DIGITAL ACCESS TO SCHOLARSHIP AT HARVARD
DASH.HARVARD.EDU

HARVARD LIBRARY
Office for Scholarly Communication

Rarefaction and Extrapolation with Hill Numbers: A Framework for Sampling and Estimation in Species Diversity Studies

Citation
Chao, Anne, Nicholas J. Gotelli, T. C. Hsieh, Elizabeth L. Sander, K. H. Ma, Robert K. Colwell, and Aaron M. Ellison. "Rarefaction and extrapolation with Hill numbers: a framework for sampling and estimation in species diversity studies." Ecological Monographs 84, no. 1 (2014): 45-67.

Published Version
doi:10.1890/13-0133.1

Permanent link
http://nrs.harvard.edu/urn-3:HUL.InstRepos:11148822

Terms of Use
This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. Submit a story .

Accessibility

Rarefaction and extrapolation with Hill numbers: a framework for sampling and estimation in species diversity studies

Anne Chao,[1] Nicholas J. Gotelli,[2] T. C. Hsieh,[1] Elizabeth L. Sander,[2] K. H. Ma,[1]

Robert K. Colwell,[3,4] and Aaron M. Ellison[5]

[1] Institute of Statistics, National Tsing Hua University, Hsin-Chu 30043, Taiwan

[2] Department of Biology, University of Vermont, Burlington, VT 05405, USA

[3] Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT 06269, USA

[4] University of Colorado Museum of Natural History, Boulder, CO 80309, USA

[5] Harvard University, Harvard Forest, 324 North Main Street, Petersham, MA 01366, USA

Abstract. Quantifying and assessing changes in biological diversity are central aspects of many ecological studies, yet accurate methods of estimating biological diversity from sampling data have been elusive. Hill numbers, or the effective number of species, are increasingly used to characterize the taxonomic, phylogenetic or functional diversity of an assemblage. However, empirical estimates of Hill numbers, including species richness, tend to be an increasing function of sampling effort and thus tend to increase with sample completeness. Integrated curves based on sampling theory that smoothly link rarefaction (interpolation) and prediction (extrapolation) standardize samples on the basis of sample size or sample completeness and facilitate the comparison of biodiversity data. Here we extend previous rarefaction and extrapolation models for species richness (Hill number ${}^{q}D$, where $q = 0$) to measures of taxon diversity incorporating relative abundance (i.e., for any Hill number ${}^{q}D$, $q > 0$) and present a unified approach for both individual-based (abundance) data and sample-based (incidence) data. Using this unified sampling framework, we derive both theoretical formulas and analytic estimators for seamless rarefaction and extrapolation based on Hill numbers. Detailed examples are provided for the first three Hill numbers: $q = 0$ (species richness), $q = 1$ (the exponential of Shannon's entropy index) and $q = 2$ (the inverse of Simpson's concentration index). We develop a bootstrap method for constructing confidence intervals around Hill numbers, facilitating the comparison of multiple assemblages of both rarefied and extrapolated samples. The proposed estimators are accurate for both rarefaction and short-range extrapolation. For long-range extrapolation, the performance of the estimators depends on both the value of q and on the extrapolation range. We tested our methods on simulated data generated from species abundance models and on data from large species inventories. We also illustrate the formulas and estimators using empirical datasets from biodiversity surveys of temperate forest spiders and tropical ants.

Key words: abundance data, diversity, extrapolation, Hill numbers, incidence data, interpolation, prediction, rarefaction, sample coverage, species richness

INTRODUCTION

The measurement and assessment of biological diversity (biodiversity) is an active research focus of ecology (Magurran 2004, Magurran and McGill 2011) and a central objective of many monitoring and management projects (Groom et al. 2005, CBD 2012). The simplest and still the most frequently used measure of biodiversity is the species richness of an assemblage. Species richness features prominently in foundational models of community ecology (MacArthur and Wilson 1967, Connell 1978, Hubbell 2001), and is a key metric in conservation biology (Brook et al. 2003, May 1988) and historical biogeography (Wiens and Donoghue 2004). In spite of its intuitive and universal appeal, however, species richness is a problematic index of biodiversity for two reasons related to sampling intensity and the species abundance distribution.

First, observed species richness is highly sensitive to sample size (the *sampling problem*). Because most species in an assemblage are rare, biodiversity samples are usually incomplete, and undetected species are a common problem. As a consequence, the observed number of species in a well-defined biodiversity sample (species density, *sensu* Gotelli and Colwell 2001) is known to be a biased underestimate of true species richness, and is highly sensitive to the area surveyed, the number of individuals counted, and the number of samples scored for species occurrence (incidence; Colwell and Coddington 1994). Thus, from a statistical perspective, species richness is very difficult to estimate accurately from a finite sample.

A second problem with species richness as a measure of biodiversity is that it does not incorporate any information about the relative abundance of species (the *abundance problem*).

By counting all species equally, species richness weights rare species the same as common ones. If two assemblages have identical species richness, it seems intuitive that any subjective sense of "diversity" should be higher in the assemblage with more-equal abundances among all the component species, whereas diversity should be lower in the assemblage that is dominated by the abundance of one or a few common species (Pielou 1975). Incorporating abundance into a biodiversity index is critical for studies of many (but not all) aspects of ecosystem function, because rare species usually make little contribution to important measures of ecosystem function such as biomass, productivity, or nutrient retention (Schwartz et al. 2000). On the other hand, rare species sometimes play key roles in ecosystem function (e.g. top predators, Terborgh et al. 2001) and are generally of greater conservation and management concern than are common ones (May 1988, Holsinger and Gottlieb 1991; but see Gaston and Fuller 2008).

An extensive literature addresses both of these issues. For the sampling problem, standardized comparisons of species richness can be made after interpolation with rarefaction (Tipper 1979) to a common level of abundance (Sanders 1968, Hurlbert 1971, Simberloff 1972, Gotelli and Colwell 2001, 2011), sampling effort (Colwell et al. 2004), or sample completeness (Alroy 2010, Jost 2010, Chao and Jost 2012). Alternatively, biodiversity data can be used to estimate an asymptotic estimator of species richness that is relatively independent of additional sampling effort. Methods for obtaining asymptotic richness estimators include estimating the area beneath a smoothed curve of a parametric species-abundance distribution (Fisher et al. 1943, Connolly and Dornelas 2011), extending the species accumulation curve by fitting parametric functions (Soberón and Llorente 1993), or using non-parametric asymptotic richness estimators (Chao 1984, Colwell and Coddington 1994) that are based on the frequency of rare species in a sample. Although many ecologists still publish analyses of raw species density data, rarefaction

and asymptotic estimators based on statistical sampling theory are becoming standard tools in biodiversity analysis (Gotelli and Ellison 2012).

Colwell et al. (2012) recently unified the interpolation and extrapolation procedures for species richness. They showed that a single, smooth sampling curve (with an expectation and an unconditional variance), derived from a *reference sample* – a collection of individuals (or sampling units) that would be gathered in a typical biodiversity survey – can be interpolated (rarefied) to smaller sample sizes or extrapolated to a larger sample size, guided by an estimate of asymptotic richness. Thus, rigorous statistical comparison of species richness can be performed not only for rarefied subsamples but also for extrapolated richness values based on samples of arbitrary and equal size. Chao and Jost (2012) developed coverage-based rarefaction and extrapolation methodology to compare species richness of a set of assemblages based on samples of equal completeness (equal coverage). The Colwell et al. (2012) sample-size-based approach standardizes based on sample effort, whereas the Chao and Jost (2012) coverage-based approach standardizes based on sample completeness, an estimated assemblage characteristic. The sample-size- and coverage-based integration of rarefaction and extrapolation together represent a unified framework for estimating species richness and for making statistical inferences based on these estimates.

Like the sampling problem, the abundance problem has been recognized for decades in the ecological literature. Ecologists have introduced a plethora of diversity indices that combine species richness and the proportion of each species into a single metric (Washington 1984). These indices tend to be highly correlated with one another, are not always expressed in units that are intuitive, sensible, or that allow comparisons, and have sampling and statistical properties that have been poorly studied (Ghent 1991). Hill numbers are a mathematically

unified family of diversity indices (differing among themselves only by an exponent q) that incorporate relative abundance and species richness and overcome many of these shortcomings. They were first used in ecology by MacArthur (1965), developed by Hill (1973), and recently reintroduced to ecologists by Jost (2006, 2007).

Hill numbers offer five distinct advantages over other diversity indices. First, Hill numbers obey an intuitive *replication principle or doubling property*. Hill (1973) proved a weak version of the doubling property: if two completely distinct assemblages (i.e., no species in common) have identical relative abundance distributions, then the Hill number doubles if the assemblages are combined with equal weights. Chiu et al. (2013, their Appendix B) recently proved a strong version of the doubling property: if two completely distinct assemblages have identical Hill numbers of order q (relative abundance distributions may be different, unlike the weak version), then the Hill number of the same order doubles if the two assemblages are combined with equal weights. Species richness is a Hill number (with $q = 0$) and obeys both versions of the doubling property, but most other diversity indices do not obey even the weak version.

A second advantage of Hill numbers is that they are all, expressed in units of *effective numbers of species*: the number of equally abundant species that would be needed to give the same value of a diversity measure. Third, key diversity indices proposed in the literature, including the widely used Shannon entropy and the Gini-Simpson index, can be converted to Hill numbers by simple algebraic transformations. Fourth, Hill numbers can be effectively generalized to incorporate taxonomic, phylogenetic, and functional diversity and thus provide a unified framework for measuring biodiversity (Chao et al. 2010, Gotelli and Chao 2013). Fifth, in the comparison of multiple assemblages, there is a direct link between Hill numbers and species compositional similarity (or differentiation) among assemblages (Jost 2007). This

property unites diversity and similarity (or differentiation).

Although species richness is one of the Hill numbers, the literature on Hill numbers and on sampling models for species richness have developed independently. The recent literature generally fails to emphasize that Hill numbers other than species richness (those with $q > 0$) are also sensitive to the number of individuals or samples collected, although the under-sampling bias is progressively less severe for Hill numbers of higher orders of q. In theory, simple rarefaction curves can be constructed for any diversity index by resampling (Walker et al. 2008, Ricotta et al. 2012), although only recently has this been done explicitly for Hill numbers (Colwell 2013, Gotelli and Ellison 2012). Asymptotic estimators for Hill numbers with $q = 1$ are closely related to the well-known entropy estimation (see Paninski 2003, and Chao et al. 2013 for reviews). For $q = 2$ and any integer > 2, nearly unbiased estimators exist (Nielsen et al. 2003, Gotelli and Chao, 2013).

In this paper, we unify the two fundamental frameworks used for the measurement and estimation of species diversity: rarefaction/extrapolation and Hill numbers. Specifically, we generalize the sample-size-based approach of Colwell et al. (2012) and the coverage-based approach of Chao and Jost (2012) to the entire family of Hill numbers. We provide asymptotic estimators for Hill numbers and use them to link analytic estimators for rarefaction and extrapolation from an empirical reference sample. To characterize the species diversity of an assemblage, we propose using three integrated rarefaction/extrapolation curves based on the first three Hill numbers: species richness, the exponential of Shannon entropy (which we refer to as *Shannon diversity*), and the inverse Simpson concentration (which we refer to as *Simpson diversity*). The formulas and estimators are tested with simulated data generated from species abundance models/inventories and applied to several empirical data sets. Finally, we highlight

the close theoretical links between Hill numbers and expected species accumulation curves (Hurlbert 1971, Dauby and Hardy 2011). With this expanded framework, ecologists will be able to effectively use Hill numbers for a host of problems in biodiversity estimation, including comparison of the species diversity of different assemblages in time or space, with reliable statistical inferences about these comparisons.

TWO TYPES OF DATA AND MODELS

To describe model parameters and sample data, we adopt the notation and terminology of Colwell et al. (2012) and Gotelli and Chao (2013). Consider a species assemblage consisting of N total individuals, each belonging to one of S distinct species. The total abundance of species i is N_i, where $i = 1, 2, \ldots, S$, $N_i > 0$, and $N = \sum_{i=1}^{S} N_i$. Let $p_i = N_i/N$ denote the true relative abundance of species i, so that $\sum_{i=1}^{S} p_i = 1$. We emphasize that the quantities N, S, $(N_1, N_2, ..., N_S)$ and $(p_1, p_2, ..., p_S)$ are the *parameters* representing, respectively, the true (albeit unknown) underlying assemblage size, the complete species richness of the assemblage, and the species absolute and relative abundance sets. We consider two sampling data structures for reference samples.

Individual-based (abundance) data and model

In most biological surveys, a sample of n individuals is taken with replacement from the assemblage, and a total of S_{obs} ($\leq S$) species are observed. (If individuals are sampled without replacement, we need to assume that the assemblage size N is much larger than the sample size n). Let X_i be the number of individuals of the ith species that are observed in the sample, $i = 1, 2, \ldots, S$; we refer to X_i as the *sample species frequency*. Let f_k be the number of species represented by exactly k individuals in the sample, $k = 0, 1, ..., n$; we refer to f_k as the *abundance frequency*

counts. From these definitions, $n = \sum_{i=1}^{S} X_i = \sum_{k \geq 1} k f_k$, and $S_{obs} = \sum_{k \geq 1} f_k$. In particular, f_1 is the number of species represented by exactly one individual (*singletons*) in the sample, and f_2 is the number of species represented by exactly two individuals (*doubletons*). The unobservable frequency f_0 denotes the number of species present in the entire assemblage but are not observed in the sample.

The multinomial probability distribution is the most widely used model for the observed species sample frequencies $(X_1, X_2, \ldots, X_S)$ for given S and $(p_1, p_2, \ldots, p_S)$:

$$P(X_1 = x_1, \ldots, X_S = x_S) = \frac{n!}{x_1! \ldots x_S!} p_1^{x_1} p_2^{x_2} \ldots p_S^{x_S} . \qquad (1)$$

Note that undetected species, i.e., $X_i = 0$, do not contribute to this distribution. In this model, the detection probability for the ith species is simply the true relative abundance $p_i = N_i / N$. In this case, the sample size n is fixed. Thus, the number of individuals represented by any single species is at most n, which is fixed by the sampling design.

Alternatively, abundance data can also be collected by sampling a fixed area or by applying a fixed sampling effort, rather than a fixed sample size. With this sampling protocol, the sample size is a random variable and thus cannot be fixed in advance, implying that the number of individuals represented by any single species can be large, without any particular limit. A commonly used area-based model is the Poisson product model, which assumes that individuals of the ith species accumulate in the reference sample according to a Poisson process (Ross 1995). This model can be traced to Fisher et al. (1943) and forms the basis for Coleman et al.'s (1982) random sampling model for species-area relationships. As shown by Colwell et al. (2012), the Poisson product model produces results that are virtually indistinguishable from those based on a multinomial model. A statistical reason for this is that the Poisson product model is closely

related to a multinomial model; see Chao and Chiu (2013) for details. Therefore, we consider only the multinomial model, which can accommodate both individual-based and area-based abundance data.

Sample-based (incidence) data and model

When the sampling unit is not an individual, but a trap, net, quadrat, plot, or timed survey, it is these *sampling units*, not the individual organisms that are sampled randomly and independently. Because it is not always possible to count individuals within a sampling unit, estimation can be based on a set of sampling units in which only the incidence (presence) of each species is recorded. The reference sample for such incidence data consists of a set of T sampling units. The presence or absence (technically, non-detection) of each species within each sampling unit is recorded to form a species-by-sampling-unit incidence matrix (W_{ij}) with S rows and T columns. The value of the element W_{ij} of this matrix is 1 if species i is recorded in the jth sampling unit, and 0 if it is absent. The row sum of the incidence matrix $Y_i = \sum_{j=1}^{T} W_{ij}$ denotes the *incidence-based frequency* of species i, i = 1, 2, ..., S. Here, Y_i is analogous to X_i in the individual-based frequency vector. Species present in the assemblage but not detected in any sampling unit yield $Y_i = 0$. The total number of species observed in the reference sample is S_{obs} (only species with $Y_i > 0$ contribute to S_{obs}).

Following Colwell et al. (2012), we adopt a *Bernoulli product model*, which assumes that the ith species has its own unique *incidence probability* π_i that is constant for any randomly selected sampling unit. Each element W_{ij} in the incidence matrix is a Bernoulli random variable (since $W_{ij} = 0$ or $W_{ij} = 1$), with probability π_i that $W_{ij} = 1$ and probability $1-\pi_i$ that $W_{ij} = 0$. The probability distribution for the incidence matrix is

$$P(W_{ij} = w_{ij}) = \prod_{i=1}^{S} \prod_{j=1}^{T} \pi_i^{w_{ij}} (1-\pi_i)^{1-w_{ij}} = \prod_{i=1}^{S} \pi_i^{y_i} (1-\pi_i)^{T-y_i} \ . \qquad (2a)$$

The model is equivalent to a binomial product model for the observed row sums $(Y_1, Y_2, \ldots, Y_S)$:

$$P(Y_i = y_i, i = 1, 2, \ldots, S) = \prod_{i=1}^{S} \binom{T}{y_i} \pi_i^{y_i} (1-\pi_i)^{T-y_i} \ . \qquad (2b)$$

Here the probability of incidence (occurrence) π_i is the probability that species *i* is detected in a sampling unit. In Appendix A, we describe a more general case (quadrat sampling) to interpret the model and explain how this model can incorporate spatial aggregation.

Let Q_k denote the *incidence frequency counts*, the number of species that are detected in exactly *k* sampling units, $k = 0, 1, \ldots, T$, i.e., Q_k is the number of species each represented exactly $Y_i = k$ times in the incidence matrix sample. Here Q_k is analogous to f_k in the abundance data. The total number of incidences *U* recorded in the *T* sampling units is analogous to *n* in the abundance data. Here *U* is a random variable and can be expressed as $U = \sum_{k=1}^{T} kQ_k = \sum_{i=1}^{S} Y_i$, and the number of observed species is $S_{obs} = \sum_{k=1}^{T} Q_k$. Here, Q_1 represents the number of *unique* species (those that are each detected in only one sampling unit) and Q_2 represents the number of *duplicate* species (those that are each detected in exactly two sampling units). The unobservable zero frequency count Q_0 denotes the number of species among the *S* species present in the assemblage that are not detected in any of the *T* sampling units.

HILL NUMBERS

Abundance data

Hill (1973) integrated species richness and species abundances into a class of diversity measures later called Hill numbers, or effective numbers of species, defined for $q \neq 1$ as

$$ {}^{q}D = \left(\sum_{i=1}^{S} p_i^q \right)^{1/(1-q)} \qquad (3a) $$

in which S is the number of species in the assemblage and the ith species has relative abundance p_i, $i = 1, 2, \ldots, S$. The parameter q determines the sensitivity of the measure to the relative frequencies. When $q = 0$, the abundances of individual species do not contribute to the sum in Eq. 3a. Rather, only presences are counted, so that ${}^{0}D$ is simply species richness. For $q = 1$, Eq. 3a is undefined, but its limit as q tends to 1 is the exponential of the familiar Shannon index, referred to here as Shannon diversity:

$$ {}^{1}D = \lim_{q \to 1} {}^{q}D = \exp\left(-\sum_{i=1}^{S} p_i \log p_i \right). \qquad (3b) $$

${}^{1}D$ weighs species in proportion to their frequency. When $q = 2$, Eq. 3a yields Simpson diversity, the inverse of the Simpson concentration:

$$ {}^{2}D = 1 / \sum_{i=1}^{S} p_i^2, \qquad (3c) $$

which places more weight on the frequencies of abundant species and discounts rare species. Investigators using Hill numbers should report, at least, the diversity of all species ($q = 0$), of "typical" species ($q = 1$), and of dominant species ($q = 2$). For a general order of q, if ${}^{q}D = x$, then the diversity is equivalent to that of an idealized assemblage with x equally abundant species, which is why Hill numbers are referred to as effective numbers of species or as species equivalents.

A complete characterization of the species diversity of an assemblage with S species and *relative abundances* $(p_1, p_2, \ldots, p_S)$ is conveyed by a diversity profile – a plot of ${}^{q}D$ versus q – from $q = 0$ to $q = 3$ or 4 (beyond this it changes little) (see Tóthmérész 1995). Although Hill numbers for $q < 0$ can be calculated, they are dominated by the frequencies of rare species and

have poor statistical sampling properties. We thus restrict ourselves to the case $q \geq 0$ throughout the paper. An example of a diversity profile is shown in Fig. 1a.

Hill numbers can be regarded as the theoretical or asymptotic diversities at a sample *size of infinity* for which the true relative abundances $\{p_1, p_2, ..., p_S\}$ of each of *i* species are known. When sample size is relevant for discussion, we use the notation ${}^qD(\infty)$ to denote the (asymptotic) Hill numbers. Throughout the paper, we use ${}^qD$ and ${}^qD(\infty)$ interchangeably, i.e., ${}^qD = {}^qD(\infty)$.

Incidence data

As far as we are aware, Hill numbers have been discussed only for abundance data and have not previously been defined for sample-based incidence data. Here, we propose the following Hill numbers for sample-based incidence data, based on the Bernoulli product model (Eq. 2a) or equivalently, the binomial product model (Eq. 2b). With either of these two models, $\sum_{i=1}^{S} \pi_i$ may be greater than 1. So we first normalize each parameter π_i (i.e. divide each π_i by the sum $\sum_{i=1}^{S} \pi_i$) to yield the *relative incidence* of the *i*th species in the assemblage. This relative incidence is assumed to be the same for any randomly selected sampling unit. Hill numbers of order *q* for incidence data are defined as:

$$ {}^q\Delta = \left(\sum_{i=1}^{S} \left[\frac{\pi_i}{\sum_{j=1}^{S} \pi_j} \right]^q \right)^{1/(1-q)}, \; q \geq 0, \; q \neq 1. \qquad (4a) $$

As with abundance data, Hill numbers for incidence data also represent the theoretical or asymptotic diversities when the number of sampling units is infinity. So we also use ${}^q\Delta$ and ${}^q\Delta(\infty)$ interchangeably, i.e., ${}^q\Delta = {}^q\Delta(\infty)$. Hill numbers ${}^qD$ for abundance data are based on

relative abundances (Eq. 3a), whereas Hill numbers ${}^{q}\Delta$ for incidence data are based on relative incidence in the assemblage (Eq. 4a). The parameter q in Eq. 4a determines the sensitivity of ${}^{q}\Delta$ to the relative incidences. If all the incidence probabilities $(\pi_1, \pi_2, ..., \pi_S)$ are identical, then Hill numbers of all orders become species richness. The Hill number ${}^{q}\Delta$ for incidence data is interpreted as the effective number of *equally frequent* species in the assemblage from which the sampling units are drawn. That is, if ${}^{q}\Delta = y$, then the diversity of the assemblage is the same as that of an idealized assemblage with y species all of equal probability of incidence.

Equation 4a yields species richness for incidence data when $q = 0$. As with Eq. 3b, the limit of ${}^{q}\Delta$ as q tends to 1 exists and gives

$$ {}^{1}\Delta = \lim_{q \to 1} {}^{q}\Delta = \exp\left(-\sum_{i=1}^{S} \frac{\pi_i}{\sum_{j=1}^{S} \pi_j} \log \frac{\pi_i}{\sum_{j=1}^{S} \pi_j} \right), \qquad (4b) $$

which is equal to the Shannon diversity for incidence data, i.e., the exponential of Shannon entropy based on the relative incidences in the assemblage. When $q = 2$, Eq. 4a becomes

$$ {}^{2}\Delta = 1 / \sum_{i=1}^{S} \left(\frac{\pi_i}{\sum_{j=1}^{S} \pi_j} \right)^2 \qquad (4c) $$

which is the Simpson diversity for incidence data, i.e., the inverse Simpson concentration based on relative incidences. By analogy to the case for abundance data, a plot of ${}^{q}\Delta$ versus q completely characterizes the species diversity of an assemblage with S species and incidence probabilities $(\pi_1, \pi_2, ..., \pi_S)$.

Diversity accumulation curve

It is well known that empirical species richness varies with sampling effort and thus also varies with sample completeness (as measured by sample coverage; see below). Therefore, we

can plot the expected species richness as a function of sample size (this plot is the familiar species accumulation curve) or as a function of sample coverage. The asymptote of this curve as sample size tends to infinity is the species richness in the entire assemblage. We now extend the concept of species accumulation curve to the concept of a *diversity accumulation curve*.

As discussed for abundance data, the Hill number of a fixed order q defined in Eq. 3a represents the asymptotic diversity at a sample size of infinity. For non-asymptotic diversity, we define the expected diversity ${}^{q}D(m)$ for a finite sample size m as the Hill numbers based on expected abundance frequency counts for a sample of size m, for which data are formed by averaging out samples of size m taken from the entire assemblage. Mathematical formulas and statistical estimation are derived in later sections. See *Discussion* for the advantages of our approach over the alternative approach that defines the non-asymptotic Hill numbers as the average Hill numbers over many samples of size m taken from the entire assemblage. Note that for species richness and expected sample completeness (see below), these two approaches give identical formulas. Our definition similarly can be extended to define the expected diversity ${}^{q}\Delta(m)$ for m sampling units under the model for incidence data.

Based on the above definition, our goal is to construct a diversity accumulation curve as a function of sample size (the number of individuals for abundance data or the number of sampling units for incidence data) or sample completeness. For example, in the model for abundance data, we consider the following focal questions:

(1) When a sample of finite size m drawn at random from the entire assemblage, what are the theoretical formulas for the expected diversity of order q, ${}^{q}D(m)$, for this sample? The plot ${}^{q}D(m)$ as a function of m is the *sample-size-based diversity accumulation curve*. As m tends to

infinity, these expected diversities approach ${}^{q}D = {}^{q}D(\infty)$ as given in Eq. 3a. An example of sample-size-based diversity accumulation curve is given in Fig. 1b.

(2) For a sample of size m, what is the expected sample completeness, $C(m)$, for this sample? The plot ${}^{q}D(m)$ as a function of $C(m)$ is the *coverage-based diversity accumulation curve*. As $C(m)$ tends to unity (complete coverage), these expected diversities also approach ${}^{q}D = {}^{q}D(\infty)$. An example of coverage-based diversity accumulation curve is given in Fig. 1c.

(3) Given the data for a reference sample of size n, what are the analytic estimators for ${}^{q}D(m)$ and $C(m)$? Rarefaction (interpolation) refers to the case $m < n$, whereas prediction (extrapolation) refers to the case $m > n$. The integrated sample-size- or coverage-based rarefaction/extrapolation sampling curve represents the estimated diversity accumulation curve based on the reference sample.

(4) When there are multiple assemblages, how do we compare their diversities based on the rarefaction/extrapolation sampling curves?

To answer the above questions, we derive in later sections the theoretical formulas of ${}^{q}D(m)$ of any finite sample size m and the corresponding analytic estimators, along with their variances and confidence intervals. Thus, sample-size- and coverage-based diversity accumulation curves can be estimated and compared across multiple assemblages. For incidence data, similar questions and the estimation of the diversity accumulation curve can be formulated.

RAREFACTION AND EXTRAPOLATION OF ABUNDANCE DATA USING HILL NUMBERS

A new perspective

The extension of the now well-understood rarefaction and extrapolation of species richness (for a refresher, see Appendix B for abundance data, and Appendix C for incidence data) to the general case of Hill numbers is not direct, and it requires a new perspective, based on abundance frequency counts with a different statistical framework. We first extend the notation f_k (abundance frequency counts of the reference sample of size n) to a more general case. We define the abundance frequency count $f_k(m)$ for *any* $m \geq 1$ as the number of species represented by exactly k individuals in a sample of size m. The expected value of the abundance frequency count $f_k(m)$ can be expressed as (See Appendix D for a proof):

$$E[f_k(m)] = \sum_{i=1}^{S} \binom{m}{k} p_i^k (1-p_i)^{m-k}, \; k = 0, 1, \ldots, m. \qquad (5)$$

Note that $E[f_0(m)] = \sum_{i=1}^{S} (1-p_i)^m$ is the expected number of undetected species in a sample of size m. For the reference sample of size n, the frequency $f_k(n)$ is simply denoted as f_k, as we defined in *Abundance data*.

To derive the theoretical formula of ${}^qD(m)$, we first describe the frequency counts expected in a sample of size m. Suppose a random sample of m individuals is taken from the entire assemblage; we obtain a set of abundance frequency counts for this sample, $\{f_k(m); k = 1, \ldots, m\}$. After an infinite number of samples of size m have been taken, the average of $f_k(m)$ for each $k = 1, 2, \ldots, m$ tends to $E[f_k(m)]$, as derived in Eq. 5. The frequency counts expected in a sample of size m are thus $\{E[f_k(m)]; k = 1, \ldots, m\}$. According to our formulation, the expected diversity of size m, ${}^qD(m)$, are the Hill numbers based on these expected frequencies. Note that, for a sample size of m, the relative abundances of species are simply $1/m$ (there are $E[f_1(m)]$ such species), $2/m$ (there are $E[f_2(m)]$ such species), …, m/m (there are $E[f_m(m)]$ such species). Thus Hill numbers of order q for a sample of size m are:

$$ {}^{q}D(m) = \left(\sum_{k=1}^{m} \left(\frac{k}{m} \right)^{q} \times E[f_k(m)] \right)^{\frac{1}{1-q}}, \; m \geq 1, \; q \neq 1. \qquad (6) $$

This formula is valid for any sample size *m*, which can be either less than the reference sample size *n* or greater than *n*. Therefore, throughout the paper, the theoretical formulas for rarefaction and extrapolation of Hill numbers refer to Eq. 6. The first column in Table 1 summarizes the formulas for the special cases of $q = 0, 1, 2$ and in general for $q > 2$.

Analytic rarefaction and extrapolation estimators for Hill numbers of order q

Based on a reference sample of size *n* with sample frequency X_i for the *i*th species and the observed frequency counts $f_k = f_k(n)$, we derive here the analytic estimators ${}^{q}\hat{D}(m)$ for ${}^{q}D(m)$ given in Eq. 6. The notation "hat" on a diversity, e.g., ${}^{q}\hat{D}(m)$, means an estimator of that diversity based on the reference sample. The observed Hill numbers, ${}^{q}D_{obs}$, for the reference sample are simply ${}^{q}\hat{D}(n)$. That is,

$$ {}^{q}\hat{D}(n) = {}^{q}D_{obs} = [\textstyle\sum_{X_i \geq 1} (X_i / n)^q]^{1/(1-q)} = [\textstyle\sum_{j=1}^{n} (j/n)^q f_j]^{1/(1-q)}. \qquad (7) $$

Our approach to deriving rarefaction formulas is based on statistical estimation theory for the expected frequency counts. The minimum variance unbiased estimator for $E[f_k(m)]$ is:

$$ \hat{f}_k(m) = \sum_{X_i \geq k} \frac{\binom{X_i}{k}\binom{n - X_i}{m - k}}{\binom{n}{m}} = \sum_{j \geq k} \frac{\binom{j}{k}\binom{n - j}{m - k}}{\binom{n}{m}} f_j, \;\; m < n, \; k \geq 1. \qquad (8) $$

See Appendix D for a proof. Here, $\binom{a}{b} \equiv 0$ if $a < b$. We use this conventional definition throughout this paper and the appendices. By substitution (from Eq. 6), we can obtain the following analytic estimators of the expected diversity of an interpolated sample of size *m*:

$$ {}^{q}\hat{D}(m) = \left(\sum_{k=1}^{m} \left(\frac{k}{m} \right)^{q} \times \hat{f}_k(m) \right)^{\frac{1}{1-q}}, \; m < n. \qquad (9a) $$

Equation 9a is the general, nearly unbiased, rarefaction formula for Hill numbers of any order q. (An estimator is nearly unbiased if its bias tends to zero when the reference sample size n is large.) The analytic estimator for the rarefaction of Hill numbers for each of the orders q = 0, 1, and 2 is thus obtained by replacing $E[f_k(m)]$ with $\hat{f}_k(m)$ in the specific formulas provided in Table 1.

The extrapolation of Hill numbers of any order q is a prediction of the expected diversity ${}^{q}D(n+m^*)$ for an augmented sample of size $m = n + m^*$. Although the general formula in Eq. 6 for ${}^{q}D(m)$ also holds for any sample size $m > n$, our estimator $\hat{f}_k(m)$ in Eq. 8 is valid only for $m < n$. Therefore, we cannot simply replace $f_k(m)$ by $\hat{f}_k(m)$ as we did for rarefaction. For each q, we need to develop a different approach by means of an estimate of species richness (for q = 0), Shannon diversity (for q = 1), Simpson diversity (for q = 2), and higher orders (for $q > 2$).

(1) *Species richness (q = 0)*

From Eqs. 6 and 9a, the species richness (q = 0) for a sample of size $m < n$ is

$$ {}^{0}\hat{D}(m) = \sum_{k=1}^{m} \hat{f}_k(m), \; m < n. \qquad (9b) $$

In Appendix D we show that this estimator is identical to the traditional individual-based rarefaction estimator (Hurlbert 1971, Smith and Grassle 1977). Our new perspective, however, offers a simpler, alternative approach to traditional individual-based rarefaction of species richness. Equation 9b shows that the traditional rarefaction estimator of the expected species richness for a sample size of m is simply the sum of the estimated frequency counts. This idea can be extended easily to Hill numbers of any orders, as we next illustrate.

The extrapolated species richness estimator for a sample of $n+m^*$ used in this paper is reviewed in Appendix B and the formula (originally derived by Shen et al. 2003) is shown in Table 1. This approach requires an estimated asymptote of species richness. Any proper species richness estimator can be used. Colwell et al. (2012) suggested using the Chao1 estimator (Chao 1984) or ACE (Chao and Lee 1992) and noted that extrapolation gives reliable estimates only up to approximately double or triple the reference sample size. This limitation is primarily a consequence of the fact that the asymptotic estimator is only a lower bound (Chao 1984).

(2) *Shannon diversity (q = 1)*

From Eqs. 6 and 9a, we have the following nearly unbiased interpolation estimator for the Hill number q =1 (Shannon diversity):

$$ {}^1\hat{D}(m) = \exp\left(\sum_{k=1}^{m} \left(-\frac{k}{m} \log \frac{k}{m} \right) \hat{f}_k(m) \right), \quad m < n. \tag{10a} $$

For our new extrapolation formula, we need an estimator for the asymptote of Shannon diversity. Chao et al. (2013) derived the following nearly unbiased estimator, $\hat{H}(\infty)$, of Shannon entropy $H = H(\infty) = -\sum_{i=1}^{S} p_i \log p_i$:

$$ \hat{H}(\infty) = \sum_{k=1}^{n-1} \frac{1}{k} \sum_{1 \le X_i \le n-k} \frac{X_i}{n} \frac{\binom{n-X_i}{k}}{\binom{n-1}{k}} + \frac{f_1}{n} (1-A)^{-n+1} \left\{ -\log(A) - \sum_{r=1}^{n-1} \frac{1}{r} (1-A)^r \right\}, \tag{10b} $$

where $A = 2f_2 / [(n-1)f_1 + 2f_2]$. As a result, the asymptotic estimator for Shannon diversity is ${}^1\hat{D}(\infty) = \exp[\hat{H}(\infty)]$. The extrapolated estimator for Shannon diversity of a sample of size $n+m^*$ is:

$$ {}^{1}\hat{D}(n+m^{*}) = \exp\left(\frac{n}{n+m^{*}} \sum_{i=1}^{S} (-\frac{X_i}{n} \log \frac{X_i}{n}) + \frac{m^{*}}{n+m^{*}} \hat{H}(\infty) \right). \qquad (10c) $$

Details of the derivation are provided in Appendix E. Extensive simulations (Chao et al. 2013) suggest that the asymptotic Shannon estimator in Eq. 10b is nearly unbiased, implying the extrapolation provided by Eq. 10c is valid for a wide prediction range. This extrapolation can be safely extended to the asymptote.

(3) *Simpson diversity (q = 2)*

The general formula for the expected Simpson diversity for any sample size m (for both $m < n$ and $m > n$):

$$ {}^{2}D(m) = \frac{1}{\sum_{k=1}^{m} \left(\frac{k}{m} \right)^2 \times E[f_k(m)]} = \frac{1}{\frac{1}{m} + \frac{m-1}{m} \sum_{i=1}^{S} p_i^2}, \; m \geq 1. \qquad (11a) $$

(See Appendix D for proofs). A minimum variance unbiased estimator of $\sum_{i=1}^{S} p_i^2$ is $\sum_{i=1}^{S} X_i(X_i-1)/[n(n-1)]$ (Good, 1953), implying that an estimator for the asymptotic Simpson diversity is ${}^{2}\hat{D} = {}^{2}\hat{D}(\infty) = n(n-1)/\sum_{X_i \geq 2} X_i\,(X_i-1)$. An interpolated estimator ($m < n$) from Eq. 11a can be expressed in two equivalent forms:

$$ {}^{2}\hat{D}(m) = \frac{1}{\sum_{k=1}^{m} \left(\frac{k}{m} \right)^2 \hat{f}_k(m)} = \frac{1}{\frac{1}{m} + \frac{m-1}{m} \sum_{i=1}^{S} \frac{X_i(X_i-1)}{n(n-1)}}. \qquad (11b) $$

We can apply Eq. 11a to an augmented size of $n + m^{*}$ and obtain the following extrapolated estimator:

$$ {}^{2}\hat{D}(n+m^{*}) = \frac{1}{\frac{1}{n+m^{*}} + \frac{n+m^{*}-1}{n+m^{*}} \sum_{i=1}^{S} \frac{X_i(X_i-1)}{n(n-1)}}. \qquad (11c) $$

The rarefaction, extrapolation and asymptotic estimators are all nearly unbiased. This means that for $q = 2$, the extrapolation can be safely extended to the asymptote.

(4) *Diversity of integer order q > 2*

For Hill numbers of integer order $q > 2$, a nearly unbiased interpolation estimator is given in Eq. 9a. A general extrapolation estimator ${}^{q}\hat{D}(n+m^{*})$ is quite complicated and is shown in the last column in Table 1 (see Appendix E for derivation details). A nearly unbiased estimator of the true asymptotic value ${}^{q}D = {}^{q}D(\infty)$ for any integer $q > 2$ is ${}^{q}\hat{D}(\infty) = [\sum_{X_i \geq q} X_i^{(q)} / n^{(q)}]^{1/(1-q)}$ (Gotelli and Chao, 2013), where $x^{(j)}$ denotes the falling factorial $x\ (x–1)\ ...\ (x–j+1)$.

Table 1 summarizes, for abundance data, all theoretical formulas and analytic estimators for rarefaction and extrapolation of Hill numbers of order $q = 0, 1, 2$ and any integer order $q > 2$. (The last row of Table 1 also gives the formulas for sample-completeness as a function of sample size; see below). We tested our estimators on simulated data generated from several species abundance models and on data from large empirical datasets (Appendix F). The results show that the proposed analytic rarefaction and extrapolation estimators match perfectly with the corresponding theoretical values for rarefied and extrapolated samples up to double the reference sample size. However, when the extrapolated sample size is more than double the reference sample size, the performance of our predictors depends on extrapolated range and the order q (see *Discussion*).

There are two kinds of variance associated with an interpolated or extrapolated estimator. A variance that is *conditional* on the reference sample measures only the variation in diversity that would arise from repeatedly resampling (without replacement) the given reference sample. This conditional variance approaches zero as m approaches n because the diversity of sample size of n is fixed (i.e., there is only one combination of all individuals or all sampling units). An

unconditional variance measures the variation in diversity that would arise if another *new* sample of size *m* were taken from the entire assemblage (rather than from the original reference sample). Therefore, the unconditional variance does not approach 0 when sample size tends to *n*, and all associated confidence intervals are symmetric, which reflects the uncertainty of the *new* sample. In deriving an "unconditional" variance, the number of undetected species must be estimated because those undetected species also affect the variation of a new sample. In most applications, unconditional variance is more useful because inferences are not restricted to the reference sample.

Colwell et al. (2012) obtained an unconditional analytic variance estimator for rarefied and extrapolated species richness estimators. However, extending this analytic approach for variance estimators to a general order of *q* becomes mathematically intractable. Therefore, we suggest a simpler, bootstrap method (Appendix G), to obtain unconditional variances and confidence intervals for all rarefied and extrapolated estimators. In the proposed procedure, we follow Colwell et al. (2012) and use the Chao1 (for abundance data) or Chao2 (for incidence data) to estimate the number of undetected species in the reference sample (Chao 1984, 1987), although any other proper estimators can also be used. The examples below illustrate our proposed sampling curves and the associated confidence intervals based on the unconditional variance from our proposed bootstrap method.

Sample-size- and coverage-based rarefaction and extrapolation

In comparing diversities among multiple assemblages, samples can be standardized by sample size or by sample completeness. Our proposed sample-size-based sampling curve for Hill numbers of each specific order *q* includes the rarefaction part (which plots ${}^{q}\hat{D}(m)$ as a function of *m*, where $m < n$, see Table 1) and the extrapolation part (which plots ${}^{q}\hat{D}(n+m^{*})$ as a function

of $n+ m^*$ for $m^* > 0$; see Table 1) and yields a smooth sampling curve, the two parts of which join smoothly at the point of the reference sample (n, ${}^qD_{obs}$). To fully incorporate the effect of relative abundance on diversity estimation, we suggest plotting curves for at least the first three Hill numbers (q = 0, 1, 2).

When there are many "invisible" species—species with extremely small relative abundance that are almost undetectable in normal sampling schemes—our intuition is that the number of undetected species in samples (or equivalently, species richness in the entire assemblage) is very hard to estimate; see Colwell et al. (2012) and Gotelli and Chao (2013) for a review. On the other hand, and contrary to intuition, one notion of sample completeness can be accurately and efficiently estimated using only information contained in the reference sample itself. Sample completeness can be measured by *sample coverage* (or simply *coverage*), a concept originally developed by the founder of modern computer science, Alan Turing, and I. J. Good (Good 1953, 2000, Good and Toulmin 1956; according to Good [2000], Turing never published this work, but gave permission to Good to publish it). Coverage is defined as the total relative abundances of the observed species, or equivalently, the proportion of the total number of individuals in an assemblage that belong to species represented in the sample. Turing and Good (Good 1953, 2000, Robbins 1968, Esty 1983, 1986) derived a simple coverage estimator (of the reference sample of size n) as one minus the proportion of singletons. Robbins (1968) showed that the average squared error of this estimator $\approx 1/n$. A tiny percentage of coverage can contain an infinite number of rare species. The estimated complement of coverage is not an estimate of the number of unseen species, but rather it estimates the proportion of the total individuals in the assemblage that belong to undetected species. For this reason, extremely rare, undetected species do not make a significant contribution to that proportion, even if there are many such species. This

intuitively explains why the estimation of species richness in highly diverse assemblages is a statistically difficult issue, whereas sample coverage can be accurately estimated.

Alroy (2010) and Jost (2010) independently proposed that samples be standardized to a common level of sample completeness (as measured by sample coverage), and developed algorithmic approaches for comparing rarefied samples. For species richness, Chao and Jost (2012) were the first to derive an analytic method for seamless coverage-based rarefaction and extrapolation. Chao and Jost (2012) suggested plotting rarefaction and extrapolation curves with respect to sample coverage rather than with respect to sample size because the expected species richness for equal sample coverage satisfies a replication principle or doubling property, which the expected species richness for equal sample size does not obey. Similar conclusions are valid for the expected diversity of any order q; see Appendix D for details. This property makes it possible to quantify ratio comparisons or any other comparisons between the magnitudes of the diversities of the assemblages.

For individual-based abundance data, Chao and Jost (2012) used a more accurate sample coverage estimate for the reference sample:

$$\hat{C}_{ind}(n) = 1 - \frac{f_1}{n}\left[\frac{(n-1)f_1}{(n-1)f_1 + 2f_2}\right]. \qquad (12)$$

They also derived an interpolated coverage estimator $\hat{C}_{ind}(m)$ for any rarefied sample of size $m < n$ and extrapolated coverage estimator $\hat{C}_{ind}(n+m^*)$ for any augmented sample of size $n+m^*$; see Table 1 (last row) for their formulas. The extrapolated coverage estimator is reliable if $m^* < n$.

As with sample-size-based curves, for any specific order q, the coverage-based interpolation (which plots ${}^q\hat{D}(m)$ with respect to $\hat{C}_{ind}(m)$) and extrapolation (which plots ${}^q\hat{D}(n+m^*)$ with

respect to $\hat{C}_{ind}(n+m^*)$) join smoothly at the reference point ($\hat{C}_{ind}(n)$, ${}^qD_{obs}$). The confidence intervals of expected diversity based on the bootstrap method also join smoothly.

Bridging sample-size- and coverage-based approaches

The sample-size-based approach plots the estimated diversity as a function of sample size, whereas the corresponding coverage-based approach plots the same diversity with respect to sample coverage. Therefore, these two approaches can be bridged by the relationship between coverage and sample size. Using the coverage estimators in Table 1 (the last row in each table), we can construct a *sample completeness curve*, which reveals sample completeness for a given sample size. From the original reference sample, this curve estimates sample completeness for smaller rarified samples, as well as for larger extrapolated samples. This curve also provides an estimate of the sample size needed to achieve a fixed degree of completeness.

An optimal stopping theory derived by Rasmussen and Starr (1979) specifies that sampling stops when sample coverage reaches a predetermined value. The sample completeness curve thus provides information about whether we should continue or stop sampling. If multiple assemblages are to be sampled and compared, Chao and Jost (2012) suggested that ecologists should sample each assemblage to the same degree of completeness. Such equally-complete samples from different assemblages can be compared directly, without any need for rarefaction or extrapolation. See *Worked Examples* for illustration.

RAREFACTION AND EXTRAPOLATION OF INCIDENCE DATA USING HILL NUMBERS

For incidence data, parallel derivations to those for abundance data yield equations for the theoretical expected diversities for any sample size *t* (the first column in Table 2). Here, "sample

size" for incidence data means "number of sampling units." The analytic estimators for rarefied samples and analytic estimators for extrapolated samples are also given in Table 2. The asymptotic estimator for each order q of Hill numbers (q = 0, 1, 2) is provided in the footnotes of Table 2. Full derivation details along with a replication principle appear in Appendix H; here we highlight the following differences from the models and estimators for abundance data.

First, for abundance data, our derivation was based on a model in which the species frequency X_i follows a binomial distribution characterized by n and the true relative abundance p_i. In contrast, for incidence data, we assume the species incidence-based frequency Y_i follows a binomial distribution characterized by T (the total number of sampling units) and incidence probability π_i. With abundance data, $\sum_{i=1}^{S} p_i = 1$, but with incidence data, $\sum_{i=1}^{S} \pi_i$ can exceed 1.

Second, the total number of individuals n in a reference sample of abundance data is fixed by design. In contrast, total number of incidences in a reference sample of T sampling units is a random variable U, with expectation $E(U) = T\sum_{i=1}^{S} \pi_i$. Therefore, $\sum_{i=1}^{S} \pi_i$ can be accurately estimated by U/T. For abundance data, the number of individuals in any rarefied or extrapolated sample size is fixed, whereas for incidence data, the number of incidences in any t sampling units is random with expectation $E(U_t) = t\sum_{i=1}^{S} \pi_i$, which is estimated by $\hat{U}_t = tU/T$.

Third, for abundance data, the primary derivations of our estimators are based on frequency counts $f_k(m)$, the number of species represented by exactly k individuals (or observed k times) in a sample of size m. The corresponding incidence frequency count is $Q_k(t)$, the number of species recorded in exactly k sampling units in a sample of t sampling units.

The sample completeness curve as a function of abundance, developed by Chao and Jost (2012), is reviewed in Appendix B, and the formulas appear in the last row of Table 1. In Appendix C we derive, for the first time, the corresponding sample completeness curve as a

function of sampling units for incidence data. With such a curve, ecologists can objectively quantify the sample completeness for any incomplete abundance or incidence datasets. These curves help determine a sample size needed in a designing a survey. All formulas are summarized in the last row of Table 2.

Based on the formulas in Table 2, for each order q of the Hill numbers ${}^{q}\Delta$ for incidence data, we can obtain an integrated rarefaction/extrapolation sampling curve with confidence intervals. Statistical inference theory implies that the proposed interpolated estimator for diversity is unbiased for $q = 0$ and nearly unbiased for $q = 1$ and 2. We support these claims with simulation tests (Appendix F). As with the abundance data, the performance of our extrapolated estimators depends on the order of Hill numbers and the prediction range of extrapolation. Simulation tests provide some general usage guidelines (see *Discussion*). In Tables 1 and 2, we summarize the properties and performance of each index based on theory and analyses of empirical and simulated data sets.

WORKED EXAMPLES: COMPARISON OF ASSEMBLAGES

Example 1: Abundance data – comparing spider species diversity in two treatments

Sackett et al. (2011) provided species abundance data for samples of spiders from four experimental forest canopy-manipulation treatments at the Harvard Forest. The treatments were established to study the long-term consequences of loss of the dominant forest tree, eastern hemlock (*Tsuga canadensis*), caused by a non-native insect, the hemlock woolly adelgid (*Adelges tsugae*) (Ellison et al. 2010). Data from two treatments are used here for illustration: (1) the Hemlock Girdled treatment, in which bark and cambium of hemlock trees were cut and the trees left in place to die to mimic tree mortality by adelgid infestation; and (2) the Hemlock Logged

treatment, in which hemlock trees were cut and removed from the plots (Ellison et al. 2010). The abundance frequency data for the two treatments (summed over two plots per treatment) are tabulated in Table 3, and the rank-abundance distributions are shown in Fig. 2. We used the data from these two treatments to illustrate the construction of two types (sample-size- and coverage-based) of rarefaction and extrapolation curves of Hill numbers. The constructed sampling curves were then used to compare spider species diversities between the two treatments.

The reference sample size (number of individual spiders) for the Girdled treatment is 168, and the observed species richness, Shannon diversity and Simpson diversity (i.e., Hill numbers for $q = 0, 1, 2$) for this reference sample size are respectively 26, 12.06, and 7.84 (solid points in Figs. 3a and 3b). The sample size for the Logged treatment is 252, and the corresponding observed Hill numbers for $q = 0, 1, 2$ are 37, 14.42 and 6.76. Thus, judging from the unstandardized raw data (the reference samples), the Logged treatment appears to have higher observed species richness and Shannon diversity, but lower Simpson diversity than the Girdled treatment.

Step (1): Compare sample-size-based sampling curves up to a base sample size (Fig. 3).

We first construct, for each of the two treatments, the integrated sample-size-based rarefaction and extrapolation curves for Hill numbers of $q = 0, 1, 2$. In Fig. 3a, we show these sample-size-based curves with 95% confidence intervals based on a bootstrap method. We extrapolated up to double the reference sample size (i.e., up to size 336 for the Girdled treatment and size 504 for the Logged treatment). In each plot, except for initial, small sample sizes, none of the confidence intervals for the three curves intersect, and the rank order of diversity is species richness > Shannon diversity > Simpson diversity. For any fixed sample size or completeness in the comparison range, if the 95% confidence intervals do not overlap, then significant differences at a level of 5% among the

expected diversities (whether interpolated or extrapolated) are guaranteed. However, overlapped intervals do not guarantee non-significance (Schenker and Gentleman 2001). The curve for species richness ($q = 0$) increases steeply with sample size in both treatments, but the curves for Shannon and Simpson diversity ($q = 1$ and $q = 2$) level off beyond the reference sample, illustrating that higher-order Hill numbers are increasingly dominated by the frequencies of the more common species and are therefore less sensitive to sampling effects.

To compare diversities between the Girdled and Logged treatments, we show in Fig. 3b, for each fixed value of q ($q = 0$, 1, and 2), the sample-size-based rarefaction and extrapolation of these two plots with 95% confidence intervals up to a *base sample size*. We suggest the base sample size to be double the smallest reference sample size or the maximum reference sample size, whichever is larger (the reason for our suggestion will become clearer in the second example, below). See Table 4 for systematic steps to determine a base sample size. In this example, the base sample size is 336 (double the smaller reference sample size). The estimated Hill numbers can then be compared across assemblages for any sample size less than the base size. In a traditional rarefaction, the data from the Logged treatment would be rarefied to a sample size of 168 individuals to match the abundance in the Girdled treatment. For this rarefied sample, the Hill numbers of $q = 0$, 1, 2 are estimated to be 31.71, 13.83 and 6.68. The proposed integrated sampling curve allows reliable comparisons for any sample size up to an abundance of 336. Across this range of abundance, Fig. 3b reveals that the Logged treatment is more diverse for all but the smallest sample sizes for species richness ($q = 0$) and Shannon diversity ($q = 1$), although the confidence intervals overlap. In contrast, for Simpson diversity ($q = 2$), the Girdled treatment is more diverse, although again the two confidence intervals overlap.

Step (2): Construct a sample completeness curve to link sample-size- and coverage-based sampling curves (Fig. 4)

Based on Eq. 12, the coverage for the Girdled treatment is estimated as 93% for the reference sample of size 168 individuals, and the coverage for the Logged treatment is 94% for the reference sample of 252 individuals. It is informative to examine how the sample completeness varies with sample size (see the formulas in the last row in Table 1). In Fig. 4, we plot the sample completeness curve as a function of sample size for each of the two treatments, up to double the reference sample size. For any sample size less than 168, the curve shows that the sample completeness for the Girdled treatment is estimated to be higher than that in Logged treatment, although the confidence intervals overlap. When sample size is larger than 168, the estimates of sample coverages for the two treatments differ little. If we apply a traditional rarefaction approach to standardize sample coverage, a sample size of approximately 168 individuals in the Logged treatment gives a sample coverage of 93%. Thus the diversity ordering of the two treatments for 93% of the assemblage individuals is the same as that for a standardized sample of 168 individuals as discussed above. The sample completeness curve figure provides a bridge between sample-size- and coverage-based sampling curves, as will be explained in the next step.

Step (3): Compare coverage-based sampling curves up to a "base coverage" (Fig. 5)

From the sample completeness curve (Fig. 4), when sample size in the Girdled treatment is doubled from 168 to 336 individuals, the sample coverage is increased from 93% to 96%. In the Logged treatment, when sample size is doubled from 252 to 504 individuals, the coverage is increased from 94% to 97%. In Fig. 5a, we present, for each treatment, the corresponding coverage-based rarefaction and extrapolation curves with 95% confidence intervals for diversity

of $q = 0, 1, 2$ when the coverage is extrapolated to the value for a doubling of each reference sample size.

In Fig. 5b, we compare the coverage-based diversities of the two treatments for $q = 0$ (left panel), $q = 1$ (middle panel) and $q = 2$ (right panel) up to the coverage of 96%. This is our "*base coverage*" (the lowest coverage for doubled reference sample sizes or the maximum coverage for reference samples, whichever is larger). See Table 4 for suggestions on the choice of base coverage. Because the increase in coverage for the extrapolation is small, and the estimated diversity for $q = 1$ and 2 hardly change beyond the reference samples, the extrapolation parts in Fig. 5b are nearly invisible for these two orders of q. Since the two confidence bands do not intersect for species richness ($q = 0$) if coverage exceeds 50% (Fig. 5b, left panel), species richness in the Logged treatment is significantly higher than in the Girdled treatment for any standardized sample coverage between 50% and 96%. For Shannon diversity ($q = 1$), the Logged treatment is more diverse but the confidence bands overlap. For Simpson diversity ($q = 2$), when coverage is less than 70%, both treatments have almost the same diversity, but when coverage is greater than 70%, the Simpson diversity for the Girdled treatment is slightly higher.

Comparing Figs. 3b and 5b, we see that the sample-size- and coverage-based curves for $q = 0$ and $q = 1$ exhibit consistent diversity orderings between the two treatments. However, for $q = 2$, the sample-size-based curves do not intersect (Fig. 3b), but the coverage-based curves have two crossing points (Fig. 5b). See *Discussion* for more comparisons of the two types of curves.

Example 2: Incidence data – comparing species diversity of tropical ants among five sites

We use the tropical ant species data collected by Longino and Colwell (2011) from five elevations on the Barva Transect, a 30 km continuous gradient of wet forest on Costa Rica's Atlantic slope. The five sites are, respectively, at elevations 50m, 500m, 1070m, 1500m, and

2000m. Species presence or absence was recorded in each sampling unit, which consisted of all worker ants extracted from a $1m^2$ forest floor plot. See Longino and Colwell (2011) for sampling and data details. A sample-by-species incidence matrix was produced for each of the 5 sites. The incidence frequency counts are given in Colwell et al. (2012, their Table 6). The plots for rank-frequency distributions of the five sites are shown by Longino and Colwell (2011, their Fig. 3). An integrated rarefaction and extrapolation curve for species richness was presented by Colwell et al. (2012, their Fig. 4b). They concluded that species richness among the five sites was significantly different (none of the confidence intervals intersect, except for very small sizes), and that richness has the ordering: 500m > 50m > 1070m > 1500m > 2000m.

Step (1): Compare sample-size-based sampling curves up to a base sample size (Fig. 6)

For each of the five sites, Fig. 6a shows the observed Hill numbers and sample-size-based rarefaction and extrapolation plots with 95% confidence intervals for three sampling curves (Hill numbers of $q = 0, 1, 2$) up to double the reference sample size. To compare diversity among the five elevations, we first determine the base sample size. The reference sample sizes T (number of sampling units) for each elevation (50m, 500m, 1070m, 1500m, and 2000m) are, respectively, 599, 230, 150, 200, and 200. The base sample size would be 599 (which is larger than $2 \times 150 = 300$, double the smallest reference sample); see Table 4 for the choice of this base sample size. An advantage of this choice of base sample size is that no data are excluded from our analysis. However, a drawback is that the extrapolation range for some samples could exceed their doubled reference sample sizes. For Shannon and Simpson diversities, the prediction biases are minimal beyond the double reference sample sizes, but for species richness in such cases we should be cautious about the prediction bias. See *Discussion* for suggestions on extrapolation range.

Next for each specific order of q, we plot the sample-size-based interpolation and extrapolation curves with 95% confidence bands for these five elevations together in the same figure, as illustrated in Fig. 6b. Extrapolations are extended to the base sample size of 599 for all sites. Our plot of $q = 0$ corresponds to Figure 4b in Colwell et al. (2012). We here extend their approach to include curves for $q = 1$ and $q = 2$, and also include coverage-based plots. For the three orders of q, diversity of the sites is consistently ordered as 500m > 50m > 1070m > 1500m > 2000m (Fig. 6b). All confidence intervals are non-overlapping (except for very small sizes), implying the diversity of any order $q = 0, 1, 2$ is significantly different among the five elevations for any fixed sample size up to 599 sampling units.

Step (2): Construct a sample completeness curve to link sample-size- and coverage-based sampling curves (Fig. 7)

The sample coverages for the five sites (50m, 500m, 1070m, 1500m, and 2000m) are estimated as 99.18%, 97.60%, 98.39%, 98.89% and 99.64% respectively, indicating that sampling is nearly complete for all sites. A summary of coverage estimators for incidence data appear in Table 2. See Appendix C for estimation details. For any fixed sample size less than 300 sampling units, the sample coverage for the two lower elevations (50m and 500m) is significantly lower than coverage at higher elevations (1070m, 1500m, 2000m). When sample size is greater than 300, the pattern persists, but the 95% confidence bands begin to overlap.

Step (3): Compare coverage-based sampling curves up to a base coverage (Fig. 8)

Figure 8a shows, for each plot, the corresponding coverage-based rarefaction and extrapolation curves for Hill numbers of $q = 0, 1, 2$ when the coverage is extrapolated to the value for a doubling of each reference sample size. From the sample completeness curve (Fig. 7), when the sample size in each site is doubled, the sample coverage increases very slightly for all

sites. There is little change in ant diversity for $q = 1$ and 2. Thus, the extrapolated portions of the curves in Fig. 8a are nearly invisible, as we also noted in Fig. 5.

When all sample sizes are doubled, the minimum value of the coverage values of these doubled sample sizes among the five sites is 99.08% (for 500m elevation). However, it is less than the coverage 99.64% of the reference sample for 2000m elevation. In order to use all data, we select our base coverage to be 99.64% (Table 4). Figure 8b compares coverage-based rarefaction and extrapolation curves up to the base coverage of 99.64%. All three coverage-based diversities show the same ordering by elevation as in the sample-size-based comparison. None of the confidence intervals overlap except at very small coverage values, implying significant differences in ant diversity among the five elevational transects at comparable coverage.

DISCUSSION

We have developed a new, comprehensive statistical framework for the analysis of biodiversity data based on Hill numbers. We also advocate the use of sample coverage (or simply *coverage*), developed by Turing and Good (Good 1953) to quantify sample completeness. To characterize the species diversity of an assemblage, we propose constructing two types of integrated rarefaction and extrapolation curves (sample-size- and coverage-based) as illustrated in Figs. 3a and 5a for Example 1, and Figs. 6a and 8a for Example 2. For each type of curve, we suggest plotting three rarefaction/extrapolation curves (with confidence intervals) corresponding to three orders ($q = 0, 1, 2$) of Hill numbers. These curves are then used to compare multiple assemblages, as illustrated in Figs. 3b and 5b of Example 1 and 6b and 8b of Example 2. The sample-size- and coverage-based curves are linked by a sample completeness curve (Figs. 4 and 7), which reveals the relationship between sample size (number of individuals or number of

sampling units) and sample completeness. This curve illustrates how much sampling effort is needed to achieve a pre-determined level of sample completeness.

The proposed estimators work well for rarefaction and short-range extrapolation in which the extrapolated sample size is up to twice the reference sample size. For rarefaction, our proposed estimator is unbiased for $q = 0$ and nearly unbiased for $q = 1$ and 2. For short-range extrapolation, the prediction bias with respect to the expected diversity is often limited. When the extrapolated sample size is more than double the reference sample size, the prediction bias depends on the extrapolated range and the order q. The magnitude of the prediction bias generally increases with the prediction range. For $q \geq 1$, the extrapolated estimator is nearly unbiased for all extrapolated sample sizes, so the extrapolation can be safely extended to the asymptote. However, for $q = 0$, extrapolation is reliable up to no more than double the reference sample size. Beyond that, the predictor for $q = 0$ may be subject to some bias because our asymptotic estimator for species richness (Chao1 for abundance data and Chao2 for incidence data) is a lower bound only (Chao 1984, 1987).

To compare the diversities of multiple assemblages, Table 4 gives guidelines to choose a base sample size and base coverage for comparing sample-size- and coverage-based curves. With the suggested base sample size and base coverage, all data are used for comparisons. Based on the integrated sample-size- and coverage-based rarefaction and extrapolation curves, ecologists can efficiently use all available data to make more robust and detailed inferences about the sampled assemblages for any standardized samples with sample size less than the base sample size, and for any equally complete samples with coverage less than the base coverage. However, Example 2 provides an example in which we extrapolate a sample beyond a doubling of its reference sample size, based on the suggested base sample size. For those samples, we

should be cautious in estimating quantitative differences in species richness ($q = 0$) among assemblages, although inferences about diversities of $q \geq 1$ are reliable.

In our formulation of a diversity accumulation curve, we define the expected diversity of a finite sample of size m as the Hill numbers based on the expected abundance frequency counts $\{E[f_k(m)]; k = 1, \ldots, m\}$. Our proposed theoretical formula is given in Eq. 6. An alternative definition would be the average Hill numbers over many samples of size m taken from the entire assemblage. Although the two approaches generally yield very close numerical values, our approach has two main advantages. We have shown (see summaries in Tables 1 and 2) that accurate estimators via estimation of frequency counts can be obtained for our approach. However, it is difficult to accurately estimate the alternative formula of the expected Hill numbers; usually algorithmic methods are needed. Another advantage is that all transformations between diversity measures are valid for any size m under our formulation. For example, Hill number of order 2 for any sample size m is exactly the inverse of the Simpson concentrations for the same size when all are based on the same expected frequencies. This is important because all diversity measures give consistent comparisons. If we use the alternative approach, then such transformations will not be exactly valid, and different measures may produce different comparative results. As proved in Propositions D1 in Appendix D, the two approaches are identical for species richness, and the same conclusion is valid for expected sample coverage.

Rarefaction and extrapolation aim to make fair comparisons among incomplete samples. Sample-size-based rarefaction and extrapolation, in which the samples are all standardized to an equal size, provide useful sampling information for a range of sizes. Coverage-based rarefaction and extrapolation, in which all samples are standardized to an equal coverage, ensure that we are comparing samples of equal completeness over a range of coverages. Taken together, these two

types of curves allow us to make more robust and detailed inferences about the sampled assemblages. Our approach provides a unifying sampling framework for species diversity studies and allows for objective comparisons of multiple assemblages.

For species richness ($q = 0$), if the expected sample-size-based species accumulation curves of two assemblages do not cross for any finite sample size > 1, then the expected coverage-based species accumulation curves for these two assemblages also do not cross at any finite coverage < 1 beyond the base point (Chao and Jost 2012). The reverse is also true. Thus, the two types of curves for species richness always give the same qualitative ordering of species richness. If crossing occurs, then the sample-size- and coverage-based curves have exactly the same number of crossing points. However, for species richness, the coverage-based method is always more efficient (requiring smaller sample sizes in each assemblage) than the traditional method for detecting any specific crossing point (Chao and Jost 2012). The two types of curves can exhibit different patterns and yield different diversity ordering for $q = 1$ (Shannon diversity) and $q = 2$ (Simpson diversity). An example of the case of $q = 2$ is illustrated in Figs. 3b and 5b. The sample-size-based curves for $q = 2$ in Fig. 3b do not intersect, but the coverage-based curves for $q = 2$ in Fig. 5b cross twice. Appendix J gives an example for the case of $q = 1$. There is another difference between the sample-size- and coverage-based standardization methods. As proved in Appendix D, the expected diversity of any order obeys a replication principle only when coverage is standardized.

In biodiversity studies, ecologists are interested in measuring not only diversity but also evenness and inequality (Ricotta 2003). Jost (2010) used partitioning theory to derive Hill's (1973) useful class of evenness measures, the ratios of Hill numbers ${}^{q}D$ and species richness, ${}^{q}D / S$ for $q > 0$, and he showed that the ratio of the logarithms of Hill numbers and logarithm of

richness, $\log({}^{q}D)/\log(S)$, including Pielou's (1975) $J' = \log({}^{1}D)/\log(S)$, express the corresponding relative evenness. These two classes of measures have been difficult to accurately estimate statistically from samples due to their strong dependence on species richness, and thus on sample size. Jost (2010) suggested estimating both S and Hill numbers at fixed coverage to obtain meaningful estimates of evenness and inequality indices. Based on the theory developed in this paper, we are now able to analytically estimate evenness and inequality indices at fixed sample size or sample coverage. This will be an important application of our proposed theory; see Tables 1 and 2 for a summary of our analytic formulas.

In addition to Hill numbers, there are two other widely used classes of measures, Renyi and Tsallis generalized entropies (Patil and Taillie 1979, 1982). These measures are simple transformations of Hill numbers; see Jost (2007). Hurlbert (1971) suggested another unified class of species indices, defined as the expected number of species in a sample of m individuals selected at random from an assemblage. The relationship between Hill numbers and Hurlbert's indices has not been clear to ecologists (Dauby and Hardy 2011). In Appendix I, we show that these two classes of infinity orders are mathematically equivalent, in the sense that they contain the same information about biodiversity. Moreover, given a reference sample, sample-size-based rarefaction and extrapolation formulas (Colwell et al. 2012) for species richness provide estimates of Hurlbert's indices. Thus, our proposed sample-size- and coverage-based rarefaction/extrapolation sampling framework for Hill numbers includes the information and estimators of all Hurlbert's indices and provides a unified approach to quantifying species diversity.

The slope of a sample-size-based expected species accumulation curve or a rarefaction/extrapolation curve also provides important information. The slope at the base point

in the species accumulation curve or rarefaction curve is closely related to the Simpson diversity and to Hurlbert's (1971) Probability of an Interspecific Encounter (PIE) measure (Olszewski 2004). The slope at any other point is closely related to the complement of coverage (Chao and Jost 2012). For coverage-based curves, see Appendix I for similar findings. In Appendix K, we consider different sampling schemes and discuss the relationship between the expected species accumulation curve, Simpson diversity, and PIE.

For Hill numbers, only species *relative* abundances are involved. Species *absolute* abundances play no role in traditional diversities. From the perspective of measuring ecosystem function, Ricotta (2003) argued that if two assemblages have the same relative abundances, the one with larger absolute abundances should be considered more diverse. We are currently working on extending Hill numbers to include absolute abundances of species. The associated rarefaction and extrapolation functions for absolute-abundance Hill numbers also merit further research. Finally, this paper has focused on traditional Hill numbers, which do not take species evolutionary history into account. Chao et al. (2010) generalized Hill numbers to a class of measures that incorporate phylogenetic distances between species. It is worthwhile to extend this work to rarefaction and extrapolation of phylogenetic and functional diversity measures (Walker et al. 2008, Ricotta et al. 2012).

All the rarefaction and extrapolation estimators proposed in this paper are featured in the online freeware application iNEXT (iNterpolation/EXTrapolation) http://chao.stat.nthu.edu.tw/inext/. The R scripts for iNEXT have been posted in *Ecological Archives*. Sample-size-based rarefaction and extrapolation estimators for richness ($q = 0$, row 1 of Tables 1 and 2) are computed by EstimateS Version 9 (Colwell 2013) http://purl.oclc.org/estimates.

ACKNOWLEDGEMENTS

The authors thank the Subject Matter Editor (Helmut Hillebrand), Andres Baselga, and one anonymous reviewer for very thoughtful and helpful comments and suggestions. A. Chao, T. C. Hsieh and K. H. Ma were supported by Taiwan National Science Council under Contract 100-2118-M007-006-MY3. N. J. Gotelli was supported by US NSF awards DEB-026575 and DEB-027478, and the US Department of Energy award DE-FG02-08ER64510. R. K. Colwell was supported by US NSF award DBI-0851245. A. M. Ellison's work on this project was supported by US NSF awards DEB 04-52254 and DEB-0620443, and Department of Energy award DE-FG02-08ER64510. This is a contribution from the Harvard Forest Long Term Ecological Research site.

LITERATURE CITED

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

Brook, B. W., N. S. Sodhi, and P. K. L. Ng. 2003. Catastrophic extinctions follow deforestation in Singapore. Nature 424:420-426.

CBD. 2012. Convention on Biological Diversity. http://www.cbd.int/

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

Chao, A. 1987. Estimating the population size for capture-recapture data with unequal catchability. Biometrics 43:783-791.

Chao, A., and C.-H. Chiu. 2013. Estimation of species richness and shared species richness. To appear as an article in Handbook of Methods and Applications of Statistics in the Atmospheric and Earth Sciences. Wiley, New York, New York, USA.

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

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

Chao, A. and S.-M. Lee. 1992. Estimating the number of classes via sample coverage. Journal of the American Statistical Association 87:210-217.

Chao, A., Y. T. Wang, and L. Jost. 2013. Entropy and the species accumulation curve: a nearly unbiased estimator of entropy via discovery rates of new species. Under revision, Methods in Ecology and Evolution.

Chiu, C.-H., L. Jost, and A. Chao. 2013. Phylognetic beta diversity, similarity, and differentiation measures based on Hill numbers. To appear in Ecological Monographs.

Coleman, B. D., M. A. Mares, M. R. Willig, and Y. H. Hsieh. 1982. Randomness, area, and species richness. Ecology 63:1121-1133.

Colwell, R. 2013. EstimateS: Statistical estimation of species richness and shared species from samples. Version 9. User's Guide and application published at: http://purl.oclc.org/estimates.

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 of London. Series B: Biological Sciences 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.

Connell, J. H. 1978. Diversity in tropical rain forests and coral reefs. Science 199:1302-1310.

Connolly, S. R., and M. Dornelas. 2011. Fitting and empirical evaluation of models for species abundance distributions. Pages 123-140 in A. E. Magurran and B. J. McGill, editors. Biological Diversity: Frontiers in Measurement and Assessment. Oxford University Press, New York, New York, USA.

Dauby, G., and O. J. Hardy. 2011. Sampled-based estimation of diversity sensu stricto by transforming Hurlbert diversities into effective number of species. Ecography 35:661-672.

Ellison, A. M., A. A. Barker-Plotkin, D. R. Foster, and D. A. Orwig. 2010. Experimentally testing the role of foundation species in forests: the Harvard Forest Hemlock Removal Experiment. Methods in Ecology and Evolution 1:168-179.

Esty, W. W. 1983. A normal limit law for a nonparametric estimator of the coverage of a random sample. The Annals of Statistics 11:905-912.

Esty, W. W. 1986. The efficiency of Good's nonparametric coverage estimator. The Annals of Statistics 14:1257-1260.

Fisher, R. A., A. S. Corbet, and C. B. Williams. 1943. The relation between the number of species and the number of individuals in a random sample of an animal population. The Journal of Animal Ecology 12:42-58.

Gaston, K. J., and R. A. Fuller. 2008. Commonness, population depletion and conservation biology. Trends in Ecology and Evolution 23:14-19.

Ghent, A. W. 1991. Insights into diversity and niche breadth analyses from exact small-sample tests of the equal abundance hypothesis. American Midland Naturalist 126: 213-255.

Good, I. J. 1953. The population frequencies of species and the estimation of population parameters. Biometrika 40:237-264.

Good, I. J. 2000. Turing's anticipation of empirical Bayes in connection with the cryptanalysis of the naval enigma. Journal of Statistical Computation and Simulation 66:101-111.

Good, I. J., and G. Toulmin. 1956. The number of new species, and the increase in population coverage, when a sample is increased. Biometrika 43:45-63.

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. The Encyclopedia of Biodiversity, 2nd Edition, Vol. 5. Academic Press, Waltham, MA, 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., and A. M. Ellison. 2012. A primer of ecological statistics. 2nd edition. Sinauer Associates, Inc., Sunderland, Massachusetts, USA.

Groom, M. J., G. K. Meffe, and C. R. Carroll. 2005. Principles of conservation biology. 3rd edition. Sinauer Associates, Inc., Sunderland, Massachusetts, USA.

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

Holsinger, K. E., and L. D. Gottlieb. 1991. Conservation of rare and endangered plants: principles and prospects. Pp. 195-208 in Falk, D. A. and Holsinger, K. E. (editors) Genetics and conservation of rare plants. Oxford University Press, New York, New York, USA.

Hubbell, S. P. 2001. A unified theory of biodiversity and biogeography. Princeton University Press, Princeton, New Jersey, USA.

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

Jost, L. 2006. Entropy and diversity. Oikos 113:363-375.

Jost, L. 2007. Partitioning diversity into independent alpha and beta components. Ecology 88:2427-2439.

Jost, L. 2010. The relation between evenness and diversity. Diversity 2:207-232.

Longino, J. T., and R. K. Colwell. 2011. Density compensation, species composition, and richness of ants on a neotropical elevational gradient. Ecosphere 2:art29.

MacArthur, R. H. 1965. Patterns of species diversity. Biological Reviews 40:510-533.

MacArthur, R. H., and E. O. Wilson. 1967. The theory of island biogeography. Princeton University Press, Princeton, New Jersey, USA

Magurran, A. E. 2004. Measuring biological diversity. Blackwell, Oxford, UK.

Magurran, A. E., and B. J. McGill, editors. 2011. Biological Diversity: Frontiers in Measurement and Assessment. Oxford University Press, Oxford, UK.

May, R. M. 1988. How many species are there on earth? Science 241:1441-1449.

Nielsen, R., D. Tarpy, and H. Reeve. 2003. Estimating effective paternity number in social insects and the effective number of alleles in a population. Molecular ecology 12:3157-3164.

Olszewski, T. D. 2004. A unified mathematical framework for the measurement of richness and evenness within and among multiple communities. Oikos 104:377-387.

Paninski, L. 2003. Estimation of entropy and mutual information. Neural Computation 15:1191-1253.

Patil, G. P., and C. Taillie. 1979. An overview of diversity. Pages 3-27 in J. F. Grassle, G. P. Patil, W. Smith, and C. Taillie, editors. Ecological diversity in theory and practice. International Cooperative Publishing House, Fairfield, Maryland, USA.

Patil, G. P., and C. Taillie. 1982. Diversity as a concept and its measurement. Journal of the American Statistical Association 77:548-561.

Pielou, E. C. 1975. Ecological Diversity. Wiley, New York, New York, USA.

Rasmussen, S. L., and N. Starr. 1979. Optimal and adaptive stopping in the search for new species. Journal of the American Statistical Association 74:661-667.

Ricotta, C. 2003. On parametric evenness measures. Journal of Theoretical Biology 222:189-197.

Ricotta, C., S. Pavoine, G. Bacaro, and A. T. R. Acosta. 2012. Functional rarefaction for species abundance data. Methods in Ecology and Evolution 3:519-525.

Robbins, H. E. 1968. Estimating the total probability of the unobserved outcomes of an experiment. The Annals of Mathematical Statistics 39:256-257.

Ross, S. M. 1995. Stochastic Processes. Wiley, New York, New York, USA.

Sanders, H. L. 1968. Marine benthic diversity: a comparative study. American Naturalist 102:243-282.

Sackett, T. E., S. Record, S. Bewick, B. Baiser, N. J. Sanders, and A. M. Ellison. 2011. Response of macroarthropod assemblages to the loss of hemlock (*Tsuga canadensis*), a foundation species. *Ecosphere* 2: art74.

Schenker, N., and J. F. Gentleman. 2001. On judging the significance of differences by examining the overlap between confidence intervals. The American Statistician 55:182-186.

Schwartz, M. W., C. A. Brigham, J. D. Hoeksema, K. G. Lyons, M. H. Mills, and P. J. Van Mantgem. 2000. Linking biodiversity to ecosystem function: implications for conservation ecology. Oecologia 122:297-305.

Shen, T. J., A. Chao, and C. F. Lin. 2003. Predicting the number of new species in further taxonomic sampling. Ecology 84:798-804.

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

Smith, W., and J. F. Grassle. 1977. Sampling properties of a family of diversity measures. Biometrics 33:283-292.

Soberón M., and J. B. Llorente. 1993. The use of species accumulation functions for the prediction of species richness. Conservation Biology 7:480-488.

Terborgh, J., L. Lopez, P. Nuñez, M. Rao, G. Shahabuddin, G. Orihuela, M. Riveros, R. Ascanio, G. H. Adler, and T. D. Lambert. 2001. Ecological meltdown in predator-free forest fragments. Science 294:1923-1926.

Tóthmérész, B. 1995. Comparison of different methods for diversity ordering. Journal of Vegetation Science 6:283-290.

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

Walker, S. C., M. S. Poos, and D. A. Jackson. 2008. Functional rarefaction: estimating functional diversity from field data. Oikos 117:286-296.

Washington, H. G. 1984. Diversity, biotic and similarity indices: a review with special relevance to aquatic ecosystems. Water Research 18: 653-694.

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

Appendix A

A binomial product model can incorporate spatial aggregation for quadrat sampling

Appendix B

Rarefaction and extrapolation for species richness (abundance data)

Appendix C

Rarefaction and extrapolation for species richness (incidence data)

Appendix D

Proof details for some formulas (Eqs. 5, 8, 9b, 11a and 11b of the main text) and a replication principle

Appendix E

Extrapolation formulas for Hill numbers of $q = 1$ and $q \geq 2$ based on abundance data

Appendix F

Using simulation to test the proposed analytic estimators

Appendix G

A bootstrap method to construct an unconditional variance estimator for any interpolated or extrapolated estimator

Appendix H

Rarefaction and extrapolation of Hill numbers for incidence data

Appendix I

Hill numbers and Hurlbert's indices

Appendix J

An example: sample-size- and coverage-based Shannon diversity curves may exhibit inconsistent patterns

Appendix K

PIE and rarefaction

Supplement

R code for the analysis of individual-based (abundance) and sample-based (incidence) species diversity data

Table 1. The theoretical formulas and analytic estimators for rarefaction and extrapolation of abundance-based Hill numbers of order q = 0 (first row of equations), q =1 (second row), q = 2 (third row) and any integer order q> 2 (fourth row), given a reference sample[1] with the observed Hill numbers ${}^{q}D_{obs}$ and estimated coverage $\hat{C}_{ind}(n)$. The last row gives equations for sample completeness as a function of sample size. It also gives the corresponding coverage estimators for rarefied samples and extrapolated samples for coverage-based rarefaction and extrapolation curves.

Theoretical Formula [2] (for all $m > 0$)	Interpolation Estimator [3] (for $m < n$)	Extrapolation Estimator [4, 5] (for a sample of size $n+ m^*$)
${}^{0}D(m) = S - E[f_0(m)] = \sum_{k=1}^{m} E[f_k(m)]$	${}^{0}\hat{D}(m) = \sum_{k=1}^{m} \hat{f}_k(m) = S_{obs} - \sum_{X_i \geq 1} \frac{\binom{n-X_i}{m}}{\binom{n}{m}}$ (minimum variance unbiased)	${}^{0}\hat{D}(n+m^*) = S_{obs} + \hat{f}_0 \left[1 - \left(1 - \frac{f_1}{n\hat{f}_0 + f_1} \right)^{m^*} \right]$ (reliable if $m^* < n$)
${}^{1}D(m) = \exp\left(\sum_{k=1}^{m} \left(-\frac{k}{m} \log \frac{k}{m} \right) \times E[f_k(m)] \right)$	${}^{1}\hat{D}(m) = \exp\left(\sum_{k=1}^{m} \left(-\frac{k}{m} \log \frac{k}{m} \right) \times \hat{f}_k(m) \right)$ (nearly unbiased)	${}^{1}\hat{D}(n+m^*) = \exp\left(\frac{n}{n+m^*} \sum_{i=1}^{S} \left(-\frac{X_i}{n} \log \frac{X_i}{n} \right) + \frac{m^*}{n+m^*} \hat{H}(\infty) \right)$ (nearly unbiased)
${}^{2}D(m) = \frac{1}{\sum_{k=1}^{m} \left(\frac{k}{m} \right)^2 \times E[f_k(m)]}$	${}^{2}\hat{D}(m) = \frac{1}{\sum_{k=1}^{m} \left(\frac{k}{m} \right)^2 \times \hat{f}_k(m)}$	${}^{2}\hat{D}(n+m^*) = \frac{1}{\frac{1}{n+m^*} + \frac{n+m^*-1}{n+m^*} \sum_{i=1}^{S} \frac{X_i(X_i-1)}{n(n-1)}}$

	(nearly unbiased)	(nearly unbiased)
$^{q}D(m)=\left(\sum_{k=1}^{m}\left(\frac{k}{m}\right)^{q}\times E[f_k(m)]\right)^{\frac{1}{1-q}}$	$^{q}\hat{D}(m)=\left(\sum_{k=1}^{m}\left(\frac{k}{m}\right)^{q}\times \hat{f}_k(m)\right)^{\frac{1}{1-q}}$ (nearly unbiased)	$^{q}\hat{D}(n+m^{*})=\left(\sum_{j=1}^{q}\frac{\psi(q,j)(n+m^{*})^{(j)}}{(n+m^{*})^{q}}\sum_{X_i\geq j}\frac{X_i^{(j)}}{n^{(j)}}\right)^{\frac{1}{1-q}}$ (nearly unbiased)
Expected sample completeness (coverage) as a function of sample size: $C_{ind}(m)=1-\sum_{i=1}^{S}p_i(1-p_i)^{m}$	$\hat{C}_{ind}(m)=1-\sum_{X_i\geq 1}\frac{X_i}{n}\frac{\binom{n-X_i}{m}}{\binom{n-1}{m}}$ (minimum variance unbiased)	$\hat{C}_{ind}(n+m^{*})=1-\frac{f_1}{n}\left[\frac{(n-1)f_1}{(n-1)f_1+2f_2}\right]^{m^{*}+1}$ (reliable for $m^{*}<n$)

[1] For the reference sample, the observed Hill number of order q is $^{q}\hat{D}(n)=\ ^{q}D_{obs}\ =[\sum_{X_i\geq 1}(X_i/n)^{q}]^{1/(1-q)}$. The coverage of the reference sample is estimated by $\hat{C}_{ind}(n)=1-\frac{f_1}{n}\left[\frac{(n-1)f_1}{(n-1)f_1+2f_2}\right]$; see Eq.12.

[2] The frequency count $f_k(m)$ is defined as the number of species represented by exactly k individuals/times in a sample of size m. The formula for $E[f_k(m)]$ is given in Eq. 5.

[3] An unbiased estimator $\hat{f}_k(m)$ for $E[f_k(m)]$ exists for $m < n$ and is given in Eq. 8.

[4] When m^* tends to infinity, each predictor tends to the estimator of the asymptotic diversity: ${}^0\hat{D}(\infty) = S_{obs} + \hat{f}_0$, where $\hat{f}_0$ is a predictor for f_0 (Chao 1984); see Eq. (B.5) in Appendix B. ${}^1\hat{D}(\infty) = \exp[\hat{H}(\infty)]$, where $\hat{H}(\infty)$ is an entropy estimator developed by Chao et al. (2013); see Eq. 10b. For an integer $q \geq 2$, ${}^q\hat{D}(\infty) = [\sum_{X_i \geq q} X_i^{(q)} / n^{(q)}]^{1/(1-q)}$, where $x^{(j)} = x(x-1)....(x-j+1)$ denotes the falling factorial; see Gotelli and Chao (2013).

[5] $\psi(q, j)$: the Stirling number of the second kind is defined by the coefficient in the expansion $x^q = \sum_{j=1}^{q} \psi(q, j) x^{(j)}$.

Table 2: The theoretical formulas and analytic estimators for rarefaction and extrapolation of Hill numbers based on incidence data for $q = 0$ (first row of equations), $q = 1$ (second row), $q = 2$ (third row) and any integer order $q > 2$ (fourth row), given a reference sample[1] with the observed Hill numbers ${}^{q}\Delta_{obs}$ and estimated coverage $\hat{C}_{sample}(T)$. The last row gives equations for sample completeness as a function of sample size, and the corresponding coverage estimators for rarefied samples and extrapolated samples. See Appendix C (for $q = 0$) and Appendix H (for $q > 0$) for notation and all derivation details.

Theoretical Formula [2] for all $t > 0$	Interpolation Estimator [3] ($t < T$)	Extrapolation Estimator [4] (for $T + t^*$ sampling units)
$${}^{0}\Delta(t) = S - E[Q_0(t)] = \sum_{k=1}^{t} E[Q_k(t)]$$	$${}^{0}\hat{\Delta}(t) = S_{obs} - \sum_{Y_i \geq 1} \frac{\binom{T - Y_i}{t}}{\binom{T}{t}}$$ (minimum variance unbiased)	$${}^{0}\hat{\Delta}(T + t^*) = S_{obs} + \hat{Q}_0 \times \left[1 - \left(1 - \frac{Q_1}{T\hat{Q}_0 + Q_1} \right)^{t^*} \right]$$ (reliable if $t^* < T$)
$${}^{1}\Delta(t) = \exp\left(\sum_{k=1}^{t} \left(-\frac{k}{U_t} \log \frac{k}{U_t} \right) \times E[Q_k(t)] \right)$$	$${}^{1}\hat{\Delta}(t) = \exp\left(\sum_{k=1}^{t} \left(-\frac{k}{\hat{U}_t} \log \frac{k}{\hat{U}_t} \right) \times \hat{Q}_k(t) \right)$$ (nearly unbiased)	$${}^{1}\hat{\Delta}(T + t^*) = \exp\left(\frac{T}{T + t^*} \sum_{i=1}^{S} \left(-\frac{Y_i}{U} \log \frac{Y_i}{U} \right) + \frac{t^*}{T + t^*} \hat{H}_{sample}(\infty) \right)$$ (nearly unbiased)
$${}^{2}\Delta(t) = \frac{1}{\sum_{k=1}^{t} \left(\frac{k}{U_t} \right)^2 \times E[Q_k(t)]}$$	$${}^{2}\hat{\Delta}(t) = \frac{1}{\sum_{k=1}^{t} \left(\frac{k}{\hat{U}_t} \right)^2 \times \hat{Q}_k(t)}$$	$${}^{2}\hat{\Delta}(T + t^*) = \frac{1}{\frac{1}{T + t^*} \times \frac{1}{U/T} + \frac{T + t^* - 1}{T + t^*} \sum_{Y_i > 0} \frac{Y_i(Y_i - 1)}{U^2(1 - 1/T)}}$$ (nearly unbiased)

	(nearly unbiased)	
$${}^{q}\Delta(t)=\left(\sum_{k=1}^{t}\left(\frac{k}{U_t}\right)^{q}\times E[Q_k(t)]\right)^{\frac{1}{1-q}}$$	$${}^{q}\hat{\Delta}(t)=\left(\sum_{k=1}^{t}\left(\frac{k}{\hat{U}_t}\right)^{q}\times \hat{Q}_k(t)\right)^{\frac{1}{1-q}}$$ (nearly unbiased)	$${}^{q}\hat{\Delta}(T+t^*)=\left(\frac{1}{(U/T)^q}\sum_{j=1}^{q}\frac{\psi(q,j)(T+t^*)^{(j)}}{(T+t^*)^q}\sum_{Y_i\geq j}\frac{Y_i^{(j)}}{T^{(j)}}\right)^{\frac{1}{1-q}}$$ (nearly unbiased)
Expected sample completeness (coverage) as a function of sample size: $$C_{sample}(t)=1-\frac{\sum_{i=1}^{S}\pi_i(1-\pi_i)^t}{\sum_{i=1}^{S}\pi_i}$$	$$\hat{C}_{sample}(t)=1-\sum_{Y_i\geq 1}\frac{Y_i}{U}\frac{\binom{T-Y_i}{t}}{\binom{T-1}{t}}$$ (nearly unbiased)	$$\hat{C}_{sample}(T+t^*)=1-\frac{Q_1}{U}\left[\frac{(T-1)Q_1}{(T-1)Q_1+2Q_2}\right]^{t^*+1}$$ (reliable for $t^* < T$)

[1] For the reference sample, the observed Hill number of order q is ${}^{q}\Delta_{obs}=[\sum_{k=1}^{T}(k/U)^q Q_k]^{1/(1-q)}$. The coverage of the reference sample is estimated by $\hat{C}_{sample}(T)=1-\frac{Q_1}{U}\left[\frac{(T-1)Q_1}{(T-1)Q_1+2Q_2}\right]$. $U=\sum_{Y_i>0}Y_i=\sum_{j=1}^{T}jQ_j$ denotes the total number of incidences in T samples.

[2] For any sample size of t, $Q_k(t)$ is defined as the number of species detected in exactly k sampling units. U_t is defined as the expected total number of incidences in t sampling units: $U_t=\sum_{j=1}^{t}jE[Q_j(t)]=t\sum_{i=1}^{S}\pi_i$.

[3] An unbiased estimator $\hat{Q}_k(t)$ for $E[Q_k(t)]$ exists for $t < T$ and is given in Eq. (H.5) in Appendix H. An unbiased estimator for U_t is $\hat{U}_t = \sum_{j=1}^{t} j\hat{Q}_j(t) = tU/T$.

[4] When t^* tends to infinity, each predictor tends to the estimator of the asymptotic diversity: ${}^0\hat{\Delta}(\infty) = S_{obs} + \hat{Q}_0$, where $\hat{Q}_0$ is a predictor for Q_0 (Chao 1987); see Eq. (C.5) in Appendix C. ${}^1\hat{\Delta}(\infty) = \exp[\hat{H}_{sample}(\infty)]$, where $\hat{H}_{sample}(\infty)$ is an entropy estimator for incidence data; see Eq. (H.7) in Appendix H. For an integer $q \geq 2$, ${}^q\hat{\Delta}(\infty) = [\sum_{Y_i \geq q} T^q Y_i^{(q)} / (U^q T^{(q)})]^{1/(1-q)}$, where $x^{(j)} = x(x-1)....(x-j+1)$. See Table 1 for the definition of the Stirling number of the second kind $\psi(q, j)$.

Table 3. Spider species abundance frequency counts in two canopy manipulation treatments (Ellison et al. 2010, Sackett et al. 2011). For the Girdled treatment, S_{obs} = 26 species, n = 168 individuals; for the Logged treatment, S_{obs} = 37 species, n = 252 individuals.

Girdled		Logged	
i	f_i	i	f_i
1	12	1	14
2	4	2	4
4	1	3	4
6	2	4	3
8	1	5	1
9	1	7	3
15	2	8	2
17	1	10	1
22	1	13	1
46	1	15	1
		16	1
		22	1
		88	1

Table 4. Systematic steps to determine base sample size for the sample-size-based rarefaction and extrapolation, and base coverage for the coverage-based rarefaction/extrapolation. Example 2 is used to illustrate each step. The reference sample size for the ith sample is denoted by n_i, i = 1, 2, …, k, and the corresponding sample coverage estimate is denoted by $C(n_i)$. For abundance data, sample size refers to the number of individuals; for incidence data, sample size refers to the number of sampling units.

Step 0: Set the *maximum extrapolated ratio r*, equal to the ratio of the extrapolated sample size and the reference sample size. For making inferences about species richness (q = 0), we suggest the maximum extrapolated size should be double the reference sample size, that is, $r = 2$. For inferences for diversity of $q \geq 1$, r can be any positive number, i.e., it is statistically safe to extrapolate to the asymptote.

Sample-size-based rarefaction/extrapolation

Step 1. Compute the maximum reference sample size, $n_a = \max\{n_1, n_2, \ldots, n_k\}$.

(In Example 2, $n_a = \max\{599, 230, 150, 200, 200\} = 599$.)

Step 2. Compute the minimum r times reference sample sizes, $n_b = \min\{rn_1, rn_2, \ldots, rn_k\}$.

(In Example 2 for $r = 2$, $n_b = \min\{1198, 460, 300, 400, 400\} = 300$.)

Step 3. The suggested base sample size is the maximum of n_a and n_b, $n_{base} = \max\{n_a, n_b\}$.

(In Example 2 for $r = 2$, $n_{base} = \max\{599, 300\} = 599$.)

Coverage-based rarefaction/extrapolation

Step 1. Compute the maximum coverage of reference sample sizes, $C_a = \max\{C(n_1), C(n_2), \ldots, C(n_k)\}$.

(In Example 2, $C_a = \max\{0.9918, 0.976, 0.9839, 0.9889, 0.9964\} = 0.9964$.)

Step 2. Compute the minimum coverage of r times reference sample sizes, $C_b = \min\{C(rn_1),$

$C(rn_2)$, …, $C(rn_k)$}.

(In Example 2 for $r = 2$, $C_b = \min\{0.9968, 0.9908, 0.9949, 0.9940, 0.9999\} = 0.9908$.)

Step 3. The suggested base coverage is the maximum of C_a and C_b, $C_{base} = \max\{C_a, C_b\}$.

(In Example 2 for $r = 2$, $C_{base} = \max\{0.9964, 0.9908\} = 0.9964$.

FIGURE LEGENDS

Fig. 1: (a) A diversity profile curve, which plots Hill numbers ${}^{q}D(\infty)$ as a function of order q, $0 \le q \le 3$. Hill numbers are calculated for a Zipf-Mandelbrot model (Magurran 2004) including 100 species with species relative abundance $p_i = c/i$ where c is a constant such that $\sum_{i=1}^{100} p_i = 1$. The three solid dots denote Hill numbers for order q = 0, 1 and 2. The diversity profile curve is a non-increasing function of q. The slope of the curve reflects the unevenness of species relative abundances. The more uneven the distribution of relative abundances, the more steeply the curve declines. For completely even relative abundances, the curve is a constant at the level of species richness. (b) Sample-size-based diversity accumulation curve, which plots the expected diversity ${}^{q}D(m)$ as a function of size m, q = 0, 1 and 2. As sample size m tends to infinity, each curve approaches ${}^{q}D(\infty)$. (c) Coverage-based diversity accumulation curve, which plots the expected diversity ${}^{q}D(m)$ as a function of expected coverage, q = 0, 1 and 2. As sample coverage tends to unity, each curve approaches ${}^{q}D(\infty)$.

Fig. 2: Rank-abundance distributions for spider data from the Girdled and Logged treatments. The proportional abundance on the y-axis (on a logarithmic scale) is calculated as the proportion of the maximum abundance.

Fig. 3: (a) (upper panel) Sample-size-based rarefaction (solid lines) and extrapolation (dotted lines, up to double the reference sample size) of spider species diversity based on the Hill numbers (q = 0, 1, 2) for the Girdled treatment and the Logged treatment. The 95% confidence intervals are obtained by a bootstrap method based on 200 replications. Reference samples are denoted by solid dots. The numbers in parentheses are the sample size and the observed Hill numbers for each reference sample. (b) (lower panel) Comparison of sample-size-based rarefaction (solid lines) and extrapolation (dotted curves), up to the base sample size of 336

individuals (i.e., double the smaller reference sample size) of spider species diversity for Hill numbers of order $q = 0$ (left panel), $q = 1$ (middle panel), and $q = 2$ (right panel). Reference samples in each treatment are denoted by solid dots. The numbers in parentheses are the sample size and the observed Hill numbers for each reference sample.

Fig. 4: Plot of sample coverage for rarefied samples (solid line) and extrapolated samples (dashed line) as a function of sample size for spider samples from the Girdled and Logged treatments. The 95% confidence intervals were obtained by a bootstrap method based on 200 replications. Reference samples are denoted by solid dots. Each of the two curves was extrapolated up to double its reference sample size. The numbers in parentheses are the sample size and the estimated sample coverage for each reference sample.

Fig. 5: (a) (Upper panel) Coverage-based rarefaction (solid line) and extrapolation (dashed line) plots with 95% confidence intervals for spider species diversity based on Hill numbers ($q = 0, 1, 2$) for the Girdled and Logged treatments. Reference samples are denoted by solid dots. In the Girdled treatment, the coverage is extrapolated to 96%, and in the Logged treatment, the coverage is extrapolated to 97% (i.e., the coverage value for a doubling of each reference sample size). The numbers in parentheses are the sample coverage and the observed Hill numbers for each reference sample. (b) (Lower panel) Comparison of the coverage-based rarefaction (solid lines) and extrapolation (dotted curves), up to the base coverage 96% (i.e., lower coverage of the doubled reference sample sizes) of spider diversity using Hill numbers of order $q = 0$ (left panel), $q = 1$ (middle panel), and $q = 2$ (right panel). Reference samples in each treatment are denoted by solid dots. Note that species richnesses (left panel) in the two treatments is significantly different when sample coverage is between 50% and 96%, as the two

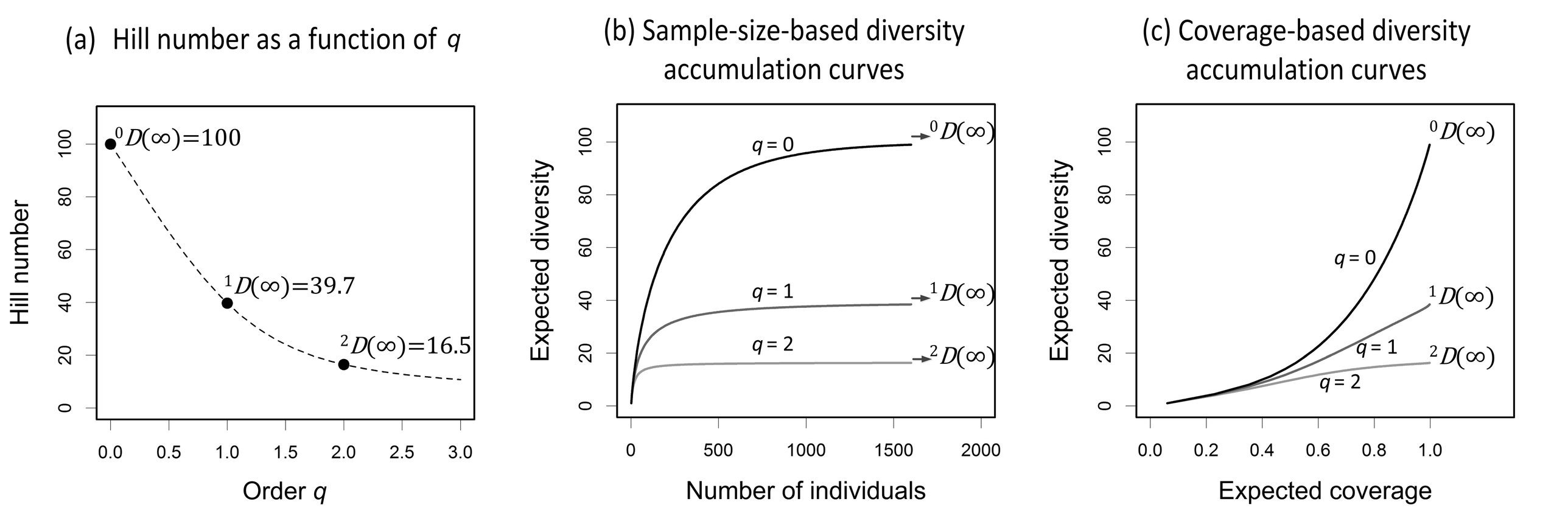
confidence bands do not intersect in this range of coverage values. The numbers in parentheses are the sample coverage and the observed Hill numbers for each reference sample.

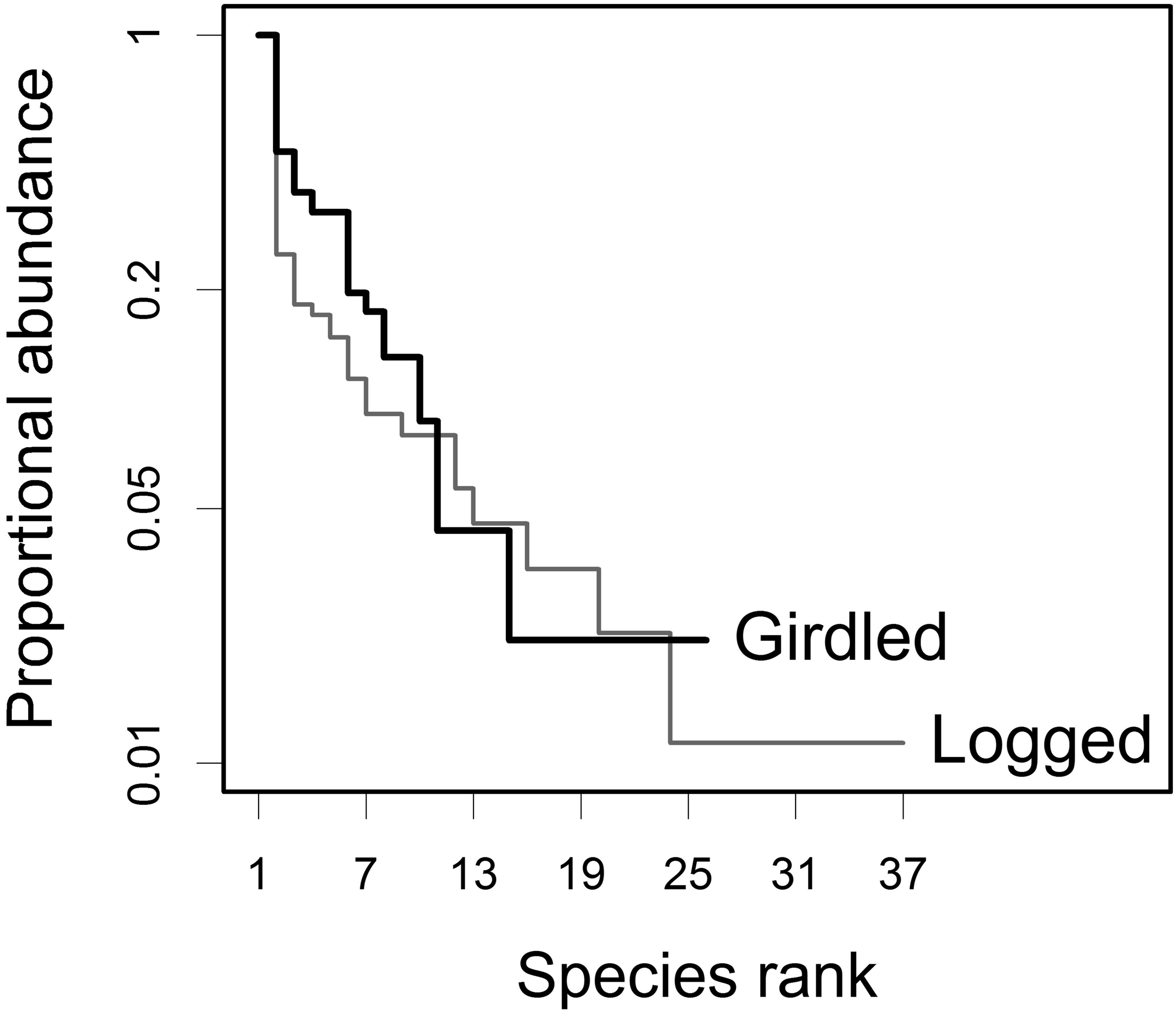
Fig. 6: (a) (Upper and middle panels) Sample-size-based rarefaction (solid lines) and extrapolation (dotted lines, up to double the reference sample size) of tropical ant diversity from Costa Rica for Hill numbers (q = 0, 1, 2) for each of the five elevations. The 95% confidence intervals are obtained by a bootstrap method based on 200 replications. Reference samples are denoted by solid dots. The numbers in parentheses are the sample size and the observed Hill numbers for each reference sample. (b) (Bottom panel) Comparison of sample-size-based rarefaction (solid line) and extrapolation (dashed line) curves with 95% confidence intervals for Hill numbers q = 0 (left panel), q = 1 (middle panel), and q = 2 (right panel). All curves are extrapolated up to the base sample size of 599. Reference samples are denoted by solid dots. The numbers in parentheses are the sample size and the observed Hill numbers for each reference sample.

Fig. 7: Plot of sample coverage for rarefied samples (solid line) and extrapolated samples (dashed line) with 95% confidence intervals for tropical ants sampled from five sites (data from Longino and Colwell 2011). Each curve was extrapolated up to a doubling of its reference sample size (the extrapolated curve for 500 m site is cut off at 700 and thus is not completely shown). Reference samples are denoted by solid dots.

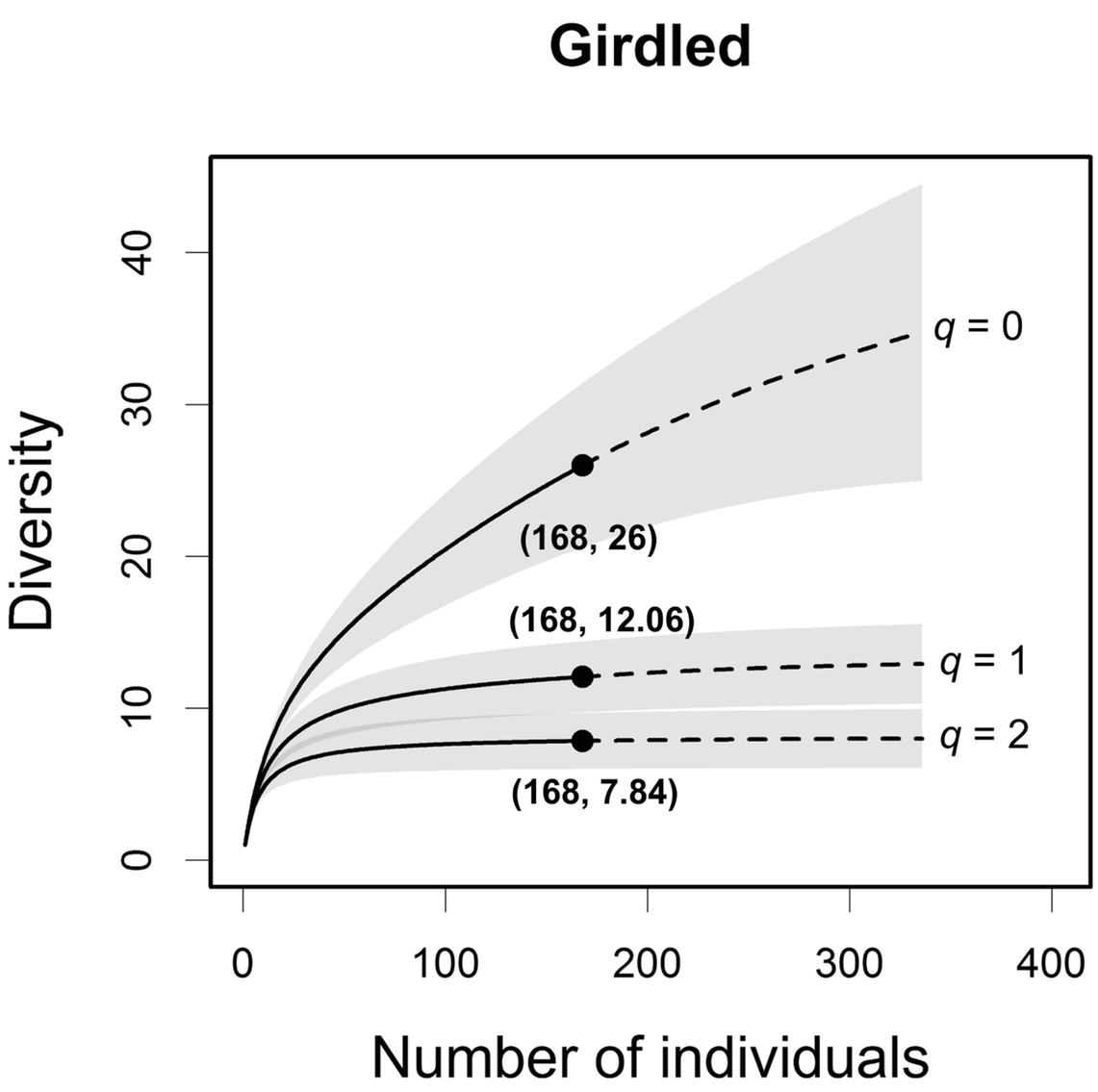
Fig. 8: (a) (Upper and middle panels) Coverage-based Rarefaction (solid line) and extrapolation (dashed line) plots with 95% confidence intervals for tropical ant diversity based on Hill numbers (q = 0, 1, 2) for five elevations. Reference samples are denoted by solid dots. The extrapolation is extended to the coverage value for a doubling of the size of each reference

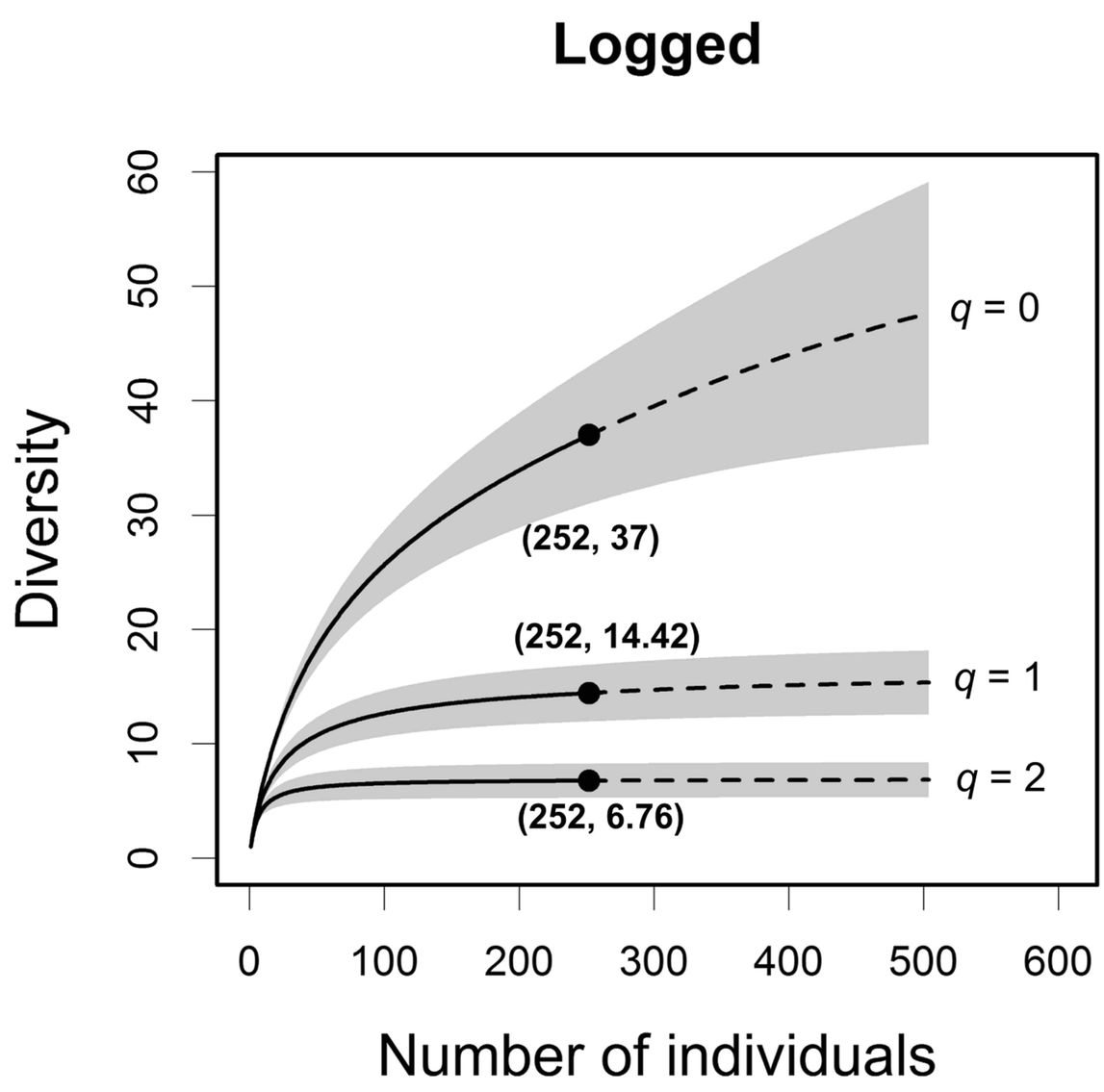
sample. The numbers in parentheses are the sample coverage and the observed Hill numbers for each reference sample. (b) (Bottom panel): Comparison of the coverage-based rarefaction (solid lines) and extrapolation (dotted curves), up to a base coverage of 99.64% for tropical ant diversity samples using Hill numbers of order $q = 0$ (left panel), $q = 1$ (middle panel), and $q = 2$ (right panel), with 95% confidence intervals based on 200 bootstrap replications. Reference samples in each plot are denoted by solid dots. The numbers in parentheses are the sample coverage and the observed Hill numbers for each reference sample. Note that some confidence intervals in (a) and (b) are very narrow so that they are almost invisible.


(a) Hill number as a function of q
Hill number
Order q
0
20
40
60
80
100
0.0
0.5
1.0
1.5
2.0
2.5
3.0
$^0D(\infty)=100$
$^1D(\infty)=39.7$
$^2D(\infty)=16.5$
(b) Sample-size-based diversity accumulation curves
Expected diversity
Number of individuals
0
500
1000
1500
2000
q = 0
q = 1
q = 2
$^0D(\infty)$
$^1D(\infty)$
$^2D(\infty)$
(c) Coverage-based diversity accumulation curves
Expected diversity
Expected coverage
0.0
0.2
0.4
0.6
0.8
1.0
q = 0
q = 1
q = 2
$^0D(\infty)$
$^1D(\infty)$
$^2D(\infty)$

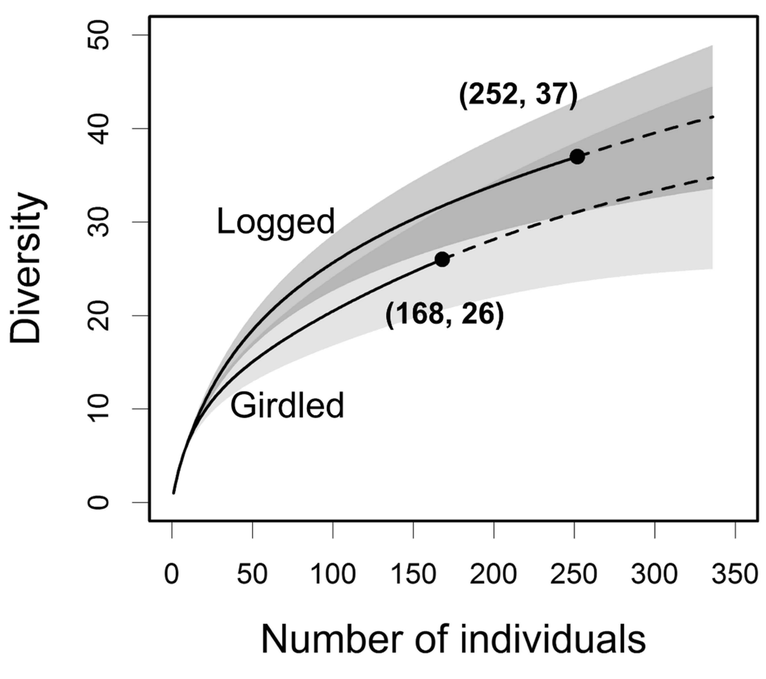

Proportional abundance
1
0.2
0.05
0.01
Girdled
Logged
1
7
13
19
25
31
37
Species rank

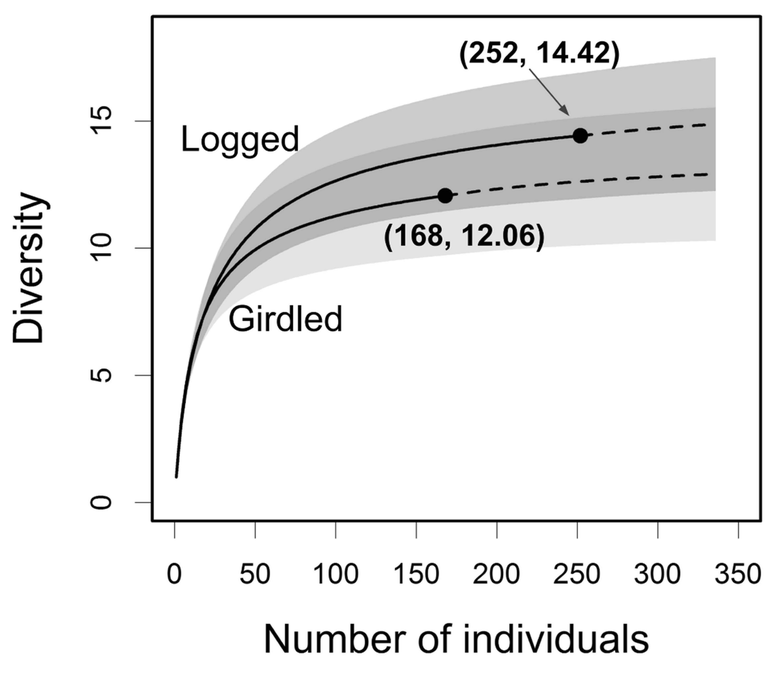
(a) Sample-size-based rarefaction and extrapolation curves for each treatment

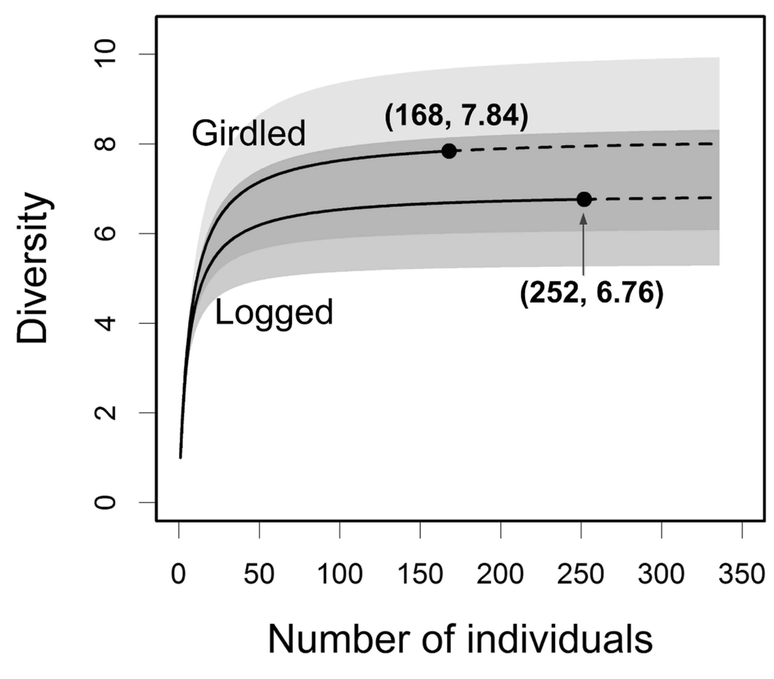




(b) Comparisons of two treatments for each order of *q*







Sample coverage

1
0.9
0.8
0.7
0.6

Girdled

Logged

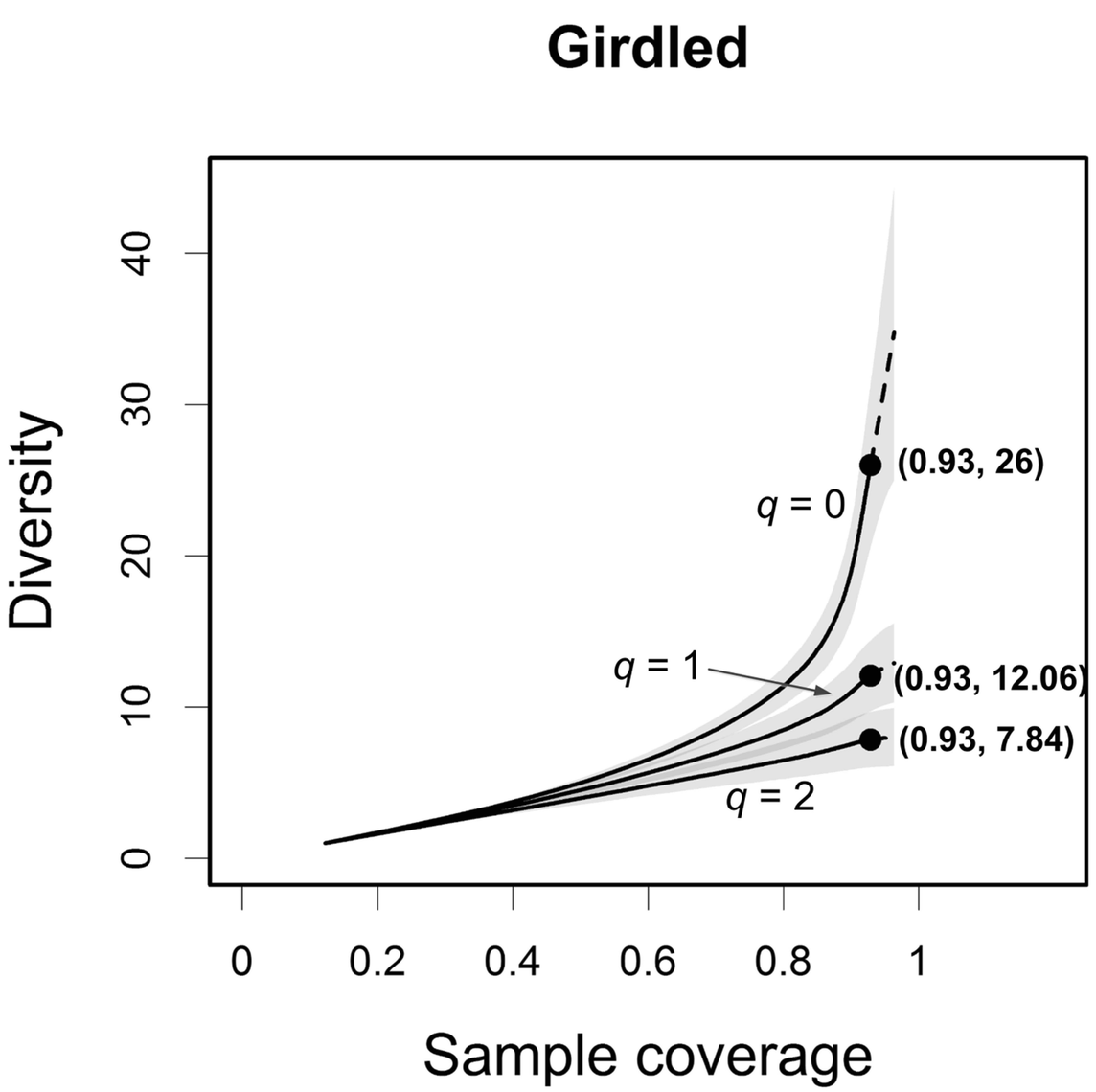
(168, 0.93)

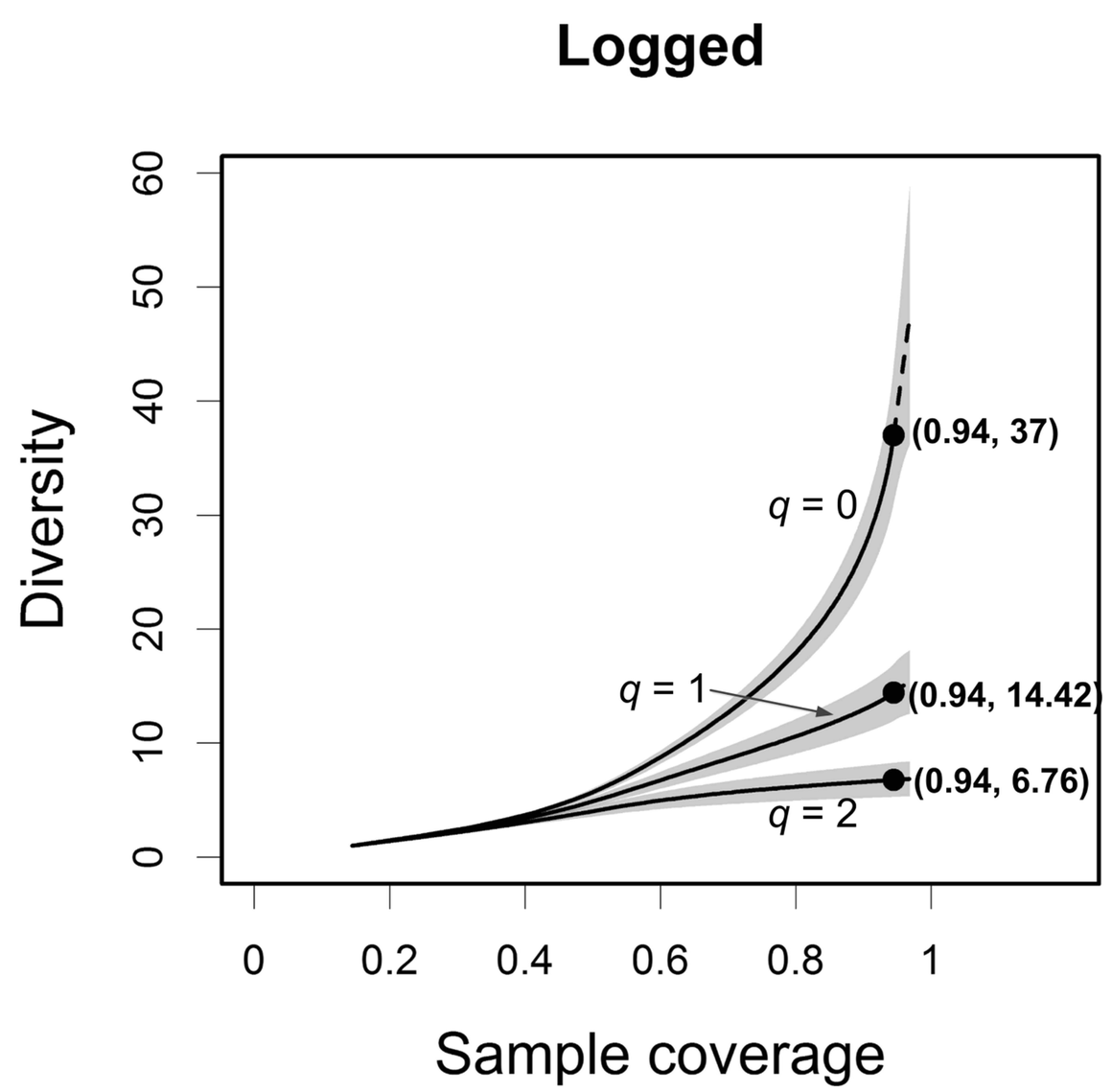
(252, 0.94)

0
100
200
300
400
500

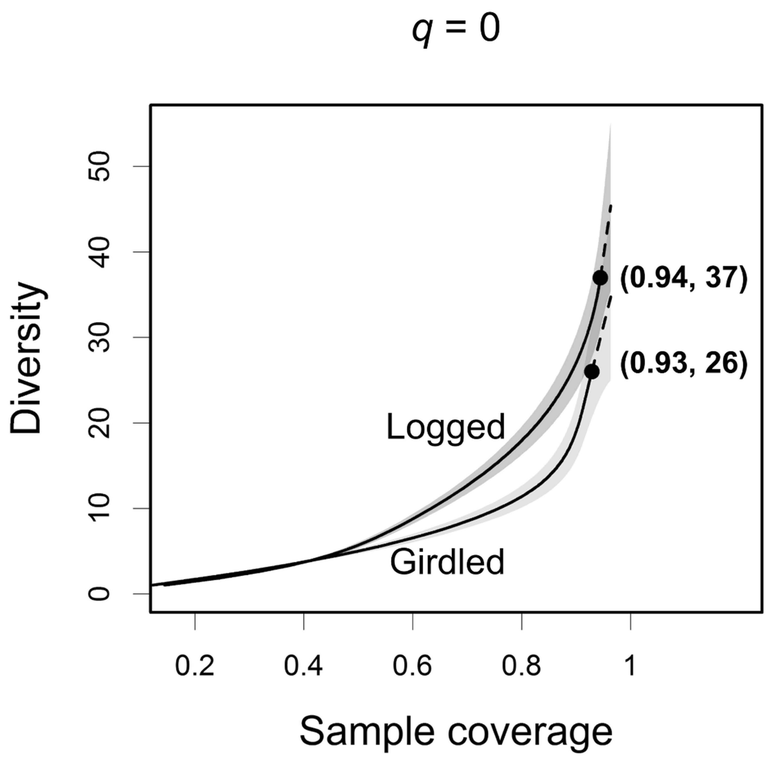
Number of individuals

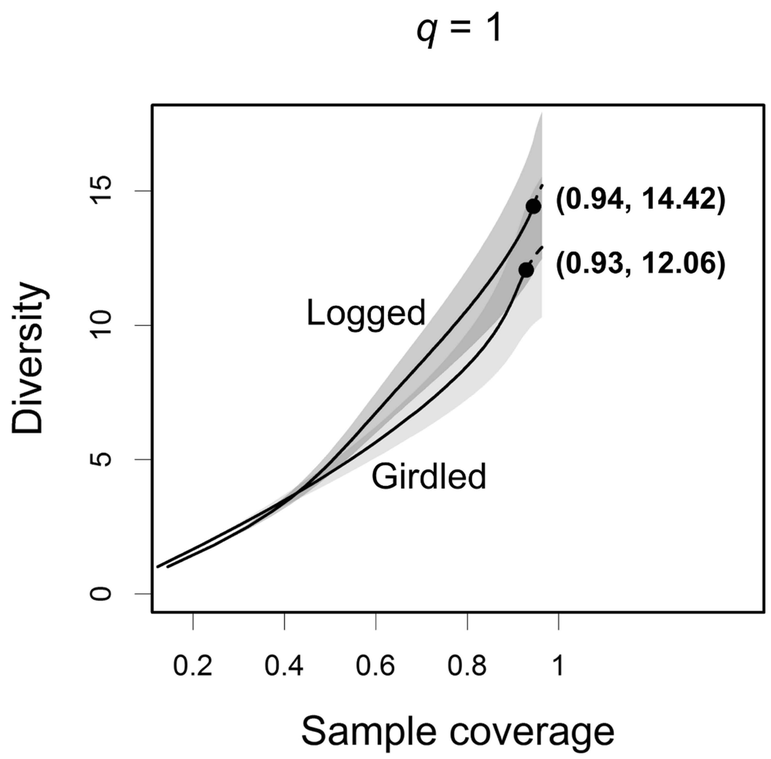
(a) Coverage-based rarefaction and extrapolation curves for each treatment

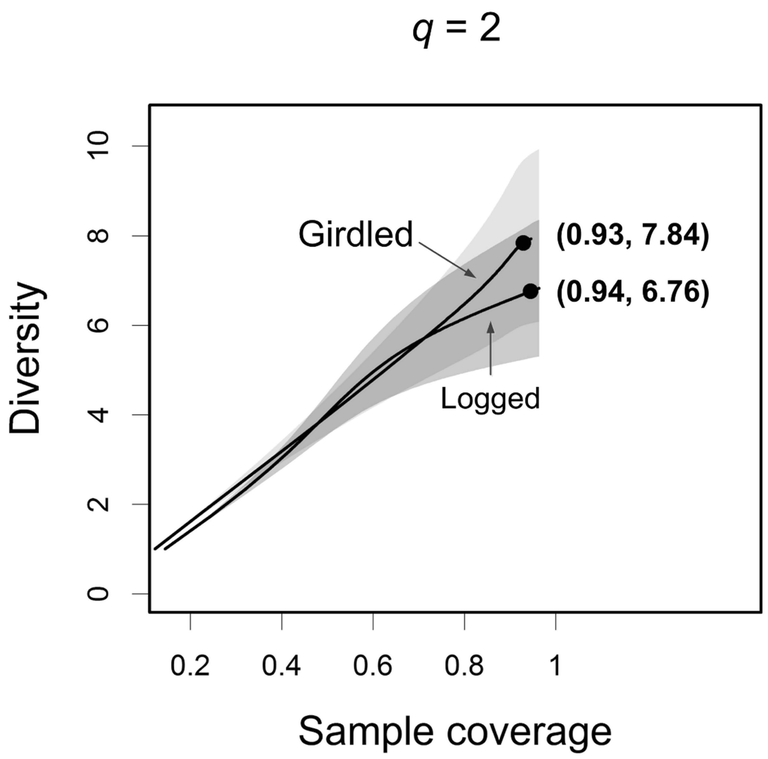




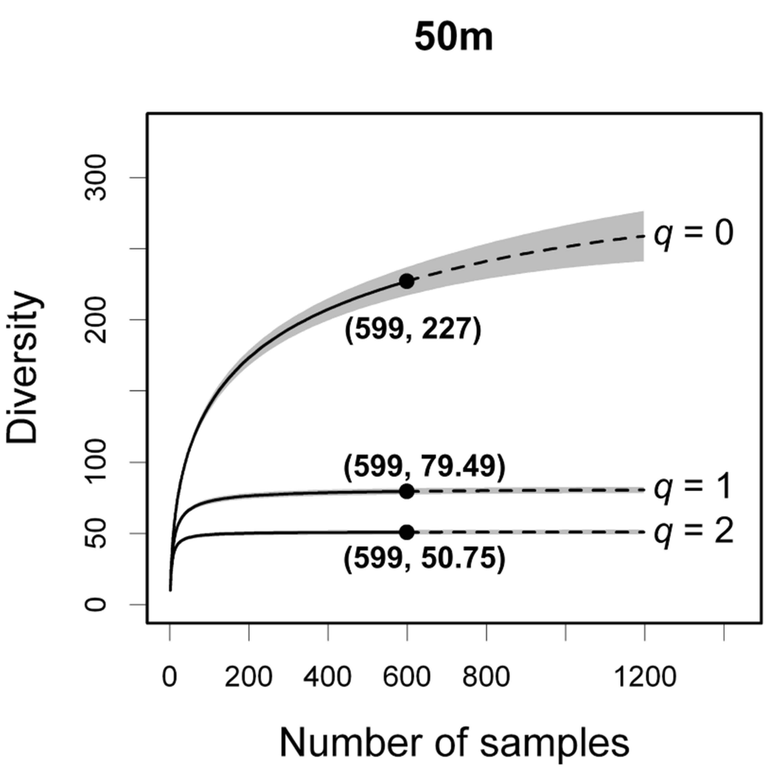
(b) Comparisons of two treatments for each order of *q*

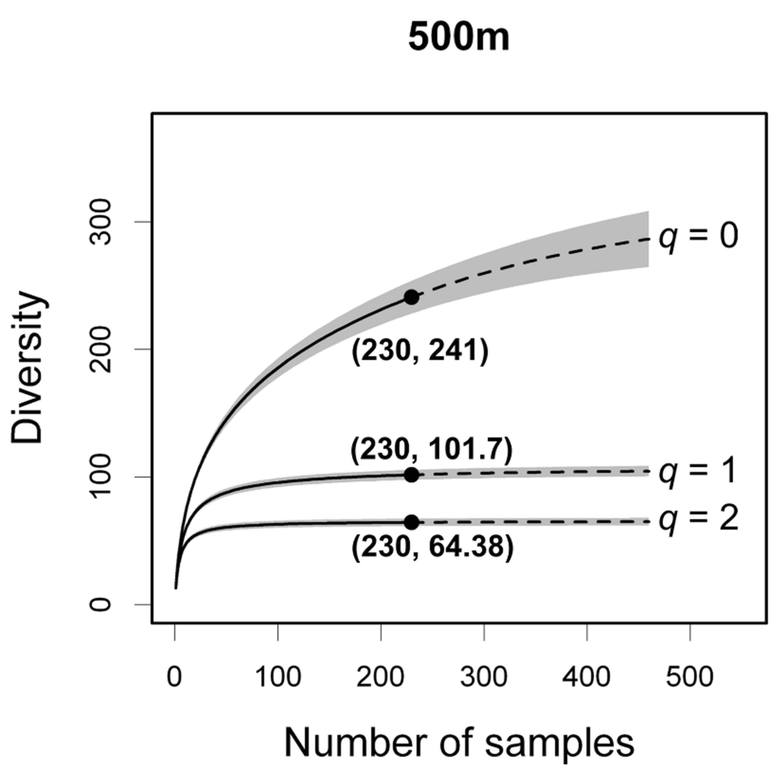


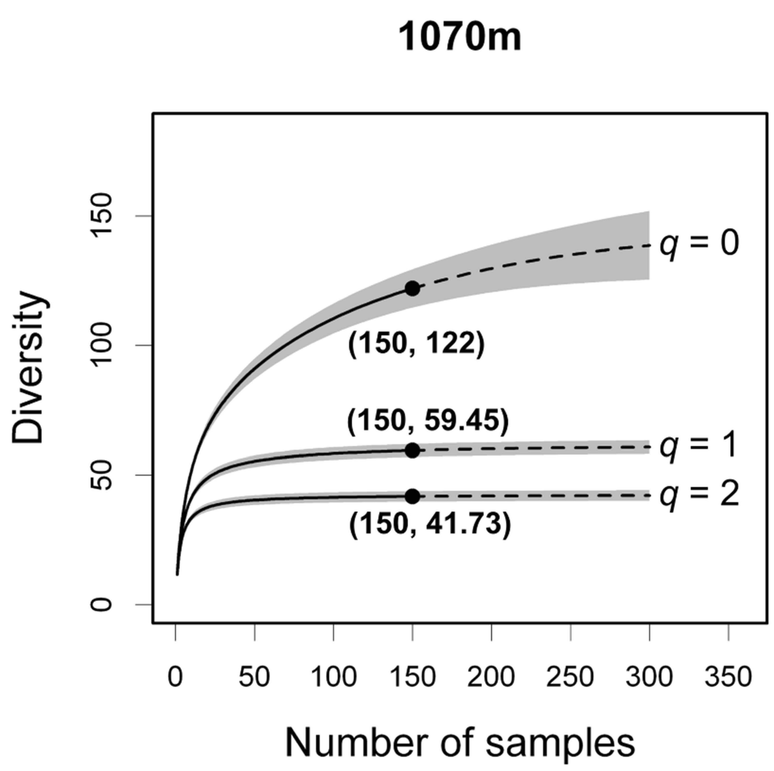


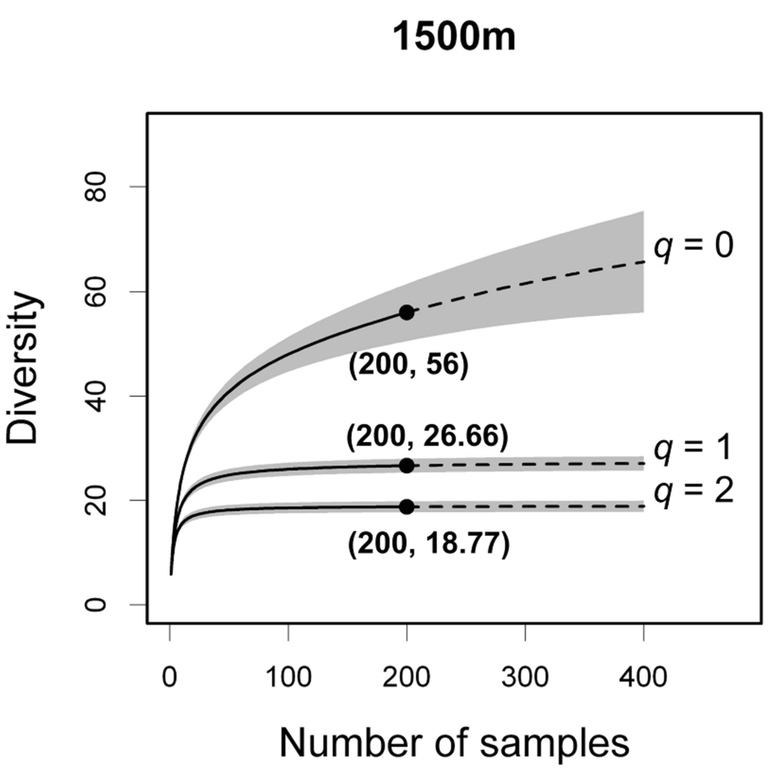


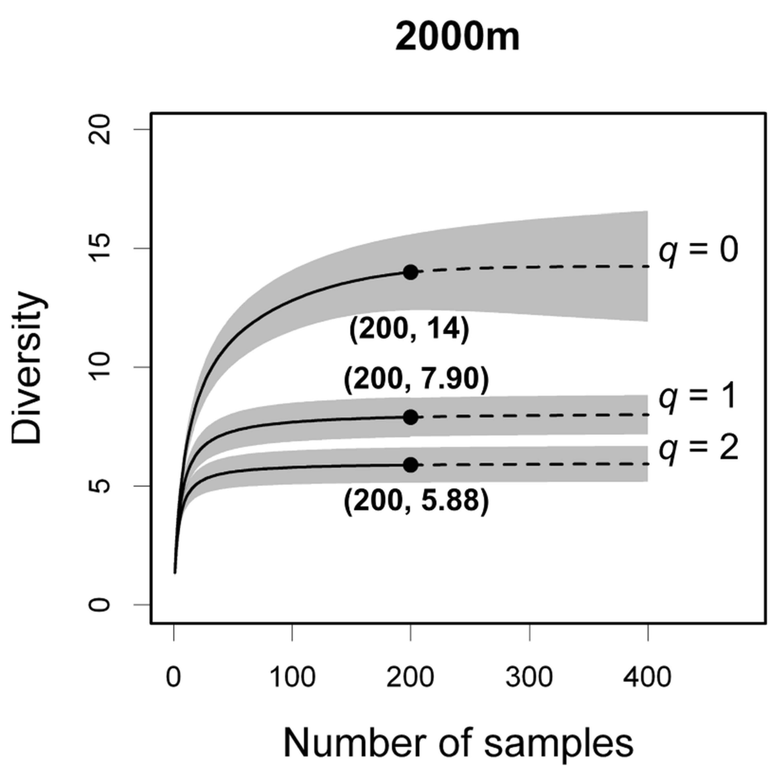
(a) Sample-size-based rarefaction and extrapolation curves for each site



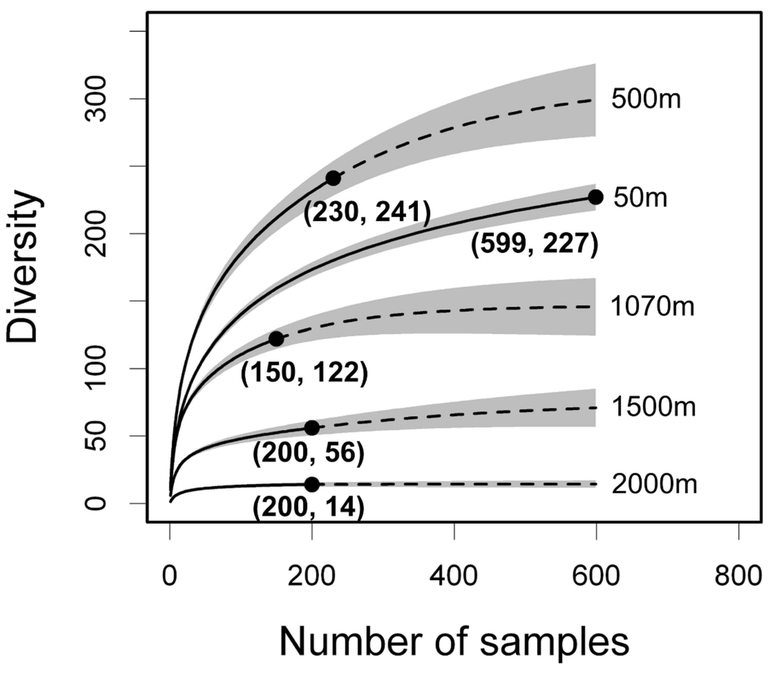


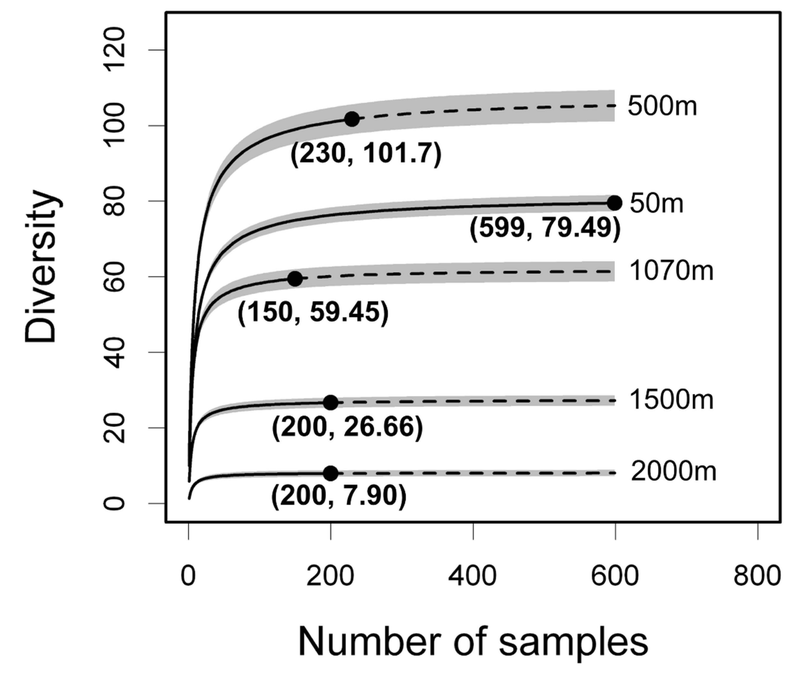


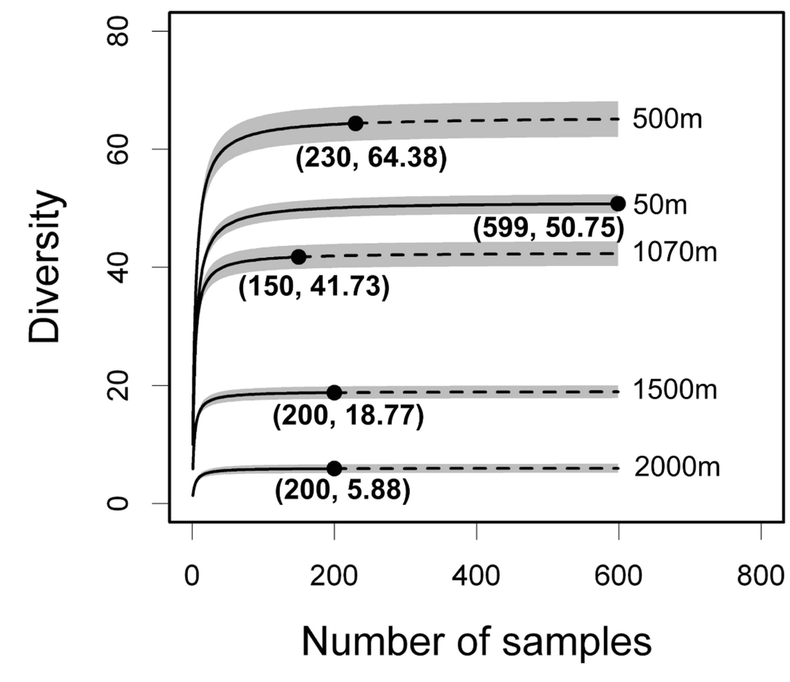


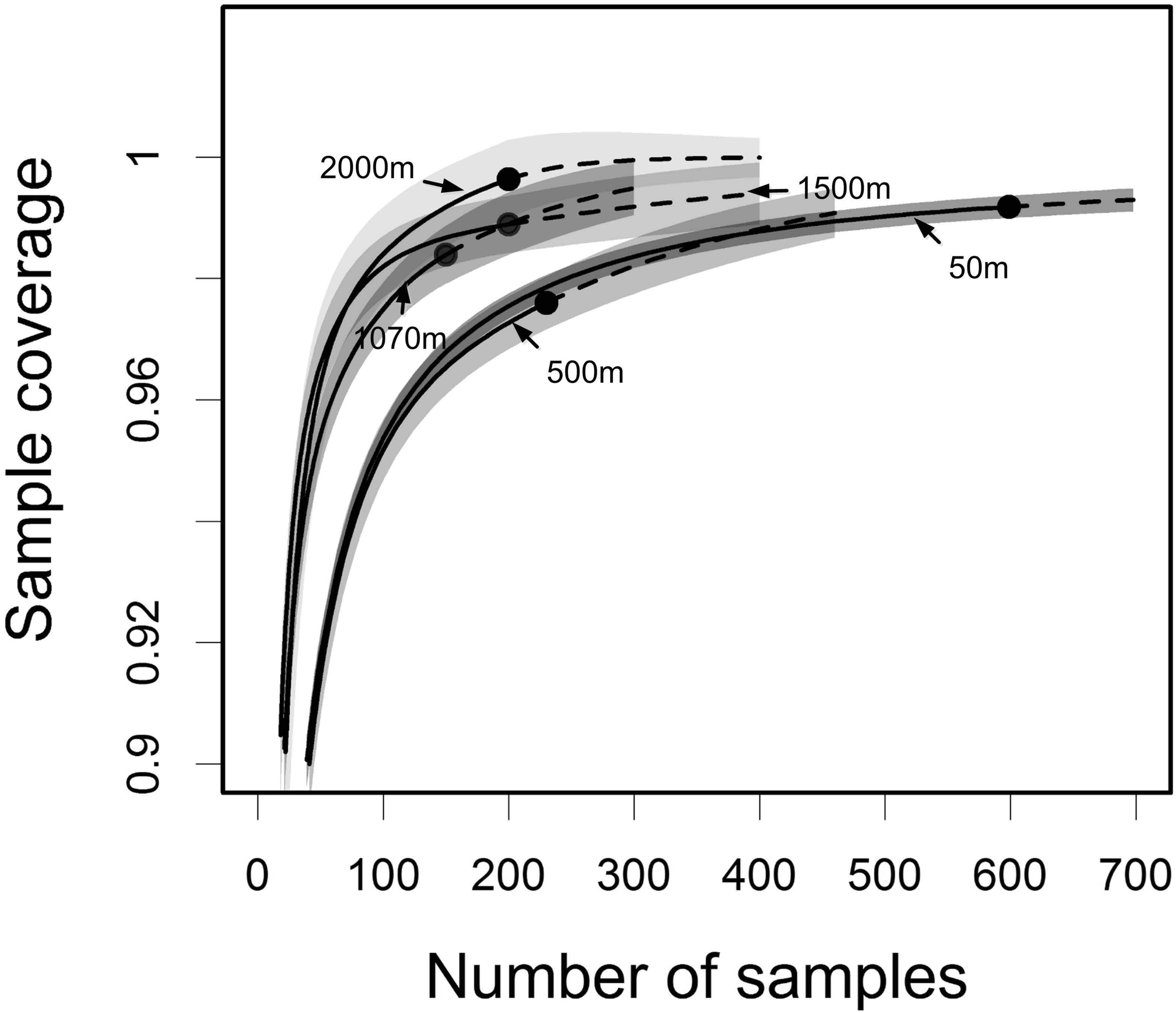


(b) Comparisons of five sites for each order of *q*

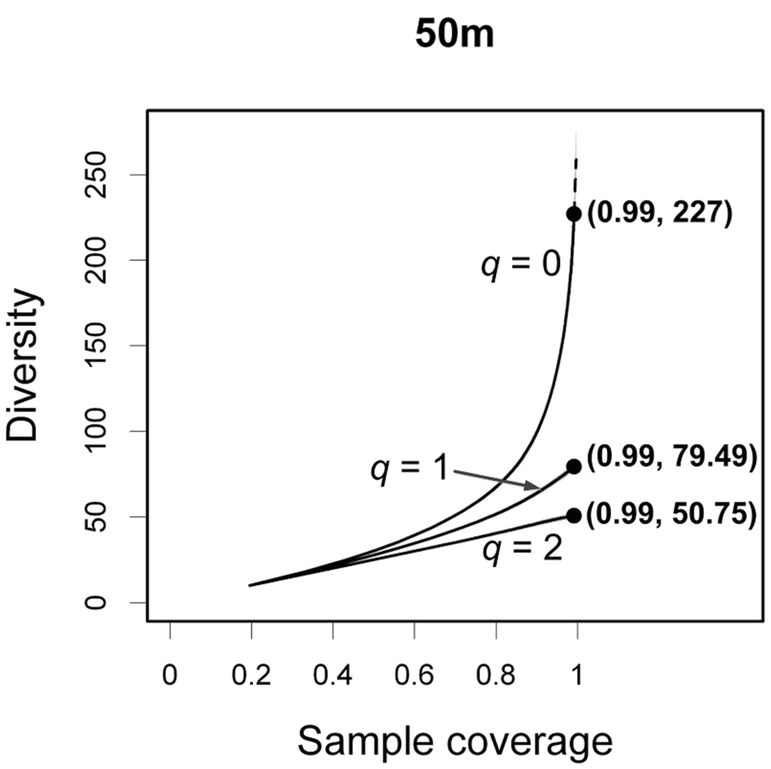


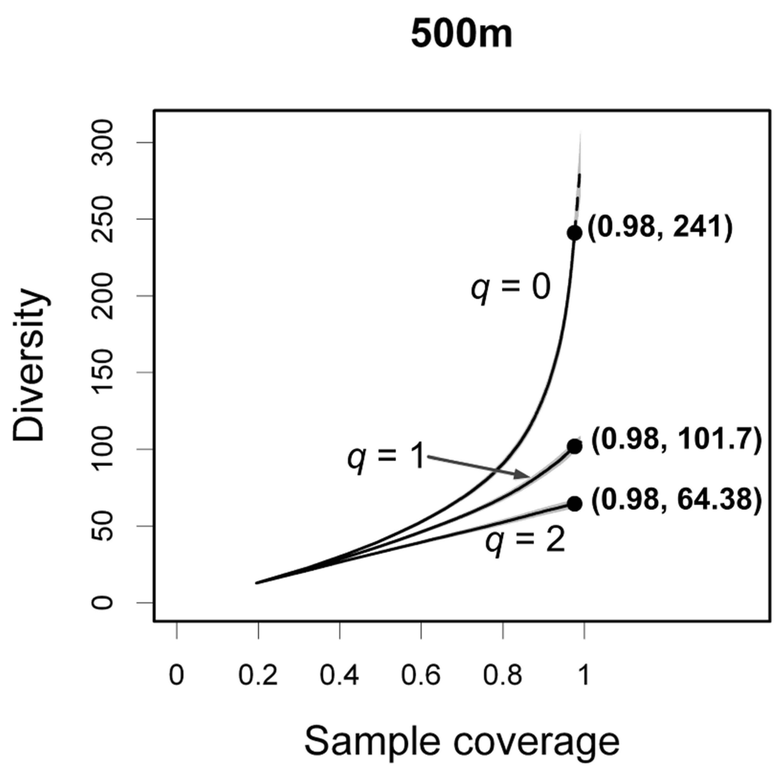


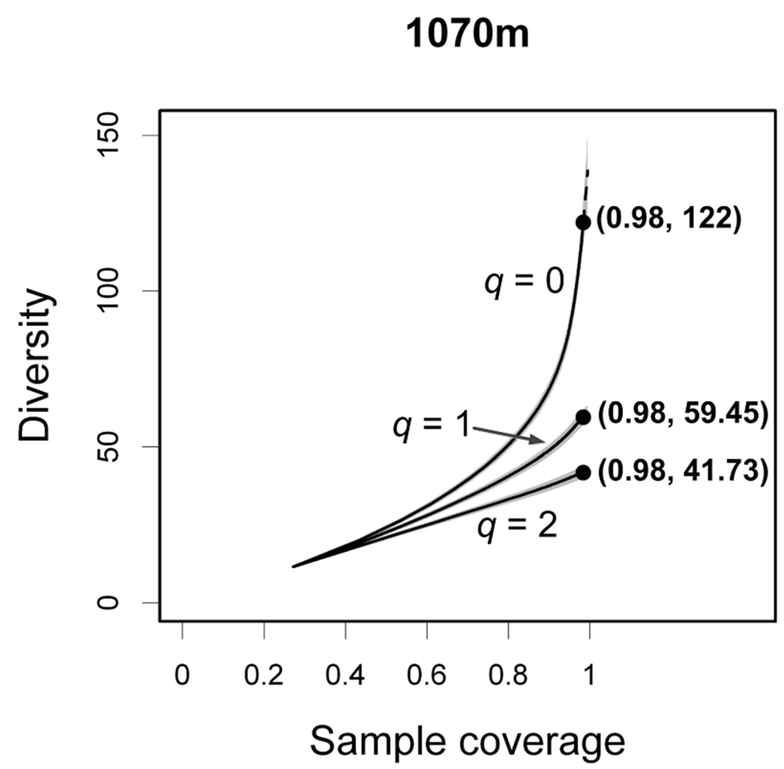


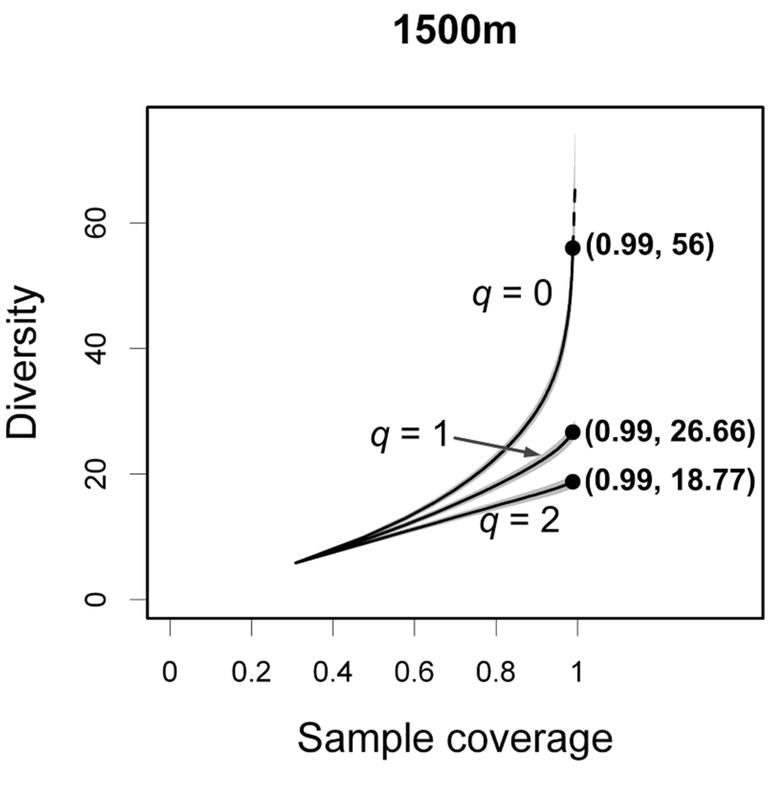

Sample coverage
1
0.96
0.92
0.9
2000m
1500m
50m
1070m
500m
0
100
200
300
400
500
600
700
Number of samples

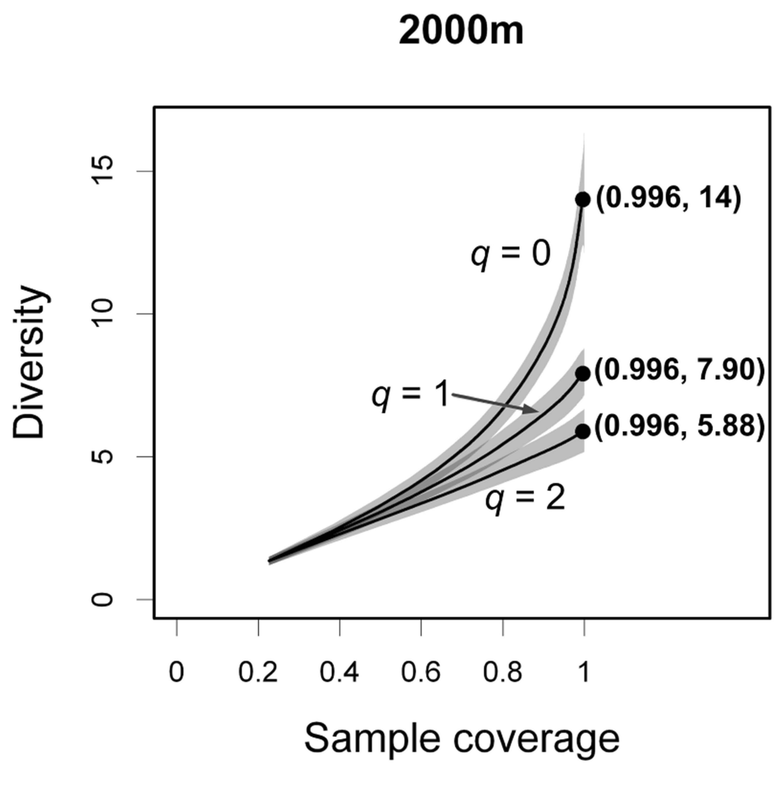
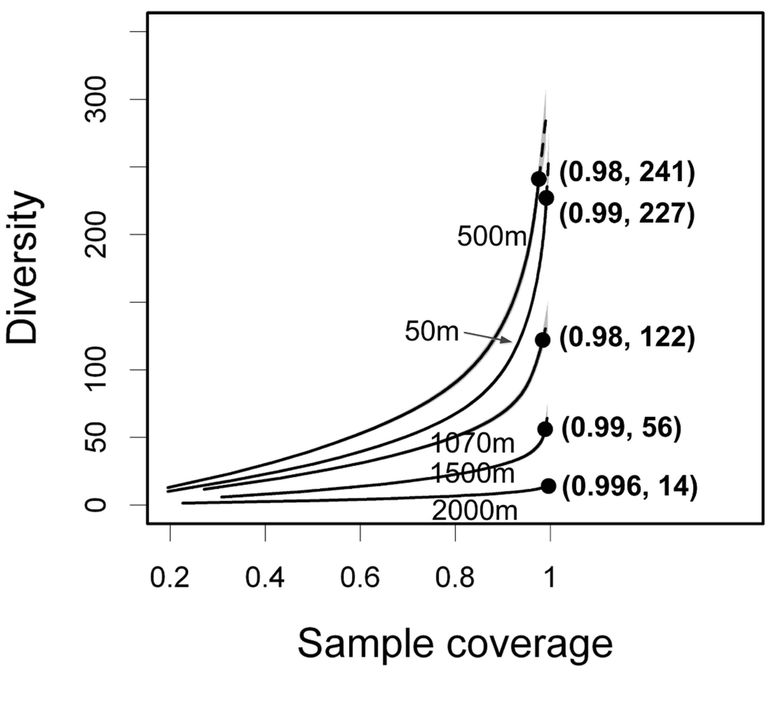
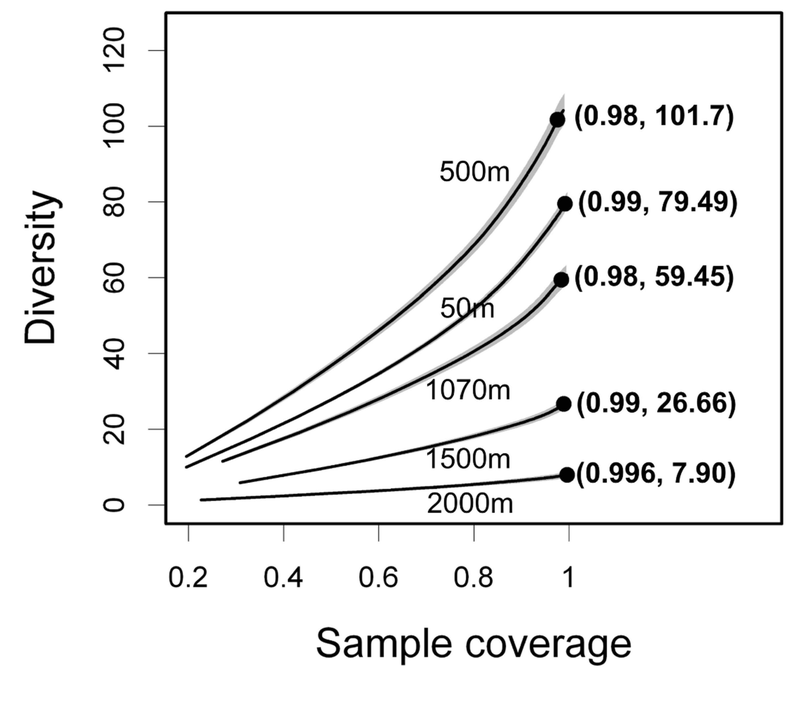
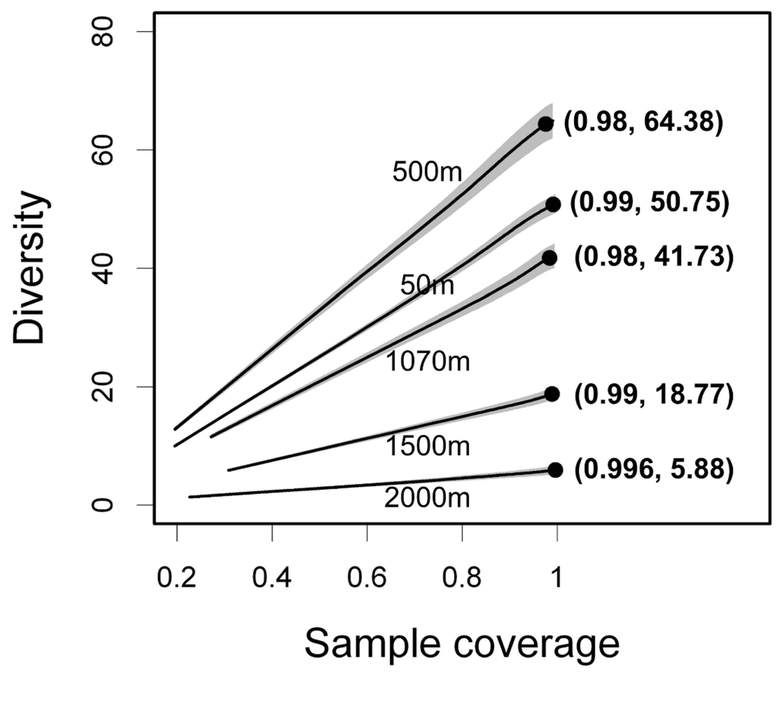
(a) Coverage-based rarefaction and extrapolation curves for each site











(b) Comparisons of five sites for each order of *q*