8-1-2015

Ecological and biogeographic null hypotheses for comparing rarefaction curves

Luis Cayuela
Universidad Rey Juan Carlos

Nicholas J. Gotelli
University of Vermont

Robert K. Colwell
University of Connecticut

Recommended Citation

Cayuela L, Gotelli NJ, Colwell RK. Ecological and biogeographic null hypotheses for comparing rarefaction curves. Ecological Monographs. 2015 Aug;85(3):437-55.

Ecological Monographs, 85(3), 2015, pp. 437–455

Ecological and biogeographic null hypotheses for comparing rarefaction curves

Luis Cayuela,[1,5] Nicholas J. Gotelli,[2] and Robert K. Colwell[3,4]

[1]*Departamento de Biología y Geología, Universidad Rey Juan Carlos, c/ Tulipán s/n, E-28933 Móstoles, Madrid, Spain*
[2]*Department of Biology, University of Vermont, Burlington, Vermont 05405 USA*
[3]*Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, Connecticut 06269 USA*
[4]*University of Colorado Museum of Natural History, Boulder, Colorado 80309 USA*

Abstract. The statistical framework of rarefaction curves and asymptotic estimators allows for an effective standardization of biodiversity measures. However, most statistical analyses still consist of point comparisons of diversity estimators for a particular sampling level. We introduce new randomization methods that incorporate sampling variability encompassing the entire length of the rarefaction curve and allow for statistical comparison of $i \geq 2$ individual-based, sample-based, or coverage-based rarefaction curves. These methods distinguish between two distinct null hypotheses: the ecological null hypothesis ($H_{0\text{eco}}$) and the biogeographical null hypothesis ($H_{0\text{biog}}$).

$H_{0\text{eco}}$ states that the i samples were drawn from a single assemblage, and any differences among them in species richness, composition, or relative abundance reflect only sampling effects. $H_{0\text{biog}}$ states that the i samples were drawn from assemblages that differ in their species composition but share similar species richness and species abundance distributions. To test $H_{0\text{eco}}$, we created a composite rarefaction curve by summing the abundances of all species from the i samples. We then calculated a test statistic Z_{eco}, the (cumulative) summed areas of difference between each of the i individual curves and the composite curve. For $H_{0\text{biog}}$, the test statistic Z_{biog} was calculated by summing the area of difference between all possible pairs of the i individual curves. Bootstrap sampling from the composite curve ($H_{0\text{eco}}$) or random sampling from different simulated assemblages using alternative abundance distributions ($H_{0\text{biog}}$) was used to create the null distribution of Z, and to provide a frequentist test of $Z \mid H_0$. Rejection of $H_{0\text{eco}}$ does not pinpoint whether the samples differ in species richness, species composition, and/or relative abundance.

In benchmark comparisons, both tests performed satisfactorily against artificial data sets randomly drawn from a single assemblage (low Type I error). In benchmark comparisons with different species abundance distributions and richness, the tests had adequate power to detect differences among curves (low Type II error), although power diminished at small sample sizes and for small differences among underlying species rank abundances.

Key words: biogeography; community ecology; Hill numbers; rarefaction; relative abundance; species composition; species diversity; statistical test.

Introduction

Quantifying biodiversity and comparing diversity among samples is a key activity in ecology and conservation biology (Magurran and McGill 2011), as well as in emerging "-omics" subdisciplines (i.e., genomics, proteomics, metabolomics; Gotelli et al. 2012). Biodiversity metrics typically reflect species richness and relative abundance, but many indices can be extended to encompass measures of phylogenetic (Chao et al. 2010), functional (Villeger et al. 2008), or trait (Violle et al. 2007) diversity. Most diversity indices are sensitive to sampling effort, and will continue to increase, albeit more and more slowly, as more samples or individuals are collected (Gotelli and Colwell 2001). Even when standardized sampling protocols are used, the abundance of organisms per sample can often differ substantially (Stevens and Carson 1999), which complicates the direct comparison of diversity among samples (James and Wamer 1982).

In theory, these sampling effects can be overcome by collecting additional samples or individuals until the species accumulation curve reaches an asymptote; additional sampling beyond the asymptote will not add more species and will not change the relative abundance distribution (Gotelli and Chao 2013). In practice, most biodiversity samples lie substantially below the asymptote (Chao et al. 2009), with many rare species in the underlying assemblage missing from the sample (Coddington et al. 2009). Sampling effects can be controlled for by randomly subsampling biodiversity data to a standardized sampling effort (rarefaction; Hurlbert

Manuscript received 30 June 2014; revised 30 January 2015; accepted 13 February 2015. Corresponding Editor: B. D. Inouye.

[5] E-mail: luis.cayuela@urjc.es

1971), or by extrapolating biodiversity metrics toward a theoretical asymptote (nonparametric extrapolation [Colwell et al. 2012, Chao et al. 2014] and nonparametric asymptotic estimators [Colwell and Coddington 1994]). Other methods, such as curve-fitting to the species accumulation curve (Soberon and Llorente 1993) or fitting parametric models to the species abundance distribution (Connolly et al. 2009), have usually not performed as well as rarefaction, nonparametric extrapolation, or asymptotic estimators (Brose et al. 2003, Walther and Moore 2005).

Colwell et al. (2012) recently unified the statistical framework for rarefaction, extrapolation, and asymptotic estimators, and showed that a single curve (with an expectation and an unconditional variance) represents the statistical expectation of the accumulation curve of species richness, for both rarefaction and extrapolation. Chao et al. (2014) extended these results to other diversity metrics in the family of Hill numbers (Hill 1973). Biodiversity sample data from different assemblages can then be effectively compared based on a standardized number of individuals (Gotelli and Colwell 2001), a standardized number of samples (Colwell et al. 2004), or a standardized coverage level (Chao and Jost 2012).

In spite of these recent advances, most diversity comparisons are made at a single level of sampling effort. For rarefaction, curves are typically compared at the abundance of the smallest sample in the collection, whereas asymptotic estimators represent, by definition, the diversity that would be expected at 100% sample coverage. Point comparisons of rarefaction curves can be problematic because diversity of all samples diminishes and converges at small sample sizes (Tipper 1979). At the other extreme, point comparisons of asymptotic estimators can be problematic because extrapolated estimators often have very large variances (Colwell et al. 2012), especially for species richness and other metrics that are sensitive to contributions from rare species (Chao et al. 2014). Moreover, some rarefaction and extrapolation curves may cross one or more times, so that the rank order of diversity measured among samples could change depending on the sampling effort that is used for standardized comparisons (Chao and Jost 2012). For these reasons, simple point comparisons of diversity at particular sample levels, which often use parametric statistics and assume a symmetric Gaussian distribution (Payton et al. 2003), may not be satisfactory.

In this study, we develop simple randomization tests for comparing the overall shape of two or more individual- or sample-based rarefaction curves. In the diversity literature, there are actually two distinct null hypotheses that have not always been clearly distinguished. The ecological null hypothesis $H_{0\mathrm{eco}}$ is that two or more samples were drawn randomly from the same underlying assemblage. If this hypothesis is true, then heterogeneity among the samples in their composition, species richness, and relative abundance is no greater than would be expected by random sampling from a single assemblage. This null hypothesis is appropriate for samples collected at a relatively small spatial scale that should potentially share most species. Deviations from the ecological null hypothesis might reflect spatial beta diversity (Anderson et al. 2011), the influence of species interactions (Chase and Leibold 2003), habitat filtering (Baldeck et al. 2013), mass effects (Amarasekare and Nisbet 2001), abiotic gradients (Wilson and Tilman 1991), and other community assembly rules (Weiher and Keddy 1999) or metacommunity processes (Leibold et al. 2004). If the ecological null hypothesis is true, differences among samples in species composition should be relatively modest, and occur mostly among the rarer species.

The biogeographical null hypothesis $H_{0\mathrm{biog}}$ is that two or more samples were drawn from assemblages that all share a common underlying species richness and relative abundance profile, regardless of their species composition. Therefore, heterogeneity among the samples in their species richness or relative abundance (but not in their species composition) is no greater than would be expected by random sampling from a single relative abundance distribution. This null hypothesis is appropriate for samples collected at larger spatial scales (relative to the organisms), from local habitat gradients or patches with partially shared biotas, to regions or even whole continents, whose floras and faunas may have evolved in relative isolation and share few or no species, but may have been exposed to relatively similar abiotic conditions. Deviations from the biogeographical null hypothesis might reflect the influence of distinctive local conditions (Qian and Ricklefs 2008), or unique historical events, such as natural or anthropogenic extinctions (Alroy 2010), adaptive radiation (Losos 2010), or the emergence and breakdown of biogeographic barriers (Wiens and Donoghue 2004). If the biogeographical null hypothesis is true, species richness and relative abundance should be relatively similar among samples, regardless of differences in species composition.

Axiomatic relationships exist between the ecological and the biogeographical null hypotheses (Fig. 1). However, the ecological null hypothesis, tested by itself, only reveals whether samples are more different than would be expected if they were drawn from a single underlying assemblage. In order to understand whether assemblages differ in species richness, species composition, or relative abundance, it is necessary to test both the ecological and the biogeographical null hypotheses, and to carefully compare the results of both tests. If $H_{0\mathrm{eco}}$ is not rejected, the same result should be obtained with $H_{0\mathrm{biog}}$ (Fig. 1a). However, when $H_{0\mathrm{eco}}$ is rejected, the samples from different habitats or regions may (Fig. 1b) or may not (Fig. 1c) exhibit rarefaction curves with statistically indistinguishable profiles. Similarly, if $H_{0\mathrm{biog}}$ cannot be rejected, we may infer that the compared

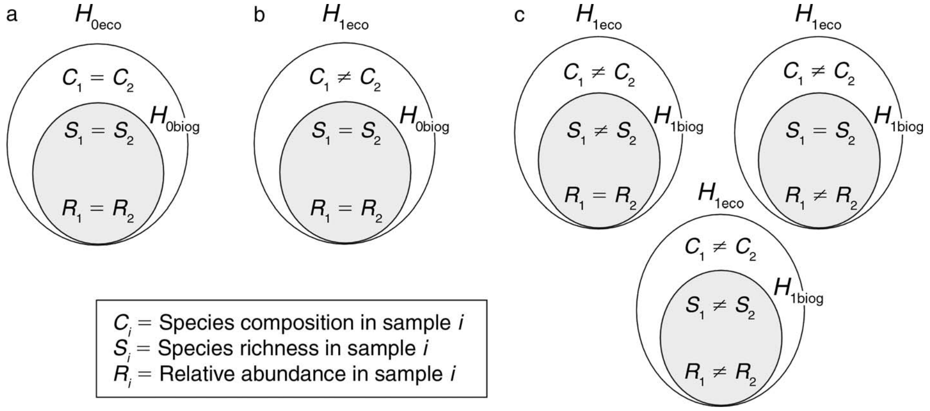


FIG. 1. Relationships between the ecological (H_{0eco}) and the biogeographical (H_{0biog}) null hypotheses. Note that the H_{0eco} encompasses three properties of ecological communities: species composition, richness, and relative abundance (outer circle), whereas the H_{0biog} only comprises two out of these three properties, namely species richness and relative abundance (inner shaded circle). (a) If two or more samples are drawn from the same assemblage, their species composition, richness, and relative abundance will be similar, and both the H_{0eco} and the H_{0biog} will be accepted. (b) If samples are drawn from two assemblages with similar species richness and relative abundance but different species composition, the H_{0eco} will be rejected, whereas the H_{0biog} will not. (c) If samples are drawn from two assemblages with different species composition and either different species richness, relative abundance, or both, then both H_{0eco} and the H_{0biog} will be rejected.

samples are similar regarding richness and relative abundance, regardless of whether they share most (Fig. 1a), few, or no (Fig. 1b) species, but samples for which the biogeographic null hypothesis is rejected will also, necessarily, appear nonrandom when compared to the ecological null hypothesis (Fig. 1c). For examples, biotas from tropical rainforests of Africa and Asia might differ completely in species composition (few or no shared species), but might exhibit similar species richness and species relative abundance because of constraints imposed by similar climates. If so, samples from the two continents would reject H_{0eco}, but might not reject H_{0biog}. If both null hypotheses are rejected, then the assemblages differ in species composition, as well as in species richness and relative abundance.

In this study, we developed randomization algorithms to test both the ecological and the biogeographical null hypotheses with sample- or individual-based ecological data. This broadens the scope of rarefaction curves to test relevant null hypotheses regarding community structure. If sample sizes are equal, our results may be similar to those obtained with multivariate approaches, such as distance-based dissimilarity measurements (Legendre and Gallagher 2001, Clarke et al. 2006, De Cáceres et al. 2013). However, it is often the case that comparisons are desired for data that are unequally sampled, in which case randomization tests based on rarefaction may offer some distinct advantages.

To evaluate the performance of these algorithms and their vulnerability to Type I or Type II statistical error, we applied them to a series of artificial benchmark data sets that were either drawn from the same assemblage, or drawn from multiple assemblages that differed systematically in their species composition, richness, or relative abundance distributions. Finally, we illustrated the use of these methods in an analysis of forest tree data from six 20–52-ha plots in tropical regions around the world, and to a smaller-scale transect survey of trees in montane cloud forests sampled in three regions of Chiapas, southern Mexico.

METHODS

Notation and organization of biodiversity data

Following Colwell et al. (2012) and Gotelli and Chao (2013), we use a common set of notation to describe biodiversity sampling data. Consider a complete assemblage for which all species and their relative abundances are known. In this complete assemblage, there are $i = 1$ to S species and N^* total individuals, with N_i individuals of species i. For individual-based (abundance) data, the reference sample consists of n individuals drawn at random from N^*, with S_{obs} species present, each represented by X_i individuals. Individual-based data can be represented as a single vector of length S_{obs}, the elements of which are the observed abundances X_i.

For sample-based incidence data, the reference sample consists of a set of T standardized sampling units, such as traps, plots, transect lines, etc. Within each of these sampling units, the presence (1) or absence (0) of each species are the required data, even though abundance data may have been collected. Sample-based incidence data can be represented as a single matrix, with $i = 1$ to S_{obs} rows and $j = 1$ to T columns, and entries $W_{ij} = 1$ or $W_{ij} = 0$ to indicate the presence or absence of species i in sampling unit j.

In this study, we made extensive use of rarefaction curves for both individual- and sample-based data. In the past, rarefaction curves have been estimated by repeated subsampling, but it is no longer necessary.

Instead, we used analytical expressions for rarefaction curves recently consolidated from previous work or newly derived by Chao et al. (2014). For each reference sample (or pseudosample), equations from Tables 1 and 2 of Chao et al. (2014) were used to generate, analytically, the expected diversity and sample coverage for each level of subsampling (Appendix A).

Rarefaction curves and diversity indices

We present results for standard rarefaction curves, in which the *x*-axis is either the abundance (individual-based) or number of samples (sample-based). In addition, we also carried out all analyses using the estimated coverage of either abundance or number of samples as the *x*-axis in the sampling curve (Chao and Jost 2012). Coverage is defined as the proportion of total individuals or samples from the complete assemblage that is represented by the species present in the sample or subset of samples. Rescaling the data to estimated coverage may provide a more powerful comparison of rarefaction curves (Chao and Jost 2012). Coverage analyses were conducted for both individual- and sample-based rarefaction.

We present results for all tests using species richness as the diversity index. Although species richness is the most popular diversity index, it is quite sensitive to sample size (Gotelli and Colwell 2001), and does not incorporate information on species abundances. Species richness, itself, is part of a mathematical series of diversity indices known as Hill numbers (Hill 1973), which can be algebraically transformed into familiar diversity indices, but have better statistical properties (Chao et al. 2014). The order q of the Hill number determines the weighting given to more common species, with species richness defined by $q = 0$. In the supplementary material (Appendix B), we present results of parallel tests for all analyses of Hill numbers $q = 1$ (exponential Shannon index), and $q = 2$ (the "inverse" Simpson index).

All of the tests described have been implemented in the accompanying "rareNMtests" package (Supplement; Cayuela and Gotelli 2014) for R (R Development Core Team 2013).

Ecological null hypothesis

In early studies on rarefaction, the original null hypothesis was that species richness in a small collection (a subsample of a specified size) could be viewed as a random subset of a larger collection (the reference sample; Simberloff 1979). However, the null hypothesis that ecologists usually want to ask is whether two or more reference samples (or subsamples of them) could be viewed as random draws from a single assemblage (Gotelli and Colwell 2011). This comparison is more challenging because it requires some estimate of the unconditional variance associated with sampling from the true assemblage (Colwell et al. 2004), rather than just the conditional variance associated with subsampling from the largest sample in the collection.

In this study, the ecological null hypothesis $H_{0\text{eco}}$ is that two (or more) reference samples, represented by either abundance or incidence data, were both drawn from the same assemblage of N^* individuals and S species. Therefore, any differences among the samples in species composition, species richness, or relative abundance reflect only random variation, given the number of individuals (or sampling units) in each collection. The alternative hypothesis, in the event that $H_{0\text{eco}}$ cannot be rejected, is that the sample data were drawn from different assemblages. If $H_{0\text{eco}}$ is true, then pooling the samples should give a composite sample that is also a (larger) random subset of the complete assemblage. It is from this pooled composite sample that we make random draws for comparison with the actual data.

EcoTest metric

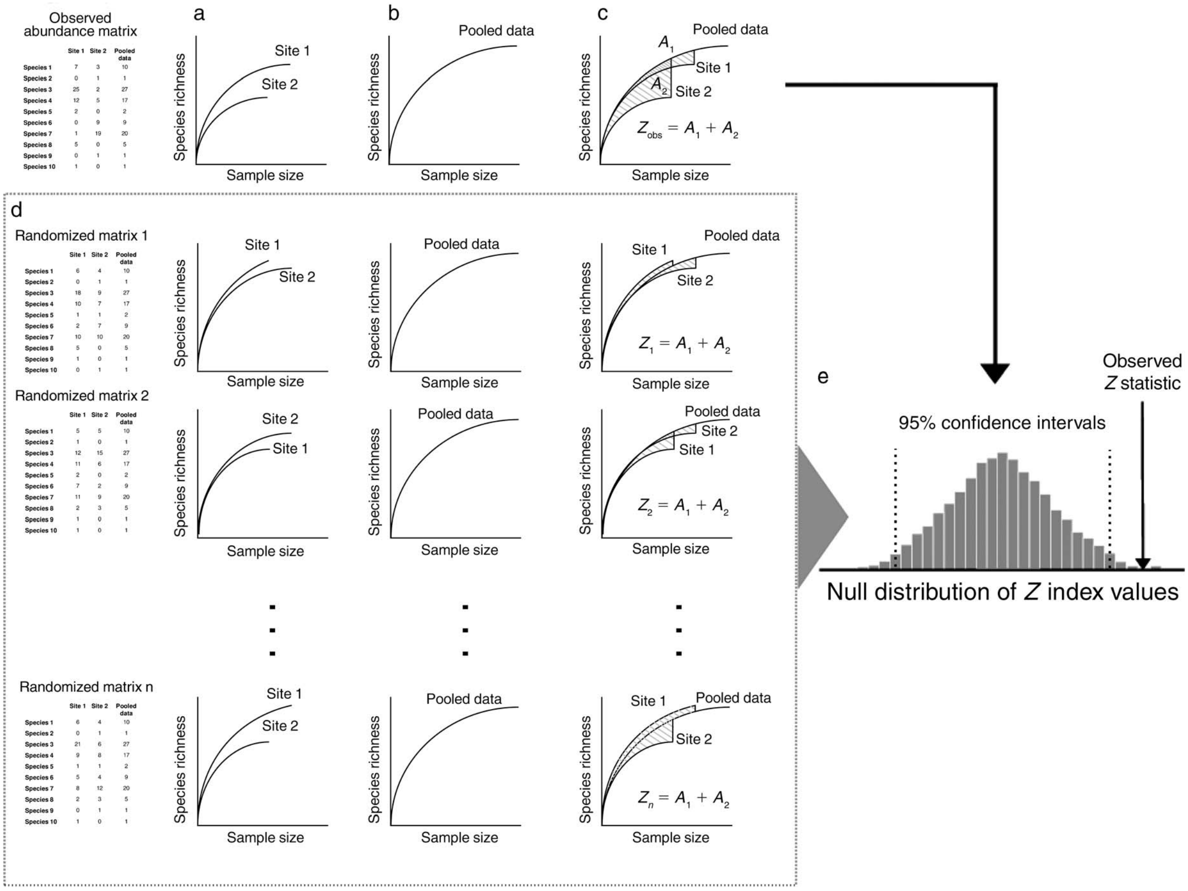
We begin by plotting the expected rarefaction curves for the individual samples (Fig. 2a) and for the pooled composite sample (Fig. 2b). Next, for each individual sample i, we calculate the cumulative area A_i between the sample rarefaction curve and the pooled rarefaction curve (Fig. 2c). For a set of $i = 1$ to K samples, we define the observed difference index

$$Z_{\text{obs}} = \sum_{i=1}^{K} A_i.$$

Note that two identically shaped rarefaction curves may nevertheless differ from the curve for the pooled sample. This difference can arise because species identities in the individual samples are retained in the pooled composite sample, which affects the shape of the pooled rarefaction curve. See Crist and Veech (2006) for a similar approach to partitioning β diversity.

EcoTest randomization algorithm

The data are next reshuffled by randomly reassigning every individual to a reference sample (for abundance data) or every sampling unit to a reference sample (for incidence data), and preserving the original sample sizes (number of individuals for abundance data, and number of sample units for incidence data). From this randomization, we again construct rarefaction curves and calculate Z_{sim} (Fig. 2d) as the cumulative area between the rarefaction curves of the randomized samples and the composite rarefaction curve. This procedure is repeated many times, leading to a distribution of Z_{sim} values and a 95% confidence interval (Fig. 1e). The position of Z_{obs} in the tail of this distribution is used as an estimate of the probability to randomly obtain this value given the null distribution of the cumulative area between the rarefaction curves of the randomized samples and the composite rarefaction curve, i.e., $p(Z_{\text{obs}} \mid H_{0\text{eco}})$. Large values of Z_{obs} relative to the null distribution imply that observed differences among samples in species composition, richness, and/or relative



Observed abundance matrix

	Site 1	Site 2	Pooled data
Species 1	7	3	10
Species 2	0	1	1
Species 3	25	2	27
Species 4	12	5	17
Species 5	2	0	2
Species 6	0	9	9
Species 7	1	19	20
Species 8	5	0	5
Species 9	0	1	1
Species 10	1	0	1

Randomized matrix 1

	Site 1	Site 2	Pooled data
Species 1	6	4	10
Species 2	0	1	1
Species 3	18	9	27
Species 4	10	7	17
Species 5	1	1	2
Species 6	2	7	9
Species 7	10	10	20
Species 8	5	0	5
Species 9	1	0	1
Species 10	0	1	1

Randomized matrix 2

	Site 1	Site 2	Pooled data
Species 1	5	5	10
Species 2	1	0	1
Species 3	12	15	27
Species 4	11	6	17
Species 5	2	0	2
Species 6	7	2	9
Species 7	11	9	20
Species 8	2	3	5
Species 9	1	0	1
Species 10	1	0	1

Randomized matrix n

	Site 1	Site 2	Pooled data
Species 1	6	4	10
Species 2	0	1	1
Species 3	21	6	27
Species 4	9	8	17
Species 5	1	1	2
Species 6	5	4	9
Species 7	8	12	20
Species 8	2	3	5
Species 9	0	1	1
Species 10	1	0	1

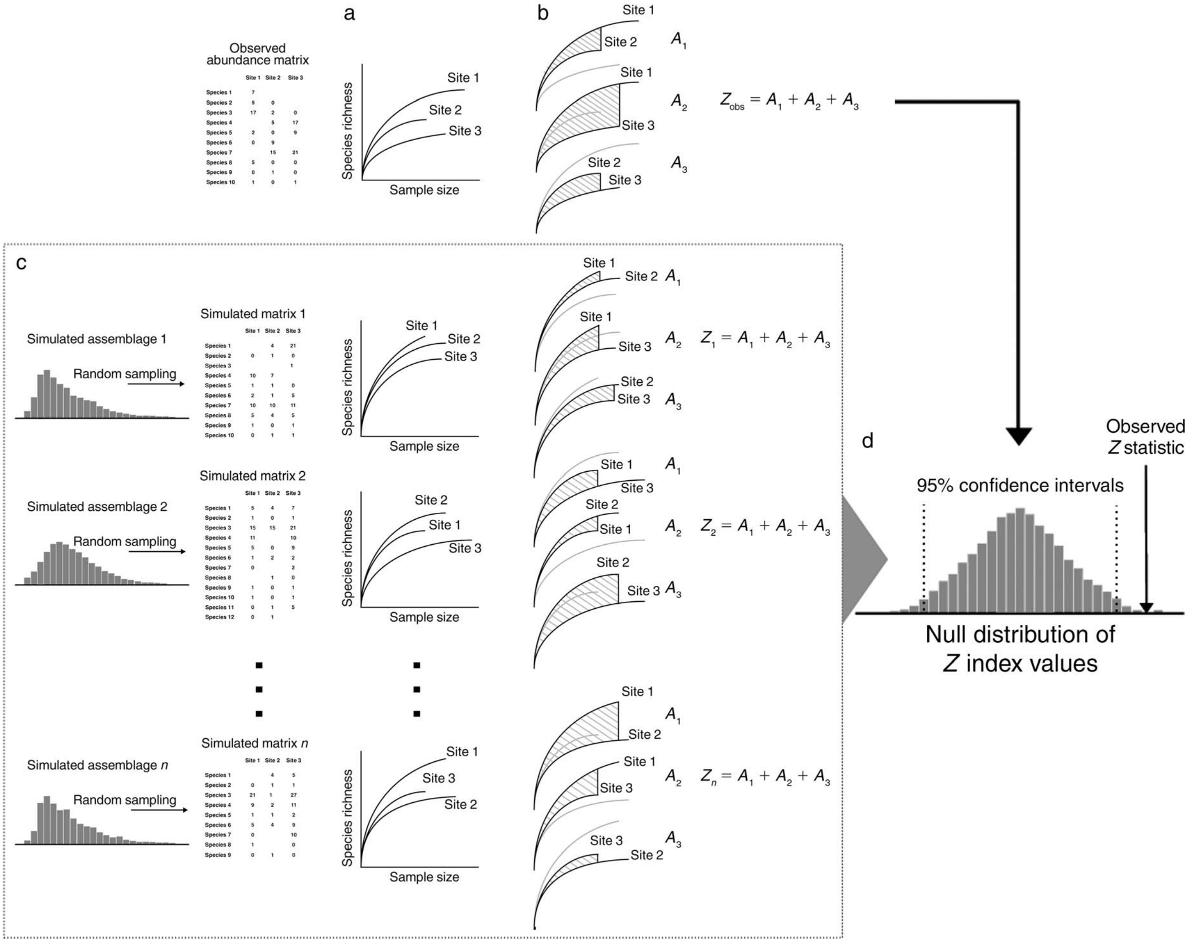
FIG. 2. Schematic representation of the EcoTest metric and algorithm: (a) rarefaction curves for the individual samples; (b) rarefaction curve of the pooled composite sample; (c) test statistic Z_{obs} is calculated as the cumulative area A_i between each individual sample rarefaction curve and the pooled rarefaction curve; (d) data are reshuffled by randomly reassigning every individual to a reference sample (for abundance data) or every sampling unit to a reference sample (for incidence data), while preserving the original sample sizes (number of individuals for abundance data and number of sample units for incidence data). From this randomization, we again construct rarefaction curves and calculate Z_{sim} as the cumulative area between the rarefaction curves of the randomized samples and the composite rarefaction curve; (e) this procedure is repeated many times, leading to a distribution of Z_{sim} values and 95% confidence intervals. The position of Z_{obs} in the tail of this distribution is used as an estimate of $p(Z_{\mathrm{obs}} \mid H_{0\mathrm{eco}})$. For the purposes of illustration, the composite curve is portrayed as extending only a small distance beyond the two reference samples. However, in the actual analyses, the x-axis for the composite sample must be the sum of the sampling effort for each of the reference samples, so it would extend much further to the right. However, the test statistic is only based on the portion of the composite curve that overlaps with the rarefaction curves of the individual samples. The tables are included solely as visual aids; all data presented is completely arbitrary.

abundance are improbable if the samples were all drawn from the same assemblage.

Biogeographical null hypothesis

In a discussion of the properties of rarefaction curves, Simberloff (1979) noted that differences in species composition can obviously lead to rejection of $H_{0\mathrm{eco}}$, even if the rarefaction curves have similar profiles: "But since rarefaction uses only the species-individuals' distributions, and not the species' names, it makes little sense to rarefy a large sample to compare it to some smaller sample if the species in the two samples are very different. If a large sample consists primarily of butterflies and a smaller one is mostly moths, we do not need rarefaction to tell us that the smaller could not reasonably be viewed as a random draw from the larger . . ."

For this reason, one of the stated assumptions of traditional rarefaction was that the species lists being compared are "taxonomically similar" (Tipper 1979, Gotelli 2008). But in many biogeographic comparisons, it is already known that the species composition is different, yet we wish to assess whether two or more reference samples differ in richness or other measures of diversity. The question of interest is not "were two or more samples randomly drawn from the same underlying species assemblage?" but "does species richness (or any other diversity metric) differ among reference samples



Observed abundance matrix

	Site 1	Site 2	Site 3
Species 1	7		
Species 2	5	0	
Species 3	17	2	0
Species 4		5	17
Species 5	2	0	9
Species 6	0	9	
Species 7		15	21
Species 8	5	0	0
Species 9	0	1	0
Species 10	1	0	1

Simulated matrix 1

	Site 1	Site 2	Site 3
Species 1		4	21
Species 2	0	1	0
Species 3			1
Species 4	10	7	
Species 5	1	1	0
Species 6	2	1	5
Species 7	10	10	11
Species 8	5	4	5
Species 9	1	0	1
Species 10	0	1	1

Simulated matrix 2

	Site 1	Site 2	Site 3
Species 1	5	4	7
Species 2	1	0	1
Species 3	15	15	21
Species 4	11		10
Species 5	5	0	9
Species 6	1	2	2
Species 7	0		2
Species 8		1	0
Species 9	1	0	1
Species 10	1	0	1
Species 11	0	1	5
Species 12	0	1	

Simulated matrix *n*

	Site 1	Site 2	Site 3
Species 1		4	5
Species 2	0	1	1
Species 3	21	1	27
Species 4	9	2	11
Species 5	1	1	2
Species 6	5	4	9
Species 7	0		10
Species 8	1		0
Species 9	0	1	0

FIG. 3. Schematic representation of the BiogTest metric and algorithm: (a) expected rarefaction curves for the individual samples; (b) test statistic Z_{obs} is calculated as the cumulative area between all unique pairs *ab* of *K* sample rarefaction curves; (c) the null distribution is constructed by creating random assemblages from a family of log-normal abundance distributions. The parameters of each of these distributions were set to specify a suite of distributions that might act as a reasonable sampling universe. Random samples are then drawn from each of the simulated assemblages, and Z_{sim} is calculated as the cumulative area between all *K* unique pairs *ab* of the randomized sample rarefaction curves; (d) this procedure is repeated many times, leading to a distribution of Z_{sim} values and a 95% confidence interval. The position of Z_{obs} in the tail of this distribution is used as an estimate of $p(Z_{\mathrm{obs}} \mid H_{0\mathrm{biog}})$. The tables are included solely as visual aids; all data presented is completely arbitrary.

after adjusting for differences in abundance or sampling effort?" The biogeographical null hypothesis ($H_{0\mathrm{biog}}$) is that, regardless of differences in species composition, the profiles of two or more rarefaction curves are similar enough that they might have been drawn from assemblages that do not differ significantly in richness or in underlying species abundance distribution.

BiogTest metric

We were not able to devise a BiogTest based on a composite sample that was strictly analogous to the EcoTest. Instead, we constructed both a different metric and a different algorithm to test $H_{0\mathrm{biog}}$. To do so, we first calculated as a test statistic the summed area between all unique pairs *ab* of the *K* sample rarefaction curves (Fig. 3a, b)

$$Z_{\mathrm{obs}} = \sum_{a=1,b=1}^{K} A_{ab}.$$

If $H_{0\mathrm{biog}}$ is true, then Z_{obs} should be relatively small because all of the rarefaction curves should have similar profiles, regardless of their species composition. In the limit, if all of the rarefaction curves had an identical profile, Z_{obs} would equal 0.

BiogTest randomization algorithm

To construct the null distribution, we created random assemblages by sampling from a presumed underlying species abundance distribution. Of the many possible distributions, including the log-series, log-normal, and broken stick, which distribution should be used? For our purposes, the log-normal distribution has some advantages: (1) for some parameter values, the log-normal generates a typical right-skewed distribution (many rare species, and a few common species) typical of well-sampled assemblages (Preston 1962*a*, *b*), (2) abundance and occurrence data collected for many taxa at widely different spatial scales often conform to an approximate log-normal distribution (Ulrich et al. 2010), (3) log-normal distributions approximate the species abundance distribution of important mechanistic models, including the neutral model (McGill et al. 2006) and stochastic versions of some niche partitioning models (Tokeshi 1993), and (4) depending on the underlying parameter values and the sample sizes, the log-normal can also generate species abundance profiles that resemble a log series or geometric series of abundances (Wilson 1993, McGill et al. 2007). Whether the log-normal distribution itself is caused by species interactions or reflects neutral processes or sampling intensity is still open to debate (McGill 2003, Sugihara et al. 2003), but is immaterial for our purposes here.

The statistical parameters of the log-normal rank abundance distribution are the number of species in the assemblage and the variance of the distribution; the latter controls the differences in abundance between common and rare species. If these underlying parameters are known, then sample size effects can be estimated by random sampling of individuals from the specified distribution. However, it is very difficult to directly estimate these parameters from a sample or set of samples (O'Hara 2005). Instead, we generated a suite of log-normal distributions that, taken together, might act as a reasonable sampling universe for comparison with a set of reference samples to test the biogeographic null hypothesis. Our strategy was to specify a distribution for each of the two parameters in the log-normal: species number and variance. As in a random-effects model (Zuur et al. 2009), each replicate of the null distribution reflects a single sample from a log-normal distribution in which the two model parameters were first determined by random assignment.

For the lower boundary of species richness, the minimum possible value cannot be smaller than the maximum number of species observed in the richest single sample among a set of samples. For the upper boundary of species richness, we calculated the upper bound of the Chao1 95% confidence interval (Chao 1984) for asymptotic species richness of each sample. The number of species in each null assemblage was then set as a random draw from a uniform distribution bounded at the low end by the maximum observed S and bounded at the high end by the maximum upper bound of the Chao1 95% confidence interval.

For the standard deviation of the log-normal, we sampled a random uniform value between 1.1 and 33 (0.1 and 3.5 on a natural logarithm [henceforth referred to as ln] scale). For empirical assemblages, standard deviations typically fall within this range (Limpert et al. 2001). Once the null assemblage was specified by selection of parameters for species richness and the standard deviation, we sampled (with replacement) the specified number of individuals for each sample in an individual-based data set (Fig. 3c). For incidence data, we sampled (without replacement) the observed number of species in each sampling unit (i.e., total number of incidences), with sample probabilities set proportional to relative abundances in the log-normal distribution. For example, if 15 species were observed in one sampling unit, the equivalent sampling unit in the simulated data set should also contain 15 species, though not necessarily the same ones. We then used the analytic formulas in Chao et al. (2014) to construct the rarefaction curves for each of the pseudosamples. The analysis from this point forward is the same as for the EcoTest. Namely, we generated a distribution of Z_{sim} and compared it to Z_{obs} to estimate $p(Z_{\mathrm{obs}} \mid H_{0\mathrm{biog}})$ (Fig. 3d).

The BiogTest can be used for other species abundance distributions (other than the log-normal) to construct the null distribution test, namely the broken stick and geometric series distributions. We used the broken stick because it is the most even of all species abundance distributions, whereas the geometric series can generate highly uneven distributions (Magurran 2003). For the broken stick, the number of species in the assemblage must first be known, then a random partition is made to define the relative abundance of each species. For the geometric series, two parameters are needed: the number of species in the assemblage and a constant ratio D ($D <$ 1), which determines the abundance of the next species in the sequence. In the geometric series, D was obtained by sampling a random uniform value between 0.1 and 1. In all cases, the number of species (S), as in the log-normal distribution, was obtained by randomly drawing from a uniform distribution that was bounded at the low end by the maximum observed S and at the high end by the maximum upper bound of the 95% confidence interval for Chao1.

To better mimic the sampling process, a negative binomial random error was added to the abundance counts every time a sample was randomly drawn from the simulated assemblage. The negative binomial distribution was used to generate realistic heterogeneity that often results from spatial clustering of individuals and other small-scale processes (Green and Plotkin 2007, Bolker 2008). The expectation μ of the negative binomial was represented by the abundance count of each species in the assemblage (see example in the Supplement). The variance of the negative binomial is

$$\mathrm{var} = \mu + \frac{\mu^2}{k}$$

where k is the dispersion parameter. For every species, k was randomly drawn from a uniform distribution between 0.01 and 25 each time a sample was drawn from the assemblage.

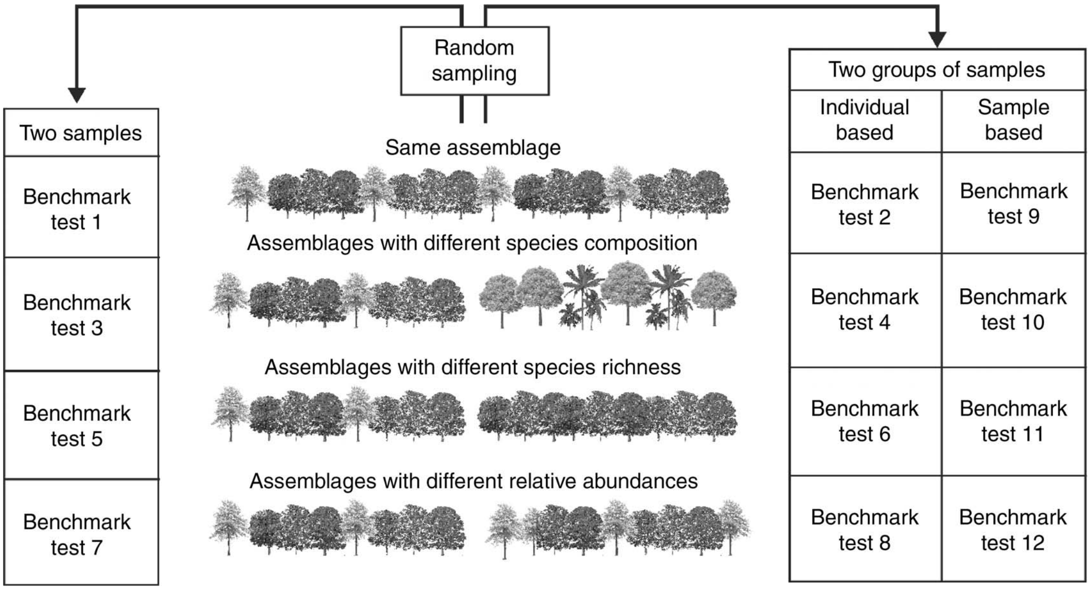


FIG. 4. Illustration of the four different groups of benchmark tests that were created to study the performance of the EcoTest and BiogTest methods: (1) samples drawn from the same assemblage; (2) samples drawn from assemblages with different species composition, but similar species richness and relative abundance; (3) samples drawn from two assemblages with different species richness, but similar composition and relative abundance; and (4) samples drawn from two assemblages with different relative abundance of species, but similar composition and species richness. For each group of scenarios, three benchmark tests were conducted depending on the sampling scheme: two individual-based rarefaction curves were compared (left column); more than two individual-based rarefaction curves were compared (middle column); and two sample-based rarefaction curves were compared (right column).

Evaluation of the algorithms

Before a new randomization method is applied to empirical data, its performance needs to be evaluated with artificial data sets that have specified characteristics (Gotelli and Ulrich 2012). Two properties are desirable for our statistical tests of differences in rarefaction curves. First, when the tests are confronted with samples that are drawn from the same assemblage, they should not reject the null hypothesis too frequently; we apply a traditional Type I error (α) criterion of 5%. Second, when these tests are confronted with samples drawn from assemblages that differ in species composition, richness, or relative abundance, they should not accept the null hypothesis too frequently, that is, the probability of committing a Type II error (β) should be low. There is no accepted standard for the level of power of a test ($1 - \beta$), but a value of 0.8 (the null hypothesis is correctly rejected 80% of the time) has been suggested (Cohen 1992). However, power analysis is rarely conducted in ecological studies (Toft and Shea 1983), because in most cases it requires specification of an alternative hypothesis and an effect size that can be detected by the test. In this case, the alternative is specified: samples were drawn from multiple assemblages. We do not explicitly define an effect size, but that size is determined by the expected area difference (Z) among rarefaction curves derived from assemblages that were defined by random parameter values (as specified earlier) for the log-normal distribution.

Simulation scenarios and benchmark tests

To study the performance of the proposed EcoTest and BiogTest, we estimated the frequency of Type I and Type II statistical errors in four different groups of scenarios (Fig. 4; Appendix B): (1) drawing random samples from the same assemblage (Benchmark tests 1, 2, 9), (2) drawing random samples from two assemblages with different species composition but similar species richness and relative abundance (Benchmark tests 3, 4, 10), (3) drawing random samples from two assemblages with different species richness, but similar composition and relative abundance (Benchmark tests 5, 6, 11), and (4) drawing random samples from two assemblages with different relative abundances of species, but similar composition and species richness (Benchmark tests 7, 8, 12). For each group of scenarios, three benchmark tests were conducted depending on the sampling scheme: (1) two individual-based rarefaction curves that were drawn either from the same assemblage (Benchmark test 1) or from two different assemblages (Benchmark tests 3, 5, 7) were compared, (2) more than two individual-based rarefaction curves that were derived from either the same assemblage (Benchmark test 2) or from two different assemblages (Benchmark tests 4, 6, 8) were compared, and (3) two sample-based rarefaction curves that were derived from either the same assemblage (Benchmark test 9) or from two different assemblages (Benchmark tests 10, 11, 12) were compared.

TABLE 1. Description of the 12 benchmark tests performed to estimate Type I and II error using simulated communities.

	Expected output		Parameters of the simulated assemblages			Extent of sampling		
Test	EcoTest	BiogTest	Assemblages	Richness	SD	*N*	*S*	Iterations
Individual based								
1	H_0	H_0	1	[10, 200]	[0.1, 3.5]	2	[50, 1000] + ε_i	750
2	H_0	H_0	1	100	2.45	[2, 15]	500 + ε_i	250
3	H_1	H_0	2	[10, 200]	[0.1, 3.5]	2	[50, 1000] + ε_i	750
4	H_1	H_0	2	100	2.45	[2, 15]	500 + ε_i	250
5								
1	H_1	H_1	1	50	[0.1, 3.5]	2	[50, 1000] + ε_i	750
2	H_1	H_{1}	1	150				
6								
1	H_1	H_1	1	50	2.45	[2, 15]	500 + ε_i	250
2	H_{1}	H_{1}	1	150				
7								
1	H_1	H_1	1	[10, 200]	[0.1, 3.5]	2	[50, 1000] + ε_i	750
2	H_{1}	H_{1}	1		[0.1, 3.5] ± 1	2		
8								
1	H_1	H_1	1	100	[0.1, 3.5]	[2, 15]	500 + ε_i	250
2	H_1	H_1	1	100	[0.1, 3.5] ± 1	[2, 15]		
Sample based								
9	H_0	H_0	1	100	[0.1, 3.5]	[10, 50]	500 + ε_i	750
10	H_1	H_0	2	100	[0.1, 3.5]	[10, 50]	500 + ε_i	750
11								
1	H_1	H_1	1	50	[0.1, 3.5]	[10, 50]	500 + ε_i	750
2	H_1	H_1	1	150	[0.1, 3.5]	[10, 50]	500 + ε_i	750
12								
1	H_1	H_1	2	100	[0.1, 3.5]	[10, 50]	500 + ε_i	750
2	H_1	H_1	2	100	[0.1, 3.5] ± 1	[10, 50]	500 + ε_i	750

Notes: For each benchmark test, the required data, the expected output of the two methods, the parameters of the simulated assemblages (number of assemblages, mean richness and standard deviation [SD; on the natural-log scale] of the underlying distribution of relative abundances), the extent of sampling from the simulated assemblages (total number of sites, *N*, number of individuals per site, *S*), and the number of simulated assemblages (iterations) is indicated. Tests 5, 6, 7, 8, 9, 11, and 12 were split into two different tests for two different assemblages. For assemblages 7, 8, and 12, assemblage 2 showed a fixed difference of the standard deviation of one unit on a natural-log scale for assemblage compared to assemblage 1. H_0, null hypothesis is accepted; H_1, null hypothesis is rejected; ε_i, random variation in the number of individuals in each sample, taken from a Poisson distribution where the parameter λ was set to sample size. Values in brackets indicate a range of potential values within which we get a random value for that particular parameter (SD, *N*, *S*).

We created sets of artificial data matrices for each scenario and sampling scheme, for which carefully selected, contrasting parameter combinations were used (Table 1). For those tests in which no rejection of the null hypothesis was expected, we estimated the probability of incorrectly rejecting H_0 as the proportion of matrices for which *P* values were below the standard 0.05 threshold, based on 200 randomizations of each matrix (i.e., Type I error). For those tests in which rejection of the null hypothesis was expected, we estimated the probability of incorrectly failing to reject H_0 as the proportion of matrices for which *P* values were above the standard 0.05 threshold, based on 200 randomizations of each matrix (i.e., Type II error). Although it is common to use 1000 or more randomizations in null model tests (Manly 2006), we found consistent results with 200 randomizations.

Parameters of the simulated artificial assemblages such as the mean species richness or evenness (measured as the standard deviation of the log-normal) were fixed in some scenarios, but in others were allowed to vary randomly among the artificial matrices (Table 1). Fixed values of richness of 50 and 150 species were used for Scenarios 5, 6, and 11, which were designed to test for differences in species richness. Note that these species richness levels refer to the assemblage from which the samples were drawn. Differences in species richness among the sampled matrices (which have fewer individuals or samples than the entire assemblage) were much smaller, with an average difference between pairs of samples of 43.7 species and a 95% confidence interval of 3–96 species. In other scenarios, species richness varied within each matrix, and was selected randomly from within a uniform range of 10–200 species. A fixed difference of the standard deviation of one unit on a ln scale was used for Scenarios 7, 8, and 12, which were designed to test for differences in the rank abundance distribution. In all other scenarios, the standard deviation of the log-normal did not vary among the assemblages, and was chosen randomly from a uniform range of 0.1–3.5 on a ln scale (values based on Limpert et al. [2001]).

The extent of sampling was also allowed to vary randomly to represent realistic sampling differences that might be found in typical biodiversity surveys (Table 1). In individual-based scenarios for which only two sites

TABLE 2. Variations of the ecological and biogeographical null model tests run on simulated matrices in each benchmark test.

Test type (x-axis), Hill number order (y-axis), and method	Distribution of null assemblages
Individual- or sample-based	
$q = 0$	
EcoTest	NA
BiogTest	log-normal
BiogTest	geometric
BiogTest	broken stick
$q = 1$	
EcoTest	NA
BiogTest	log-normal
BiogTest	geometric
BiogTest	broken-stick
$q = 2$	
EcoTest	NA
BiogTest	log-normal
BiogTest	geometric
BiogTest	broken stick
Coverage-based	
$q = 0$	
EcoTest	NA
BiogTest	log-normal
BiogTest	geometric
BiogTest	broken stick
$q = 1$	
EcoTest	NA
BiogTest	log-normal
BiogTest	geometric
BiogTest	broken stick
$q = 2$	
EcoTest	NA
BiogTest	log-normal
BiogTest	geometric
BiogTest	broken stick

Note: NA, not applicable. Hill numbers were given different orders (q); $q = 0$, (c, d) $q = 1$, and (e, f) $q = 2$; q of the Hill number determines the weighting given to more common species, with species richness defined by $q = 0$, the exponential Shannon index shown by $q = 1$, and the inverse Simpson index shown by $q = 2$.

were compared (Scenarios 1, 3, 5, 7), the number of individuals per site was chosen randomly between 50 and 1000 individuals in each artificial matrix. In individual-based scenarios for which more than two sites were compared (Scenarios 2, 4, 6, 8) and for sample-based scenarios (9, 10, 11, 12), the number of individuals was set constant at 500 per site. To include some realistic sampling variation, we added a Poisson error to the number of individuals in each sample, where the parameter λ of the Poisson distribution was set to sample size. The number of artificial matrices generated for each benchmark test was set at 250 (Scenarios 2, 4, 6, 8) or 750 (Scenarios 1, 3, 5, 7, 9–12). In all, 24 variations of benchmark tests (6 for EcoTest and 18 for BiogTest) were conducted for each artificial matrix to account for (1) individual- or sample-based vs. coverage-based rarefaction curves, (2) different Hill numbers ($q = 0$, $q = 1$, $q = 2$), (3) the use of the EcoTest and BiogTest, and (4) different distributions of null assemblages for the BiogTest (Table 2).

In summary, we set up four different groups of scenarios, with three sampling schemes each, generated either 250 or 750 artificial data matrices for each combination of scenario and sampling scheme, and applied 24 benchmark tests (Appendix B). All the analyses were conducted in R (R Development Core Team 2013) and run at the Bioportal server, a web-based portal for data analysis that allows for parallel processing. Computation time for the suite of different benchmark tests ranged from two days (for comparison of sample-based rarefaction curves) to four weeks (for comparison of multiple individual-based rarefaction curves). Overall, running all benchmark tests on the different sets of artificial matrices required more than two months of processing time at the Bioportal server. However, the analysis of a typical data set on a personal computer can be completed in reasonable amounts of time. For example, on a 64-bit Intel Core laptop with 7.8 GB of memory (Intel, Santa Clara, California, USA), a comparison with 200 iterations of two individual-based rarefaction curves for tropical trees with ~20 000 and 15 000 individuals, and 226 and 173 species, respectively, took 30 min for the EcoTest and 19 min for the BiogTest. The R code used to run all benchmark tests is available in the Supplement.

Description of case studies

In addition to the benchmark tests with artificial data sets, we also applied the methods to two empirical data sets, one with abundance data (i.e., individual-based rarefaction curves) and one with incidence data (i.e., sample-based rarefaction curves). For the individual-based case study, we used tree data from six 20–52-ha plots established at several tropical sites around the world: two plots in South America, two in Africa, and two in Southeast Asia (Fig. 5a and Table 3). Species lists and abundances for these plots are available at the Center for Tropical Forest Sciences (CTFS) website. We pooled the plot data from each site, and analyzed the abundance of individual trees exceeding 10 cm in diameter at breast height (DBH). The CTFS project uses the taxonomy of the Angiosperm Phylogeny Group for the orders and families of flowering plants (Angiosperm Phylogeny Group 2009). All species names were additionally cross-checked against The Plant List using the Taxonstand R package (Cayuela et al. 2012*b*) to avoid the use of synonyms and to correct typographical errors (The Plant List database is *available online*).[6] We compared individual-based rarefaction curves for Hill numbers $q = 0$ (i.e., species richness), $q = 1$, and $q = 2$ using either sample size (i.e., number of individuals) or sample coverage. Continental patterns may reflect constraints imposed by similar climates as well as

[6] http://www.theplantlist.org

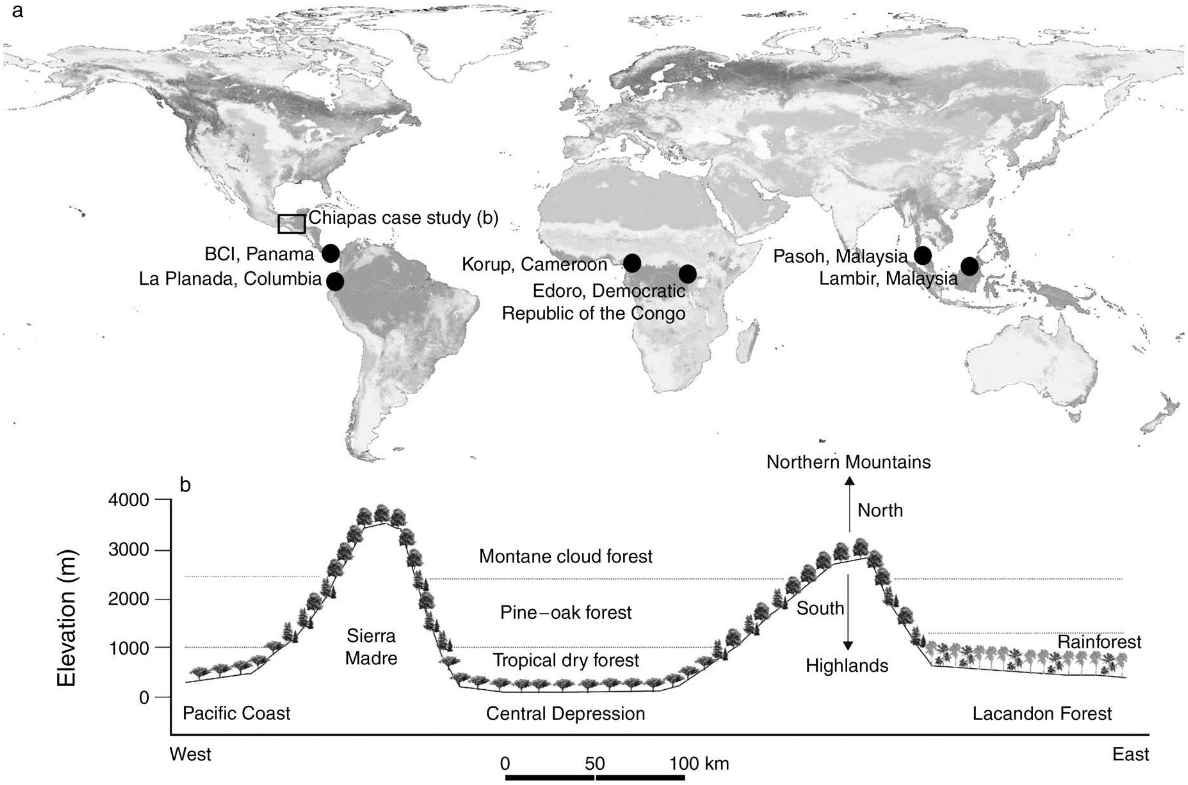


FIG. 5. (a) Location of six Center for Tropical Forest Sciences (CTFS) plots. The distribution of tropical rainforest is indicated in dark gray within tropical latitudes. BCI stands for Barro Colorado Island. Base map from http://eoimages.gsfc.nasa.gov/images/news/NasaNews/ReleaseImages/LCC/Images/lcc_global_2048.jpg. (b) Distribution of montane cloud forests and location of study sites in Chiapas, Mexico, along a vegetation catena running from west to east.

differences reflecting unique histories (Ricklefs and Schluter 1993). We used the individual-based version of the two proposed methods with 200 iterations each to test the ecological and biogeographical null hypothesis within and between continents.

For the sample-based case study, we used a set of 224 circular 0.1-ha plots from tropical montane cloud forests (see Plate 1) in three regions of the state of Chiapas, Mexico: El Triunfo Biosphere Reserve in the Sierra Madre (100 plots), the Highlands (38 plots; Cayuela et al. 2006), and the Northern Mountains (86 plots; Fig. 5b; see Supplement; Ramírez-Marcial et al. 2001). The three regions were dominated mostly by primary forests on top of the hills, with some chronic low intensity human disturbances such as selective logging, but there were also some secondary forests, particularly in the Highlands. For these plots, incidence records were based on trees exceeding 5 cm in DBH. Data were obtained from the BIOTREE-NET website (Cayuela et al. 2012*a*). As in the first case study, species

TABLE 3. Attributes of the six Center for Tropical Forest Sciences (CTFS) forest plots.

Location	Description	Coordinates	Elevation (m)	Plot size (ha)	Spp.	Inds.
Barro Colorado Island, Panama	lowland tropical moist forest	9.15° N, 79.85° W	120–160	50	226	21 198
La Planada, Colombia	tropical montane cloud forest	1.16° N, 77.99° W	1796–1891	25	173	15 013
Korup, Cameroon	evergreen tropical forest	5.07° N, 8.85° E	150–240	50	307	24 591
Edoro, Democratic Republic of Congo	evergreen tropical moist forest	1.47° N, 28.58° E	700–850	20	207	9 382
Pasoh, Malaysia	evergreen tropical moist (dipterocarp) forest	2.98° N, 102.31° E	70–90	50	671	28 279
Lambir, Malaysia	evergreen tropical moist (dipterocarp) forest	4.19° N, 114.02° E	104–244	52	990	32 611

Notes: Data were retrieved from http://www.ctfs.si.edu. Spp. indicates number of species, Inds. indicates number of individuals >10 cm diameter at breast height.

Table 4. Percentage of commission of Type I (incorrect rejection of a true null hypothesis) and Type II error (failure to reject a false null hypothesis) in 12 different benchmark tests (see Table 1 for a description of these tests) for the two proposed methods (EcoTest, BiogTest).

			BiogTest					
	EcoTest		Log-normal		Geometric		Broken stick	
Benchmark test	Type I error	Type II error	Type I error	Type II error	Type I error	Type II error	Type I error	Type II error
1	6.13		5.87		25.47		6.53	
2	7.60		5.20		45.20		5.20	
3		0.00	5.20		22.80		7.33	
4		0.00	2.93		44.35		4.18	
5		0.27		6.67		3.07		6.53
6		0.00		11.30		2.93		10.04
7		0.60		5.71		4.07		4.67
8		0.00		1.67		0.00		1.68
9	4.80		4.00		30.80		3.87	
10		0.00	8.67		32.67		10.13	
11		0.13		6.93		3.20		7.33
12		0.00		3.47		1.20		4.00

Note: For the BiogTest method, three different underlying distributions of species relative abundances for the benchmark null model communities were used (log-normal, geometric, broken stick). Cells left blank indicate non-applicable data.

names were standardized and typographical errors were corrected using The Plant List through the Taxonstand R package (Cayuela et al. 2012*b*). Distances between regions ranged from ~50 km (the Highlands and Northern Mountains) to ~250 km (Sierra Madre and Northern Mountains). Despite differences in elevation among these forests, ~53% of the total species occur in two or more of the forests. Thus, our null hypothesis was that the three regions should display similar patterns of species composition, richness, and relative abundance of species. We used the sample-based version of the two proposed methods with 200 iterations each to test the ecological and biogeographical null hypotheses between regions.

Results

Benchmark performance of EcoTest

For species richness ($q = 0$), the EcoTest had Type I errors of ~5% for both individual- (Scenarios 1 and 2) and sample-based rarefaction (Scenario 9). The EcoTest also had very low Type II errors (always less than 1%), so the null hypothesis was almost always rejected when data were generated from two different assemblages, and then pooled to generate a null distribution for testing (Table 4, Benchmark tests 3–8, 10–12).

For Hill numbers $q = 1$ and $q = 2$, Type I error rates were slightly higher (Appendix B: parts a and b), but still ranged from only 5% to 9%. Type II errors for Hill numbers $q = 1$ and $q = 2$ were very infrequent (always less than 1%). For coverage-based analyses, Type I error rates for the EcoTest with Hill numbers $q = 0$, $q = 1$, and $q = 2$ were between 4% and 14% (Appendix B: parts c–e), somewhat higher than the rates for individual- or sample-based rarefaction curves (Table 4). Type II error rates for coverage-based analyses were always less than 6% for all Hill numbers.

Benchmark performance of BiogTest

The BiogTest creates null assemblages by drawing random parameter values (species richness and standard deviation) to simulate a spectrum of log-normal distributions of species abundances. It is therefore not surprising that, for both Type I and Type II error tests for species richness ($q = 0$), the BiogTest performs well when the test matrices (which simulate empirical reference samples) themselves were also drawn from a log-normal distribution (Table 4). In such cases, the error rates in the different scenarios were less than 9% for Type I error, and less than 12% for Type II error. Similar values for Type I and Type II error rates were found when the assemblages were drawn from a broken stick distribution, and then compared to a null distribution created from a log-normal (Table 4). When the test matrices were created from a geometric series distribution, Type II error rates for species richness decreased as compared to a null distribution created from a log-normal. However, Type I error rates for the geometric series data sets increased to values between 25% and 45% for the various scenarios.

For Hill numbers $q = 1$ and $q = 2$ (Appendix B: parts a and b), the best performance for the BiogTest was also found for the log-normal and broken stick distributions, with Type I errors ranging between 0% and 8%. Type II errors were higher for these distributions and in some cases exceeded 70%. As with species richness ($q = 0$), the worst performance for Type I error occurred when samples were drawn from a geometric series distribution (Appendix B: parts a and b).

Both Type I and Type II error rates for all three Hill numbers (Appendix B: parts c–e), were usually higher for coverage-based analyses compared to sample- or individual-based analyses (Appendix B: parts c–e vs. Table 4).

Table 5. Comparison of tropical floras within (South America, Africa, Asia) and between regions with null model tests using individual-based and coverage-based rarefaction curves for Hill number orders $q = 0$, $q = 1$, and $q = 2$.

	Individual-based						Coverage-based					
	$q = 0$		$q = 1$		$q = 2$		$q = 0$		$q = 1$		$q = 2$	
Null hypothesis	Eco	Biog	Eco	Biog	Eco	Biog	Eco	Biog	Eco	Biog	Eco	Biog
Within continents												
South America	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
Africa	<0.01	<0.01	<0.01	<0.01	<0.01	0.130	<0.01	<0.01	<0.01	<0.01	<0.01	0.030
Asia	<0.01	<0.01	<0.01	<0.01	<0.01	0.570	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
Between continents												
South America vs. Africa	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
South America vs. Asia	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
Africa vs. Asia	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01

Notes: Entries in the table represent P values for EcoTest (Eco) and BiogTest (Biog). The corresponding null hypothesis is rejected if $P < 0.05$. Number of iterations for each test was 200. Null communities for the BiogTest were simulated using a lognormal distribution for relative species abundance.

Case studies

An individual-based comparison of global tropical rainforests

For all Hill numbers, there were significant differences within and between continents by the EcoTest, using both individual- and coverage-based rarefaction methods (Table 5). For the BiogTest, individual-based comparisons for Hill number $q = 2$ (exponential Shannon diversity) showed no differences between samples within Africa (dotted lines in Fig. 6e) and within Asia (dashed lines in Fig. 6e). However, Asian samples differed significantly from each other when plotted as a function of coverage (Table 5; Fig. 6e vs. 6f). All other comparisons revealed differences in species richness and/or relative abundance.

A sample-based comparison of three montane cloud forest regions of Chiapas

Based on the EcoTest, all three tree assemblages (Sierra Madre, Highlands, and Northern Mountains) are significantly different from one another (Table 6). However, based on the BiogTest, the rarefaction profiles of the Highlands and Northern Mountains did not differ from each other for sample-based rarefaction of all Hill numbers (Table 6, Fig. 7). For the coverage-based analysis, differences were detected for species richness ($q = 0$), but not for the higher-order Hill numbers ($q = 1$, $q = 2$).

Discussion

Limitations of classical rarefaction methods

Originally, statistical comparisons of rarefaction curves were made by rarefying the larger of two samples and determining whether or not the species richness for the smaller sample fell within the 95% confidence interval for the rarefaction curve (Hurlbert 1971). For comparisons of multiple samples, this approach is unsatisfying because it means that all samples must be rarefied down to the most poorly sampled collection (Chao and Jost 2012). Moreover, the original rarefaction algorithm (Hurlbert 1971, Simberloff 1972) is based on subsampling without replacement from a reference sample, which follows a hypergeometric distribution. The hypergeometric distribution generates a variance that is conditional on the observed sample, which causes the confidence interval to shrink to zero at the largest sample size (i.e., the reference sample). More recent extensions have provided an estimator of the unconditional variance for both sample- (Colwell et al. 2004) and individual-based rarefaction curves (Colwell et al. 2012), allowing for improved tests at any sampling level.

Colwell et al. (2012) recently showed that rarefaction curves can be smoothly joined with nonparametric extrapolation curves and extended out toward the species asymptote. But the question still arises, exactly what sampling level should be used for comparing rarefied or extrapolated diversity curves? Comparisons at very small sample sizes are problematic because all individual-based rarefaction curves of species richness contain less information as they converge to the point 1,1 (one individual always yields one species). Comparisons at the asymptote are also problematic because the variance, especially for species richness estimators, may be very large (but see Chao et al. [2014] for some useful guidelines for rarefying and extrapolating rarefaction curves over a "safe" range for statistically valid comparisons).

New methods to compare rarefaction curves

Our methods do not require an estimator of the unconditional variance, but they do incorporate sampling variability that encompasses the entire length of the interpolated rarefaction curve. Additionally, the proposed EcoTest and BiogTest allow for two distinct kinds of comparisons of rarefaction curves.

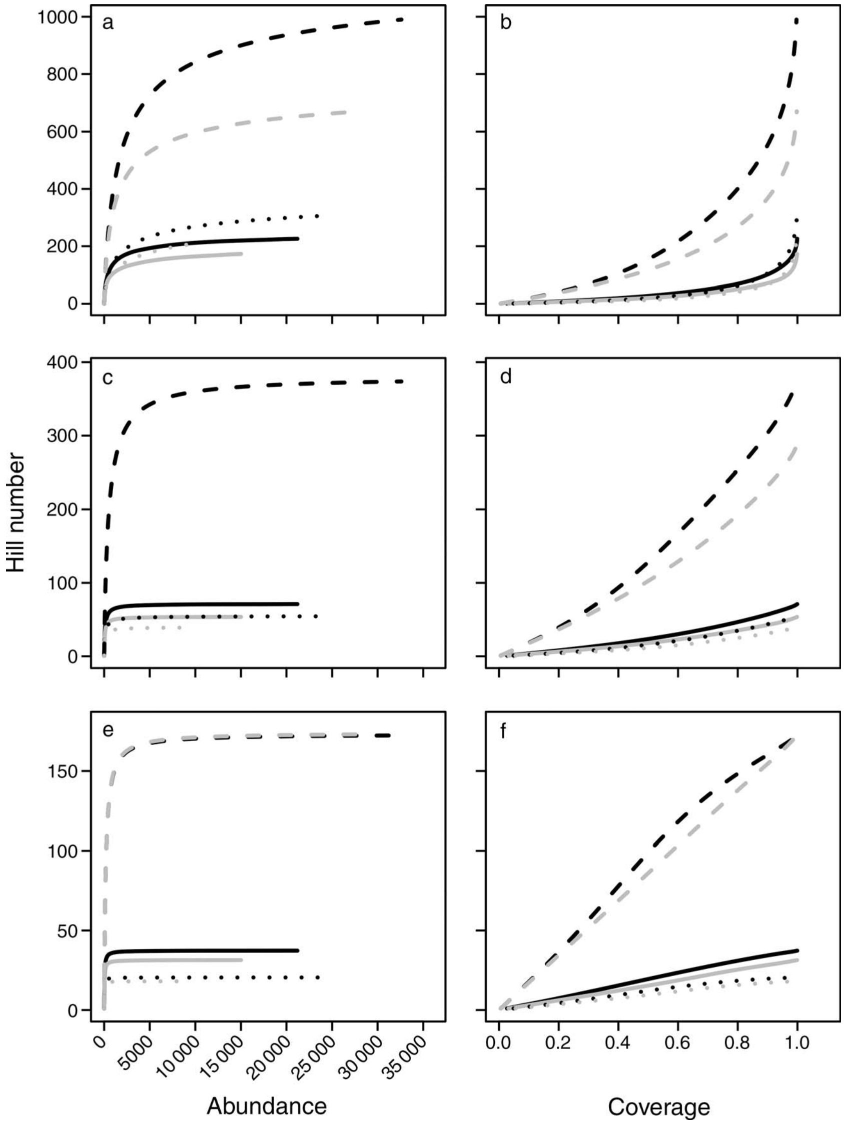


FIG. 6. Individual- (a, c, e) and coverage-based (b, d, f) rarefaction curves for the six CTFS forest plots using Hill numbers with weightings (a, b) $q = 0$, (c, d) $q = 1$, and (e, f) $q = 2$; order q of the Hill number determines the weighting given to more common species, with species richness defined by $q = 0$, the exponential Shannon index shown by $q = 1$, and the inverse Simpson index shown by $q = 2$. Plots from different continents are displayed with different line types: South America with solid lines (Barro Colorado Island [black], La Planada [gray]), Africa with dotted lines (Korup [black], Edoro [gray]), and Asia with dashed lines (Lambir [black], Pasoh [gray]). Two hundred iterations were used for each test. Null communities for the BiogTest were simulated using a log-normal distribution for relative species abundance.

The EcoTest addresses the simplest null hypothesis, which is that two samples were drawn from the same underlying assemblage, so that any differences in species richness, relative abundance, and species composition reflect only sampling effects. The EcoTest is a "distribution free" test that is based on a pooling of sample data to generate a null distribution. This test performed very well in all benchmark comparisons, consistently distinguishing samples that were created from distributions that differed in their underlying species richness (Scenarios 5, 6, and 11), relative abundance (Scenarios 7, 8, and 12), or species composition (Scenarios 3, 4, and 10; Table 4). The test also had a low Type I error rate when data were randomly sampled from a single underlying assemblage (Scenarios 1, 2, and 9)

The EcoTest successfully discriminates among samples that differ only in species composition, as expected according to Simberloff (1979). However, it is more often the case that multiple samples will contain some shared and some distinct species, as well as undetected species that may also be shared among samples (Chao et al. 2005). Regardless of the similarity of the rarefaction

TABLE 6. Comparison of tree communities from tropical montane forests in Chiapas, Mexico, using incidence sample-based and coverage-based rarefaction curves for Hill number orders $q = 0$, $q = 1$, and $q = 2$.

	Sample-based						Coverage-based					
	$q = 0$		$q = 1$		$q = 2$		$q = 0$		$q = 1$		$q = 2$	
Null hypothesis	Eco	Biog	Eco	Biog	Eco	Biog	Eco	Biog	Eco	Biog	Eco	Biog
Sierra Madre vs. Highlands	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
Sierra Madre vs. Northern Mountains	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
Highlands vs. Northern Mountains	<0.01	0.855	<0.01	0.970	<0.01	0.995	<0.01	0.005	<0.01	0.190	<0.01	0.570

Notes: Entries in the table represent P values. The corresponding null hypothesis is rejected if $P < 0.05$. Number of iterations for each test was 200. Null communities for the BiogTest were simulated using a log-normal distribution for relative species abundance.

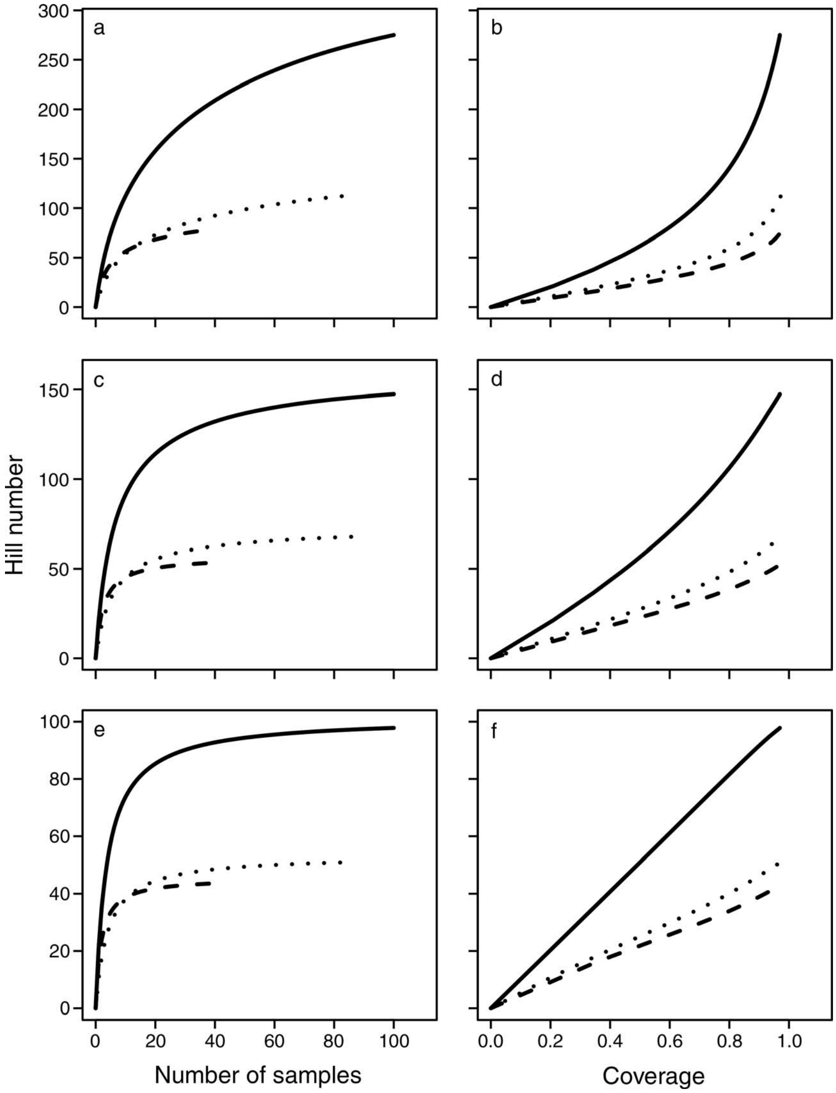


FIG. 7. Sample- (a, c, e) and coverage-based (b, d, f) rarefaction curves for three montane cloud forest regions in Chiapas, Mexico (El Triunfo [solid line], Northern Mountains [dotted line], and Highlands [dashed line]) using Hill numbers (a, b) $q = 0$, (c, d) $q = 1$, and (e, f) $q = 2$. Two hundred iterations were used for each test. Null communities for the BiogTest were simulated using a log-normal distribution for relative species abundance.

PLATE 1. Tropical montane forests represent one of the world's richest repositories of plant biodiversity, and play an important role in the provision of regulatory services such as water interception. Photo credit: L. Cayuela.

curves, differences in species composition can lead to a strong rejection of the null hypothesis for the EcoTest (Scenarios 3, 4, and 10). Therefore, it is not surprising that, for every empirical comparison of global tropical rainforests (Fig. 6) and regional montane forests (Fig. 7), the EcoTest was always strongly rejected (Tables 5, 6).

Deviations from the null hypothesis for the EcoTest can reflect everything from small-scale species interactions (which can alter relative abundances among sample plots; Chase and Leibold 2003) to regional differences in beta diversity (which can reflect turnover in species composition; Wilson and Tilman 1991, Baldeck et al. 2013) to large-scale differences in species richness (which can reflect differences in evolutionary rates; Qian and Ricklefs 2008). The sampling null hypothesis that is made explicit with the EcoTest is the starting point for comparing rarefaction curves. However, so many forces can affect this test that it is perhaps not surprising that it will be reliably rejected for well-sampled empirical assemblages such as those in Figs. 6 and 7.

The EcoTest makes use only of the expected diversity from the rarefaction curve. For species richness, Li and Mao (2012) derived a simultaneous confidence band for a species accumulation curve that can be used to evaluate differences between two accumulation curves, similar to what we have done with the cumulative area of difference between each rarefaction curve and the composite curve (Fig. 2c). The randomization algorithm for the EcoTest is identical to the method of Solow (1993), in which all partitions of the data among different samples are equally likely, and is similar to the method developed by Collins et al. (2009) to compare rank occupancy–abundance profiles (ROAPs). In that sense, the EcoTest belongs to a growing class of tests for differences in β diversity among samples (Crist and Veech 2006, Anderson et al. 2011).

In the biogeographic realm, the comparison of interest is the profile of the rarefaction curve, without regard to the underlying species composition. In biogeographic comparisons among distinct regions or continents, it is known at the outset that there are strong differences in species composition, so the EcoTest is not assessing the appropriate null hypothesis. For this question, we were not able to devise an appropriate algorithm based on pooling of samples to generate a composite distribution for random subsampling. Because the species names in the individual samples are not retained, there does not seem to be a reliable way to determine the rank order of each species in the full, pooled assemblage. For this reason, we used a family of log-normal distributions, in which the null assemblages were determined from randomly chosen parameters for species richness and the standard deviation, the two underlying parameters of the log-normal.

It has long been recognized that both parameter estimation and the goodness of fit of empirical rank abundance data are very sensitive to the choice of the underlying statistical distribution that is used (O'Hara 2005, McGill et al. 2007). For this reason, we created null distribution data sets using log-normal, geometric series, and broken stick distributions. The BiogTest had satisfactory performance in most cases, although error rates were unacceptably high when the samples were drawn from a geometric series (Table 4). In retrospect, the poor fit with the geometric series arises because the dominance fraction D was chosen from a uniform range of 0 to 1. For most of this range, the resulting rank abundance distributions fall off very steeply, which does not fit well with most log-normal distributions or with most empirical data sets. In contrast, the rank abundance profile of a broken stick series is more even, and the BiogTest with broken stick and log-normal samples performs much better than with geometric series samples.

Application to case studies

As we anticipated, the same empirical data set can give different answers when tested with EcoTest vs. BiogTest. For example, with Hill number $q = 2$, the two Malaysian forest samples (Pasoh and Lambir; dashed lines in Fig. 6) and the two African forest samples (Edoro and Korup; dotted lines in Fig. 6) differ by EcoTest, but not by BiogTest (Table 5). Visually, these pairs of curves are nearly coincident in the sample-based rarefaction plots of Fig. 6e, so it is a sensible result that the null hypothesis was not rejected by BiogTest in these cases. These results are reassuring because the sample sizes underlying these comparisons were very large (9382 to 32 611 individuals). In such cases, there is a danger that H_0 might always be rejected when the sample is large enough, even when effect sizes are very small.

When plotted against coverage, however, the BiogTest for these same comparisons is statistically significant (Table 5). Coverage-based analyses do not necessarily give the same results as traditional sample- or individual-based rarefaction (Chao and Jost 2012), as the shapes of species accumulation curves are very different when plotted against coverage rather than sample size. We note that the coverage-based rarefaction curves for these data did not overlap as did the individual-based curves (Fig. 6e vs. 6f), which corresponds to the different outcomes of the statistical tests.

For the smaller-scale comparisons of three mountain assemblages, the sample-based rarefaction curves for the Northern Mountains and the Highlands overlapped closely (Fig. 7a, c, e), and did not differ by the BiogTest for any of the Hill numbers (Table 6). For the coverage-based analyses, the rarefaction curves again diverged, although only the curves for species richness (Hill number $q = 0$) were statistically significant (Fig. 7b, d, f; Table 6). This result is not entirely unexpected because the Sierra Madre and the Chiapas Massif (which includes the Highlands and Northern Mountains) have a different geological history, and were isolated altitudinally by the Central Depression (Fig. 7b) from the Late Jurassic to the Late Cretaceous (Padilla y Sánchez 2007). In addition, the two areas have different histories of land use, with the Highlands and Northern Mountains experiencing stronger human impacts from hunting, logging, and agriculture (Ramírez-Marcial et al. 2001, Cayuela et al. 2006) than the Sierra Madre (N. Ramírez-Marcial, *personal communication*).

Further prospects

We have presented these analyses in terms of familiar diversity metrics of species richness and higher-order Hill numbers. However, the same approach could be used with any other diversity metric, including univariate measures of trait, functional, and phylogenetic diversity. The Chao et al. (2014) procedures could also be used to improve our tests, but they require asymptotic estimators of the diversity metric and its variance, and those are so far available only for Hill numbers (including species richness). Our tests can be applied to any diversity metric that can be calculated for a reference sample and for bootstrapped random subsamples.

In spite of a long history of use of rarefaction methods in ecology and evolution, new statistical developments continue to improve the estimation of biodiversity patterns and statistical inference from sample data. The distinction between the ecological and biogeographic null hypotheses may prove useful in pinpointing how differences in species composition, relative abundance, and species richness contribute to biodiversity patterns.

Acknowledgments

This study was possible thanks to a mobility grant of the Spanish Ministry of Education to L. Cayuela (José de Castillejos Programme, CAS 12/00140), and partial funding by project CGL2013-45634-P. N. J. Gotelli was supported by U.S. NSF DEB 1257625, NSF DEB 1144055, 2, and NSF DEB 1136644. R. K. Colwell was supported by Ciencia sem Fronteiras fellowship (Brazil). Mario González- Espinosa and Neptalí Ramírez-Marcial provided part of the plot data for the Chiapas data set and valuable comments on the interpretation of the results for this case study. We thank Anne Chao for comments on the manuscript and for extended correspondence on rarefaction and asymptotic diversity estimators. We are indebted to the BioPortal for enabling data analyses in their web-based portal. Note that the BioPortal no longer exists, as it has now become the LifePortal. L. Cayuela, N. J. Gotelli, and R. K. Colwell contributed equally to this paper.

Literature Cited

Alroy, J. 2010. The shifting balance of diversity among major marine animal groups. Science 329:1191–1194.

Amarasekare, P., and R. M. Nisbet. 2001. Spatial heterogeneity, source–sink dynamics, and the local coexistence of competing species. American Naturalist 158:572–584.

Anderson, M. J., et al. 2011. Navigating the multiple meanings of β diversity: a roadmap for the practicing ecologist. Ecology Letters 14:19–28.

Angiosperm Phylogeny Group. 2009. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. Botanical Journal of the Linnean Society 161:105–121.

Baldeck, C. A., et al. 2013. Habitat filtering across tree life stages in tropical forest communities. Proceedings of the Royal Society B 280:20130548.

Bolker, B. M. 2008. Ecological models and data in R. Princeton University Press, Princeton, New Jersey, USA.

Brose, U., N. D. Martinez, and R. J. Williams. 2003. Estimating species richness: sensitivity to sample coverage and insensitivity to spatial patterns. Ecology 84:2364–2377.

Cayuela, L., et al. 2012*a*. The Tree Biodiversity Network (BIOTREE-NET): prospects for biodiversity research and conservation in the tropics. Biodiversity and Ecology 4:211–224.

Cayuela, L., J. D. Golicher, J. M. Rey-Benayas, M. González-Espinosa, and N. Ramírez-Marcial. 2006. Effects of fragmentation and disturbance on tree diversity in tropical montane forests. Journal of Applied Ecology 43:1172–1182.

Cayuela, L., and N. J. Gotelli. 2014. rareNMtests: ecological and biogeographical null model tests for comparing rarefaction curves. R package version 1.0. http://cran.r-project.org/package=rareNMtests

Cayuela, L., I. Granzow de la Cerda, F. S. Albuquerque, and J. D. Golicher. 2012*b*. Taxonstand: an R package for species names standardisation in vegetation databases. Methods in Ecology and Evolution 3(6):1078–1083.

Chao, A. 1984. Nonparametric estimation of the number of classes in a population. Scandinavian Journal of Statistics 11:265–270.

Chao, A., R. L. Chazdon, R. K. Colwell, and T. J. Shen. 2005. A new statistical approach for assessing compositional similarity based on incidence and abundance data. Ecology Letters 8:148–159.

Chao, A., C. H. Chiu, and L. Jost. 2010. Phylogenetic diversity measures based on Hill numbers. Philosophical Transactions of the Royal Society B 365:3599–3609.

Chao, A., R. K. Colwell, C. W. Lin, and N. J. Gotelli. 2009. Sufficient sampling for asymptotic minimum species richness estimators. Ecology 90:1125–1133.

Chao, A., N. J. Gotelli, T. C. Hsieh, E. L. Sander, K. H. Ma, R. K. Colwell, and A. M. Ellison. 2014. Rarefaction and extrapolation with Hill numbers: a framework for sampling and estimation in species diversity studies. Ecological Monographs 84:45–67.

Chao, A., and L. Jost. 2012. Coverage-based rarefaction and extrapolation: standardizing samples by completeness rather than size. Ecology 93:2533–2547.

Chase, J. M., and M. A. Leibold. 2003. Ecological niches: linking classical and contemporary approaches. University of Chicago Press, Chicago, Illinois, USA.

Clarke, K. R., P. J. Somerfield, and M. G. Chapman. 2006. On resemblance measures for ecological studies, including taxonomic dissimilarities and a zero-adjusted Bray-Curtis coefficient for denuded assemblages. Journal of Experimental Marine Biology and Ecology 330:55–80.

Coddington, J. A., I. Agnarsson, J. A. Miller, M. Kuntner, and G. Hormiga. 2009. Undersampling bias: the null hypothesis for singleton species in tropical arthropod surveys. Journal of Animal Ecology 78:573–584.

Cohen, J. 1992. A power primer. Psychological Bulletin 112:155–159.

Collins, C. D., R. D. Holt, and B. L. Foster. 2009. Patch size effects on plant species decline in an experimentally fragmented landscape. Ecology 90:2577–2588.

Colwell, R. K., A. Chao, N. J. Gotelli, S. Y. Lin, C. X. Mao, R. L. Chazdon, and J. T. Longino. 2012. Models and estimators linking individual-based and sample-based rarefaction, extrapolation, and comparison of assemblages. Journal of Plant Ecology 5:3–21.

Colwell, R. K., and J. A. Coddington. 1994. Estimating terrestrial biodiversity through extrapolation. Philosophical Transactions of the Royal Society B 345:101–118.

Colwell, R. K., C. X. Mao, and J. Chang. 2004. Interpolating, extrapolating, and comparing incidence-based species accumulation curves. Ecology 85:2717–2727.

Connolly, S. R., M. Dornelas, D. R. Bellwood, and T. P. Hughes. 2009. Testing species abundance models: a new bootstrap approach applied to Indo-Pacific coral reefs. Ecology 90:3138–3149.

Crist, T. O., and J. A. Veech. 2006. Additive partitioning of rarefaction curves and species-area relationships: unifying alpha-, beta- and gamma-diversity with sample size and habitat area. Ecology Letters 9:923–932.

De Cáceres, M., P. Legendre, and F. He. 2013. Dissimilarity measurements and the size structure of ecological communities. Methods in Ecology and Evolution 4:1167–1177.

Gotelli, N. J. 2008. A primer of ecology. Fourth edition. Sinauer, Sunderland, Massachusetts, USA.

Gotelli, N. J., and A. Chao. 2013. Measuring and estimating species richness, species diversity, and biotic similarity from sampling data. Pages 195–211 *in* S. A. Levin, editor. Encyclopedia of biodiversity. Second edition. Volume 5. Academic Press, Waltham, Massachusetts, USA.

Gotelli, N. J., and R. K. Colwell. 2001. Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. Ecology Letters 4:379–391.

Gotelli, N. J., and R. K. Colwell. 2011. Estimating species richness. Pages 39–54 *in* A. E. Magurran and B. J. McGill, editors. Biological diversity: frontiers in measurement and assessment. Oxford University Press, New York, New York, USA.

Gotelli, N. J., A. M. Ellison, and B. A. Ballif. 2012. Environmental proteomics, biodiversity statistics and food-web structure. Trends in Ecology & Evolution 27:436–442.

Gotelli, N. J., and W. Ulrich. 2012. Statistical challenges in null model analysis. Oikos 121:171–180.

Green, J. L., and J. B. Plotkin. 2007. A statistical theory for sampling species abundances. Ecology Letters 10:1037–1045.

Hill, M. O. 1973. Diversity and evenness: a unifying notation and its consequences. Ecology 54:427–432.

Hurlbert, S. H. 1971. The non-concept of species diversity: a critique and alternative parameters. Ecology 52:577–586.

James, F. C., and N. O. Wamer. 1982. Relationships between temperate forest bird communities and vegetation structure. Ecology 63:159–171.

Legendre, P., and E. D. Gallagher. 2001. Ecologically meaningful transformations for ordination of species data. Oecologia 129:271–280.

Leibold, M. A., et al. 2004. The metacommunity concept: a framework for multi-scale community ecology. Ecology Letters 7:601–613.

Li, J., and C. X. Mao. 2012. Simultaneous confidence inference on species accumulation curves. Journal of Agricultural, Biological, and Environmental Statistics 17:1–14.

Limpert, E., W. A. Stahel, and M. Abbt. 2001. Log-normal distributions across the sciences: keys and clues. BioScience 51:341–352.

Losos, J. B. 2010. Adaptive radiation, ecological opportunity, and evolutionary determinism. American Naturalist 175:623–639.

Magurran, A. E. 2003. Measuring biological diversity. Wiley-Blackwell, Hoboken, New Jersey, USA.

Magurran, A. E., and B. J. McGill, editors. 2011. Biological diversity: frontiers in measurement and assessment. Oxford University Press, Oxford, UK.

Manly, B. F. J. 2006. Randomization, bootstrap and Monte Carlo methods in Biology. Third edition. Chapman & Hall/CRC Press, Boca Raton, Florida, USA.

McGill, B. J. 2003. Does Mother Nature really prefer rare species or are log-left-skewed SADs a sampling artefact? Ecology Letters 6:766–773.

McGill, B. J., et al. 2007. Species abundance distributions: moving beyond single prediction theories to integration within an ecological framework. Ecology Letters 10:995–1015.

McGill, B. J., B. Maurer, and M. D. Weiser. 2006. Empirical evaluation of neutral theory. Ecology 87:1411–1423.

O'Hara, R. B. 2005. Species richness estimators: how many species can dance on the head of a pin? Journal of Animal Ecology 74:375–386.

Padilla y Sánchez, R. J. 2007. Evolución geológica del sureste mexicano desde el Mesozoico al presente en el contexto regional del Golfo de México. Boletín de la Sociedad Geológica Mexicana 59(1):19–42.

Payton, M. E., M. H. Greenstone, and N. Schenker. 2003. Overlapping confidence intervals or standard intervals? What do they mean in terms of statistical significance? Journal of Insect Science 3(34):1–6.

Preston, F. W. 1962*a*. The canonical distribution of commonness and rarity: part I. Ecology 43:185–215.

Preston, F. W. 1962*b*. The canonical distribution of commonness and rarity: part II. Ecology 43:410–432.

Qian, H., and R. E. Ricklefs. 2008. Global concordance in diversity patterns of vascular plants and terrestrial vertebrates. Ecology Letters 11:547–553.

R Development Core Team. 2013. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.r-project.org/

Ramírez-Marcial, N., M. González-Espinosa, and G. Williams-Linera. 2001. Anthropogenic disturbance and tree diversity in montane rain forests in Chiapas, Mexico. Forest Ecology and Management 154(1):311–326.

Ricklefs, R. E., and D. Schluter, editors. 1993. Species diversity in ecological communities: historical and geographical perspectives. University of Chicago Press, Chicago, Illinois, USA.

Simberloff, D. 1972. Properties of the rarefaction diversity measurement. American Naturalist 106:414–418.

Simberloff, D. 1979. Rarefaction as a distribution-free method of expressing and estimating diversity. Pages 159–170 *in* J. F. Grassle and G. P. Patil, editors. Ecological diversity in theory and practice. International Cooperative Publishing House, Fairland, Maryland, USA.

Soberon, J., and J. Llorente. 1993. The use of species accumulation functions for the prediction of species richness. Conservation Biology 7:480–488.

Solow, A. 1993. A simple test for change in community structure. Journal of Animal Ecology 62:191–193.

Stevens, M. H. H., and W. P. Carson. 1999. Plant density determines species richness along an experimental fertility gradient. Ecology 80:455–465.

Sugihara, G., L. F. Bersier, T. R. E. Southwood, S. L. Pimm, and R. M. May. 2003. Predicted correspondence between species abundances and dendrograms of niche similarities. Proceedings of the National Academy of Sciences USA 100:5246–5251.

Tipper, J. C. 1979. Rarefaction and rarefiction; the use and abuse of a method in paleobiology. Paleobiology 5:423–434.

Toft, C. A., and P. J. Shea. 1983. Detecting community-wide patterns: estimating power strengthens statistical inference. American Naturalist 122:618–625.

Tokeshi, M. 1993. Species abundance patterns and community structure. Advances in Ecological Research 24:111–186.

Ulrich, W., M. Ollik, and K. I. Ugland. 2010. A meta-analysis of species–abundance distributions. Oikos 119:1149–1155.

Villeger, S., N. W. H. Mason, and D. Mouillot. 2008. New multidimensional functional diversity indices for a multifaceted framework in functional ecology. Ecology 89:2290–2301.

Violle, C., M. L. Navas, D. Vile, E. Kazakou, C. Fortunel, I. Hummel, and E. Garnier. 2007. Let the concept of trait be functional! Oikos 116:882–892.

Walther, B. A., and J. L. Moore. 2005. The concepts of bias, precision and accuracy, and their use in testing the performance of species richness estimators, with a literature review of estimator performance. Ecography 28:815–829.

Weiher, E., and P. Keddy. 1999. Ecological assembly rules: perspectives, advances, retreats. Cambridge University Press, Cambridge, UK.

Wiens, J. J., and M. J. Donoghue. 2004. Historical biogeography, ecology, and species richness. Trends in Ecology and Evolution 19:639–644.

Wilson, J. B. 1993. Would we recognize a broken-stick community if we found one? Oikos 67:181–183.

Wilson, S. D., and D. Tilman. 1991. Components of plant competition along an experimental gradient of nitrogen availability. Ecology 72:1050–1065.

Zuur, A., E. N. Ieno, N. Walker, A. A. Saveliev, and G. M. Smith. 2009. Mixed effects models and extensions in ecology with R. Springer, New York, New York USA.

Supplemental Material

Ecological Archives

Appendices A and B and the Supplement are available online: http://dx.doi.org/10.1890/14-1261.1.sm