METHOD

Rarefaction and extrapolation with beta diversity under a framework of Hill numbers: The iNEXT.beta3D standardization

Anne Chao[1] | Simon Thorn[2,3] | Chun-Huo Chiu[4] | Faye Moyes[5] |
Kai-Hsiang Hu[1] | Robin L. Chazdon[6,7] | Jessie Wu[1] |
Luiz Fernando S. Magnago[8] | Maria Dornelas[5] | David Zelený[9] |
Robert K. Colwell[7,10,11] | Anne E. Magurran[5]

[1]Institute of Statistics, National Tsing Hua University, Hsinchu, Taiwan

[2]Hessian Agency for Nature Conservation, Environment and Geology, Biodiversity Center, Gießen, Germany

[3]Czech Academy of Sciences, Biology Centre, Institute of Entomology, České Budějovice, Czech Republic

[4]Department of Agronomy, National Taiwan University, Taipei, Taiwan

[5]Centre for Biological Diversity and Scottish Oceans Institute, School of Biology, University of St Andrews, St Andrews, UK

[6]Tropical Forest and People Research Centre, University of the Sunshine Coast, Sippy Downs, Queensland, Australia

[7]Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, Connecticut, USA

[8]Centro de Formação em Ciências Agroflorestais, Universidade Federal do Sul da Bahia (UFSB), Ilhéus, Brazil

[9]Institute of Ecology and Evolutionary Biology, National Taiwan University, Taipei, Taiwan

[10]University of Colorado Museum of Natural History, Boulder, Colorado, USA

[11]Departamento de Ecologia, Universidade Federal de Goiás, Goiânia, Brazil

Correspondence
Anne Chao
Email: chao@stat.nthu.edu.tw

Abstract

Based on sampling data, we propose a rigorous standardization method to measure and compare beta diversity across datasets. Here beta diversity, which quantifies the extent of among-assemblage differentiation, relies on Whittaker's original multiplicative decomposition scheme, but we use Hill numbers for any diversity order $q \geq 0$. Richness-based beta diversity ($q = 0$) quantifies the extent of species identity shift, whereas abundance-based ($q > 0$) beta diversity also quantifies the extent of difference among assemblages in species abundance. We adopt and define the assumptions of a statistical sampling model as the foundation for our approach, treating sampling data as a representative sample taken from an assemblage. The approach makes a clear distinction between the theoretical assemblage level (unknown properties/parameters of the assemblage) and the sampling data level (empirical/observed statistics computed from data). At the assemblage level, beta diversity for N assemblages reflects the interacting effect of the species abundance distribution and spatial/temporal aggregation

of individuals in the assemblage. Under independent sampling, observed beta (= gamma/alpha) diversity depends not only on among-assemblage differentiation but also on sampling effort/completeness, which in turn induces dependence of beta on alpha and gamma diversity. How to remove the dependence of richness-based beta diversity on its gamma component (species pool) has been intensely debated. Our approach is to standardize gamma and alpha based on sample coverage (an objective measure of sample completeness). For a single assemblage, the iNEXT method was developed, through interpolation (rarefaction) and extrapolation with Hill numbers, to standardize samples by sampling effort/completeness. Here we adapt the iNEXT standardization to alpha and gamma diversity, that is, alpha and gamma diversity are both assessed at the same level of sample coverage, to formulate standardized, coverage-based beta diversity. This extension of iNEXT to beta diversity required the development of novel concepts and theories, including a formal proof and simulation-based demonstration that the resulting standardized beta diversity removes the dependence of beta diversity on both gamma and alpha values, and thus reflects the pure among-assemblage differentiation. The proposed standardization is illustrated with spatial, temporal, and spatiotemporal datasets, while the freeware iNEXT.beta3D facilitates all computations and graphics.

INTRODUCTION

Diversity varies in space and time. Spatial and temporal variation in species composition is one of the most fundamental features of the natural world. In his pioneering papers, Whittaker (1960, 1972) proposed to decompose multiplicatively the total diversity in a region (gamma) into its within-assemblage component (alpha diversity) and among-assemblage component (beta diversity). Beta diversity measures the extent of among-assemblage differentiation. Previous work has focused on spatial and temporal comparisons of the diversity of individual assemblages (e.g., Chao et al., 2021; Chase et al., 2020; Dornelas et al., 2014). As indicated in a recent workshop on "Biodiversity and Climate Change" (IPBES-IPCC, 2021), there is an urgent need to understand, monitor, and rigorously quantify species compositional shifts and reorganization of biodiversity, beyond simple comparisons of diversity. In addition, changes in beta diversity have emerged as a signature of the Anthropocene; tracking and monitoring compositional changes in ecosystems should thus be a conservation priority (Chase et al., 2018; Dornelas et al., 2014; Eriksson & Hillebrand, 2019; Hillebrand et al., 2018; Magurran et al., 2019; Mori et al., 2018; Wang & Loreau, 2016).

Measuring beta diversity among assemblages is more complicated than quantifying diversity in a single assemblage. Yet, little consensus has been reached on how to characterize and compare beta diversity based on sampling data (Ellison, 2010; Engel et al., 2021; Gotelli et al., 2022; Hillebrand et al., 2018; Jost, 2010; McGlinn et al., 2019, 2021; Xing & He, 2021, among others). In this paper, we propose a rigorous standardization method to measure and compare beta diversity across datasets. Below we briefly introduce the background of the iNEXT standardization for a single assemblage and the general framework for formulating beta diversity among N assemblages; some previous relevant approaches are outlined, followed by the introduction of our proposed methodology. Figure 1 shows a schematic flow diagram of pertinent research questions, methodology/software, and graphical output.

The iNEXT standardization for a single assemblage

In the past decade, a consensus has emerged around using Hill numbers (Hill, 1973), often referred to as the "effective number of species," for quantifying species diversity in an assemblage (Ellison, 2010 and papers that followed, as well as Moreno & Rodríguez, 2010, plus subsequent commentaries). In Hill's influential 1973 paper, he integrated species richness and species relative abundance to characterize a
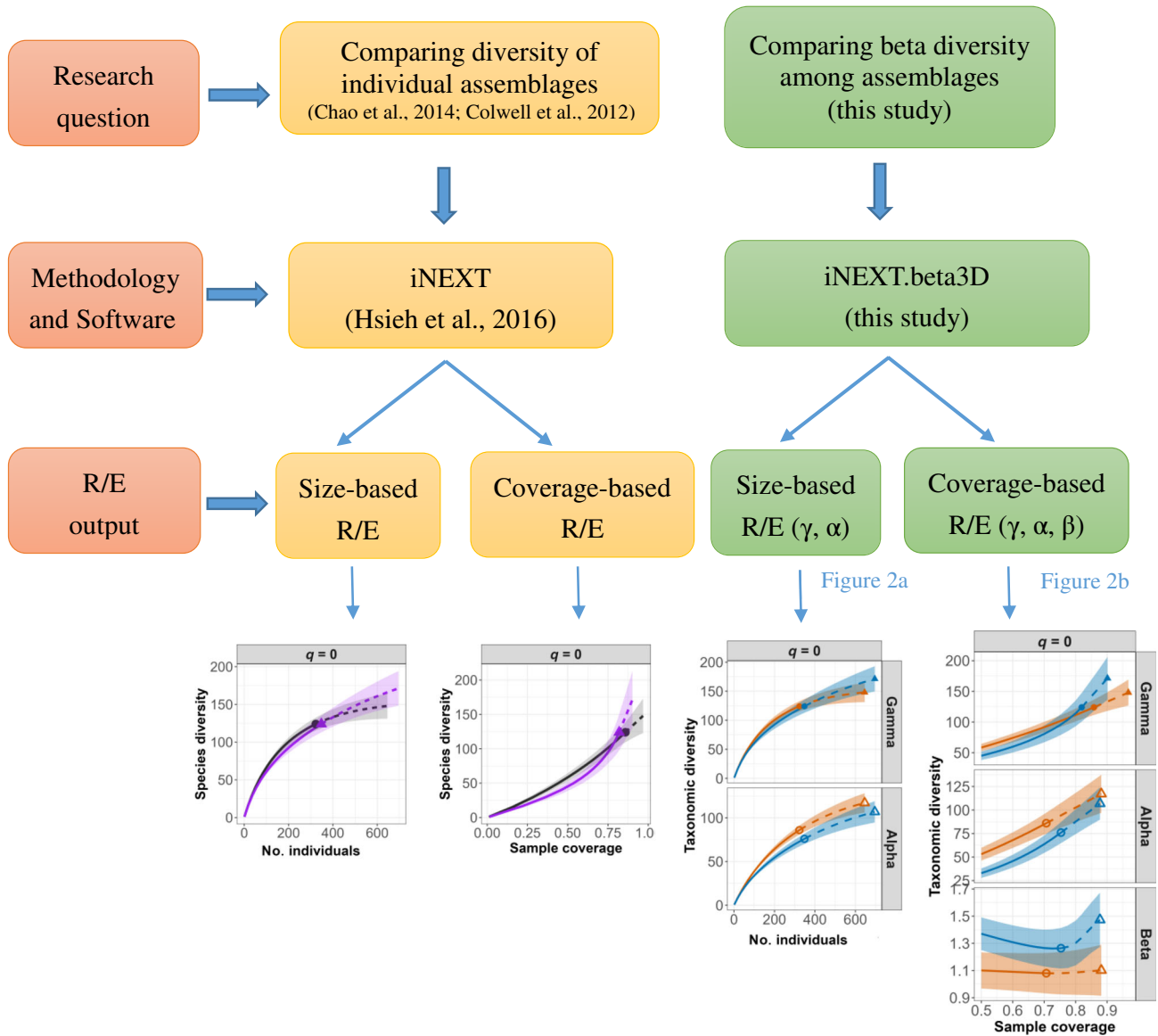
FIGURE 1 A schematic figure showing research questions, methodology/software, and rarefaction/extrapolation (R/E) output. Here, we illustrate a research question, analyzed by iNEXT, to compare species/taxonomic diversity (Hill numbers) of two rainforest fragments in Brazil (Marim and Rebio2 fragments). In this paper we use iNEXT.beta3D to compare beta diversity between the Edge and Interior habitats within each of the two fragments. For a size-based R/E curve, all samples are standardized to a fixed sample size (number of individuals), whereas for a coverage-based R/E curve, all samples are standardized to a fixed level of sample coverage (an objective measure of sample completeness). The output at the bottom only shows the R/E curves for $q = 0$; the corresponding R/E curves for $q = 1$ and $q = 2$ can also be computed; see Figure 2 for complete curves. See the text for the definitions of gamma, alpha, and beta (γ, α, β) diversity for multiple assemblages.

class of species diversity measures that became known as Hill numbers. This class of measures is parameterized by a diversity order q, which determines the measures' sensitivity to species relative abundance. Hill numbers of orders $q = 0$, 1, and 2 unify three well established indices of biodiversity: species richness ($q = 0$), Shannon diversity ($q = 1$, the exponential of Shannon entropy) and Simpson diversity ($q = 2$, the inverse of the one-complement of the Gini–Simpson index). Hill numbers can be formulated based on the species abundance distribution (SAD) for abundance data or the species incidence distribution (SID) for replicated incidence

data (frequency of occurrence among sampling units). The mathematical formulas for Hill numbers for both individual-based abundance data and replicated incidence data are briefly reviewed later.

In biodiversity analyses based on incomplete sampling data, it is important to adopt a statistical sampling-model-based approach that treats data as a representative sample from an assemblage. The approach makes a clear distinction between the theoretical assemblage level (theoretical properties/parameters for the assemblage) and the sampling data level (empirical/observed statistics computed

from data). For example, the number of species observed in a survey represents a statistic at the data level, and the number of species in the assemblage from which the data were taken represents the corresponding unknown parameter at the assemblage level; see Hurlbert (1971), Smith and Grassle (1977), Colwell and Coddington (1994), Lande (1996) and Colwell et al. (2012) for examples of distinguishing the two levels. This distinction enables researchers to transparently see the estimation target at the assemblage level, while pinpointing model assumptions at the data level. A useful ecological index should have an interpretable meaning at the assemblage level and good statistical properties at the data level. Statistical sampling models link the two levels and can be applied to make inferences about the assemblage based on sampling data (Colwell et al., 2012).

When individuals are distributed in space, an assemblage is characterized by the SAD and how all the individuals are distributed over space (spatial aggregation or clustering). At the assemblage level, species diversity (Hill numbers) is formulated to reflect the SAD information. At the data level, the observed species diversity depends not only on the two unknown assemblage-level properties (SAD and aggregation), but also on sampling effort/completeness. It is generally difficult to disentangle the SAD effect from any diversity measure at the data level, mainly because spatial aggregation is a complicating factor (Fortin et al., 2012). A statistical strategy involves (1) conducting an independent sampling to obviate the need to control for spatial aggregation (see below for reasons), and then (2) applying rarefaction and extrapolation to standardize or control for sampling effort/completeness. Classic rarefaction with species richness (Gotelli & Colwell, 2011) and the iNEXT standardization via interpolation/rarefaction and extrapolation with Hill numbers (Chao et al., 2014; Colwell et al., 2012; Hsieh et al., 2016) represent two methodologies based on this strategy. In the iNEXT standardization, all samples are standardized to a fixed sample size or a fixed level of sample coverage (or simply "coverage," an objective measure of sample completeness); more details on sample coverage are provided in the later review of the iNEXT standardization.

A basic assumption for classic rarefaction and the iNEXT standardization is that individuals (or other sampling units) are independently selected from the assemblage, that is, the outcome of any particular sampled individual does not affect or reveal any information about the outcome of the other sampled individuals; the sample is thus referred to as an independent sample. Such independent sampling can be done by randomly sampling mobile organisms, or by sampling sedentary organisms (e.g., plants) in small quadrats or along narrow transects. With independent samples, statistical sampling theory implies that observed diversity,

and both the size-based and coverage-based standardized diversity from the iNEXT method, are not influenced by the degree of spatial aggregation or interspecific association in the assemblage. That is, individuals in the entire assemblage could be highly aggregated without affecting standardized diversity at the data level, thus obviating the need to control for aggregation effects in sample standardization. Consequently, the difference in the standardized diversity estimate reflects the pure difference due to SADs. We use simulation results (in Appendix S1) to numerically demonstrate the above statistical sampling theory and theoretically prove its validity under a negative binomial distribution, a widely used model for spatial aggregation.

Note that another assumption often made for classic rarefaction is that individuals should be randomly placed in the assemblage (i.e., the assumption of random dispersion pattern or random placement of individuals). Under independent sampling, this unrealistic assemblage-level assumption of random dispersion is unnecessary and superfluous (Smith & Grassle, 1977; Tipper, 1979). Random dispersion is an assemblage-level property that is unknown and uncontrolled by samplers, whereas the sampling method and design are controlled by samplers.

One may argue that collecting truly independent samples at the data level is infeasible in most ecological surveys. Nevertheless, the iNEXT method is quite robust to departures from fully independent sampling. We provide in Appendix S1 some simulation results to numerically demonstrate that the iNEXT method is valid when sampled individuals are dependent to some extent. When the dependence among individuals in data is strong and thus cannot be ignored (e.g., counting individuals of plant species in relatively large quadrats or plots), the recommended iNEXT approach is to convert abundance to incidence data, in appropriate sampling units; incidence data in sampling units are generally only weakly dependent on, and therefore much less sensitive to aggregation than the corresponding abundance data. Incidence data in the resulting sampling units can be deemed as nearly independent, and thus the use of the incidence-based iNEXT standardization is justified. Such a sampling strategy was first proposed by Shinozaki (1963), and theoretically proved by Smith et al. (1985) and by Chao et al. (2014, their appendix A). Colwell et al. (2004, 2012) and Chao and Colwell (2017) demonstrated that replicated incidence data support statistical approaches to biological inference that are as powerful as the corresponding abundance-data-based approaches.

Beta diversity for N assemblages

Our goal in this paper is to assess beta diversity for quantifying biodiversity change and species compositional

differentiation over a geographical area, across time, or along an environmental gradient. A wide range of beta diversity concepts has been proposed at the assemblage level; this list seems endless (e.g., see Anderson et al., 2011; Chao & Chiu, 2016; Tuomisto, 2010). In this paper, we apply Whittaker's original richness-based multiplicative decomposition scheme, but use Hill numbers for any diversity order $q \geq 0$. Under this general framework, beta diversity or species compositional differentiation refers to the change/turnover of species identity/abundance among assemblages. Richness-based beta diversity ($q = 0$) quantifies the extent of species identity shift, whereas abundance-based ($q > 0$) beta diversity also quantifies the extent of change in the species pattern of commonness and rarity. To extend the iNEXT standardization to beta diversity, we briefly review in later sections the theoretical formulas for gamma, alpha, and beta diversity as well as three essential properties required for beta diversity.

To simplify the presentation and discussion, unless otherwise stated, we will mainly focus on spatial beta diversity among N assemblages in a region. However, all arguments and derivations can be directly applied to temporal beta diversity. We adopt a statistical-sampling-based model and assume that individuals (or other sampling units) are selected independently from the pooled assemblage. We refer to the SAD in the pooled assemblage as the regional SAD. At the assemblage level, beta diversity is characterized by the species-by-assemblage abundance matrix, or equivalently, the regional SAD and spatial aggregation of individuals over the N assemblages. Thus, beta diversity reflects the interacting effect of the regional SAD and spatial aggregation.

Under independent sampling, observed beta diversity (as the ratio of observed gamma and alpha diversity) depends on among-assemblage differentiation and sampling effort/completeness, which in turn induces dependence of beta on alpha and gamma diversity (Chase et al., 2011; Kraft et al., 2011). How to remove the dependence of richness-based beta diversity on its gamma component (species pool) has been intensely debated; see the next subsection. To compare beta diversity meaningfully across datasets, our approach described later is to standardize sample coverage to control for sample completeness and prove that the resulting standardized beta diversity can remove alpha and gamma dependence. Then the resulting standardized beta diversity can purely reflect the among-assemblage differentiation.

Some previous approaches

Chase et al. (2011) proposed the use of null models to control for the alpha dependence of beta diversity. Kraft et al. (2011) applied a different individual-based null-model approach to control for the species pool (gamma diversity of order $q = 0$) dependence of beta diversity and proposed the use of a statistic called beta deviation to compare species compositional differences among sites along latitudinal and elevational gradients. Although the null-model approach has been applied by ecologists (e.g., Mori et al., 2015; Myers et al., 2013; Xing & He, 2019), its validity and usefulness for measuring species compositional change have been questioned (Bennett & Gilbert, 2016; Qian et al., 2013; Ulrich et al., 2017, 2018). The issue of which methods can be used to remove the dependence on alpha and gamma has sparked intense debate and stimulated new developments in the ecological literature (Chase et al., 2011, 2018; Engel et al., 2021; Hillebrand et al., 2018; McGlinn et al., 2019; Xing & He, 2021 among others).

A research issue closely related to beta diversity is the assessment of the SAD effect and spatial-aggregation effect based on rarefaction curves. McGlinn et al. (2019, 2021) compared three rarefaction curves (spatial sample-based, nonspatial sample-based, and individual-based) to assess the effects attributed to: (1) the SAD, (2) the total abundance, and (3) spatial aggregation. All three rarefaction curves are computed from a single spatially explicit dataset. Engel et al. (2021) applied a coverage-based rarefaction method with beta diversity to measure the degree of spatial aggregation. However, the arguments of Ulrich et al. (2017) imply that it is impossible to disentangle SAD and spatial-aggregation effects using the species-by-assemblage matrix from a single dataset. Some of the previous approaches will be briefly compared in the *Discussion* section, along with our perspectives on how to disentangle the spatial-aggregation effect.

Very few previous approaches are based on statistical sampling models; the targets of estimation and the data are thus not distinguished. In addition, these previous works mainly focus on species richness-based beta diversity, which only quantifies the extent of a species identity shift. For beta diversity based on the framework of Hill numbers of any order $q \geq 0$, a statistical sampling-model-based standardization method that fulfills some essential requirements (including independence of beta diversity on alpha and gamma) is still lacking. The seemingly endless debates and divergent opinions mentioned above suggest that a consensus on how to compare beta diversity across datasets has not yet been achieved.

The proposed iNEXT.beta3D approach

Here, we generalize the iNEXT method to iNEXT.beta3D standardization via rarefaction and extrapolation with Hill-number-based beta diversity. We first adapt the iNEXT standardization to alpha and gamma diversities.

Then we propose the use of coverage-based standardization for beta diversity by assessing alpha and gamma diversity at the same standardized level of sample coverage, to formulate standardized, coverage-based beta diversity. After standardizing sample coverage to control for sampling effects, the resulting beta diversity provides a statistical removal of the dependence of beta diversity on both gamma and alpha values, and thus purely reflects among-assemblage differentiation.

This extension from iNEXT to beta diversity requires novel concepts and theories. The novel aspects, in addition to the use of a sampling model, include the following: (1) Extending species-richness-based beta diversity to Hill-number-based beta diversity and thus allowing the extent of change in species abundance to be assessed. (2) Using a recently developed alpha diversity so that a mathematically tractable rarefaction and extrapolation of beta diversity can be developed; traditional abundance-sensitive alpha diversity does not work. (3) Proving the coverage-based standardized beta diversity satisfies all essential properties of a differentiation measure including the independence of beta diversity on alpha and gamma diversity. The iNEXT.beta3D provides a rigorous and meaningful approach to comparing beta diversity across datasets.

The paper is organized as follows. We first present our model assumptions and theoretical methodologies. Readers interested in applications may skip this material and move to the application sections, where the proposed standardization is illustrated with spatial, temporal and spatiotemporal datasets. In the applications, we also specifically use a dataset to illustrate how to analyze data characterized by intraspecific aggregation. The freeware iNEXT.beta3D (an expansion of iNEXT; Chao, 2023) has been developed to facilitate all computations and graphics for the proposed methodologies (see *Data availability statement*).

A SINGLE ASSEMBLAGE: THE iNEXT STANDARDIZATION

In this section, we briefly review Hill numbers and their replication principle property at the assemblage level, followed by the iNEXT standardization and the replication principle at the data level; see Part (a) of Table 1 for a summary.

Theoretical formulas/properties at the assemblage level

Assume that there are S species in an assemblage, indexed by $i = 1, 2, ..., S$. Let z_i represent the true raw or absolute abundance (number of individuals) of species i,

or other metrics such as biomass or spatial coverage of corals. The total abundance in the assemblage, or assemblage size, is expressed as $z_+ = \sum_{i=1}^{S} z_i$. The SAD is characterized by the species abundance vector $(z_1, z_2, ..., z_S)$. The relative abundance of species i in the assemblage is defined by $p_i = z_i/z_+$ so that $\sum_{i=1}^{S} p_i = 1$. The Hill number of order q is defined as the following function in terms of species richness and relative abundances (Hill, 1973):

$$^qD = \left(\sum_{i=1}^{S} \left(\frac{z_i}{z_+} \right)^q \right)^{1/(1-q)}$$

$$= \left(\sum_{i=1}^{S} p_i^q \right)^{1/(1-q)}, \; q \geq 0, \; q \neq 1, \qquad (1)$$

which is interpreted as the effective number of species. Hill numbers encompass all information in the SAD.

The parameter q determines the sensitivity of the measure to the relative abundance of species. When $q = 0$, ^0D is simply species richness, which counts species equally without regard to their relative abundances. When $q = 1$, Equation (1) is undefined, but its limit as q approaches 1 is the exponential of Shannon entropy. The measure ^1D (Shannon diversity) can be interpreted as the effective number of common/abundant species. The measure for $q = 1$ counts individuals equally and weighs each species in proportion to its abundance. The Hill number $q = 2$ reduces to the reciprocal of Simpson concentration index, and disproportionately favors individuals of abundant species; the measure ^2D (Simpson diversity) is thus interpreted as the effective number of very abundant species.

Hill numbers fulfill an important property called the replication principle, or equivalently, the doubling property (Hill, 1973). Assume that Assemblage II consists of K "copies" of Assemblage I. Each copy has the same number of species and the same species abundances as Assemblage I, but with completely different species, unique to each copy. The replication principle states that the diversity of Assemblage II will be K times the diversity of Assemblage I. Shannon entropy (in units of information) and the Gini–Simpson index (i.e., the probability that two randomly selected individuals belong to different species) have been widely used in various disciplines. However, neither of these two classic abundance-sensitive indices satisfies the replication principle, a deficiency that causes inconsistent or counterintuitive interpretations (Jost, 2007). When the two indices are converted to Hill numbers, all transformed diversities are in units of "species" equivalents and satisfy the replication principle, resolving the interpretational problems related to the original forms of these indices.

TABLE 1 A summary table of parameters at the assemblage level (first column), sample statistics at the data level (second column), and the iNEXT standardization (third column); S denotes the number of species, N denotes the number of assemblages, and R/E = rarefaction/extrapolation.

Assemblage level	Sampling data level	iNEXT standardization
a. A single assemblage		
The species abundance distribution (SAD) with true abundance vector: $(z_1, z_2, ..., z_S)$	Reference sample: observed species abundance vector $(X_1, X_2, ..., X_S)$; sample size $n = X_+ = \sum_{i=1}^{S} X_i$	Sample-size-based R/E curve: plot standardized diversity with respect to sample size. Coverage-based R/E curve: plot standardized diversity with respect to sample coverage
b. N assemblages		
True $S \times N$ (species $\times$ assemblage) abundance matrix: $[z_{ik}]$	Reference sample matrix: observed abundance matrix $[X_{ik}]$; sample size $n = X_{++}$	
Gamma diversity based on the pooled assemblage		
The regional SAD with true abundance vector $(z_{1+}, z_{2+}, ..., z_{S+})$	Gamma reference sample: observed abundance vector $(X_{1+}, X_{2+}, ..., X_{S+})$; sample size $n = X_{++}$	Gamma diversity R/E curve: applying the iNEXT formulas (in Table 2) to the gamma reference sample in the pooled assemblage
Alpha diversity based on the joint assemblage		
The joint SAD with tree abundance vector $(z_{11}, ..., z_{S1}, z_{12}, ..., z_{S2}, ..., z_{SN})$	Alpha reference sample: observed abundance vector $(X_{11}, ..., X_{S1}, X_{12}, ..., X_{S2}, ..., X_{SN})$; sample size $n = X_{++}$	Alpha diversity R/E curve: applying the iNEXT formulas (in Table 2) to the alpha reference sample in the joint assemblage
Beta diversity = gamma/alpha		
See Equation (5)	See Equation (6)	Beta diversity R/E curve: ratio of gamma and alpha; both gamma and alpha are assessed at the same standardized coverage

MacArthur was the first to convert the two complexity measures (Shannon entropy and the Gini–Simpson index) to the concept of an effective number of species, that is, the number of equally abundant species that would be needed to yield the same value of the diversity measure (MacArthur, 1965).

The formula in Equation (1) is based on species abundance data, for which the sample unit is an individual. In many applications, the sampling unit is often a trap, net, quadrat, plot, or timed survey, and only species' incidence or occurrence (detection and nondetection) in each sampling unit is recorded. Analyses based on incidence data are less sensitive to the effects of clustering or aggregation of individuals, compared with those based on abundance data (Colwell et al., 2004, 2012). For incidence-based occurrence data among multiple sampling units, Chao et al. (2014) defined the corresponding species diversity of order q as the Hill number based on species relative detection probabilities for any occurrence record. That is, the probability p_i in Equation (1) is replaced by the relative detection probability of species i in any sampling unit (i.e., the probability that any detected occurrence/incidence is classified as species i). Hill numbers for incidence data quantify the effective number of equally frequent species. For $q = 0$, this

measure reduces to species richness, and the measures of $q = 1$ and $q = 2$ can be respectively interpreted as the effective number of frequent and highly frequent species in the assemblage. All properties and inferences for abundance data can be readily extended to incidence data (Chao & Colwell, 2017).

The iNEXT standardization and model assumptions

Under the assemblage-level framework presented in the preceding subsection, assume a reference sample of n individuals is selected from the assemblage. Let $(X_1, X_2, ..., X_S)$ denote the corresponding sample abundance or frequency vector in the reference sample; only species with $X_i > 0$ are observed. Under independent sampling, the sample abundance vector follows a multinomial distribution with cell total n and cell probabilities $(p_1, p_2, ..., p_S)$, $n = X_+ = \sum_{i=1}^{S} X_i$. This model links the assemblage and the reference data, so that the expected diversity in rarefied and extrapolated can be evaluated.

As indicated in the *Introduction* section and verified by the simulations in Appendix S1, observed and standardized species diversity for an independent

reference sample depends on the SAD and sampling effort/completeness, but not on spatial aggregation. Therefore, the issue of controlling for spatial aggregation is obviated, and only sampling effort/completeness needs to be standardized or controlled for. To compare diversity across more than one assemblage, the iNEXT method was developed (Chao et al., 2014), through interpolation (rarefaction) and extrapolation with Hill numbers, to standardize samples by sample size or sample coverage, which is defined as the fraction of the individuals in the assemblage (including individuals of undetected species) belonging to species detected in the sample. The concept of sample coverage was originally

developed by Alan Turing in his cryptographic work during WWII (Good, 1953) and has been applied to many disciplines.

For a hypothetical sample of size $m \geq 1$, let $^qD(m)$ be the expected diversity (Hill numbers) that would be observed in the sample, and $C(m)$ be the expected sample coverage. For $q = 0, 1, 2$, and a general order q, the theoretical formulas for $^qD(m)$ and $C(m)$ for any size $m \geq 1$ are given in the first column of Table 2. See Chao et al. (2014) for details. In the iNEXT standardization based on the reference sample, the estimators of $^qD(m)$ and $C(m)$, denoted as $^q\widehat{D}(m)$ and $\widehat{C}(m)$ respectively, are provided for rarefied samples (with $m < n$) and

TABLE 2 The iNEXT standardization and estimation based on a reference sample with species abundance vector $(X_1, X_2, ..., X_S)$, taken from an assemblage with species abundance distribution characterized by species abundance vector $(z_1, z_2, ..., z_S)$ and relative abundance vector $(p_1, p_2, ..., p_S)$, where $p_i = z_i/z_+ = z_i/\sum_{k=1}^{S} z_k$.

Theoretical diversity formulas[a]	Interpolation estimator (for sample of size $m < n$)	Extrapolation estimator[b] (for sample size $n + m^* > n$)
$q = 0$ $$^0D(m) = \sum_{k=1}^{m} E[f_k(m)] = S - \sum_{i=1}^{S}(1-p_i)^m$$	$$^0\widehat{D}(m) = \sum_{k=1}^{m}\widehat{f}_k(m) = {}^0D_{obs} - \sum_{X_i>0}\frac{\binom{n-X_i}{m}}{\binom{n}{m}}$$	$$^0\widehat{D}(n+m^*) = {}^0D_{obs} + \left[{}^0\widehat{D}(\infty) - {}^0D_{obs}\right]$$ $$\times \left[1 - \left(1 - {}^0\widehat{\beta}\right)^{m^*}\right]$$ $$= {}^0D_{obs} + \widehat{f}_0\left[1 - \left(1 - \frac{f_1}{n\widehat{f}_0 + f_1}\right)^{m^*}\right]$$
$q = 1$ $$^1D(m) = \exp\left(\sum_{k=1}^{m}\left(-\frac{k}{m}\log\frac{k}{m}\right)\times E[f_k(m)]\right)$$	$$^1\widehat{D}(m) = \exp\left(\sum_{k=1}^{m}\left(-\frac{k}{m}\log\frac{k}{m}\right)\times\widehat{f}_k(m)\right)$$	$$^1\widehat{D}(n+m^*) = {}^1D_{obs} + \left[{}^1\widehat{D}(\infty) - {}^1D_{obs}\right]$$ $$\times\left[1 - \left(1 - {}^1\widehat{\beta}\right)^{m^*}\right]$$
$q = 2$ $$^2D(m) = \left[\sum_{k=1}^{m}\left(\frac{k}{m}\right)^2\times E[f_k(m)]\right]^{-1}$$	$$^2\widehat{D}(m) = \left[\sum_{k=1}^{m}\left(\frac{k}{m}\right)^2\times\widehat{f}_k(m)\right]^{-1}$$	$$^2\widehat{D}(n+m^*) = \left(\frac{1}{n+m^*} + \sum_{i=1}^{S}\frac{(n+m^*-1)}{n+m^*}\frac{X_i(X_i-1)}{n(n-1)}\right)^{-1}$$
General order q $$^qD(m) = \left(\sum_{k=1}^{m}\left(\frac{k}{m}\right)^q\times E[f_k(m)]\right)^{\frac{1}{1-q}}$$	$$^q\widehat{D}(m) = \left(\sum_{k=1}^{m}\left(\frac{k}{m}\right)^q\times\widehat{f}_k(m)\right)^{\frac{1}{1-q}}$$	$$^q\widehat{D}(n+m^*) = {}^qD_{obs} + \left[{}^q\widehat{D}(\infty) - {}^qD_{obs}\right]$$ $$\times\left[1 - \left(1 - {}^q\widehat{\beta}\right)^{m^*}\right]$$
Sample coverage $$C(m) = 1 - \sum_{i=1}^{S}p_i(1-p_i)^m$$	$$\widehat{C}(m) = 1 - \sum_{i=1}^{S}\frac{X_i}{n}\frac{\binom{n-X_i}{m}}{\binom{n-1}{m}}$$	$$\widehat{C}(n+m^*) = 1 - \frac{f_1}{n}\left[\frac{(n-1)f_1}{(n-1)f_1+f_2}\right]^{m^*+1}$$

Note: The table gives the theoretical formulas and analytical estimators for rarefaction and extrapolation of Hill numbers of orders $q = 0, 1, 2$, and any general order $q \geq 0$, given a reference sample of size n with observed Hill numbers $^qD_{obs}$ and observed sample coverage C_{obs}. The last row gives the theoretical formulas and analytical estimators of sample coverage for rarefied and extrapolated samples. The extrapolation formulas are guided by Chao and Jost's (2015) asymptotic diversity estimator $^q\widehat{D}(\infty)$. Here the original extrapolation formula for $q = 1$ in Chao et al. (2014) is replaced by an improved one based on a unified formula.

[a]$f_k(m)$ = the number of species represented by exactly k individuals in a hypothetical sample of size m taken from the assemblage. The corresponding count for the reference sample is denoted as $f_k = f_k(n)$. $\widehat{f}_0 = {}^0\widehat{D}(\infty) - {}^0D_{obs}$ denotes the estimated undetected species richness. See Chao et al. (2014) for the formulas of $E[f_k(m)]$ and its estimator $\widehat{f}_k(m)$.

[b]See Chao et al. (2014) for the formula of $^q\widehat{\beta}$.

extrapolated samples (with $m = n + m^* > n$); see the second and third columns of Table 2 for $q = 0$, 1, 2. For rarefied samples, diversity estimators are nearly unbiased; for extrapolated samples, the performance of diversity estimators depends on the order q, as detailed below. The asymptotic diversity estimators for Hill numbers of orders $q = 0$, 1, and 2 based on Chao and Jost (2015) are also provided in the iNEXT standardization. These asymptotic diversity estimates are used to guide the extrapolation formulas.

Based on the formulas in Table 2, the two types of rarefaction and extrapolation sampling curves produced through the iNEXT standardization are summarized below.

The traditional (sample-)size-based rarefaction/extrapolation curve

This type of curve depicts the estimated Hill numbers as a function of sample size. The size-based sampling curve for Hill numbers includes the rarefaction part (which plots $^q\widehat{D}(m)$ as a function of m, where $m < n$) and the extrapolation part (which plots $^q\widehat{D}(n + m^*)$ as a function of $n + m^* > n$). These two parts join smoothly at the point of the reference sample, and the confidence intervals (CIs) based on the bootstrap method also join smoothly. For the measure $q = 0$, the size can be extrapolated, at most, to double the reference sample size (Chao et al., 2014). For the measures with $q = 1$ and $q = 2$, if data are not sparse, the extrapolation can be reliably (for $q = 1$) or unbiasedly (for $q = 2$) extended to infinity to attain the estimated asymptote.

This size-based sampling curve can be used to visually determine whether data are sufficient to provide reliable asymptotic diversity estimates (Chao et al., 2020). In many applications, the curve becomes stable and levels off for Shannon diversity ($q = 1$) and Simpson diversity ($q = 2$). In these two cases, the bias of our asymptotic estimator is limited or negligible and thus the true diversity can be inferred from the sampling data. In contrast, the curve for species richness ($q = 0$) is typically still increasing at double the reference sample size, signifying that the asymptotic estimator represents only a lower bound.

The coverage-based rarefaction/extrapolation curve

This type of curve depicts the estimated Hill numbers as a function of sample coverage. When data do not contain sufficient information to infer the diversity of an entire assemblage, that is, when asymptotic estimators are subject to bias, we can nonetheless infer and compare diversity for a standardized fraction of the assemblage's

individuals. The coverage-based sampling curve includes rarefaction (which plots $^q\widehat{D}(m)$ with respect to $\widehat{C}(m)$) and extrapolation (which plots $^q\widehat{D}(n + m^*)$ with respect to $\widehat{C}(n + m^*)$), joining smoothly at the reference sample point. The CIs based on the bootstrap method also join smoothly.

For the rarefaction and extrapolation curves with $q = 1$ and $q = 2$, if data are not sparse, the extrapolation can often be extended to the coverage of 100% to attain the estimated asymptote. However, the curve for species richness can be extended only up to a maximum value (i.e., the coverage value of an extrapolated sample with twice the reference sample size).

Replication principle for coverage-based standardization

The replication principle for Hill numbers is an important property not only for diversity measures at the assemblage level but also for diversity estimators computed from sampling data. However, biodiversity sampling or survey data are subject to sampling uncertainty/errors. It is thus infeasible to directly examine the properties that a measure possesses based on sampling data; one must prove that the expected diversity measure fulfills the replication principle via statistical models (Chao & Jost, 2012).

A major difference between the size-based and coverage-based standardizations is that only the expected coverage-based standardized diversity estimator fulfills the replication principle. This property for species richness was proved by Chao & Jost (2012, their appendix A) and later proved for Hill numbers (Chao et al., 2014, their appendix D). Consider that Assemblage II consists of K "copies" of Assemblage I at the assemblage level, as described earlier. The replication principle implies the diversity of Assemblage II for any order q is K times more diverse than Assemblage I at the assemblage level. At the data level, we assume a random sample of m individuals is taken from Assemblage I. Under the multinomial model, the sample size needed in Assemblage II to attain the same expected sample coverage is larger by a factor of approximately K (i.e., Km). In other words, K-times sample size is required for a K-times-diverse assemblage. Then for any order $q \geq 0$, the expected diversity in Assemblage II for the sample with a common standardized level of coverage is approximately K times the expected diversity of Assemblage I (Chao et al., 2014; Chao & Jost, 2012), establishing the replication principle for the coverage-based method.

Chao and Jost (2012) used numerical examples to demonstrate that the traditional species richness estimates for

samples with a standardized size do not satisfy the essential replication principle. A consequence is that the magnitude of the difference in species richness among assemblages is much compressed (Chao & Jost, 2012, their table 1). In contrast, a species richness estimator based on rarefaction for standardized sample coverage satisfies the replication principle. Similar conclusion is also valid for abundance-sensitive ($q > 0$) Hill numbers (Chao et al., 2014). For this reason, Chao et al. (2020) and Chazdon et al. (2022) recommend the use of coverage-based rarefaction and extrapolation to meaningfully compare diversity across multiple assemblages, and the use of size-based curves to determine whether the asymptotic estimates are unbiased or subject to negative bias.

ASSEMBLAGE LEVEL: THEORY FOR N ASSEMBLAGES

In this section, we briefly review the concepts and formulas for diversity decomposition (gamma, alpha, and beta) needed to present our proposed methodology in the next section.

Gamma and alpha diversity

Assume that in the species pool of N assemblages, there are S species, indexed by 1, 2, ..., S. Let $[z_{ik}] \geq 0$ be an $S \times N$ (species-by-assemblage) raw abundance matrix, where z_{ik} represents species raw or absolute abundance (i.e., the true number of individuals or other abundance proxies) of the i-th species in the k-th assemblage, $i = 1, 2, ..., S$, $k = 1, 2, ..., N$. The total abundance in the matrix is denoted by $z_{++} = \sum_{k=1}^{N} \sum_{i=1}^{S} z_{ik}$.

Gamma diversity is defined as the diversity of the pooled assemblage in which the abundance of any species is obtained by directly pooling its abundances over the N assemblages. That is, the pooled assemblage comprises S species and the regional SAD is characterized by the abundances $(z_{1+}, z_{2+}, ..., z_{S+})$, where $z_{i+} = \sum_{k=1}^{N} z_{ik}$ denotes the total abundance of species i in the entire pooled assemblage; see Part (b) of Table 1 for a summary of the framework. Gamma diversity at the assemblage level is computed from the regional SAD in the pooled assemblage:

$$^{q}D_{\gamma} = \left(\sum_{i=1}^{S} \left(\frac{z_{i+}}{z_{++}} \right)^{q} \right)^{1/(1-q)} = \left(\sum_{i=1}^{S} p_{i+}^{q} \right)^{1/(1-q)}, q \neq 1,$$

(2)

which is interpreted as the effective number of species in the pooled assemblage.

While the formulation of gamma diversity is unequivocal, there have been at least three different formulations of alpha diversity proposed for N assemblages based on Hill numbers. In conventional formulations, alpha diversity is expressed as a weighted mean of the diversities of individual assemblages. For example, Routledge (1979) and Jost (2007) derived different types of assemblage weights under different criteria. Here we adopt the formulation of Chiu et al. (2014) for the reasons given below: (1) When $q = 0$, Chiu et al.'s (2014) alpha diversity leads to a beta formula that is identical to Whittaker's original richness-based definition. (2) The Chiu et al. (2014) formula avoids the problem that other measures of alpha diversity sometimes yield a beta value less one (i.e., gamma < alpha) or greater than N. (3) Chao and Chiu (2016) showed that only through this alpha formulation can a bridge be built between the two major approaches to beta diversity, that is, the variance framework (Legendre & De Cáceres, 2013) and the diversity decomposition approach (presented below). (4) Most importantly, only with Chiu et al.'s formula can a simple and mathematically tractable rarefaction and extrapolation with alpha diversity for all $q \geq 0$ be formulated; conventional alpha diversity does not work.

In Chiu et al.'s (2014) alpha diversity, each individual is associated with two classifications, namely species identity and assemblage affiliation. The $S \times N$ cells of the raw abundance matrix are treated as if each cell were a "species" in the framework of Hill numbers. Each cell represents a species–assemblage combination. We define a joint assemblage that consists of all species–assemblage combinations (as if they were "species"). Although all combinations are arranged in a two-dimensional species × assemblage matrix, for alpha diversity, they are treated as if they form a single assemblage, the joint assemblage. For simplicity, we refer to the joint abundance distribution of species–assemblage combinations as the joint SAD in the joint assemblage, although "species" refer to cells/combinations. In other words, the joint SAD is characterized by the "species" abundances $(z_{11}, z_{21}, ..., z_{S1}, z_{12}, z_{22}, ..., z_{S2}, ..., z_{SN})$. Some z_{ik} may be 0; these nonexisting combinations do not have any effect on diversity computations. The relative "species" abundance in the joint assemblage is defined as $p_{ik} = z_{ik}/z_{++}$. A numerical example is given in Appendix S2 to illustrate a joint assemblage along with the corresponding pooled assemblage.

Joint diversity, a term first proposed by Gregorius (2010), is defined as the effective number of species–assemblage combinations, that is, the Hill numbers for the joint assemblage based on the joint SAD:

$$
\begin{aligned}
{}^{q}D_{\text{joint}} &= \left\{ \sum_{i=1}^{S} \sum_{k=1}^{N} \left(\frac{z_{ik}}{z_{++}} \right)^{q} \right\}^{\frac{1}{1-q}} \\
&= \left\{ \sum_{i=1}^{S} \sum_{k=1}^{N} p_{ik}^{q} \right\}^{\frac{1}{1-q}}, \; q \neq 1. \quad (3)
\end{aligned}
$$

See Appendix S2 for an illustrative numerical example. Chiu et al. (2014) proposed the following alpha formula in terms of mean joint diversity:

$$
{}^{q}D_{\alpha} = \frac{{}^{q}D_{\text{joint}}}{N} = \frac{1}{N} \left\{ \sum_{i=1}^{S} \sum_{k=1}^{N} \left(\frac{z_{ik}}{z_{++}} \right)^{q} \right\}^{\frac{1}{1-q}}, \; q \neq 1. \quad (4)
$$

Chiu et al.'s alpha diversity quantifies the effective number of species–assemblage combinations per assemblage; see Gregorius (2020) for further justification of this alpha formula. Based on the gamma (Equation 2) and alpha formulas (Equation 4), beta diversity is defined as the ratio of gamma to alpha diversities:

$$
{}^{q}D_{\beta} = \frac{{}^{q}D_{\gamma}}{{}^{q}D_{\alpha}}, \; q \geq 0. \quad (5)
$$

Thus, beta diversity is computed from two SADs (i.e., the regional SAD for gamma and the joint SAD for alpha). It follows from Equations (2) and (4) that beta diversity is not influenced by the total abundance because only the species relative abundances of each SAD are involved. Beta diversity is interpreted as the effective number of assemblages based on the raw abundance matrix $[z_{ik}]$. A simple numerical example is given in Appendix S2 to demonstrate how to compute gamma, alpha, and beta diversity.

Three properties for beta diversity at the assemblage level

Researchers from different perspectives have proposed many criteria for a beta diversity and a differentiation measure (e.g., Barwell et al., 2015; Jost et al., 2011; Koleff et al., 2003; Legendre & De Cáceres, 2013; Wilson & Shmida, 1984). Chao and Chiu (2016) showed that beta diversity defined in Equation (5) satisfies all the essential properties listed in Legendre & De Cáceres (2013; their properties P1–P9). Among those properties, we focus only on the following three criteria that are needed to assure the independence of beta diversity on alpha and gamma at the assemblage level:

1. The fixed minimum criterion: If all N assemblages are identical in terms of species identity and raw abundance, beta diversity attains a fixed minimum value of one. That is, effectively, there is only one assemblage.

2. The fixed maximum criterion: When no species are shared among N assemblages (i.e., complete species turnover), beta diversity should attain a fixed maximum value of N. That is, effectively, N assemblages are needed to find out all species.

3. The replication invariance principle: When all species are replicated (by adding species new to each assemblage) $K \geq 2$ times, both gamma and alpha are increased, but beta diversity should remain unchanged. That is, based on an original species-by-assemblage ($S \times N$) abundance matrix, an expanded matrix ($KS \times N$) merging K copies of the original matrix should have the same beta diversity as the original one. A numerical example is given in Appendix S2 to explain the replication invariance principle. When $K \geq 2$ copies of an original species-by-assemblage abundance matrix are merged, the replication principle for Hill numbers implies that both gamma and alpha diversity of the merged matrix are K times the corresponding gamma and alpha of the original matrix. Thus, beta diversity (Equation 5) for all $q \geq 0$ satisfies the species replication invariance principle at the assemblage level; see the next section for the corresponding data-level properties.

Because beta diversity takes values in the range of $[1, N]$ for all $q \geq 0$, it can be monotonically transformed into four classes of dissimilarity measures in the range $[0, 1]$. The four classes of measures that encompass the classic Jaccard and Sørensen dissimilarity measures ($q = 0$) and the abundance-sensitive Horn and Morisita–Horn dissimilarity measures as special cases; see Chao, Chiu, Wu, et al. (2019, the middle column of their table 1) for a summary. However, the commonly used Bray–Curtis dissimilarity (Bray & Curtis, 1957) is not included. It is generally difficult to develop rarefaction/extrapolation and robust estimation for the Bray–Curtis dissimilarity. Thus, the use of empirical Bray–Curtis dissimilarity for analyzing incomplete sampling data is not statistically justified and may be subject to substantial bias.

DATA LEVEL: THE iNEXT.beta3D STANDARDIZATION

Under the framework described at the assemblage level, assume an independent sample of n individuals

is selected from a region consisting of N assemblages. A sample species-by-assemblage raw abundance matrix $[X_{ik}]$ (the reference sample matrix) was recorded, where X_{ik} denotes the observed raw abundance/frequency of the i-th species in the k-th assemblage, $i = 1, 2, ..., S, k = 1, 2, ..., N$. Only those species with row total abundance ≥ 1 (i.e., species that were each observed in at least one assemblage) are detected in the sample; those species with row total abundance $= 0$ in the sample remain undetected and therefore do not affect diversity computation. Let X_{i+} (total abundance of the i-th species in the sample) and X_{++} (total sample abundance) be defined in the same manner here as in the assemblage level. Replacing the true species abundances by the observed data in gamma, alpha, and beta formulas (Equations 2, 4 and 5), we obtain the corresponding observed or empirical versions; see Part (b) of Table 1 for the notational distinction between the assemblage and the data levels.

As indicated in the *Introduction*, at the data level, observed beta diversity (as the ratio of observed gamma and alpha diversity) under independent sampling depends on among-assemblage differentiation and sampling effort/completeness, as well as alpha and gamma diversity. To compare beta diversity across datasets, sample coverage should be controlled for by proper standardization. Below we first adapt the iNEXT method to gamma and alpha diversity, and to beta diversity. Then we prove that standardized beta diversity is independent of alpha and gamma diversity.

The iNEXT standardization for gamma and alpha diversity

Rarefaction and extrapolation for gamma diversity

Since gamma diversity is defined as the Hill numbers of the pooled assemblage, we can simply apply the iNEXT estimation to the observed data from the pooled assemblage. That is, consider the regional SAD characterized by the abundance vector $(z_{1+}, z_{2+}, ..., z_{S+})$ in the pooled assemblage, and apply the iNEXT standardization based on the observed abundance data $(X_{1+}, X_{2+}, ..., X_{S+}$; referred to as the gamma reference sample, with size $n = X_{++})$. Then the size-based and coverage-based rarefaction and extrapolation curves, along with the estimated asymptote for gamma diversity, can be derived. As with the single-assemblage case, the standardized gamma diversity purely reflects the regional SAD information, regardless of the degree of spatial aggregation.

Rarefaction and extrapolation for alpha diversity

Our alpha diversity (in Equation 4) is a mean joint diversity, where joint diversity is defined as the Hill numbers of the joint assemblage. Consider the joint SAD characterized by the abundance vector $(z_{11}, z_{21}, ..., z_{S1}, z_{12}, z_{22}, ..., z_{S2}, ..., z_{SN})$ in the joint assemblage. Then apply the iNEXT standardization to this single joint assemblage based on the data $(X_{11}, X_{21}, ..., X_{S1}, X_{12}, X_{22}, ..., X_{S2}, ..., X_{SN}$; referred to as the alpha reference sample, with size $n = X_{++})$. Then the size-based and coverage-based rarefaction and extrapolation curves, along with the estimated asymptote for alpha diversity, can be derived. The standardized alpha diversity purely reflects the joint SAD information, regardless of the degree of spatial aggregation. If we adopted the conventional definition of alpha diversity (i.e., the weighted mean of diversities of individual assemblages), then multiple assemblages would be involved in the standardization, which would result in the corresponding rarefaction and extrapolation of alpha diversity becoming mathematically intractable. Our approach is elegantly simplified because only one single joint assemblage is involved, so that the iNEXT standardization can be readily applied.

The iNEXT.beta3D standardization for beta diversity

For any specified coverage value C, both alpha and gamma diversity are first assessed at the same given level of sample coverage, as described in the preceding two subsections. The resulting gamma and alpha diversity estimates are denoted as ${}^{q}\widehat{D}_{\gamma}(C)$ and ${}^{q}\widehat{D}_{\alpha}(C)$, respectively. Let this common coverage value correspond to sample size $m_{\gamma}(C)$ in the pooled assemblage and sample size $m_{\alpha}(C)$ in the joint assemblage. These two sample sizes vary with coverage value C, but for notational simplicity, we suppress the use of C in the notation of $m_{\gamma}(C)$ and $m_{\alpha}(C)$ and simply use $m_{\gamma} \equiv m_{\gamma}(C)$ and $m_{\alpha} \equiv m_{\alpha}(C)$ assuming that C is the same for both sample sizes. The two sizes generally differ for a given common coverage. We define the standardized beta diversity at the coverage level of $C > 0$ as:

$$ {}^{q}\widehat{D}_{\beta}(C) = \frac{{}^{q}\widehat{D}_{\gamma}(C)}{{}^{q}\widehat{D}_{\alpha}(C)} = \frac{{}^{q}\widehat{D}_{\gamma}(m_{\gamma})}{{}^{q}\widehat{D}_{\alpha}(m_{\alpha})}. \tag{6} $$

A bootstrap method is used to approximate the uncertainties of both rarefied and extrapolated beta diversity estimates and to construct the associated CIs. For any fixed coverage value, the standardized beta diversity is

not affected by the total abundance in the assemblage. One may wonder whether we can formulate the corresponding size-based standardization. We show that standardization to a common size fails to satisfy essential requirements for a differentiation measure; see *Discussion* for details.

Our proposed coverage-based rarefaction and extrapolation curve of beta diversity depicts $^q\widehat{D}_\beta(C)$ as a function of sample coverage $C > 0$. As with the iNEXT standardization, for gamma and alpha diversity of orders $q = 1$ and $q = 2$, the coverage-based extrapolation generally can be reliably extended to asymptotes (i.e., coverage $= 1$) if data are not too sparse. Therefore, we can compare beta diversity obtained from Hill numbers of $q = 1$ and $q = 2$ across datasets at any standardized level of coverage up to 100%.

For richness-based ($q = 0$) gamma and alpha diversity, the extrapolation can often be extended only up to the coverage value for samples extrapolated to twice the reference sample size. We denote the sample coverage of the alpha and gamma observed (reference) sample as $C_{\text{obs},\alpha}$ and $C_{\text{obs},\gamma}$, respectively; the sample size for both gamma and alpha reference samples is $n = X_{++}$. Also, we denote the two coverage values for twice the reference sample size as $C_{2n,\alpha}$ and $C_{2n,\gamma}$. The four coverage values typically satisfy either $C_{\text{obs},\alpha} < C_{2n,\alpha} < C_{\text{obs},\gamma} < C_{2n,\gamma}$ or $C_{\text{obs},\alpha} < C_{\text{obs},\gamma} < C_{2n,\alpha} < C_{2n,\gamma}$. The performance of our standardized beta diversity for $q = 0$ depends on the standardized coverage value C:

1. When $C \le C_{\text{obs},\alpha}$, both $^q\widehat{D}_\gamma(C)$ and $^q\widehat{D}_\alpha(C)$ are obtained from rarefied samples. In this range, the standardized beta diversity $^q\widehat{D}_\beta(C)$ of $q = 0$ is nearly unbiased.
2. When $C_{\text{obs},\alpha} < C \le C_{2n,\alpha}$, $^q\widehat{D}_\alpha(C)$ is obtained from an extrapolated sample up to a size less than double the reference sample size, and $^q\widehat{D}_\gamma(C)$ is obtained from either a rarefied sample or an extrapolated sample up to a size less than double the reference sample size. In either case, the standardized beta diversity of $q = 0$ is generally reliable.
3. When $C_{2n,\alpha} < C$, extrapolation to sizes greater than double the reference sample size is required to obtain $^q\widehat{D}_\alpha(C)$. The corresponding standardized beta diversity of $q = 0$ thus may be subject to some bias.

We suggest that the sampling curve for beta diversity of $q = 0$ be extended up to the sample coverage value of $C_{2n,\alpha}$ to avoid substantial bias (i.e., the extrapolation limit for beta diversity is the same as alpha diversity). In other words, $C_{2n,\alpha}$ represents the maximum coverage value that we can reliably infer for beta diversity with $q = 0$ for a given dataset. We provide the rarefaction and extrapolation formulas for gamma and alpha diversity for $q = 0$, 1, and 2 in Appendix S2. All the formulas for $q = 0$,

1, and 2 are parallel to those given in Table 2. As with the assemblage level, the coverage-based rarefaction and extrapolation curves of beta diversity at the data level can be transformed into four classes of dissimilarity measures in [0, 1].

Properties of the proposed coverage-based standardized beta diversity

At the assemblage level, we highlighted three essential properties for our beta diversity measure. This section demonstrates that, at the data level (based on the observed species × abundance matrix [X_{ik}]), our expected standardized beta diversity for any fixed value of sample coverage value also satisfies the three properties.

The fixed minimum criterion is approximately fulfilled at the data level: When N assemblages are identical at the assemblage level, the joint assemblage represents N copies of the pooled assemblage. Therefore, at the data level, the sample size needed in the joint assemblage to attain the same expected sample coverage is approximately N times that in the pooled assemblage, that is, we have $m_\alpha = Nm_\gamma$ in Equation (6). The expected joint diversity (i.e., N times alpha diversity) is then approximately N times the expected gamma diversity, implying the expected standardized beta diversity approaches the minimum value of one for any coverage value.

The fixed maximum criterion is satisfied at the data level: When no species are shared among N assemblages at the assemblage level, the pooled assemblage is identical to the joint assemblage. The sample size needed in the pooled assemblage and in the joint assemblage should be the same to attain the same sample coverage. That is, we have $m_\gamma = m_\alpha$ for any given coverage value. Consequently, the expected gamma diversity is identical to the expected joint diversity, yielding a maximum value of N for the expected standardized beta diversity for any coverage value. In this case, any standardized beta based on the observed data also correctly yields the true maximum beta diversity value of N. Thus, under the special case of no shared species, the fixed maximum criterion is satisfied not only for the expected standardized beta diversity, but also for the empirical beta values.

The principle of species replication invariance is approximately valid at the data level. As described at the assemblage level, we assume that K copies of an original species-by-assemblage ($S \times N$) matrix are merged to obtain an expanded ($KS \times N$) matrix. All copies have identical abundance matrices (and thus identical beta values), but no species are shared among copies. Note that, at the assemblage level, the pooled assemblage in

the expanded matrix consists of K copies of the pooled assemblage of the original table, and the joint assemblage in the expanded table consists of K copies of the original joint assemblage. Assume that, in the original matrix, sample sizes m_γ (for gamma diversity) and m_α (for alpha diversity) are required to attain a common value of expected sample coverage. Building on the iNEXT theory, the sample sizes needed in the expanded matrix for gamma and alpha to attain the same expected sample coverage are approximately Km_γ and Km_α. Based on the replication principle for gamma and alpha diversity, the expected gamma (and alpha) diversity of any order $q \geq 0$ in the expanded matrix is approximately K times that of gamma (and alpha) in the original matrix. Consequently, the expected beta diversity with a common level of standardized coverage is approximately invariant when all species are replicated K times.

The above three properties for any coverage value demonstrate that our standardization, based on the same level of coverage at the data level, provides a statistical solution that removes the dependence of beta diversity on both gamma and alpha values. We further conducted some simulations to support our theory; see Appendix S3 for our simulation results.

APPLICATIONS TO SPATIAL DATA

In this section, we illustrate the proposed iNEXT.beta3D standardization with the tree species data collected by Magnago and colleagues between 2011 and 2012 from rainforest fragments in Espírito Santo State, Brazil (Magnago et al., 2014). The fragmentation of tropical rainforests is one of the major drivers of the loss of global biodiversity. Thus, it is important to understand how species composition changes between fragment edges and interiors. We selected 16 transects (eight Edge transects and the corresponding eight Interior transects) from fragments with area >100 ha. A summary of the observed data in the selected 16 transects is shown in Appendix S4: Table S1.

Each transect comprised ten 10×10 m plots at 20-m intervals. Each Edge transect was placed about 5 m inside the fragment and parallel to the forest edge, and the corresponding Interior transect was located at least 300 m from the nearest edge. Within each transect, every living tree with a diameter at breast height (DBH) >4.8 cm and 1.3 m height was recorded and identified to species. The goal here is to assess how beta diversity between the Edge habitat and the Interior habitat varies with fragment area/size. Because trees are surveyed in relatively small plots, the spatial dependence

between two individual trees is weak; the use of the iNEXT.beta3D standardization is justified.

The iNEXT.beta3D standardization for two fragments (Figure 2)

We first selected two fragments to illustrate the proposed iNEXT.beta3D standardization procedure within each fragment. We specifically selected a small fragment (Marim, 104.71 ha) and a large fragment (Rebio2, 20417.38 ha). One reason for selecting these two fragments was because they have the same number of species (124 species each) in the pooled Edge and Interior data, that is, the observed richness-based ($q = 0$) gamma is identical in the two fragments.

In the Marim fragment, there were 88 species ($n = 168$ individuals) in the Edge transect, and 84 species ($n = 155$) in the Interior transect. In the Rebio2 fragment, there were 75 species ($n = 173$ individuals) in the Edge transect, and 77 species ($n = 175$) in the Interior transect. Thus, the larger fragment has a lower average number of species (i.e., lower alpha richness) and thus a higher observed richness-based beta diversity than the smaller fragment.

As indicated in the *Introduction*, observed beta cannot be used for comparing the beta diversity of the two fragments because, within each fragment, the gamma and alpha reference samples have different coverage values; see the legend of Figure 2. In addition, the coverage values for the two gamma (and also alpha) reference samples differ between the two fragments. To make meaningful comparisons of beta diversity for the two fragments, sample coverage for both gamma and alpha should be standardized not only within the individual fragments but also between the fragments. The complete iNEXT.beta3D standardization comprises the following two procedures:

1. Assessment of the sample-size-based rarefaction and extrapolation curves for gamma and alpha diversity up to double the reference sample size (Figure 2a).

Figure 2a shows the size-based rarefaction and extrapolation curves for gamma and alpha diversity for each fragment up to double the reference sample size. These curves can be used to determine whether data contain sufficient information to infer true gamma and alpha diversity. For $q = 0$, neither the gamma nor the alpha sampling curve for either plot levels off, implying that there were many undetected rare species and sampling data do not contain sufficient information to accurately infer true gamma and alpha richness; fair
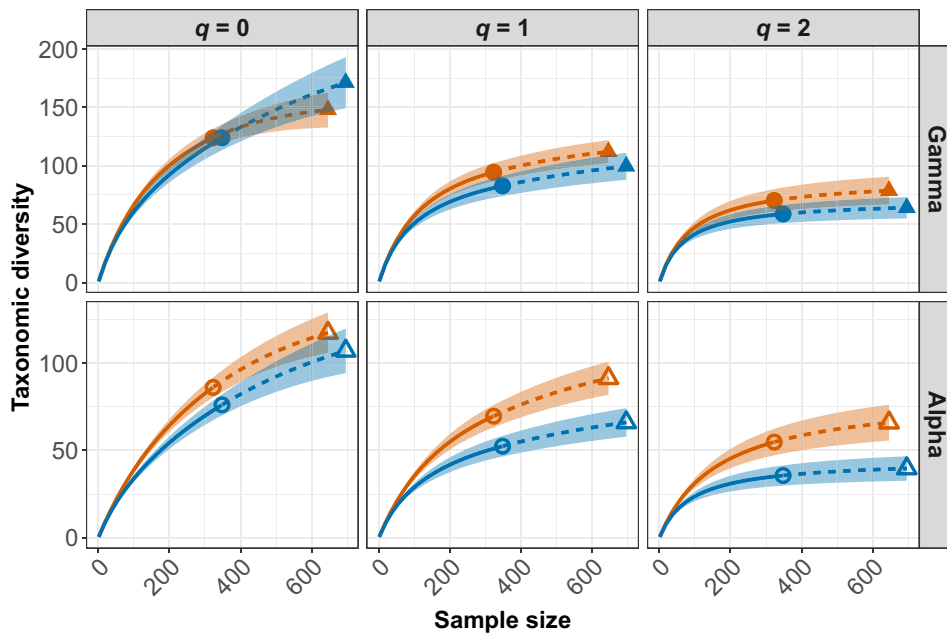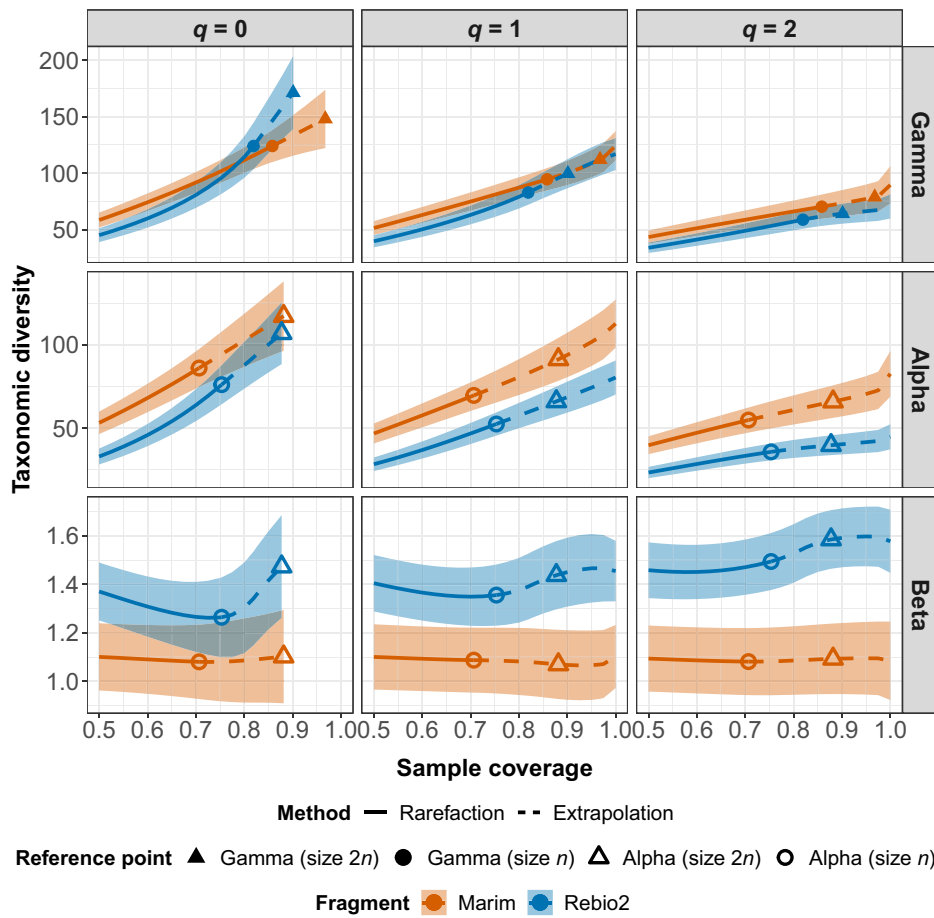
FIGURE 2 Legend on next page.

comparisons of richness-based beta diversity thus can only be made by standardizing the sample coverage up to a limited coverage value, as described later.

For Shannon-diversity-based beta ($q = 1$) and Simpson-diversity-based beta ($q = 2$), both the alpha and gamma sampling curves in Figure 2a tend to stabilize, implying that bias in our asymptotic gamma (and alpha) estimator is limited for $q = 1$ and 2. The resulting asymptotic beta diversity (for complete coverage) is compared in the next procedure, for coverage-based standardization. Note that in Figure 2a the plot for size-based beta diversity (in which both gamma and alpha are assessed at the same sample size) is not shown, because the size-based approaches for beta diversity lack theoretical justification and do not lead to legitimate differentiation measures; see *Discussion* for reasoning.

2. Assessment of coverage-based rarefaction and extrapolation curves for alpha, gamma, and beta diversity up to complete coverage (i.e., coverage = 1) for $q = 1$ and 2, and up to a limited coverage value for $q = 0$ (Figure 2b).

Figure 2b shows the coverage-based rarefaction and extrapolation curves for gamma, alpha, and beta diversity. For gamma and alpha diversities of $q = 1$ and $q = 2$, extrapolation can be extended to complete coverage. Thus, the corresponding beta diversity can be computed for any level of coverage up to 100%.

For $q = 0$, as shown in Figure 2a, data are not sufficient for extrapolation to extend to asymptotes. In this case, the extrapolation for gamma and alpha can be extended only to the coverage value corresponding to double each reference sample size. We suggest that the rarefaction and extrapolation curve for beta diversity be extended to the same limit as alpha diversity, which is $C_{2n,\alpha} = 88.16\%$ in the Marim fragment and 87.74% in the Rebio2 fragment. Therefore, extrapolation for beta diversity of $q = 0$ can be extended up to the lower coverage value of the two extrapolated samples, that is, only up to about 88%.

For any specified coverage value up to 88% (for $q = 0$) and to 100% (for $q = 1$ and 2), Figure 2b reveals that the two fragments have approximately the same standardized gamma diversity, whereas the standardized alpha diversity in the Marim fragment is significantly higher than that in the Rebio2 fragment, leading to a higher standardized beta diversity in the Rebio2 fragment. For $q = 1$ and 2, the difference in standardized beta diversity is statistically significant because the CIs for the two fragments do not overlap for any coverage value. These two fragments provide an example of the ranking of two fragments being different in alpha and beta diversity (i.e., the smaller fragment has higher alpha diversity, but the larger fragment has higher beta diversity).

We use the observed richness to intuitively explain why the larger fragment exhibits higher beta diversity between the Edge and Interior habitats. In the Marim fragment, there were 48 species shared by the Edge and Interior transects, but only 28 species shared in the Rebio2 fragment (Appendix S4: Table S1), despite the fragment totals being equal. The number of species in the Edge and Interior transects in the Marim fragment are, respectively, higher than in the corresponding Edge and Interior transects in the Rebio2 fragment. However, over half of the species in each transect in the Marim fragment were shared, while more than 60% of the species in each transect in the Rebio2 fragment were unique species. That is, a higher proportion of the unique species in the Edge transect in the Marim fragment were lost and replaced by shared species than in the Rebio2 fragment. Our analysis demonstrates that a similar conclusion is valid not only for the observed beta diversity but also for any standardized beta diversity, regardless of focus on rare species ($q = 0$), abundant species ($q = 1$) or very

FIGURE 2 iNEXT.beta3D standardization for beta diversity between the Edge and Interior habitats for two fragments. (a) Sample-size-based rarefaction (solid curves) and extrapolation (dashed curves) with gamma diversity and alpha diversity, and (b) coverage-based rarefaction (solid curves) and extrapolation (dashed curves) with gamma, alpha, and beta diversity based on the tree species of the Marim fragment (orange curves) and the Rebio2 fragment (blue curves) for diversity orders $q = 0$ (left panels), $q = 1$ (middle panels), and $q = 2$ (right panels). Here, size-based *beta* diversity is not a legitimate differentiation measure and thus is omitted; see *Discussion*. The two gamma reference samples are denoted by solid circles and alpha reference samples by hollow circles. The two extrapolated gamma and alpha samples with double the reference sample size ($2n$) are denoted, respectively, as solid triangles and hollow triangles. In (a), sample size in each plot is extrapolated up to $2n$. In (b), the coverage values in the Marim fragment for observed alpha, extrapolated alpha with size $2n$, observed gamma, and extrapolated gamma with size $2n$ are, respectively: $C_{obs,\alpha} = 70.67\%$, $C_{2n,\alpha} = 88.16\%$, $C_{obs,\gamma} = 85.82\%$, and $C_{2n,\gamma} = 96.77\%$; in the Rebio2 fragment, the corresponding four coverage values are: $C_{obs,\alpha} = 75.34\%$, $C_{2n,\alpha} = 87.74\%$, $C_{obs,\gamma} = 81.93\%$, and $C_{2n,\gamma} = 90.13\%$. In (b), for $q = 1$ and $q = 2$, the extrapolation for alpha, beta, and gamma can be extended to their asymptotes (i.e., coverage value of 100%). For $q = 0$, the extrapolation for gamma and alpha can be extended to the coverage value corresponding to size $2n$; the extrapolation limit for beta diversity is the same as alpha diversity. All 95% CIs (shaded areas) were obtained by a bootstrap method based on 200 replications. Comparisons of beta diversity between the two habitats should be based on a standardized value of sample coverage.

abundant ($q = 2$) species. Detailed biological explanations and interpretations are given in the next subsection.

Beta diversity increases with fragment size

We performed the same iNEXT.beta3D standardization procedures as those presented in the preceding section for the other six fragments. For each of the other six fragments, we obtained coverage-based rarefaction and extrapolation curves for alpha, gamma, and beta diversity, as we did for the Marim and Rebio2 fragments as shown in Figure 2b. Since it is not feasible to simultaneously compare eight curves, we extracted standardized beta diversity for six levels of standardized coverage values (60%, 70%, 80%, 90%, 95%, 100%); here standardized diversity corresponding to a coverage value of 100% represents the asymptotic estimate, which may be subject to some bias for $q = 0$. Figure 3 shows the beta diversity of the order $q = 0$, 1 and 2 for the six selected coverage values (rows 2 to 7), along with the observed patterns (row 1).

The plots and fits based on the iNEXT.beta3D standardization in Figure 3 revealed that the standardized beta diversity between the Edge and Interior habitats increases with fragment size for any standardized coverage level up to 100%, regardless of species rareness or dominance. Nearly all the fits based on standardized beta values are significant. A similar increasing (but not significant) pattern is also valid for the observed beta diversity. In contrast, the corresponding gamma and alpha diversity tend to decline slowly and nonsignificantly with fragment size (Appendix S4: Figure S1). The rate of decline is steeper for alpha diversity, leading to an increasing trend for beta diversity.

Edge effects transform forest microclimate conditions, reducing the air humidity while increasing the air temperature and light intensity (Magnago et al., 2015, 2017). These effects usually have an edge penetration of ~100 m, but can reach more than 300 m for some variables, such as wind increase (Laurance et al., 2002, 2006). Because the distance from forest edges to their interiors in smaller fragments is relatively shorter, the environmental conditions in the Edge and Interior habitats become more similar (Magnago et al., 2017). This homogenization in microclimate is one major cause of lower beta diversity in smaller fragments.

In addition, as also revealed in the preceding subsection, fragment edges lost (between Edge and Interior) some unique species that were generally replaced by shared species, and this loss/replacement rate is higher in smaller fragments. Due to changes in the forest microclimate, edge habitats and small fragments are rapidly colonized by pioneer tree species (high light demand species; Laurance et al., 2006); these species are considered generalists and capable of easily colonizing and dispersing throughout agricultural matrices (Magnago et al., 2014). Thus, in fragment edges and small fragments, the loss/replacement of unique tree species by shared tree species was probably caused by the severe conditions created by edge habitats. These edge conditions could select from a small pool of species that are more adapted to intense light conditions, higher temperatures, and less humidity. Therefore, our results were as expected in accordance with niche theory, as applied to fragmentation studies (Püttker et al., 2015). That is, reduced environmental differentiation promoted by fragmentation (see results of microclimate homogenization for the fragments studied by Magnago et al., 2017) could reduce the beta diversity between Edge and Interior in small fragments, where colonized trees are selected from a small species pool.

APPLICATIONS TO TEMPORAL DATA

We first illustrate the iNEXT.beta3D procedures with beetle abundance data for two time periods, followed by an application to tree abundance data for long-term time series.

Temporal beta diversity for two time periods

To illustrate iNEXT.beta3D procedures for temporal data, we use saproxylic beetle data collected from 2008 to 2011 in a large experiment on the effects of salvage logging on biodiversity (Thorn et al., 2014, 2016). The underlying data are available in the BioTIME database (Dornelas et al., 2018). Saproxylic beetles refer to those that depend on the fungal decay of wood during some part of their life cycle (Alexander, 2008). The experiment was conducted in the Bavarian Forest National Park in southeastern Germany, a forested area subjected to different kinds of natural disturbances (Thorn et al., 2017). In January 2007, a mature Norway spruce forest of ~1000 ha was felled by the windstorm Kyrill. From those 1000 ha, an area of roughly 200 ha remained unlogged whereas storm-felled trees in the remaining 800 ha were removed by experimental salvage logging. Afterward, 44 permanent plots (22 in logged and 22 in unlogged areas) were established. Saproxylic beetles were trapped using 44 flight-interception traps, one at the centroid of each plot. Traps were emptied monthly and abundance data were pooled over plots within each area and year. All beetles were identified to the species
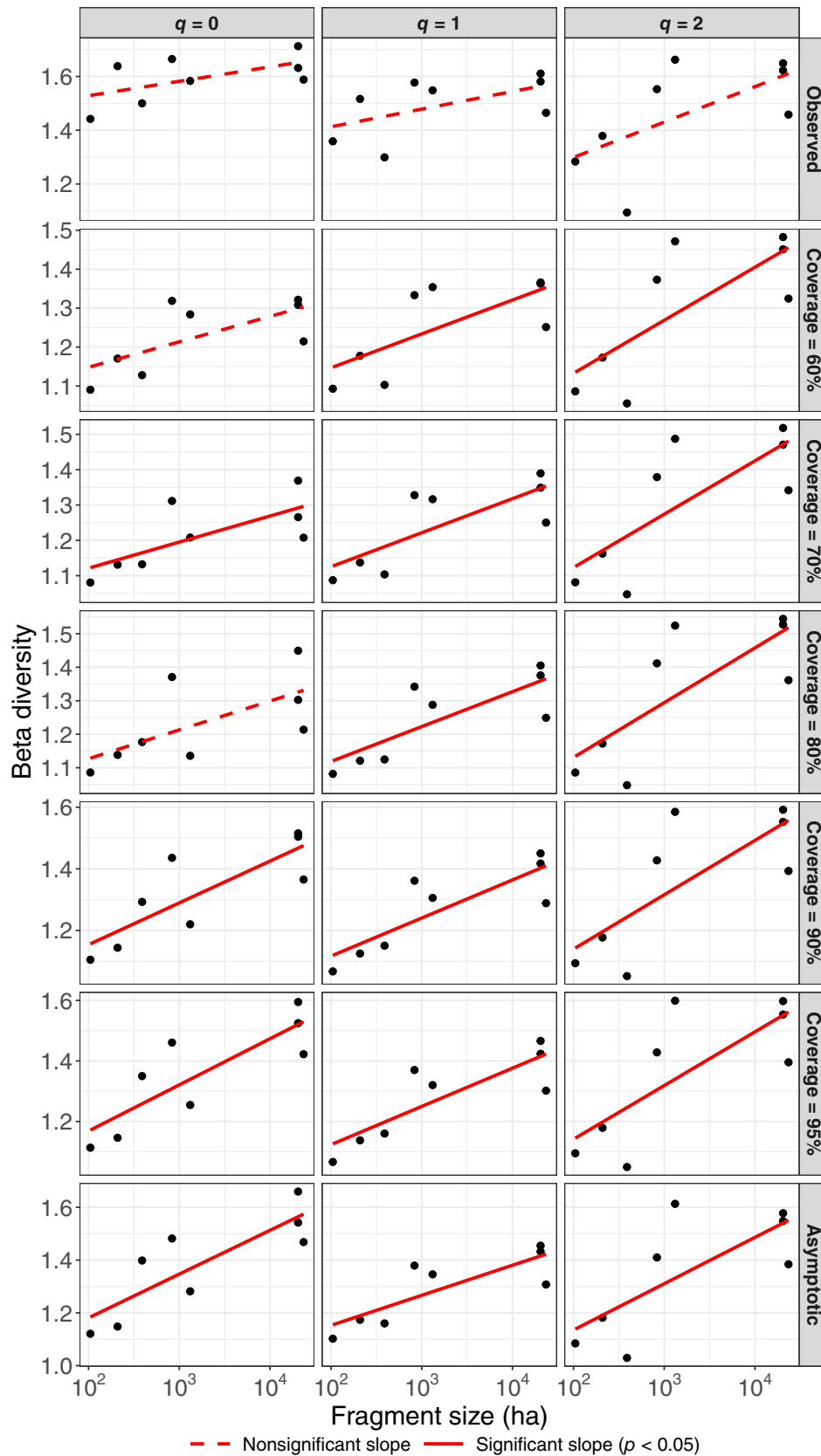
FIGURE 3 Beta diversity between the Edge and Interior habitats as a function of fragment size based on data from eight fragments. Row 1 shows observed beta diversity, and rows 2 to 7 show the standardized beta diversity under six coverage levels: 60%, 70%, 80%, 90%, 95%, and 100% (asymptotic) for order $q = 0$ (column 1), $q = 1$ (column 2) and $q = 2$ (column 3). The fitted curves were made using a linear trend; a solid red line fit indicates significance ($p < 5\%$) and a dotted line indicates nonsignificance.

level; see Thorn et al. (2014) for sampling details. We assume that the data collected in each of the 44 plots are representative of the entire assemblage in that plot.

During the first stages of wood decay following natural disturbances, the abundances of a few bark beetles, the so-called "pest species," such as the European spruce bark beetle, *Ips typographus* and associated species, increase within a short time period (Wermelinger, 2002). This peak is particularly significant in unlogged, naturally disturbed plots because of the larger amount of deadwood (Thorn et al., 2014). Thus, the large temporal variation in pest species abundance was observed during the first 2 years (2008 to 2009). To obtain more stable diversity patterns, we pooled the data from 2008 and 2009 (referred to as the first time period) and the data from 2010 and 2011 (referred to as the second time period); see Appendix S5: Table S1 for some summary statistics. Our goal was to assess beta diversity between the two time periods separately for logged and unlogged areas. Summary statistics and some detailed numerical values for observed, asymptotic, and standardized gamma, alpha, and beta diversity of $q = 0$, 1, and 2 are provided in Appendix S5: Table S2. Since all the iNEXT.beta3D procedures for temporal data with two time periods are parallel to those for spatial beta diversity with two localities, we omit the description of most procedures; see Appendix S5 for details.

Figure 4a shows the size-based rarefaction and extrapolation for the gamma and alpha diversity for $q = 0$, 1, and 2 in each area. The reference sample size, that is, the total number of individuals observed over the two time periods, is much higher in the unlogged area compared with the logged area, due to the larger amount of deadwood. However, contrary to most studies, this larger sample size led to lower observed gamma (and alpha) diversity. Consequently, for any standardized sample size, gamma (and alpha) diversity of the unlogged area is lower than the logged area, regardless of the diversity order (Figure 4a). Here, the size-based beta diversity is not a legitimate differentiation measure and thus the corresponding plot is not shown.

As in most studies, gamma and alpha rarefaction and extrapolation sampling curves for $q = 0$ for each area do not level off, whereas the curves for $q = 1$ and $q = 2$ stabilize. This pattern implies that we can compare alpha, beta, and gamma diversity of $q = 1$ and 2 for any standardized coverage up to their asymptotes; for $q = 0$, their respective asymptotic values cannot be accurately estimated, and comparison can only be made for any standardized coverage up to a maximum, as shown below.

Figure 4b shows the coverage-based gamma, alpha, and beta rarefaction and extrapolation curves separately for logged and unlogged areas. For gamma (and alpha), the observed sample coverage value and the corresponding coverage value for the extrapolated sample with twice the reference sample size are all very close to 100% (see Appendix S5: Table S1). Because the increase in coverage for the extrapolation is small, the estimated diversity hardly changes beyond the reference samples, so the extrapolation portions in Figure 4b are nearly invisible for all diversity orders.

If we focus on abundant ($q = 1$) and very abundant ($q = 2$) species for gamma and alpha, the nonoverlapping CIs in Figure 4b reveal that the logged area has significantly higher alpha and gamma diversity, but the unlogged area has significantly higher beta diversity. These orderings are valid for any standardized sample coverage up to 100% (complete coverage). For richness-based ($q = 0$) measures, a similar conclusion is valid but only up to a coverage value of about 95% (row 3, column 1 in Figure 4b), due to insufficient information about extremely rare species.

Within the logged area, Figure 4b also reveals that beta diversity stays at a constant level with order q, and all beta values differ from one only to a limited extent. This result signifies that beta diversity over time in the logged area is low and roughly the same for rare, common, and very abundant species. However, beta diversity within the unlogged area is increasing with increasing order q, that is, beta diversity is relatively high for very abundant species, moderate for common/abundant species, and relatively low for rare species. In other words, the change in species abundance in the unlogged area was the main driver for compositional change between the two time periods.

For our beetle data, the logged area generally has higher alpha and gamma diversity than the unlogged area for any standardized sample coverage, but the ordering is reversed for beta diversity. This pattern may be caused by a comparably low amount of deadwood in logged areas, such as fine woody debris, logging residuals etc., which are quickly decomposing. Available deadwood additionally develops into moist and shady microclimates; these changes support higher gamma and alpha diversity but limit beta diversity over time. In unlogged areas, in contrast, a large amount of deadwood (in the form of trunks, large stumps, and branches) remains much longer. This longer persistence of deadwood tends to facilitate species turnover over time, especially for more abundant species. Despite the substantial difference in the amount of deadwood between the logged and unlogged areas, our iNEXT.beta3D method provides robust and statistically justified standardization comparisons of temporal beta diversity for the two areas.

(a) Sample-size-based rarefaction and extrapolation for gamma and alpha diversity (2008−2009 vs. 2010−2011)
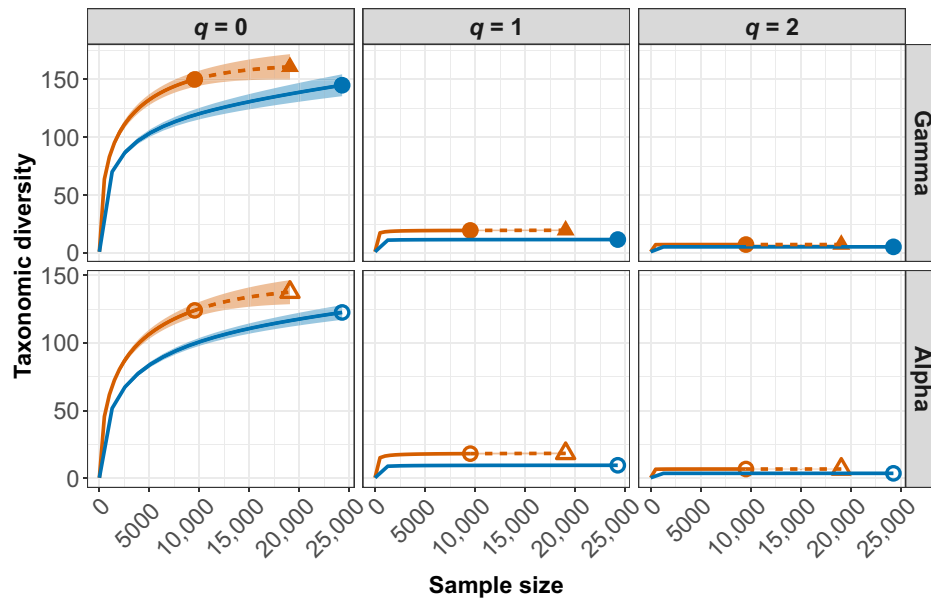
(b) Coverage-based rarefaction and extrapolation for gamma, alpha, and beta diversity (2008−2009 vs. 2010−2011)
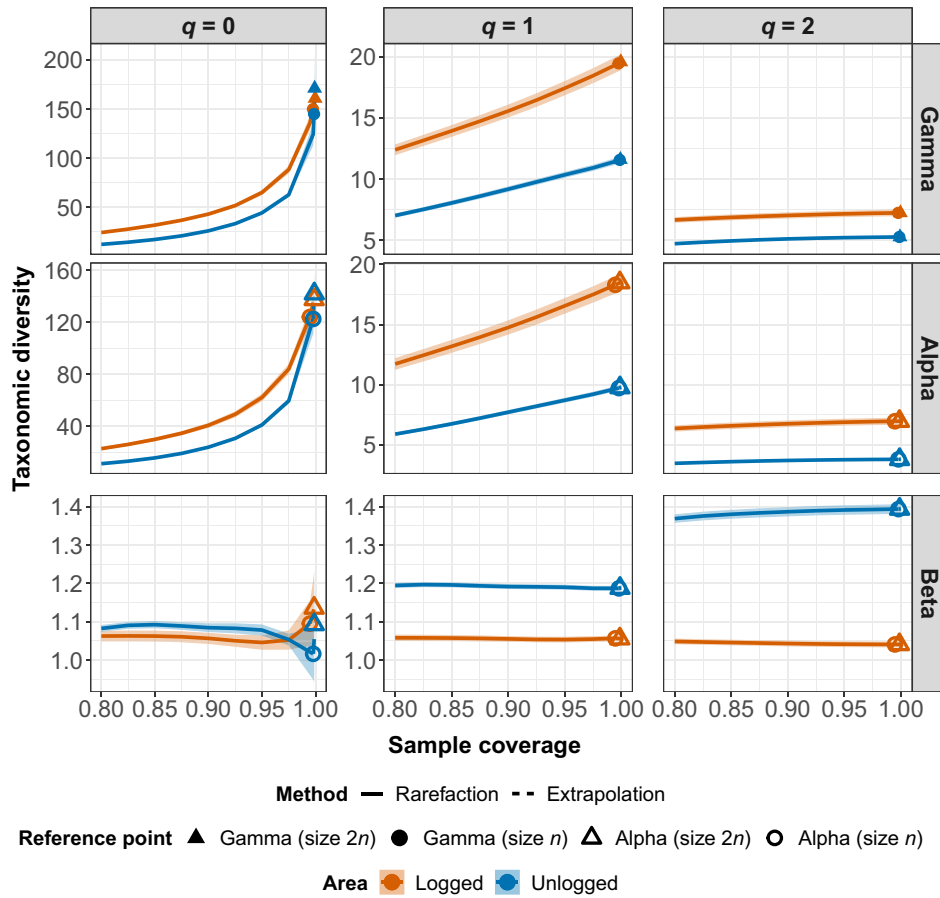

FIGURE 4 Legend on next page.

Temporal beta diversity for long-term time series data

We apply the iNEXT.beta3D standardization procedure to assess beta diversity over time for the yearly tree species data collected from six 1-ha (50 m × 200 m) second-growth rainforests in Costa Rica. The six forests, with their abbreviated names and ages in 2005 (the selected reference time; see below) are: Finca el Bejuco (FEB, 10 years), Juan Enriquez (JE, 10 years), Lindero Sur (LSUR, 20 years), Tirimbina (TIR, 23 years), Lindero el Peje (LEP, 28 years), and Cuatro Rios (CR, 33 years). The species abundance data for all stems ≥5 cm in DBH were recorded annually from 1997 to 2017 for four older second-growth forests (CR, LEP, TIR, and LSUR), and from 2005 to 2017 for the two youngest ones (FEB and JE); see Chazdon et al. (2022) for sampling details and summary statistics. All data are available through Dryad (Chazdon, 2021).

Since all abundance data were recorded within each 1-ha forest plot, individual trees of some species may exhibit intraspecific aggregation and thus may not be modeled as independent sampling units. In this case, it is statistically preferable to first convert species abundance records in each forest to occurrence or incidence (detection/nondetection) data. We thus further divided each 1-ha forest into 100 subplots (each with 0.01 ha) and recorded only species' incidence records in each subplot, computing the incidence frequency for a species as the number of subplots in which that species was detected. By treating the incidence frequency of each species among subplots as a "proxy" for its abundance, the iNEXT.beta3D standardization can be adapted to deal with spatially aggregated data and to avoid the effect of intraspecific aggregation. See *Discussion* for more details.

Because data for all six forests are available from 2005 to 2017, we chose 2005 as a base year, and temporal beta diversity was computed between 2005 and any subsequent year within each forest. Based on species incidence frequency data, our goal was to assess the trajectories of species compositional change (relative to the base year), as a function of forest age, during the process of forest succession from 2005 to 2017. That is, within each forest, we computed gamma, alpha and temporal beta diversity for 12 pairs of years (2005 vs. 2006, 2005 vs. 2007, ..., 2005 vs. 2017).

The same iNEXT.beta3D standardization procedures (but based on incidence data) as those presented for two time periods were applied to each of the 12 pairs of years. Coverage-based rarefaction and extrapolation curves for alpha, gamma and beta diversity, such as those in Figure 4b, were first obtained. As it is not possible to summarize the results based on simultaneously comparing 12 rarefaction and extrapolation curves, we extract and present the standardized gamma, alpha, and beta diversity for four levels of standardized coverage values: 80%, 90%, 95%, and 100% (asymptotic diversity). Figure 5 shows the trajectories of temporal beta diversity for these four coverage values (rows 2 to 5) with respect to forest age, along with the observed beta diversity (row 1). The trajectories for alpha and gamma diversity appear in Appendix S5: Figure S1. In Appendix S5, we also report the corresponding trajectories for beta diversity based on abundance data (Appendix S5: Figure S2); all patterns are generally consistent with those summarized below for incidence data.

First, note that, within each forest, no matter what the estimation methods (observed, standardized, or asymptotic diversity) and diversity order q, beta diversity generally increases with time. That is, in comparison with the base year, the extent of species compositional differentiation increases as time elapses. We next examine how temporal beta diversity and its slope behave, as a function of forest age separately for the following two cases:

FIGURE 4 iNEXT.beta3D standardization for two time periods. (a) Sample-size-based rarefaction (solid curves) and extrapolation (dashed curves) for gamma diversity and alpha diversity, and (b) coverage-based rarefaction (solid curves) and extrapolation (dashed curves) with gamma, alpha, and beta diversity, based on the saproxylic species abundance data collected in two time periods (2008–2009 vs. 2010–2011), separately for the logged area (orange curves) and unlogged area (blue curves) for diversity orders $q = 0$ (left panels), $q = 1$ (middle panels), and $q = 2$ (right panels). Here, size-based beta diversity is not a legitimate differentiation measure and thus is omitted; see *Discussion*. The two gamma reference samples are denoted by solid circles and alpha reference samples by hollow circles. The two extrapolated gamma and alpha samples, with double the reference sample size ($2n$), are denoted respectively by solid triangles and hollow triangles. In (a), sample size in each plot is extrapolated up to $2n$, but the extrapolation part for the unlogged area is not shown, for clarity. In (b), in both areas, the coverage values for alpha, extrapolated alpha with size $2n$, gamma, and extrapolated gamma with size $2n$ are all very close to 1 (all >99.5%). The increment in coverage from size n to $2n$ is so low that extrapolation in each panel is almost invisible. In (b), for $q = 1$ and $q = 2$, the extrapolation for alpha, beta, and gamma can be extended to the asymptote (i.e., coverage value of 100%). For $q = 0$, the extrapolation for gamma and alpha can be extended to the coverage value corresponding to double each reference sample size; the extrapolation limit for beta diversity is the same as alpha diversity. All 95% CIs (shaded areas) were obtained by a bootstrap method based on 200 replications. Some CIs are very narrow, so that they are hardly seen.
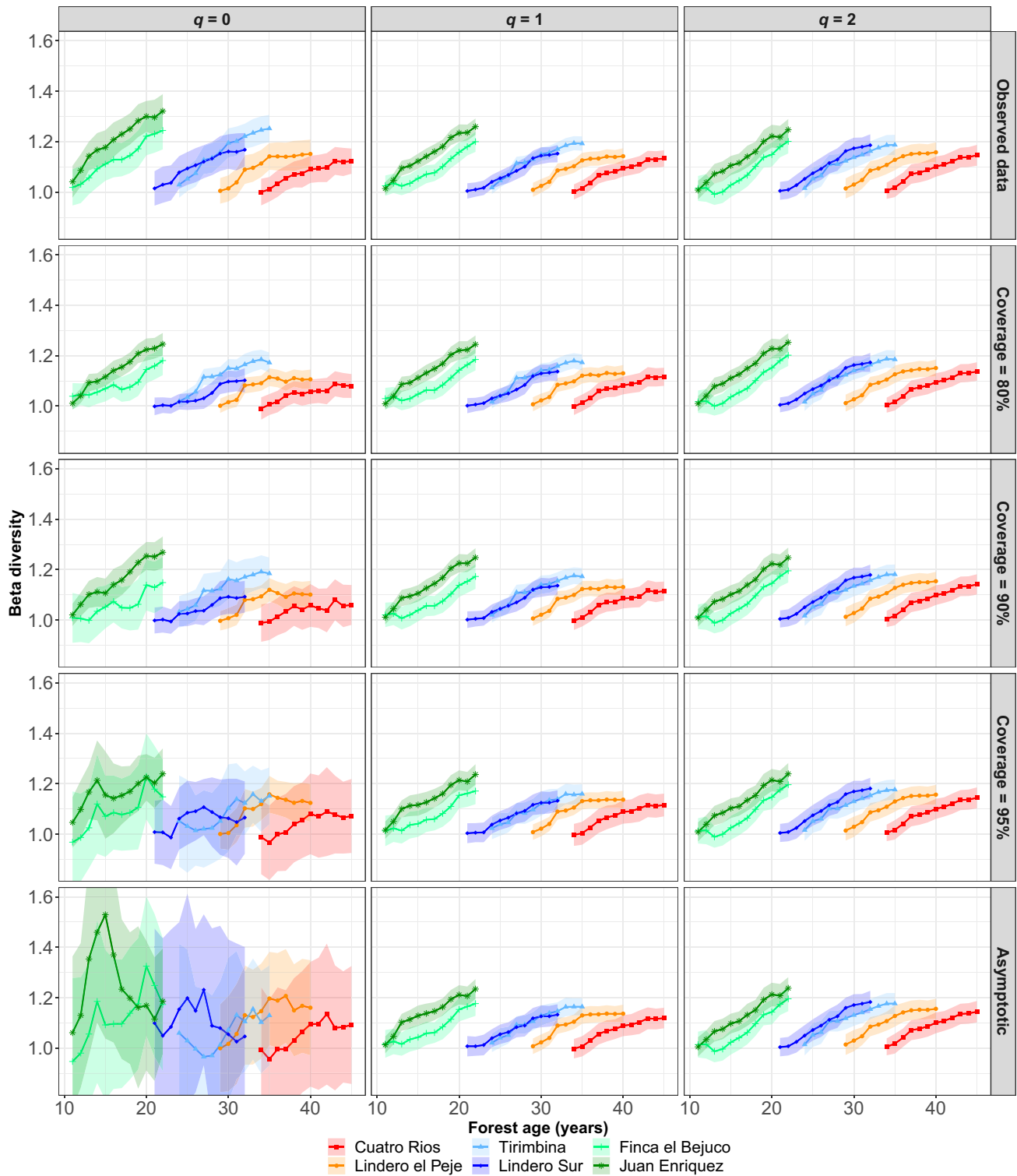
FIGURE 5 Trajectories of temporal beta diversity over time for tree species incidence data among 100 subplots in six second-growth rainforests. All temporal beta diversity within each forest was computed based on the data in 2005 (as the reference base time) and the data of any subsequent year. Row 1 shows the observed beta diversity, and rows 2 to 5 show the standardized beta diversity under four coverage levels (80%, 90%, 95%, and 100%) for beta diversity of order $q = 0$ (left panels), $q = 1$ (middle panels) and $q = 2$ (right panels). For standardized coverage levels of 80%, 90%, and 95%, standardized beta diversity was computed from interpolated samples for both gamma and alpha, to depict the trajectories graphically. Asymptotic beta diversity was computed based on "long-range extrapolated" (i.e., extrapolated size is greater than double the reference sample size) samples for both gamma and alpha; thus, only asymptotic beta diversity for $q = 1$ and $q = 2$ are reliable. The asymptotic beta diversity for $q = 0$ may be subject to long-range extrapolation bias. All 95% CIs (shaded areas) were obtained by a bootstrap method based on 200 replications. The corresponding plots based on abundance data are shown in Appendix S5: Figure S3.

1. Diversity order $q = 1$ and $q = 2$. If we focus on frequent species ($q = 1$, middle panels in Figure 5) or very frequent species ($q = 2$, right panels in Figure 5), a clear pattern is revealed: regardless of estimation method, temporal beta diversity (between the base year and any fixed subsequent year) declines very slowly as forest age is increased; the two youngest forests of the same age (JEB and JE) exhibit slightly higher beta diversity than the other four forests. A similar slow declining pattern with forest age is also revealed when comparing the rate of compositional change (i.e., the slope of beta diversity). The two youngest forests have slightly higher shift rates than the other four forests.

2. Diversity order $q = 0$. Except for asymptotic beta diversity (which is subject to some bias and fluctuation due to long-range extrapolation), observed and standardized beta diversity curves roughly follow similar patterns as described above for $q = 1$ and $q = 2$.

In the two youngest sites (FEB and JE), the canopy was nearly closed by 2005 (there were still some open areas in JE). But it is still at a young stage with continued recruitment of early successional trees into the canopy and establishment of shade-tolerant trees in small tree-size classes. This early successional stage is the time of maximum dynamics in tree species composition. Thus, as expected, the trajectories of gamma and alpha diversity increase over time in the two youngest forests (Appendix S5: Figure S1), regardless of diversity order q and standardized coverage value. The rate of increase in gamma diversity is much higher than for alpha diversity, due to rapid species turnover, yielding relatively high standardized beta diversity and high shift rate (high slope of changing beta diversity) for the two youngest forests.

As time progresses (>20 years) and the canopy closes, the understory becomes more shaded in older forests, some shade-intolerant species are eliminated, and colonizing shade-tolerant species thrive and establish slowly. The conditions are poor for the establishment of shade-intolerant species and many of the abundant species show high rates of mortality as stands are thinning (decreasing in total stem abundance but increasing in the basal area). Except for the TIR forest, standardized gamma and alpha diversity in the other three, older forests (LEP, LSUR, and CR) decline slowly with time (see Appendix S5: Figure S1) for any standardized coverage value. In these older forests, early successional species are dropping out and the rate of establishment of new, shade-tolerant species is slow. Although beta diversity is still increasing in the four older forests, both the magnitude of beta diversity and the slope of change

in beta are lower than the corresponding values in the two youngest sites. Our iNEXT.beta3D standardization provides insights into the trajectories of temporal beta diversity over time during forest succession that are driven by both species replacement and changes in species incidence frequency. Appendix S5: Figure S2 exhibits the corresponding trajectories based on abundance data; the patterns based on abundance data are generally consistent with those based on incidence data, showing, in this case, that the effect of spatial aggregation was not too strong.

APPLICATIONS TO SPATIOTEMPORAL DATA

We applied the iNEXT.beta3D standardization procedure to assess both temporal and spatial beta diversity change for fish species data based on annual surveys of the west coast of Scotland. Data were downloaded from the ICES (International Council for the Exploration of the Seas) portal, that is, DATRAS (2021) Fish Survey Data "Scottish West Coast International Bottom Trawl Survey" (SWC-IBTS) for commercial fish species in the first quarter of each year from 1985 to 2010. All data are available at https://datras.ices.dk; see Magurran et al. (2015) for sampling details and some pertinent analyses.

Each sample in the data represents a discrete trawl occurring at a specific geographical point within an ICES rectangle; each rectangle represents a 30′ latitude by 1° longitude grid cell. Species abundances were measured by catch per unit effort (CPUE, i.e., the number of individuals per species caught during a 1-h trawl). All rectangles in the original data roughly cover the area between 53° N to 60° N, with a total of 124 species captured between 1985 and 2010. We mainly focused on data covering rectangles above 55° N latitude (a total of 120 species) because trawls were conducted below 55° N only from 1997 to 2006.

To obtain stable spatial gamma, alpha, and beta diversity patterns, we pooled four latitudinal bands (55.5°, 56°, 56.5°, and 57° N) to form the South area, and pooled another four latitudinal bands (58°, 58.5°, 59°, and 59.5° N) to form the North area. The data in the middle (with latitudinal band 57.5° N) were excluded to focus on the contrast between the northern and southern spatial patterns. There were no species unique to this middle latitudinal band, that is, all observed species in this band were caught in the South and/or in the North area. Within a particular year, data from all rectangles in each area were pooled. To avoid using a very low number of standardized samples in assessing temporal gamma, alpha, and beta diversity, we also pooled annual data

into 2-year periods, that is, we considered a total of 13 time periods/intervals: 1985–1986, 1987–1988, ..., 2009–2010. Thus, the following analyses are based on spatiotemporal data for two areas (North and South) and 13 time periods.

The number of trawls/samples varies by rectangle and thus by area and time period. Within each area, the number of samples along with observed species richness in 13 time periods are shown in Appendix S6: Table S1. The minimum number of samples is 28, among all time periods in the two areas. Thus, the sampling effort was standardized by randomly selecting 28 samples in each time period within any given area; the iNEXT.beta3D standardization was applied to the pooled species abundance data (in terms of CPUE) based on these randomly selected samples. Temporal gamma, alpha, and beta diversity were computed between the first time period (1985–1986, base time period) and each subsequent period. Spatial gamma, alpha, and beta diversity were computed between the South and North areas within each time period. The random selection procedure of 28 samples was repeated 200 times. The following plots are all based on the average values of 200 replications.

For each randomly selected set of 28 samples, the estimated sample coverage for gamma and alpha reference samples in each time period and each fixed area was generally very high (>99%), for both temporal and spatial data. For $q = 1$ (focusing on abundant species) and $q = 2$ (focusing on very abundant species), the data contain sufficient information to infer asymptotic diversity values. However, there were some time periods for which data do not contain sufficient information to infer asymptotic gamma or alpha for total species richness ($q = 0$, focusing on rare species), because a tiny fraction of an assemblage's individuals may represent a very large number of vanishingly rare species. Thus, for species richness, coverage-based rarefaction and extrapolation are needed for both spatial and temporal data. We computed temporal and spatial gamma, alpha, and beta using four estimation methods: (1) observed diversity, (2) standardization to a coverage value of 99%, (3) standardization to a coverage value of 99.9%, and (4) asymptotic estimation, that is, extrapolation to complete coverage.

Figure 6a shows the plots of temporal (left three panels) and spatial (right three panels) beta diversity for $q = 0$, 1, and 2 under the four estimation methods. Regardless of data type (spatial or temporal) and diversity order, beta diversity values revealed generally consistent patterns among the four estimation methods, except for the asymptotic richness-based ($q = 0$) temporal beta diversity. The exception arises because the estimated asymptotic richness-based beta diversity may be subject to some bias, due to insufficient data for rare species. To help interpret the results, Figure 6b shows temporal (left three panels) and spatial (right three panels) gamma, alpha, and beta diversity of orders $q = 0$, 1, and 2 for a standardized coverage value of 99.9%. The corresponding plots of gamma and alpha diversity for the other three estimation methods exhibit similar patterns; see Appendix S6 for details.
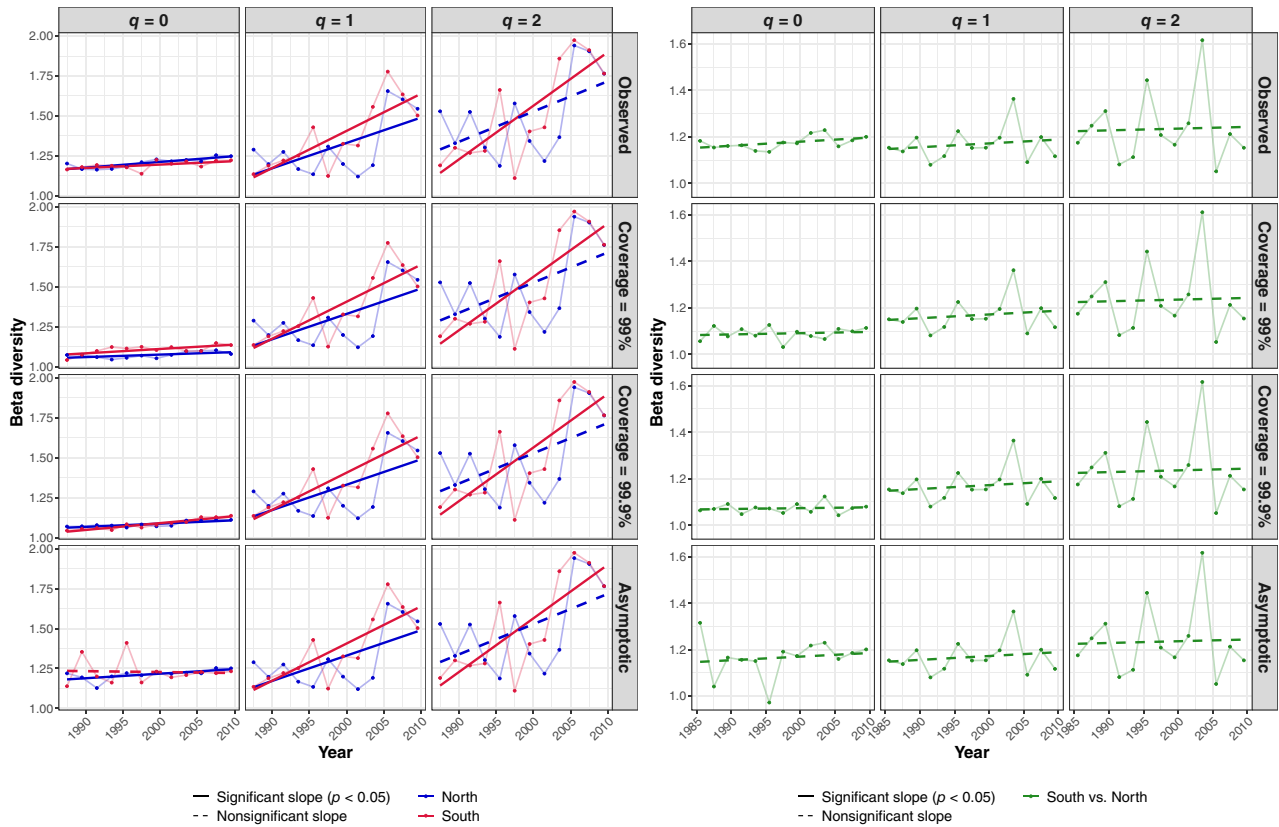
Temporal beta diversity

Figure 6a clearly demonstrates for both areas that temporal beta diversity increases relative to the initial period (i.e., similarity decreases) from 1985 to 2010; most of the linear fitted trends are statistically significant. Note that within each area, both temporal gamma and alpha diversity generally tend to decline (Figure 6b; Appendix S6: Figure S1) over time, but the rate of decline is steeper for alpha diversity, leading to an increasing trend for beta. For both areas, temporal beta diversity for abundant and very abundant species ($q = 1$ and 2) is much higher than for rare species ($q = 0$). In addition, the rate of species compositional change (i.e., the slope of beta diversity) for abundant and very abundant species ($q = 1$ and 2) is also higher than for rare species ($q = 0$). These stronger effects for $q = 1$ and 2 imply that temporal beta diversity and its rate of change were principally attributable to abundant and very abundant species, that is, linked to a shift in the SAD, rather than to a change in species richness. Whether temperature, salinity and/or other environmental variables can be used to explain the significant changes in temporal beta diversity rate merits further investigation.

When comparing the temporal beta diversity of the two areas, for abundant and very abundant species ($q = 1$ and 2), we find that the South area generally had higher beta diversity and higher rates of compositional shift rate than the North area; for rare species ($q = 0$) both areas had comparable beta diversity values and shift rates.

Spatial beta diversity

Within each fixed time period, Figure 6a shows that spatial beta diversity between the North and the South did not exhibit a significant change over time for each fixed diversity order q. Figure 6b implies that spatial gamma and alpha diversity both tended to decline over time at almost the same rate for a fixed value of q, leading to a nearly constant level for spatial beta diversity over time (i.e., shift rate in beta over time is virtually 0 for all $q = 0$,

(a) Observed, standardized (at coverage values of 99% and 99.9%) and asymptotic temporal (left) and spatial (right) beta diversity

(b) Standardized temporal and spatial gamma, alpha and beta diversity specifically for a rarefied coverage value 99.9%
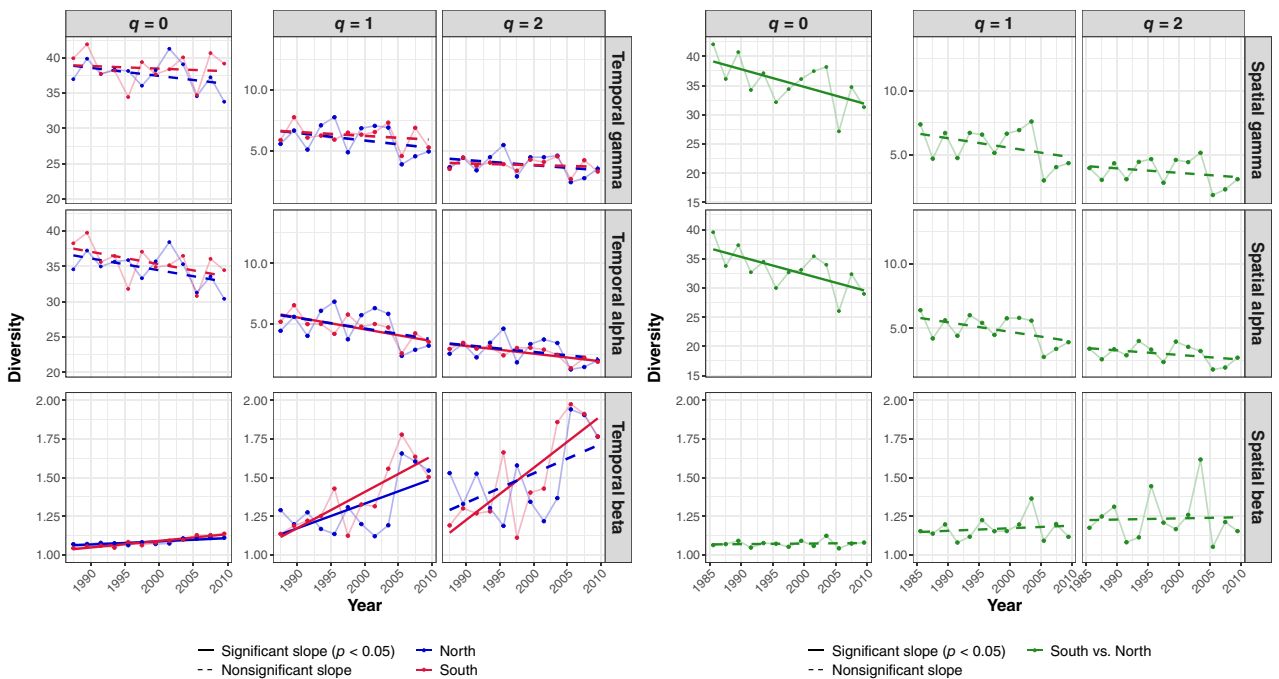

FIGURE 6 Legend on next page.

1, and 2), with the level increasing with order q. Consequently, the change of abundance among species can nonetheless explain spatial beta diversity, but not its rate of change.

CONCLUSIONS

Under the framework of Hill numbers, we have generalized the iNEXT standardization (in Table 2) for a single assemblage to alpha, beta, and gamma diversity of order $q \geq 0$ for multiple assemblages. In our approach, the iNEXT standardization for gamma diversity is applied to the gamma reference sample in the pooled assemblage, whereas the corresponding standardization for alpha diversity is applied to the alpha reference sample in the joint assemblage; see Table 1 for a summary and Appendix S2 for an illustrative example. Our proposed standardized beta diversity (Equation 6), for any given level of sample coverage, is expressed as the ratio between gamma and alpha diversity, where both alpha and gamma diversity are standardized to the same level of sample coverage. Sample coverage is an objective measure of sample completeness and can be very accurately estimated from gamma and alpha reference samples. We have proved that coverage-based standardized beta diversity estimates fulfill the three essential properties that justify the independence of our standardized beta diversity from gamma and alpha diversity. The rarefaction and extrapolation formulas for alpha, beta, and gamma diversities ($q = 0$, 1, and 2) appear in Appendix S2.

A novel contribution of our proposed iNEXT.beta3D strategy based on sampling data lies in its statistical sampling-model-based approach, which makes a clear distinction between the theoretical assemblage level and the sampling data level. For spatial data collected from a region consisting of N assemblages, we summarize some key points and pertinent model assumptions:

1. At the assemblage level, beta diversity (i.e., the extent of among-assemblage differences) represents the interacting effect of the regional SAD and spatial aggregation.

2. As with classic rarefaction, one basic assumption for the iNEXT.beta3D standardization using rarefaction and extrapolation for Hill numbers is that an independent sample of n individuals (or other sampling units) is selected from the region. Simulation results (in Appendix S1) demonstrate that our data-level standardization is not affected by assemblage-level spatial aggregation and is quite robust to departures from fully independent sampling.

3. At the data level, observed beta diversity depends not only on among-assemblage differentiation but also on sample effort/completeness, which in turn induces dependence on alpha and gamma diversity.

4. Under independent sampling, our iNEXT.beta3D approach standardizes sample coverage, which removes the dependence of beta on alpha and gamma diversity. Our coverage-based standardized beta diversity purely reflects the among-assemblage differentiation.

5. When independent sampling is not feasible, such as the large-area-based forest data in our applications, we suggest using appropriate sampling units to convert abundance data to incidence data, which are generally only weakly dependent on, and therefore much less sensitive to aggregation than the corresponding abundance data. Thus, the rarefaction and extrapolation method can be applied to replicated incidence data.

DISCUSSION

In this section, for sampling data we discuss why comparisons across datasets should not use observed beta diversity and size-based, standardized beta diversity. We also compare our iNEXT.beta3D method with three previous approaches, highlighting the ways in which it helps solve existing challenges, and suggest a method for assessing the spatial-aggregation effect, which is a known complicating factor in biodiversity assessment.

FIGURE 6 Temporal and spatial gamma, alpha and beta diversity for fish species based on SWC-IBTS data from 1985 to 2010. (a) Temporal beta (left three panels) and spatial beta (right three panels) diversity for orders $q = 0$, 1, and 2, based on four estimation methods, that is, observed, standardized (at two coverage values, 99% and 99.9%) and asymptotic beta values. (b) Temporal (left three panels) and spatial (right three panels) gamma, alpha, and beta diversity for orders $q = 0$, 1, and 2 specifically for a standardized coverage value 99.9%. Temporal gamma, alpha, and beta diversity were computed between the first time period (1985–1986, base time period) and any subsequent period, separately for the South area (red dots) and the North area (blue dots). Spatial gamma, alpha, and beta diversity (green dots) were computed between the South and North areas within each time period. Sampling effort was standardized at 28 samples across the two areas and 13 time periods. The random selection procedure of 28 samples was repeated 200 times, and all plots are based on the average values of the 200 replications. A linear trend was fitted to data points in each panel; statistical significance ($p < 0.05$) for a fitted linear trend is denoted by a solid line and nonsignificance is denoted by a dashed line. The corresponding plots of gamma and alpha diversity for the other three estimation methods are provided in Appendix S6: Figure S1.

The bias of the observed richness-based beta diversity

Here "bias" means a systematic tendency that leads to differences between an estimator/statistic and the true parameter. It is well known that the observed number of species in sampling data exhibits negative bias. The bias could be serious for data from a hyperdiverse assemblage (e.g., Coddington et al., 2009). The observed richness-based ($q = 0$) beta diversity and dissimilarity measures, including the classic Jaccard and Sørensen dissimilarity measures, generally exhibit positive bias (Chao et al., 2005), but negative bias or no bias could also arise, under certain circumstances:

1. When no species are shared among assemblages, observed richness-based beta diversity always correctly yields the maximum value of N; in this case, observed beta diversity is unbiased and standardized beta diversity for any level of coverage also correctly attains the maximum value of N.
2. In the special case in which N assemblages are identical, Lande (1996) was the first to notice that positive bias exists for richness-based dissimilarity even for very large sample sizes. In this case, our standardized beta diversity for any level of coverage approaches the correct value of 1.
3. Negative bias could arise when N assemblages only share very abundant species, while the rare ones tend to be locally endemic. At the data level, two samples might yield the same few common species, but fail to reveal rare species that would differentiate the assemblages in larger samples (Colwell & Coddington, 1994).

Why is sample-size-based, standardized beta diversity not a legitimate differentiation measure?

To standardize gamma and alpha to the same level of coverage in our iNEXT.beta3D procedure, the sample sizes needed in the joint assemblage and the pooled assemblage, that is, m_α and m_γ in Equation (6), generally differ, except for the case in which no species are shared among assemblages. The two sample sizes should satisfy the following two constraints: (1) When the data of N assemblages have no shared species, we should have $m_\gamma = m_\alpha$ for any given coverage value. (2) When N assemblages are identical, we should have $m_\alpha = Nm_\gamma$. Any fixed relationship (as defined below) for the two sample sizes will make the corresponding standardization violate either the fixed maximum requirement or the fixed minimum requirement.

Consider the two common ways to standardize sample sizes:

1. The sample size in each of the N assemblages and in the pooled assemblage are all standardized to a fixed sample size, that is, $m_\alpha = Nm_\gamma$ in Equation (6). Such a size-based standardized beta diversity fails to satisfy the fixed maximum criterion unless data are complete (i.e., coverage approaches 100%).
2. The sample size in each of the N assemblages is standardized to a fixed size, but the size in the pooled assemblage is taken to be N times the fixed number, that is, $m_\alpha = m_\gamma$ in Equation (6). Such a size-based standardized beta diversity fails to satisfy the fixed minimum criterion unless data are complete.

We thus can conclude that neither of the two size-based standardizations leads to a legitimate differentiation measure.

Extension to phylogenetic and functional beta diversity

In this paper, we have focused mainly on taxonomic beta diversity, which is based on species abundances in multiple assemblages. Thus, our procedures do not take species phylogenetic relations and species trait differences into account. A rapidly growing literature addresses phylogenetic diversity and functional diversity measures. Chao et al. (2021) extended the iNEXT methodology to the iNEXT.3D standardization, which integrates the three dimensions of biodiversity (taxonomic, phylogenetic, and functional diversity) in a unified framework. Under the decomposition framework of Hill numbers, Chiu et al. (2014) generalized taxonomic beta diversity to a phylogenetic version, and Chao, Chiu, Villéger, et al. (2019) further generalized beta to a functional version. Recently, we have extended the iNEXT.beta3D methodology to phylogenetic and functional versions so that a unified framework is established for the three dimensions of beta diversity.

Comparisons with some previous approaches

The individual-based, null-model approach to beta diversity

Kraft et al. (2011) developed an individual-based randomization procedure, under a null model, to remove species pool dependence on beta diversity. They proposed the use of a statistic called beta deviation which is formulated

as the difference between the observed "beta" and the average of expected "beta" values, divided by the standard deviation of expected values. Here "beta" is defined as $\beta_P = 1 - \overline{S}/S$ (proportional species turnover), where S denotes the species richness in the pooled assemblage (i.e., richness-based gamma) and $\overline{S}$ denotes the simple average of the species richness of individual assemblages (i.e., richness-based alpha). The expected "beta" values are calculated by randomly distributing all individuals in the observed pool to all assemblages, while preserving the SAD in the observed pool and the number of individuals in the data from each assemblage.

As Kraft et al. indicated, beta deviation provides a useful statistic for testing the null hypothesis (random spatial distribution from the observed pool of individuals), that is, whether the observed beta diversity can be obtained by random sampling from the observed pool of individuals. From a statistical perspective, the inference one can extract from this test is: if the magnitude of beta deviation exceeds some critical value, one can reject the above randomness assumption. Otherwise, data are not sufficient to support nonrandomness. The critical values for a significance level of 5% in most applications can be taken to be −2 and 2 (i.e., approximate lower and upper critical values based on a standard normal distribution). A rejection implies that the observed beta cannot be obtained from a random sampling of the observed pool of individuals. However, the observed pool deviates from the true pool for incomplete sampling data (Tuomisto & Ruokolainen, 2012; Xing & He, 2021). As with all statistical tests, the magnitude of the randomization test based on beta deviation depends on sampling effort (Bennett & Gilbert, 2016; Qian et al., 2013; Ulrich et al., 2017, 2018), as well as species pool size (Chao et al., 2023). Thus, caution is called for when using the magnitude of beta deviation to quantify the effect of nonrandom spatial distribution and among-assemblage differentiation. Xing and He (2021) proposed a modified version with a scaled sampling effort, but under the restrictive assumption that the regional SAD follows a log-series distribution.

Engel et al.'s (2021) sample-coverage-based standardization approach

Engel et al. (2021) focused on species-richness-based ($q = 0$) beta diversity and applied a coverage-based rarefaction and extrapolation method for standardization. However, their standardization is different from the iNEXT.beta3D approach. Although they standardize sample coverage across datasets (each dataset consisting of N assemblages), size-based standardization is used to obtain gamma and alpha species richness within each

dataset, that is, the sample size in each of the N assemblages and in the pooled assemblage are all standardized to a fixed sample size. As discussed earlier regarding the problem with a size-based approach, unless the sample coverage tends to one, the fixed maximum criterion will not be satisfied. Our coverage-based approach solves this problem because, when there are no shared species among N assemblages (i.e., complete turnover), standardized beta always yields the maximum value of N, for any value of sample coverage.

In addition, in the Engel et al. (2021) standardization, datasets are compared at a single targeted level of sample coverage, while in our procedures, datasets can be compared for a range of coverage values. When there are many datasets, targeted coverage levels in Engel et al.'s method can be low. We use a numerical example to illustrate the differences. Engel et al. (2021, their figure 5) applied their method to analyze Gentry's worldwide plots data on woody plant abundance records (Gentry, 1988; Phillips & Miller, 2002) with each plot consisting of 10 small subplots. In their analysis, standardized beta diversity among 10 subplots was compared only at a single target coverage level of 10% across all plots; they found no evidence for any latitudinal patterns. However, under such a low coverage value, a large amount of data in each plot was discarded and only a few species were involved in the comparison across all plots. Consequently, the data do not have enough power to detect any systematic changes. Based on the iNEXT.beta3D method, with extrapolation to twice the reference sample size in each plot, we can infer richness-based beta diversity for a range of coverage values up to the maximum. Using Gentry's data in 197 plots, we can compare 82 plots up to a coverage value of 80%, 115 plots up to 70%, 141 plots up to 60%, and 172 plots up to 50%. Data then have sufficient power to reveal significant latitudinal beta diversity patterns; see Chao et al. (2023) for analysis details.

The McGlinn et al. (2019, 2021) decomposition approach

Based on abundance data collected from georeferenced quadrats (plots or patches), McGlinn et al. (2021) decomposed individual-based and sample-based species rarefaction curves into three components attributed to: (1) the SAD, (2) the total abundance, and (3) spatial aggregation. The effect of each component is assessed by comparing three types of rarefaction curves (spatial sample-based, nonspatial sample-based, and individual-based). All three types of rarefaction curves considered by McGlinn et al. (2021) are computed from a single

spatially explicit dataset. Their assessment of the three components relies on the premise that individual-based rarefaction curves purely reflect variation in the SAD, regardless of the extent of spatial aggregation. From our simulations reported in Appendix S1, this premise is justified if spatial aggregation is not strong. However, when spatial aggregation is strong, individuals sampled in any sampling unit are no longer independent, violating the basic assumption for rarefaction theory. Consequently, individual-based rarefaction curves may not purely reflect the SAD effect. In this case, to correctly reflect the SAD effect, we recommend additional sampling, which will allow the pure effect of the SAD and the extent of spatial aggregation to be inferred, as suggested below.

Assessing spatial-aggregation effect if additional sampling is feasible

We suggest assessing the aggregation effect by comparing two statistically-valid species rarefaction and extrapolation curves. Assume that T quadrats/plots (each quadrat/plot is regarded as a sampling unit) are independently selected, and species abundance data are collected from each quadrat. We construct two rarefaction/extrapolation curves:

1. A sample (incidence)-based species rarefaction/extrapolation curve obtained by converting abundance data in each quadrat to incidence data. Since the sampling units are independently selected, the sample-based rarefaction/extrapolation curve legitimately conveys all information in the SID to purely reflect the SID effect.
2. An individual (abundance)-based species rarefaction/extrapolation curve obtained based on an additional sampling in which an approximately independent sample of n individuals is selected from the assemblage, where the sample size n is the same as the total abundance recorded from T quadrats. This individual-based rarefaction curve conveys all information in the SAD and purely reflects the SAD effect.

When there is no spatial aggregation, the two rarefaction/extrapolation curves will exactly match; for weak aggregation, the two curves deviate very little; for strong aggregation, the two curves deviate to a large extent. Therefore, the difference between the two curves can be used to assess the degree of spatial aggregation. In Appendix S7, some preliminary simulation results are reported to illustrate our suggested approach. However, the difference between the two rarefaction curves depends on sample completeness and quadrat size; more research is needed.

CONFLICT OF INTEREST STATEMENT
The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT
All R code and data are available in Zenodo from Chao (2023) at https://doi.org/10.5281/zenodo.8025870. For readers without an R background, the online software "iNEXT.beta3D" is available from https://chao.shinyapps.io/iNEXT_beta3D/ to facilitate all computation and graphics.

ORCID
Anne Chao https://orcid.org/0000-0002-4364-8101
Simon Thorn https://orcid.org/0000-0002-3062-3060
Chun-Huo Chiu https://orcid.org/0000-0002-7096-2278
Faye Moyes https://orcid.org/0000-0001-9687-0593
Robin L. Chazdon https://orcid.org/0000-0002-7349-5687
Luiz Fernando S. Magnago https://orcid.org/0000-0001-6335-754X
Maria Dornelas https://orcid.org/0000-0003-2077-7055
David Zelený https://orcid.org/0000-0001-5157-044X
Robert K. Colwell https://orcid.org/0000-0002-1384-0354
Anne E. Magurran https://orcid.org/0000-0002-0036-2795

REFERENCES

Alexander, K. N. A. 2008. "Tree Biology and Saproxylic Coleoptera: Issues of Definitions and Conservation Language." *Revue d'écologie (Terre Vie)* 63(suppl. 10): 9–13. http://hdl.handle.net/2042/55805.

Anderson, M. J., T. O. Crist, J. M. Chase, M. Velend, B. D. Inouye, A. L. Freestone, N. J. Sanders, et al. 2011. "Navigating the Multiple Meanings of β Diversity: A Roadmap for the Practicing Ecologist." *Ecology Letters* 14: 19–28.

Barwell, L. J., N. J. Isaac, and W. E. Kunin. 2015. "Measuring β-Diversity with Species Abundance Data." *Journal of Animal Ecology* 84: 1112–22.

Bennett, J. R., and B. Gilbert. 2016. "Contrasting Beta Diversity among Regions: How Do Classical and Multivariate Approaches Compare?" *Global Ecology and Biogeography* 25: 368–377.

Bray, J. R., and J. T. Curtis. 1957. "An Ordination of the Upland Forest Communities of Southern Wisconsin." *Ecological Monographs* 27: 325–349.

Chao, A. 2023. "Data from: Rarefaction and Extrapolation with Beta Diversity under a Framework of Hill Numbers: The iNEXT.beta3D Standardization." Zenodo. https://doi.org/10.5281/zenodo.8025870.

Chao, A., R. L. Chazdon, R. K. Colwell, and T. J. Shen. 2005. "A New Statistical Approach for Assessing Similarity of Species Composition with Incidence and Abundance Data." *Ecology Letters* 8: 148–159.

Chao, A., and C.-H. Chiu. 2016. "Bridging the Variance and Diversity Decomposition Approaches to Beta Diversity Via Similarity and Differentiation Measures." *Methods in Ecology and Evolution* 7: 919–928.

Chao, A., C.-H. Chiu, K.-H. Hu, and D. Zelený. 2023. "Revisiting Alwyn H. Gentry's Forest Transect Data: Latitudinal Beta-Diversity Patterns Are Revealed Using a Statistical Sampling-Model-Based Approach." *Japanese Journal of Statistics and Data Science*. https://doi.org/10.1007/s42081-023-00214-1.

Chao, A., C.-H. Chiu, S. Villéger, I.-F. Sun, S. Thorn, Y.-C. Lin, J. M. Chiang, and W. B. Sherwin. 2019. "An Attribute-Diversity Approach to Functional Diversity, Functional Beta Diversity, and Related (Dis)Similarity Measures." *Ecological Monographs* 89: e01343.

Chao, A., C.-H. Chiu, S.-H. Wu, C. L. Huang, and Y.-C. Lin. 2019. "Comparing Two Classes of Alpha Diversities and their Corresponding Beta and (Dis)Similarity Measures, with an Application to the Formosan Sika Deer Cervus Nippon Taiouanus Reintroduction Programme." *Methods in Ecology and Evolution* 10: 1286–97.

Chao, A., and R. K. Colwell. 2017. "Thirty Years of Progeny from Chao's Inequality: Estimating and Comparing Richness with Incidence Data and Incomplete Sampling." *SORT (Statistics and Operations Research Transactions)* 41: 1–52.

Chao, A., N. J. Gotelli, T. C. Hsieh, E. L. Sander, K. H. Ma, R. K. Colwell, and A. M. Ellison. 2014. "Rarefaction and Extrapolation with Hill Numbers: A Framework for Sampling and Estimation in Species Diversity Studies." *Ecological Monographs* 84: 45–67.

Chao, A., P. A. Henderson, C. H. Chiu, F. Moyes, K. H. Hu, M. Dornelas, and A. E. Magurran. 2021. "Measuring Temporal Change in Alpha Diversity: A Framework Integrating Taxonomic, Phylogenetic and Functional Diversity and the iNEXT.3D Standardization." *Methods in Ecology and Evolution* 12: 1926–40.

Chao, A., and L. Jost. 2012. "Coverage-Based Rarefaction and Extrapolation: Standardizing Samples by Completeness Rather than Size." *Ecology* 93: 2533–47.

Chao, A., and L. Jost. 2015. "Estimating Diversity and Entropy Profiles Via Discovery Rates of New Species." *Methods in Ecology and Evolution* 6: 873–882.

Chao, A., Y. Kubota, D. Zelený, C.-H. Chiu, C.-F. Li, B. Kusumoto, M. Yasuhara, et al. 2020. "Quantifying Sample Completeness and Comparing Diversities among Assemblages." *Ecological Research* 35: 292–314.

Chase, J. M., S. A. Blowes, T. M. Knight, K. Gerstner, and F. May. 2020. "Ecosystem Decay Exacerbates Biodiversity Loss with Habitat Loss." *Nature* 584: 238–243.

Chase, J. M., N. J. Kraft, K. G. Smith, M. Velend, and B. D. Inouye. 2011. "Using Null Models to Disentangle Variation in Community Dissimilarity from Variation in α-Diversity." *Ecosphere* 2(2): art24.

Chase, J. M., B. J. McGill, D. J. McGlinn, F. May, S. A. Blowes, X. Xiao, T. M. Knight, O. Purschke, and N. J. Gotelli. 2018. "Embracing Scale-Dependence to Achieve a Deeper Understanding of Biodiversity and its Change across Communities." *Ecology Letters* 21: 1737–51.

Chazdon, R. 2021. "Tree Abundance in Eight 1-ha Tropical Forest Plots in Northeastern Costa Rica from 1997–2017." Dryad. https://doi.org/10.5061/dryad.ncjsxksvr.

Chazdon, R. L., N. Norden, R. K. Colwell, and A. Chao. 2022. "Monitoring Recovery of Tree Diversity during Tropical Forest Restoration: Lessons from Long-Term Trajectories of Natural Regeneration." *Philosophical Transactions of the Royal Society B* 378: 20210069.

Chiu, C.-H., L. Jost, and A. Chao. 2014. "Phylogenetic Beta Diversity, Similarity, and Differentiation Measures Based on Hill Numbers." *Ecological Monographs* 84: 21–44.

Coddington, J., I. Agnarsson, J. Miller, M. Kuntner, and G. Hormiga. 2009. "Undersampling Bias: The Null Hypothesis for Singleton Species in Tropical Arthropod Surveys." *Journal of Animal Ecology* 78: 573–584.

Colwell, R. K., A. Chao, N. J. Gotelli, S. Y. Lin, C. X. Mao, R. L. Chazdon, and J. T. Longino. 2012. "Models and Estimators Linking Individual-Based and Sample-Based Rarefaction, Extrapolation and Comparison of Assemblages." *Journal of Plant Ecology* 5: 3–21.

Colwell, R. K., and J. A. Coddington. 1994. "Estimating Terrestrial Biodiversity through Extrapolation." *Philosophical Transactions of the Royal Society B* 345: 101–118.

Colwell, R. K., C. X. Mao, and J. Chang. 2004. "Interpolating, Extrapolating, and Comparing Incidence Based Species Accumulation Curves." *Ecology* 85: 2717–27.

DATRAS. 2021. "ICES Scottish West Coast Bottom Trawl Survey (SWC-IBTS) 1985–2010." https://datras.ices.dk.

Dornelas, M., L. H. Antão, F. Moyes, A. E. Bates, A. E. Magurran, D. Adam, A. A. Akhmetzhanova, et al. 2018. "BioTIME: A Database of Biodiversity Time Series for the Anthropocene." *Global Ecology and Biogeography* 27: 760–786.

Dornelas, M., N. J. Gotelli, B. McGill, H. Shimadzu, F. Moyes, C. Sievers, and A. E. Magurran. 2014. "Assemblage Time

Series Reveal Biodiversity Change but Not Systematic Loss." *Science* 344: 296–99.

Ellison, A. M. 2010. "Partitioning Diversity." *Ecology* 91: 1962–63.

Engel, T., S. A. Blowes, D. J. McGlinn, F. May, N. J. Gotelli, B. J. McGill, and J. M. Chase. 2021. "Using Coverage-Based Rarefaction to Infer Non-random Species Distributions." *Ecosphere* 12(9): e03745.

Eriksson, B. K., and H. Hillebrand. 2019. "Rapid Reorganization of Global Biodiversity." *Science* 366: 308–9.

Fortin, M. J., P. M. A. James, A. MacKenzie, S. J. Melles, and B. Rayfield. 2012. "Spatial Statistics, Spatial Regression, and Graph Theory in Ecology." *Spatial Statistics* 1: 100–109.

Gentry, A. H. 1988. "Changes in Plant Community Diversity and Floristic Composition on Environmental and Geographical Gradients." *Annals of the Missouri Botanical Garden* 75: 1–34.

Good, I. J. 1953. "The Population Frequencies of Species and the Estimation of Population Parameters." *Biometrika* 40: 237–264.

Gotelli, N. J., and R. K. Colwell. 2011. "Estimating Species Richness." In *Biological Diversity: Frontiers in Measurement and Assessment*, edited by A. Magurran and B. McGill, 39–54. Oxford: Oxford University Press.

Gotelli, N. J., F. Moyes, L. H. Antão, S. A. Blowes, M. Dornelas, B. J. McGill, A. Penny, et al. 2022. "Long-Term Changes in Temperate Marine Fish Assemblages Are Driven by a Small Subset of Species." *Global Change Biology* 28: 46–53.

Gregorius, H.-R. 2010. "Linking Diversity and Differentiation." *Diversity* 2: 370–394.

Gregorius, H.-R. 2020. "Factorization of Joint Metacommunity Diversity into its Marginal Components: An Alternative to the Partitioning of Trait Diversity." *Theory in Biosciences* 139: 253–263.

Hill, M. O. 1973. "Diversity and Evenness: A Unifying Notation and its Consequences." *Ecology* 54: 427–432.

Hillebrand, H., B. Blasius, E. T. Borer, J. M. Chase, J. A. Downing, B. K. Eriksson, C. T. Filstrup, et al. 2018. "Biodiversity Change Is Uncoupled from Species Richness Trends: Consequences for Conservation and Monitoring." *Journal of Applied Ecology* 55: 169–184.

Hsieh, T. C., K. H. Ma, and A. Chao. 2016. "iNEXT: An R Package for Rarefaction and Extrapolation of Species Diversity (Hill Numbers)." *Methods in Ecology and Evolution* 7: 1451–56.

Hurlbert, S. H. 1971. "The Nonconcept of Species Diversity: A Critique and Alternative Parameters." *Ecology* 52: 577–586.

IPBES-IPCC. 2021. "Biodiversity and Climate Change." Scientific Outcome. https://ipbes.net/sites/default/files/2021-06/20210609_scientific_outcome.pdf.

Jost, L. 2007. "Partitioning Diversity into Independent Alpha and Beta Components." *Ecology* 88: 2427–39.

Jost, L. 2010. "Independence of Alpha and Beta Diversities." *Ecology* 91: 1969–74.

Jost, L., A. Chao, and R. L. Chazdon. 2011. "Compositional Similarity and Beta Diversity." In *Biological Diversity: Frontiers in Measurement and Assessment*, edited by A. Magurran and B. McGill, 66–84. Oxford: Oxford University Press.

Koleff, P., K. J. Gaston, and J. J. Lennon. 2003. "Measuring Beta Diversity for Presence-Absence Data." *Journal of Animal Ecology* 72: 367–382.

Kraft, N. J. B., L. S. Comita, J. M. Chase, N. J. Sanders, N. G. Swenson, T. O. Crist, and H. V. Cornell. 2011. "Disentangling the Drivers of β Diversity along Latitudinal and Elevational Gradients." *Science* 333: 1755–58.

Lande, R. 1996. "Statistics and Partitioning of Species Diversity, and Similarity among Multiple Communities." *Oikos* 76: 5–13.

Laurance, W. F., T. E. Lovejoy, H. L. Vasconcelos, E. M. Bruna, R. K. Didham, P. C. Stouffer, C. Gascon, R. O. Bierregaard, S. G. Laurance, and E. Sampaio. 2002. "Ecosystem Decay of Amazonian Forest Fragments: A 22-Years Investigation." *Conservation Biology* 16: 605–618.

Laurance, W. F., H. E. M. Nascimento, S. G. Laurance, A. Andrade, J. E. L. S. Ribeiro, J. P. Giraldo, T. E. Lovejoy, et al. 2006. "Rapid Decay of Tree-Community Composition in Amazonian Forest Fragments." *Proceedings of the National Academy of Sciences of the United States of America* 103: 19010–14.

Legendre, P., and M. De Cáceres. 2013. "Beta Diversity as the Variance of Community Data: Dissimilarity Coefficients and Partitioning." *Ecology Letters* 16: 951–963.

MacArthur, R. H. 1965. "Patterns of Species Diversity." *Biological Reviews* 40: 510–533.

Magnago, L. F. S., D. P. Edwards, F. A. Edwards, A. Magrach, S. V. Martins, and W. F. Laurance. 2014. "Functional Attributes Change but Functional Richness Is Unchanged after Fragmentation of Brazilian Atlantic Forests." *Journal of Ecology* 102: 475–485.

Magnago, L. F. S., A. Magrach, J. Barlow, C. E. G. R. Schaefer, W. F. Laurance, S. V. Martins, and D. P. Edwards. 2017. "Do Fragment Size and Edge Effects Predict Carbon Stocks in Trees and Lianas in Tropical Forests?" *Functional Ecology* 31: 542–552.

Magnago, L. F. S., M. F. Rocha, L. Meyer, S. V. Martins, and J. A. A. Meira-Neto. 2015. "Microclimatic Conditions at Forest Edges Have Significant Impacts on Vegetation Structure in Large Atlantic Forest Fragments." *Biodiversity and Conservation* 24: 2305–18.

Magurran, A. E., M. Dornelas, F. Moyes, N. J. Gotelli, and B. McGill. 2015. "Rapid Biotic Homogenization of Marine Fish Assemblages." *Nature Communications* 6: 1–5.

Magurran, A. E., M. Dornelas, F. Moyes, and P. A. Henderson. 2019. "Temporal β Diversity—A Macroecological Perspective." *Global Ecology and Biogeography* 28: 1949–60.

McGlinn, D. J., T. Engel, S. A. Blowes, N. J. Gotelli, T. M. Knight, B. J. McGill, and J. M. Chase. 2021. "A Multiscale Framework for Disentangling the Roles of Evenness, Density, and Aggregation on Diversity Gradients." *Ecology* 102: e03233.

McGlinn, D. J., X. Xiao, F. May, N. J. Gotelli, T. Engel, S. A. Blowes, T. M. Knight, O. Purschke, J. M. Chase, and B. J. McGill. 2019. "Measurement of Biodiversity (MoB): A Method to Separate the Scale-Dependent Effects of Species Abundance Distribution, Density, and Aggregation on Diversity Change." *Methods in Ecology and Evolution* 10: 258–269.

Moreno, C. E., and P. Rodríguez. 2010. "A Consistent Terminology for Quantifying Species Diversity?" *Oecologia* 163: 282–297.

Mori, A. S., S. Fujii, R. Kitagawa, and D. Koide. 2015. "Null Model Approaches to Evaluating the Relative Role of Different Assembly Processes in Shaping Ecological Communities." *Oecologia* 178: 261–273.

Mori, A. S., F. Isbell, and R. Seidl. 2018. "β-Diversity, Community Assembly, and Ecosystem Functioning." *Trends in Ecology & Evolution* 33: 549–564.

Myers, J. A., J. M. Chase, I. Jiménez, P. M. Jørgensen, A. Araujo-Murakami, N. Paniagua-Zambrana, and R. Seidel. 2013. "Beta-Diversity in Temperate and Tropical Forests Reflects Dissimilar Mechanisms of Community Assembly." *Ecology Letters* 16: 151–57.

Phillips, O. L., and J. S. Miller. 2002. *Global Patterns of Plant Diversity: Alwyn H. Gentry's Forest Transect Data Set.* St. Louis: Missouri Botanical Garden Press.

Püttker, T., A. D. A. Bueno, P. I. Prado, and R. Pardini. 2015. "Ecological Filtering or Random Extinction? Beta-Diversity Patterns and the Importance of Niche-Based and Neutral Processes Following Habitat Loss." *Oikos* 124: 206–215.

Qian, H., S. Chen, L. Mao, and Z. Ouyang. 2013. "Drivers of β-Diversity along Latitudinal Gradients Revisited." *Global Ecology and Biogeography* 22: 659–670.

Routledge, R. 1979. "Diversity Indices: Which Ones Are Admissible?" *Journal of Theoretical Biology* 76: 503–515.

Shinozaki, K. 1963. "Notes on the Species-Area Curve." In *Abstracts of 10th Annual Meeting of Ecological Society of Japan*. Tokyo: Ecological Society of Japan.

Smith, E. P., P. M. Stewart, and J. Cairns. 1985. "Similarities between Rarefaction Methods." *Hydrobiologia* 120: 167–170.

Smith, W., and J. F. Grassle. 1977. "Sampling Properties of a Family of Diversity Measures." *Biometrics* 33: 283–292.

Thorn, S., C. Bässler, T. Gottschalk, T. Hothorn, H. Bussler, K. Raffa, and J. Müller. 2014. "New Insights into the Consequences of Post-Windthrow Salvage Logging Revealed by Functional Structure of Saproxylic Beetles Assemblages." *PLoS One* 9: e101757.

Thorn, S., C. Bässler, M. Svoboda, and J. Müller. 2017. "Effects of Natural Disturbances and Salvage Logging on Biodiversity—Lessons from the Bohemian Forests." *Forest Ecology and Management* 388: 113–19.

Thorn, S., S. A. B. Werner, J. Wohlfahrt, C. Bässler, S. Seibold, P. Quillfeldt, and J. Müller. 2016. "Response of Bird Assemblages to Windstorm and Salvage Logging—Insights from Analyses of Functional Guild and Indicator Species." *Ecological Indicators* 65: 142–48.

Tipper, J. C. 1979. "Rarefaction and Rarefiction—The Use and Abuse of a Method in Paleoecology." *Paleobiology* 5: 423–434.

Tuomisto, H. 2010. "A Diversity of Beta Diversities: Straightening up a Concept Gone Awry. Part 1. Defining Beta Diversity as a Function of Alpha and Gamma Diversity." *Ecography* 33: 2–22.

Tuomisto, H., and K. Ruokolainen. 2012. "Comment on "Disentangling the Drivers of β Diversity along Latitudinal and Elevational Gradients"." *Science* 335: 1573.

Ulrich, W., A. Baselga, B. Kusumoto, T. Shiono, H. Tuomisto, and Y. Kubota. 2017. "The Tangled Link between β-and γ-Diversity: A Narcissus Effect Weakens Statistical Inferences in Null Model Analyses of Diversity Patterns." *Global Ecology and Biogeography* 26: 1–5.

Ulrich, W., Y. Kubota, B. Kusumoto, A. Baselga, H. Tuomisto, and N. J. Gotelli. 2018. "Species Richness Correlates of Raw and Standardized Co-occurrence Metrics." *Global Ecology and Biogeography* 27: 395–99.

Wang, S., and M. Loreau. 2016. "Biodiversity and Ecosystem Stability across Scales in Metacommunities." *Ecology Letters* 19: 510–18.

Wermelinger, B. 2002. "Development and Distribution of Predators and Parasitoids during Two Consecutive Years of an Ips Typographus (Col., Scolytidae) Infestation." *Journal of Applied Entomology* 126: 521–27.

Whittaker, R. H. 1960. "Vegetation of the Siskiyou Mountains, Oregon and California." *Ecological Monographs* 30: 279–338.

Whittaker, R. H. 1972. "Evolution and Measurement of Species Diversity." *Taxon* 12: 213–251.

Wilson, M. V., and A. Shmida. 1984. "Measuring Beta-Diversity with Presence-Absence Data." *Journal of Ecology* 72: 1055–64.

Xing, D., and F. He. 2019. "Environmental Filtering Explains a U-Shape Latitudinal Pattern in Regional β-Deviation for Eastern North American Trees." *Ecology Letters* 22: 284–291.

Xing, D., and F. He. 2021. "Analytical Models for β-Diversity and the Power-Law Scaling of β-Deviation." *Methods in Ecology and Evolution* 12: 405–414.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.