / Precipitation
0.62
-0.0001
Rarity
0.05
Individuals
0.95
0.20
Richness
BIC = 52.9
e)
Temperature / chlorophyll a
0.46
0.45
0.23
Rarity
0.07
Individuals
0.42
Richness
BIC = 1029.0
Temperature / chlorophyll a
0.46
0.16
0.23
Rarity
0.07
Individuals
0.62
0.38
Richness
BIC = 70.6
Temperature / chlorophyll a
0.46
0.23
Rarity
0.07
Individuals
0.70
0.40
Richness
BIC = 154.4

Figure 1. Competing path models for species richness, based on environmental variables (top), rarity, and the number of individuals, for each dataset: a) Barva, b) Gastropods, c) Breeding Bird Survey, d) Gentry, e) Reef Life Survey. The centre model includes all pathways, the left model eliminates the direct effect of rarity on richness, and the right model eliminates the direct effect of the environmental variable(s) on richness. The thickness of the arrows is proportional to the effect size. All models were evaluated using the Bayesian Information Criterion (BIC). The data have the highest likelihood of having been produced by the pathways shown in the path model with the lowest BIC (shown in bold).

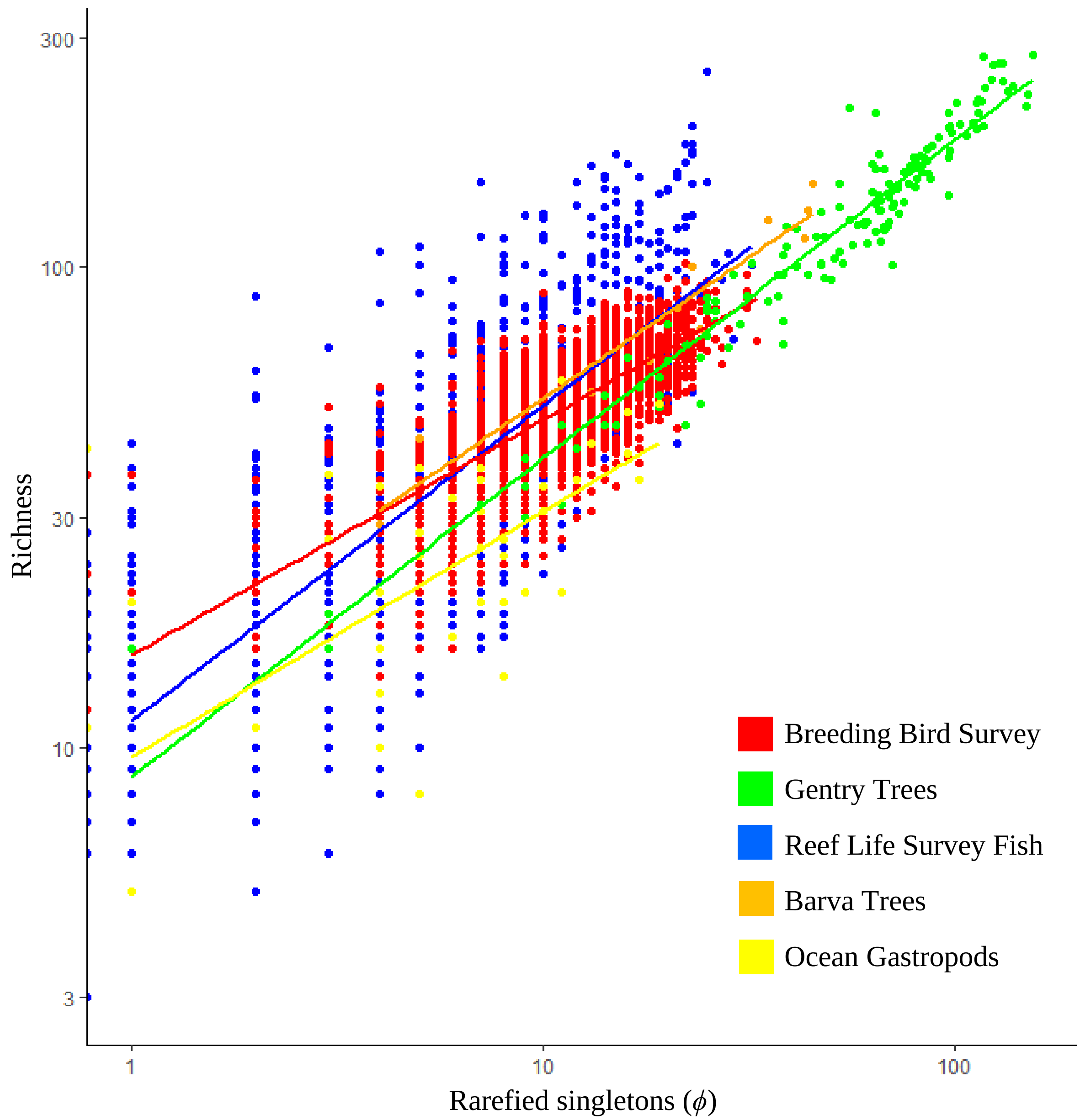

Figure 2. The relationship between rarity and species richness in all five datasets, with each point representing a single sampling unit (plot, route, site, etc.) and with each dataset distinguished by color. Regression lines are shown in matching color. Note that both axes are log-transformed.

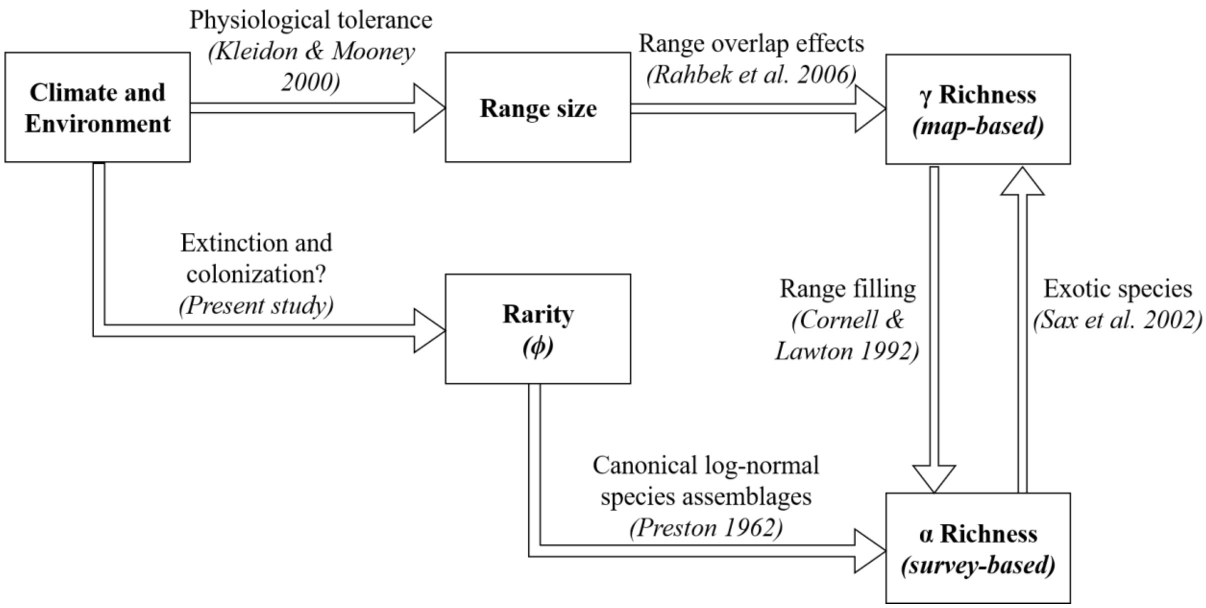

Figure 3. A proposed general framework that links the mechanism discussed in the current paper to the literature.

Table S1. Results of Komolgorov-Smirnov tests for log-normality in each of the five datasets.

Dataset	K-S statistic	p-value
Gentry	0.50	$< 2.2\times10^{-16}$
BBS	0.60	$< 2.2\times10^{-16}$
Reef Life	0.54	$< 2.2\times10^{-16}$
Gastropods	0.53	$< 2.2\times10^{-16}$
Barva	0.57	$< 2.2\times10^{-16}$

Table S2. Results of chi-square tests on two competing pathways- the sub-model of the main path models that represents our main hypothesis, and one where rarity (ϕ) and richness (N) are switched.

Dataset	E→ϕ→N χ2 p-value	E→N→ ϕ χ2 p-value
Gentry	0.070	0.133
BBS	0.000	0.225
Reef Life	0.000	0.000
Gastropods	0.065	0.118
Barva	0.056	0.024

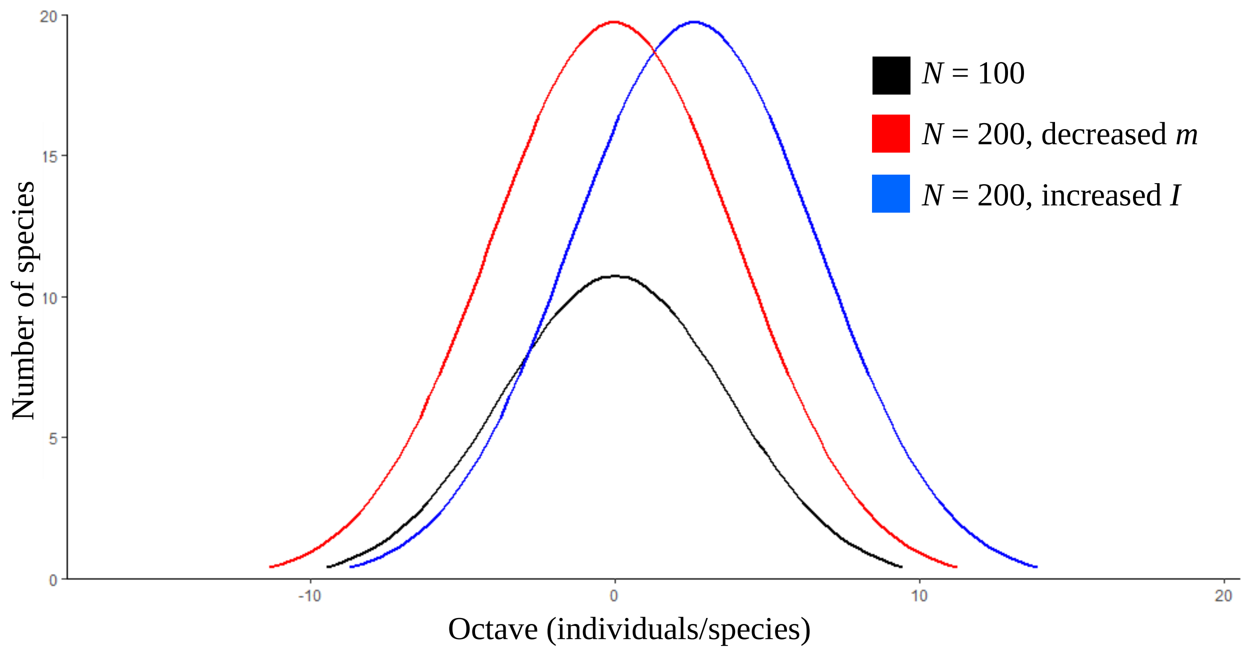

Figure S1. The effects of varying the two free parameters, *m* and *I*, on Preston's log-normal distribution. *m* represents the abundance of the rarest species in the assemblage, and *I* represents the total number of individuals. Octaves constitute twofold increases in the number of individuals per species. Comparing the black curve to the red shows the effect of increasing *N* by decreasing *m*; comparing the black curve to the blue shows the effect of increasing *N* by increasing *I*.

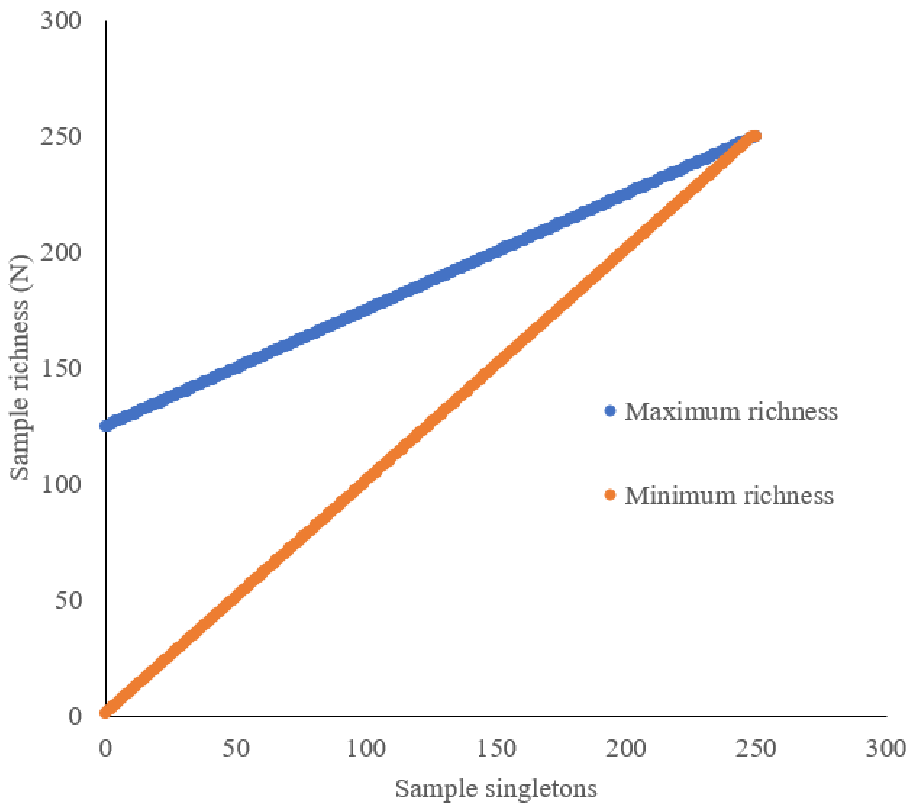

Figure S2. The bounds of the possible relationship between richness and singletons in a sample, where singletons cannot exceed richness, and richness cannot be lower than singletons. Especially for low to medium numbers of singletons, the bounds on the possible values are quite wide.

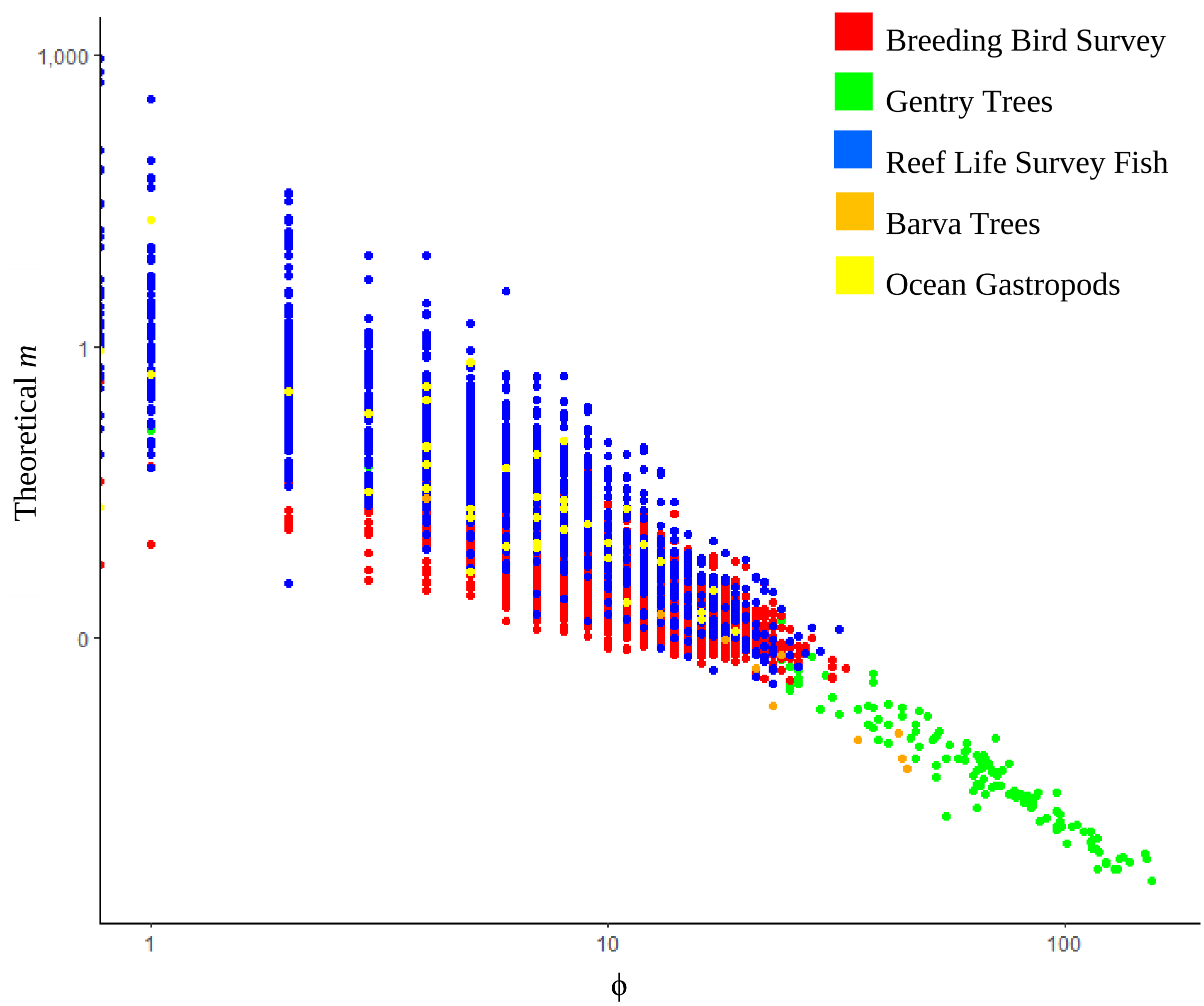

Figure S3. The relationship between a theoretically computed *m*, based on equation (1), and our proxy ϕ, in each of the five datasets. The strong correlation justifies our use of the proxy to estimate *m*. The correlation is negative because greater rarity is indicated by decreasing *m*, but by increasing ϕ.

Appendix

GENTRY

```
Call:
lm(formula = (I(log(Richness))) ~ (I(log(rare_sing))) + (I(log(Individuals))),
    data = GENrarefiedno0)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.29005 -0.05433  0.00866  0.05944  0.60760 

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)    
(Intercept)         -0.41165    0.24902  -1.653      0.1    
I(log(rare_sing))    0.64955    0.01100  59.029   <2e-16 ***
I(log(Individuals))  0.44340    0.04228  10.487   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1117 on 146 degrees of freedom
Multiple R-squared:  0.9632,  Adjusted R-squared:  0.9627 
F-statistic:  1908 on 2 and 146 DF,  p-value: < 2.2e-16
```

BBS

```
Call:
lm(formula = (I(log(Richness))) ~ (I(log(rare_sing))) + (I(log(Individuals))),
    data = BBSrarefiedno0)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.74159 -0.09317  0.01306  0.10662  0.76347 

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)    
(Intercept)         1.519751   0.047289   32.14   <2e-16 ***
I(log(rare_sing))   0.437433   0.008186   53.44   <2e-16 ***
I(log(Individuals)) 0.209994   0.007359   28.53   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1645 on 2654 degrees of freedom
Multiple R-squared:  0.6336,  Adjusted R-squared:  0.6334 
F-statistic:  2295 on 2 and 2654 DF,  p-value: < 2.2e-16
```

RLS

```
Call:
lm(formula = (I(log(Richness))) ~ (I(log(rare_sing))) + (I(log(Individuals))),
    data = RLSrarefiedno0)

Residuals:
     Min       1Q   Median       3Q      Max 
-1.28545 -0.20329  0.01185  0.21808  1.29843 

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)    
(Intercept)         0.651797   0.062909   10.36   <2e-16 ***
I(log(rare_sing))   0.589428   0.012353   47.72   <2e-16 ***
I(log(Individuals)) 0.242961   0.008019   30.30   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
Residual standard error: 0.3254 on 1128 degrees of freedom
Multiple R-squared:  0.7731,  Adjusted R-squared:  0.7727
F-statistic:  1922 on 2 and 1128 DF,  p-value: < 2.2e-16
```

GASTROPODS

```
Call:
lm(formula = (I(log(Richness))) ~ (I(log(rare_sing))) + (I(log(Individuals))),
    data = OBGrarefiedno0)

Residuals:
     Min       1Q   Median       3Q      Max
-0.57200 -0.15717  0.00919  0.18031  0.62291

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)         -0.22466    0.35016  -0.642    0.525
I(log(rare_sing))    0.69493    0.06916  10.048 7.48e-12 ***
I(log(Individuals))  0.36650    0.04837   7.578 6.98e-09 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2684 on 35 degrees of freedom
Multiple R-squared:  0.7739,  Adjusted R-squared:  0.761
F-statistic: 59.91 on 2 and 35 DF,  p-value: 5.005e-12
```

BARVA

```
Call:
lm(formula = (I(log(Richness))) ~ (I(log(rare_sing))) + (I(log(Individuals))),
    data = BVTrarefiedno0)

Residuals:
     Min       1Q   Median       3Q      Max
-0.21218 -0.08619 -0.04703  0.09897  0.22685

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)          0.72992    2.85731   0.255    0.805
I(log(rare_sing))    0.61163    0.06672   9.167 1.62e-05 ***
I(log(Individuals))  0.29028    0.43781   0.663    0.526
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.15 on 8 degrees of freedom
Multiple R-squared:  0.9305,  Adjusted R-squared:  0.9132
F-statistic: 53.59 on 2 and 8 DF,  p-value: 2.327e-05
```