Distance-Based Functional Diversity Measures and Their Decomposition: A Framework Based on Hill Numbers

Chun-Huo Chiu, Anne Chao*

Institute of Statistics, National Tsing Hua University, Hsin-Chu, Taiwan

Abstract

Hill numbers (or the "effective number of species") are increasingly used to characterize species diversity of an assemblage. This work extends Hill numbers to incorporate species pairwise functional distances calculated from species traits. We derive a parametric class of functional Hill numbers, which quantify "the effective number of equally abundant and (functionally) equally distinct species" in an assemblage. We also propose a class of mean functional diversity (per species), which quantifies the effective sum of functional distances between a fixed species to all other species. The product of the functional Hill number and the mean functional diversity thus quantifies the (total) functional diversity, i.e., the effective total distance between species of the assemblage. The three measures (functional Hill numbers, mean functional diversity and total functional diversity) quantify different aspects of species trait space, and all are based on species abundance and species pairwise functional distances. When all species are equally distinct, our functional Hill numbers reduce to ordinary Hill numbers. When species abundances are not considered or species are equally abundant, our total functional diversity reduces to the sum of all pairwise distances between species of an assemblage. The functional Hill numbers and the mean functional diversity both satisfy a replication principle, implying the total functional diversity satisfies a quadratic replication principle. When there are multiple assemblages defined by the investigator, each of the three measures of the pooled assemblage (gamma) can be multiplicatively decomposed into alpha and beta components, and the two components are independent. The resulting beta component measures pure functional differentiation among assemblages and can be further transformed to obtain several classes of normalized functional similarity (or differentiation) measures, including N-assemblage functional generalizations of the classic Jaccard, Sørensen, Horn and Morisita-Horn similarity indices. The proposed measures are applied to artificial and real data for illustration.

Introduction

Functional diversity quantifies the diversity of species traits in biological communities, and is widely regarded as a key to understanding ecosystem processes and environmental stress or disturbance [1–11]. A higher functional diversity signifies greater differences among species trait values, more distinct ecological functions, and thus potentially better functional stability to perturbations caused by human impacts or environment stresses [12–15]. Thus, it is critical to quantify functional diversity properly [16].

Functional diversity is typically quantified by using measures based on species trait values and species abundance (or any measure of species importance, e.g., cover or biomass). A wide array of functional diversity measures have been developed in the literature [4,17–22]; see [23–26] for reviews. There are three major approaches to construct functional diversity measures: trait-value-based [27], dendrogram-based [4,28–31], and distance-based [17,32–35]. For the trait-value-based approach, measures are calculated from species trait values directly. In the dendrogram-based approach, a functional dendrogram is constructed by applying a clustering algorithm to the species pairwise distance matrix. However, it has been shown that different clustering methods may lead to different conclusions [21,30,31]. An unavoidable issue in the dendrogram-based approach is how to select a clustering algorithm to construct a functional dendrogram.

This paper is focused on the distance-based approach which does not require a dendrogram. The selection of clustering algorithm can thus be avoided. A commonly used functional diversity index in the distance-based approach is *FAD* (Functional Attribute Diversity), the sum of pairwise distances between species [17]. However, *FAD* does not take into account species abundances. Rao's quadratic entropy Q and its transformations have also been extensively applied to quantify functional diversity [32–35]. The measure Q, a generalization of the traditional Gini-Simpson index, incorporates both species pairwise distances and species abundances. However, it inherits mathematical properties of the Gini-Simpson index which are inappropriate for a diversity measure [35,36–38]. The problems with interpreting Q as a diversity measure will be briefly discussed and illustrated by examples later in this paper. Ricotta and Szeidl [35] and de Bello et al. [39] resolved these problems by converting Q to "species equivalents"; see later text for details. However, we show here that their solution in its original form does not behave properly for non-ultrametric distance matrices. (A distance metric d is ultrametric if it satisfies the criterion $d(x, y) \leq \max\{d(x, z), d(y, z)\}$ for all x, y and z.)

In many applications, the distance matrices calculated from species traits do not satisfy this criterion and thus are non-ultrametric. For example, the commonly used Gower distance matrices calculated from three habitats in our real data (see *Examples and Applications*) are all non-ultrametric. Also, measures based on quadratic entropy gives common species much more weight than their population fraction. It would be more informative to have a parameter to control the sensitivity of the measure to species abundances. We were thus motivated to derive a new parametric class of measures that are valid for both ultrametric and non-ultrametric matrices.

Our framework is based on Hill numbers, a one-parameter family of diversity indices (differing among themselves only by a parameter q which determines the sensitivity to the relative abundances) that incorporate species richness and relative abundances. Hill numbers include species richness, Shannon diversity (the exponential of entropy) and Simpson diversity (inverse of the Simpson index). They were first used in ecology by MacArthur [40], developed by Hill [41], and recently reintroduced to ecologists by Jost [42,43]. A very brief description of Hill numbers is provided below.

Hill numbers are increasingly used to characterize abundance-based species diversity of an assemblage; see a series of papers in a recent forum [44]. An important advantage of using Hill numbers is that Hill numbers obey an intuitive *replication principle*, an essential mathematical property that captures biologists' intuitive notion of diversity [40,41]; see *Conclusion and Discussion* for more details. The replication principle requires that if we have N equally diverse, equally large assemblages with no species in common, the diversity of the pooled assemblage must be N times the diversity of a single assemblage. We refer to the special case of $N=2$ as a "doubling property" as defined in [36]. Hill numbers were recently extended to incorporate phylogenetic distance and dendrogram-based functional distance between species [45] while still satisfying the replication principle.

This work first generalizes Hill numbers to distance-based *functional Hill numbers*, which quantify "the effective number of equally abundant and (functionally) equally distinct species". Throughout this paper, species are *equally distinct* if all species pairwise functional distances are a fixed constant. To fully characterize distance-based functional diversity, we also need measures in units of "distance". The product of our functional Hill number and Rao's quadratic entropy (the abundance-weighted mean distance between species) quantifies the *mean functional diversity* (per species), i.e., the effective sum of functional distances between a fixed species to all other species (plus intraspecific distance if exists). The product of the functional Hill number and the mean functional diversity thus quantifies the *total functional diversity* (or simply *functional diversity*), the effective total distance between species of the assemblage. When all species are equally distinct, our functional Hill numbers reduce to ordinary Hill numbers. When species abundances are not considered or species are equally abundant, our total functional diversity reduces to *FAD* [17]. Thus our approach also extends *FAD* to incorporate species abundances. Different perspectives regarding the distance-based approaches and the replication principle can be found in [46,47,48] and [48], respectively.

When there are multiple assemblages defined by the investigator, Hill numbers can be multiplicatively partitioned into independent (or unrelated) alpha and beta components [43,49]. An advantage of our approach is that each of the three functional diversity measures we propose (functional Hill numbers, mean functional diversity and total functional diversity) can be used for complete multiplicative partitioning. The beta component can be

further transformed onto the range [0, 1] to obtain normalized measures of functional similarity (or differentiation), including N-assemblage functional generalizations of the classic Jaccard, Sørensen, Horn and Morisita-Horn similarity indices. Our framework thus unites functional diversity measures and functional similarity (or differentiation) among assemblages. Our previous work on Hill numbers covered taxonomic diversity and phylogenetic diversity. With the present development of functional Hill numbers, we now have a unified approach to quantifying and partitioning taxonomic, phylogenetic and functional diversities. Most previously-proposed diversity measures can be transformed into this framework and can be better understood through it. See Chao et al. [50] for an integrated framework.

Hill Numbers

In the traditional species diversity, only species richness and species abundances are considered. Assume there are S species in an assemblage and species are indexed by $i = 1, 2, ..., S$. Let p_i denote the relative abundance of the ith species. Hill [41] integrated species richness and species abundances into a parametric class of diversity measures later called Hill numbers, or the effective numbers of species, defined for $q \neq 1$ as

$$^qD = \left(\sum_{i=1}^{S} p_i^q \right)^{1/(1-q)}. \tag{1a}$$

The parameter q determines the sensitivity of the measure to the relative abundances. When $q = 0$, ^0D is simply species richness. For $q = 1$, Eq. 1a is undefined, but its limit as q tends to 1 is the exponential of the familiar Shannon entropy, and is referred to as Shannon diversity in [51]:

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

The measure ^1D weighs species in proportion to their abundances. When $q = 2$, Eq. 1a yields the inverse of the Simpson concentration which is referred to as Simpson diversity [51]:

$$^2D = 1 / \sum_{i=1}^{S} p_i^2. \tag{1c}$$

This measure places more weight on the abundant species and strongly discounts rare species. For all q, if $^qD = u$, the diversity (of order q) of the actual assemblage is the same as that of an idealized assemblage with u equally abundant species. This is why Hill numbers are referred to as the effective numbers of species or as species equivalents.

A complete characterization of the traditional abundance-based species diversity of an assemblage with S species and relative abundances $(p_1, p_2, ..., p_S)$ is conveyed by a diversity profile plotting ^qD versus q from $q = 0$ to $q = 3$ or 4 (beyond this it changes little) [52]. Although Hill numbers for $q < 0$ can be calculated, they are dominated by the abundances of rare species and have poor statistical sampling properties. We thus restrict ourselves to the case $q \geq 0$ throughout the paper.

Hill [41] proved a weak version of replication principle for Hill numbers: if two equally large assemblages with no species in common have identical relative abundance distributions, then the Hill number of the pooled assemblage is doubled. Chiu et al. (Appendix B of [36]) recently proved a strong version as given in

Introduction: the assumptions needed are that N assemblages with no species in common are equally large and equally diverse (relative abundance distributions may be different, unlike the weak version). Species richness is a Hill number (with $q = 0$) and obeys both versions of the replication principle, but most other traditional diversity indices do not obey even the weak version. The replication principle has been discussed for characterizing abundance-based species diversity measures [40–43,53–56]. This replication principle will be generalized to functional diversity measures in later sections.

Previous Distance-based Functional Diversity and Differentiation Measures

A large number of functional diversity measures have been proposed in the literature, and each measure quantifies a different aspect of species trait space. Here we mainly review those distance-based functional indices and differentiation measures that are related to our functional generalization of Hill numbers. The *FAD* measure is defined as [17]

$$FAD = \sum_{i=1}^{S} \sum_{j=1}^{S} d_{ij}, \qquad (2a)$$

where d_{ij} is the functional distance between the ith and jth species, $d_{ij} = d_{ji} \geq 0$. However, this measure does not take into account the abundances of the species, which may play an important role in the functioning of ecosystems; see [19,57–63].

Functional diversity measures combining both functional distance and species abundance have been proposed [20,32,33,34,64,65]. Rao's quadratic entropy for an assemblage with species relative abundances $(p_1, p_2, ..., p_S)$ is the most widely used measure [32]:

$$Q = \sum_{i=1}^{S} \sum_{j=1}^{S} d_{ij} p_i p_j. \qquad (2b)$$

The measure Q is interpreted as the mean distance between any two randomly sampled individuals in the assemblage. It can also be interpreted as the abundance-weighted mean distance between two species. Here the weighting factor for a distance d_{ij} is the product of the relative abundances, $p_i p_j$. This measure is an extension of the Gini-Simpson index. Although this measure has wide applications in many disciplines, Q, like the Gini-Simpson index, is not linear with respect to the addition of new species and thus does not obey the replication principle, causing counterintuitive results in ecological applications [35,36,66]. For an additive decomposition, another problem arises when the species functional distance matrix does not result in the concavity of Q [5,67] (e.g. for the Gower distance matrix, in general). Then Q in the pooled assemblage (gamma quadratic entropy) may be smaller than the Q of the average of local assemblages (alpha quadratic entropy), implying Q in this situation could not to be used for additive decomposition [67,68,69]. When additive decomposition is feasible, the associated differentiation measure that has been used in the literature is the quadratic entropy excess normalized by the gamma quadratic entropy [70,71]:

$$Q_{\beta}^* = \frac{Q_\gamma - Q_\alpha}{Q_\gamma}, \qquad (2c)$$

where Q_α and Q_γ denote respectively the alpha and gamma quadratic entropy. However, when alpha quadratic entropy is

high, the differentiation measure Q_{β}^* always tends to zero (implying no differentiation) regardless of distance matrices and differences in species abundances across assemblages [36]. This behavior leads to severe interpretational problems.

To fix the problems with Q, Ricotta and Szeidl [35] and de Bello et al. [39] made an advance by transforming Q to the "species equivalents", which is the effective number of equally distinct species with a constant distance d_{max} for all different-species pairs; here d_{max} denotes the maximum value in the distance matrix. Their transformation is expressed as

$$Q_e = \frac{1}{1 - Q/d_{max}}. \qquad (2d)$$

We refer to this number as "*the effective number of (equally distinct) species with maximum distance*". Equivalently, they scale all distances so they are between 0 and 1, by dividing each distance by its maximum value in the distance matrix prior to all analyses. de Bello et al. [39], Villéger et al. [72] and Escalas et al. [73] applied the above formula to gamma and alpha quadratic entropies and obtained the corresponding effective number of species for gamma (denoted here by $Q_{e,\gamma}$) and alpha (denoted by $Q_{e,\alpha}$), where the subscript "e" denotes "effective". The resulting beta based on a multiplicative decomposition is $Q_{e,\beta} = Q_{e,\gamma}/Q_{e,\alpha}$. de Bello [39] further transformed this beta diversity into a normalized differentiation measure so that the resulting measure is in the unit interval [0, 1]:

$$Q_{e,\beta}^* = (1 - 1/Q_{e,\beta})/(1 - 1/N), \qquad (2e)$$

where N denotes the number of assemblages. Villéger et al. [72] also proposed a normalized differentiation measure:

$$Q_{e,\beta}^{**} = (Q_{e,\beta} - 1)/(N - 1). \qquad (2f)$$

However, as we will show by examples, Eq. 2d and the two associated differentiation measures (given in Eqs. 2e and 2f) might yield un-interpretable results when they are applied to non-ultrametric distance matrices. This motivates our new approach which is valid for both ultrametric and non-ultrametric matrices.

Guiasu and Guiasu [38,74] proposed a class of distance-weighted Gini-Simpson index as follows:

$$GS_D = \sum_{i=1}^{S} \sum_{j=1}^{S} d_{ij} p_i p_j (1 - p_i p_j) = Q - \sum_{i=1}^{S} \sum_{j=1}^{S} d_{ij}(p_i p_j)^2. \qquad (2g)$$

They also proposed the corresponding measure for a multiplicative decomposition. We will show that the three measures (*FAD*, Q and GS_D) are closely related to our proposed measures. Leinster and Cobbold [75] derived a parametric class of measures sensitive to species similarity. Scheiner [63] also proposed a metric that integrates abundance, phylogeny, and function. Since both approaches are also based on a framework of Hill numbers, it is important to distinguish these two previous approaches from ours; see *Conclusion and Discussion* for more details. Neither Leinster and Cobbold's approach nor Scheiner's metric have been developed to construct normalized similarity (or differentiation) measures that can be applied to analyze datasets such as those discussed in *Examples and Applications*.

Proposed Functional Diversity Measures

A Simple Framework for Ordinary Hill Numbers. We first present a simple conceptual framework for ordinary Hill numbers. Then we extend it to obtain our proposed functional Hill numbers. The intuitive interpretation of the "effective number of species" implies that if an assemblage with S species and species abundance vector $(p_1, p_2, ..., p_S)$ has diversity D, then the diversity of this actual assemblage is the same as that of an idealized reference assemblage with D species and species abundance $(1/D, 1/D, ..., 1/D)$.

Now we construct the q-th power sum ($q \neq 1$) of the abundances $(p_1, p_2, ..., p_S)$ with unity weight for each species, i.e., $\sum_{i=1}^{S} p_i^q$; see Table 1. Taking the same function for the idealized reference assemblage, i.e., replacing S and $(p_1, p_2, ..., p_S)$ by D and $(1/D, 1/D, ..., 1/D)$ respectively, we obtain $\sum_{i=1}^{D} (1/D)^q = D^{1-q}$. Equating the two sums shows that D is the Hill number of order q:

$$\sum_{i=1}^{S} p_i^q = \sum_{i=1}^{D} \left(\frac{1}{D}\right)^q = D^{1-q} \Rightarrow D = {}^q D = \left(\sum_{i=1}^{S} p_i^q\right)^{1/(1-q)}.$$

This provides a simple and intuitive derivation of Hill numbers. This derivation facilitates the extension of Hill numbers to incorporate functional distances.

Functional Diversity Measures of an Assemblage. Let d_{ij} denote the functional distance between the ith and jth species, with $d_{ij} \geq 0$, and $d_{ij} = d_{ji}$. Denote the $S \times S$ symmetric pairwise distance matrix by $\Delta = [d_{ij}]$. In our approach, species functional distance, which quantifies the proximity of species in functional trait space, can be any type of symmetric matrix. To extend Hill numbers to incorporate functional distances between species, we consider a framework based on pairs of species [38,74]. That is, we consider a collection of all S^2 pairs of species: $\{(1, 1), (1, 2), (1, 3), ..., (S, S)\}$. The joint "relative abundance" or joint probability for each species-pair (i, j) is $p_i p_j$. Consider the matrix $\Omega = [p_i p_j]$, where the (i, j) element of the matrix is $p_i p_j$ (Table 1). Note that the mean distance between any two species weighted by their joint probability is Rao's quadratic entropy defined in Eq. 2b.

Analogous to the derivation of Hill numbers, we consider the q-th power sum ($q \neq 1$) of all elements of the matrix $\Omega = [p_i p_j]$ with weight d_{ij} for species pair (i, j), i.e., $\sum_{j=1}^{S} \sum_{i=1}^{S} d_{ij}(p_i p_j)^q$. A similar concept of the "effective number of equally abundant and equally distinct species" as in ordinary Hill numbers can be applied to the functional version as follows. When species are equally distinct with a constant pairwise distance, the quadratic entropy Q must be equal to this constant. An assemblage with the effective number of species D means that this assemblage has the same diversity as an idealized reference assemblage having D equally common and equally distinct species with a constant distance Q for all S^2 pairs of species. Here we have S^2 pairs because same-species pairs are included so that intraspecific variability can be considered when trait values are available at the individual level [25,76]. (If there is no intraspecific variability, then the distance for a same-species pair is set to be 0 and a common distance $Q^* = QD/(D-1)$ is set for different-species pairs; see Table 1. All measures derived in the following are still valid when intraspecific distance is zero, and all interpretations can be adapted to the case when there is no intraspecific variability.) For simplicity, our derivation and interpretations are mainly based on S^2 pairs of species.

Taking the same q-th power sum function ($q \neq 1$) for the idealized reference assemblage with a constant weight Q for all D^2 species pairs, we obtain $\sum_{i=1}^{D} \sum_{j=1}^{D} Q\left(\frac{1}{D}\frac{1}{D}\right)^q$. Equating the two sums

from the actual and the idealized reference assemblages leads to

$$\sum_{i=1}^{S} \sum_{j=1}^{S} d_{ij}(p_i p_j)^q = \sum_{i=1}^{D} \sum_{j=1}^{D} Q\left(\frac{1}{D}\frac{1}{D}\right)^q.$$

Then we can solve D and the solution given below is denoted by ${}^q D(Q)$:

$$D = {}^q D(Q) = \left[\sum_{i=1}^{S} \sum_{j=1}^{S} \frac{d_{ij}}{Q}(p_i p_j)^q \right]^{\frac{1}{2(1-q)}}, \quad q \neq 1. \qquad (3)$$

For $q = 1$, we define the following limit as our measure:

$${}^1 D(Q) = \lim_{q \to 1} {}^q D(Q) = \exp\left[-\frac{1}{2} \sum_{i=1}^{S} \sum_{j=1}^{S} \frac{d_{ij}}{Q} p_i p_j \log(p_i p_j). \right]$$

The measure ${}^q D(Q)$ is a function of the distance matrix $\Delta = [d_{ij}]$ and the joint probability matrix $\Omega = [p_i p_j]$. Here we express it as a function of the quadratic entropy Q to emphasize the important role of Q in the construction of other measures (see Eqs. 4a and 4b) and in the proof of the replication principle (discussed later). The measure ${}^q D(Q)$ is the dimension (the number of columns or rows) of the distance matrix of the idealized reference assemblage in Table 1. We refer to it as the *functional Hill number* of order q. The measure ${}^q D(Q)$ can be interpreted as "*the effective number of equally abundant and (functionally) equally distinct species*" with a constant distance Q for all species pairs. Thus if ${}^q D(Q) = v$, then the functional Hill number of order q of the actual assemblage is the same as that of an idealized assemblage having v equally abundant and equally distinct species with a constant distance Q for all species pairs; see Table 1 for illustration.

To derive measures in units of "distance", note that in the idealized reference assemblage, all columns and all rows have identical sums. We define the column (or row) sum as our proposed measure of *mean functional diversity (per species)*, ${}^q MD(Q)$, of order q:

$${}^q MD(Q) = [{}^q D(Q)] \times Q, \qquad (4a)$$

which quantifies the effective sum of pairwise distances between a fixed species and all other species (plus intraspecific distance if exists). In other words, ${}^q MD(Q)$ measures the dispersion per species in the functional trait space [18]. The product of the functional Hill numbers and the mean functional diversity thus quantifies the *total functional diversity* (or simply *functional diversity*), ${}^q FD(Q)$, in the assemblage:

$${}^q FD(Q) = {}^q D(Q) \times {}^q MD(Q) = [{}^q D(Q)]^2 \times Q$$

$$= \left[\sum_{i=1}^{S} \sum_{j=1}^{S} d_{ij} \left(\frac{p_i p_j}{Q}\right)^q \right]^{1/(1-q)}. \qquad (4b)$$

This functional diversity quantifies the effective total distance between species of the assemblage. If ${}^q FD(Q) = u$, then the effective total distance between species of the actual assemblage with quadratic entropy Q is the same as that of an idealized assemblage having $(u/Q)^{1/2}$ equally abundant and equally distinct species with a constant distance Q for all species pairs.

Table 1. A framework for Hill numbers, functional Hill numbers, mean functional diversity and (total) functional diversity of a single assemblage.

	Abundance vector/matrix	weights	q-th power sum ($q \neq 1$)	Equating the two q-th power sums
(1) Hill numbers				
Actual assemblage	S species with relative abundance vector:	Unity weight for each species	$\sum_{i=1}^{S} p_i^q$	$\sum_{i=1}^{S} p_i^q = \sum_{i=1}^{D} \left(\frac{1}{D}\right)^q = D^{1-q}$ $\Rightarrow {}^q D = \left(\sum_{i=1}^{S} p_i^q\right)^{1/(1-q)}$
	$(p_1, p_2, \ldots, p_S)$	$(1, 1, \ldots, 1)$		
Idealized reference assemblage	D equally-abundant species	Unity weight for each species	$\sum_{i=1}^{D} \left(\frac{1}{D}\right)^q = D^{1-q}$	(Hill number of order q)
	$\left(\frac{1}{D}, \frac{1}{D}, \cdots, \frac{1}{D}\right)$	$(1, 1, \ldots, 1)$		
(2) Functional Hill number, mean functional diversity and (total) functional diversity				
Actual assemblage	$S \times S$ matrix of the product of relative abundances for pairs of species $\begin{bmatrix} p_1^2 & p_1 p_2 & \cdots & p_1 p_S \\ p_2 p_1 & p_2^2 & \cdots & p_2 p_S \\ \vdots & \vdots & \ddots & \vdots \\ p_S p_1 & p_S p_2 & \cdots & p_S^2 \end{bmatrix}$	$S \times S$ distance matrix as weight $\begin{bmatrix} d_{11} & d_{12} & \cdots & d_{1S} \\ d_{21} & d_{22} & \cdots & d_{2S} \\ \vdots & \vdots & \ddots & \vdots \\ d_{S1} & d_{S2} & \cdots & d_{SS} \end{bmatrix}$	$\sum_{i=1}^{S} \sum_{j=1}^{S} d_{ij} (p_i p_j)^q$	$\sum_{i=1}^{S} \sum_{j=1}^{S} d_{ij} (p_i p_j)^q$ $= \sum_{i=1}^{D} \sum_{j=1}^{D} Q \left(\frac{1}{D} \frac{1}{D}\right)^q$ $= \sum_{i \neq j}^{D} \sum_{j=1}^{D} Q^* \left(\frac{1}{D} \frac{1}{D}\right)^q$
Idealized reference assemblage	$D \times D$ matrix of the product of equal relative abundances for pairs of species	$D \times D$ idealized distance matrix as weights	$\sum_{i=1}^{D} \sum_{j=1}^{D} Q \left(\frac{1}{D} \frac{1}{D}\right)^q$ Or $\sum_{i \neq j}^{D} \sum_{j=1}^{D} Q^* \left(\frac{1}{D} \frac{1}{D}\right)^q$	$\Rightarrow {}^q D = {}^q D(Q)$ $= \left[\sum_{i=1}^{S} \sum_{j=1}^{S} \frac{d_{ij}}{Q} (p_i p_j)^q\right]^{\frac{1}{2(1-q)}}$
	$\begin{bmatrix} \left(\frac{1}{D}\right)^2 & \left(\frac{1}{D}\right)^2 & \cdots & \left(\frac{1}{D}\right)^2 \\ & & \cdots & \\ \vdots & \vdots & \ddots & \vdots \\ \left(\frac{1}{D}\right)^2 & \left(\frac{1}{D}\right)^2 & \cdots & \left(\frac{1}{D}\right)^2 \end{bmatrix}$	$\begin{matrix} Q & Q & \cdots & Q \\ Q & Q & \cdots & Q \\ \vdots & \vdots & \ddots & \vdots \\ Q & Q & \cdots & Q \end{matrix}$		(Functional Hill number = number of rows or columns in the idealized distance matrix) ${}^q MD(Q) = [{}^q D(Q) - 1] \times Q^*$ $= [{}^q D(Q) \times Q$
		or		(Mean functional diversity = column/row sum in the idealized distance matrix)
		$\begin{bmatrix} 0 & Q^* & \cdots & Q^* \\ Q^* & 0 & \cdots & Q^* \\ \vdots & \vdots & \ddots & \vdots \\ Q^* & Q^* & \cdots & 0 \end{bmatrix}$		
		$Q^* = QD/(D-1)$		${}^q FD(Q)$ $= {}^q D(Q)[{}^q D(Q) - 1] \times Q^*$ $= {}^q D(Q) \times {}^q MD(Q)$
				(Total functional diversity = grand sum of the idealized distance matrix)

doi:10.1371/journal.pone.0100014.t001

Consider the following special cases to intuitively understand the meaning of our functional diversity measures and their relationships with *FAD* (Eq. 2a) or GS_D (Eq. 2g):

(a) When all species in the assemblage are equally distinct (i.e., $d_{ij} \equiv Q$ for all species pairs (i, j), for $i, j = 1, 2, \ldots, S$), the functional Hill number ${}^q D(Q)$ reduces to ordinary Hill number.

(b) For $q = 0$, ${}^0 D(Q) = (FAD/Q)^{1/2}$, ${}^0 MD(Q) = (FAD \times Q)^{1/2}$, and ${}^0 FD(Q) = FAD$, where *FAD* is defined in Eq. 2a. Thus, our measures have a direct link to *FAD*.

(c) If all species are equally abundant, then for any distance matrix (d_{ij}), we have ${}^q D(Q) = S$, and ${}^q FD(Q) = FAD$ for all orders of q. Therefore, when species abundances are not considered ($q = 0$) or species are equally abundant, our total functional diversity reduces to *FAD*. In the equally abundant case, we have ${}^q MD(Q) = {}^q FD(Q)/S$, implying that our mean functional diversity is conceptually similar to the Modified Functional Attribute Diversity (MFAD) proposed by Schmera et al. [18].

(d) When $q = 2$, we have the following link to the weighted Gini-Simpson index GS_D defined in Eq. 2g [38,74]:

$$^2FD(Q) = \left(\sum_{i=1}^{S} \sum_{j=1}^{S} \frac{d_{ij}}{Q^2}(p_i p_j)^2 \right)^{-1} = \frac{Q^2}{Q - GS_D}. \quad (4c)$$

As with the diversity profile for Hill numbers, a profile which plots $^qD(Q)$, $^qMD(Q)$ or $^qFD(Q)$ with respect to the order q completely characterizes the information each measure gives for an assemblage. As proved in Appendix S1, all three measures $^qD(Q)$, $^qMD(Q)$ and $^qFD(Q)$ are Schur-concave with respect to the product of relative abundances, implying these measures satisfy a functional version of "weak monotonicity" [45,77,78]. That is, if a rarest new species is added to an assemblage, then the measure $^qFD(Q)$ does not decrease regardless of distance matrices. Also, if a rarest new species is added to an assemblage such that the quadratic entropy remains unchanged, then all three measures do not decrease.

Functional Diversity Measures for a Pair of Assemblages. We next extend Rao's quadratic entropy, FAD, functional Hill number, mean functional diversity and total functional diversity to a pair of assemblages (say, I and II). Assume that there are S_1 species in Assemblage I and S_2 species in Assemblage II. Let the two sets of species relative abundances be denoted by $(p_{11}, p_{21}, ..., p_{S_1,1})$ and $(p_{12}, p_{22}, ..., p_{S_2,2})$ for Assemblage I and II respectively.

We first extend Rao's quadratic entropy to a pair of assemblages. Assume that an individual is randomly selected from each of the assemblages. Then the probability that the individual from Assemblage I belongs to species i and the individual from Assemblage II belongs to species j is $p_{i1}p_{j2}$, $i = 1, 2, ..., S_1$, $j = 1, 2, ..., S_2$. The mean distance between these two randomly selected individuals is

$$Q_{12} = \sum_{i=1}^{S_1} \sum_{j=1}^{S_2} d_{ij} p_{i1} p_{j2}. \quad (5a)$$

This measure can also be interpreted as the abundance-weighted mean distance between a species from Assemblage I and a species from Assemblage II, and the weighting factor is the product of their relative abundances. For simplicity, we refer to Q_{12} as the *mean distance between species of Assemblage I and Assemblage II*. Clearly, we have $Q_{12} = Q_{21}$. The traditional Rao's quadratic entropy for Assemblage I is simply Q_{11} for the same-assemblage pair (I, I) and the quadratic entropy for Assemblage II is simply Q_{22} for the same-assemblage pair (II, II).

We can apply a similar approach to that in Table 1 by conceptually thinking that there are two idealized assemblages, and each assemblage includes $^qD(Q_{12})$ equally abundant and equally distinct species such that the two actual assemblages and the two idealized assemblages have the same value of a given diversity measure. Replacing the joint probability matrix $\Omega = [p_i p_j]$ in Table 1 with the $S_1 \times S_2$ matrix $\Omega_{12} = [p_{i1} p_{j2}]$ and using parallel derivations, we obtain the following functional Hill number for Assemblage I and Assemblage II:

$$^qD(Q_{12}) = \left[\sum_{i=1}^{S_1} \sum_{j=1}^{S_2} \frac{d_{ij}}{Q_{12}}(p_{i1}p_{j2})^q \right]^{\frac{1}{2(1-q)}}. \quad (5b)$$

This measure is interpreted as "the effective numbers of equally abundant and equally distinct species in each of two assemblages,

with a constant distance Q_{12} between species of Assemblage I and Assemblage II". We also define the mean functional diversity of Assemblage I and II as $^qMD(Q_{12}) = {}^qD(Q_{12}) \times Q_{12}$, which quantifies the effective sum of pairwise distances between a fixed species in one assemblage and all species in the other assemblage. Then the product of $^qD(Q_{12})$ and $^qMD(Q_{12})$ quantifies the *total functional diversity* (or simply *functional diversity*) *of* Assemblage I and Assemblage II as

$$^qFD(Q_{12}) = {}^qD(Q_{12}) \times {}^qMD(Q_{12}) = \left[\sum_{i=1}^{S_1} \sum_{j=1}^{S_2} d_{ij} \left(\frac{p_{i1}p_{j2}}{Q_{12}} \right)^q \right]^{\frac{1}{1-q}} \quad (5c)$$

In the special case of $q = 0$, the above total functional diversity reduces to the total sum of all pairwise distances between species of Assemblage I and Assemblage II. Since Q_{12} is not involved in the measure for $q = 0$, we denote $^0FD(Q_{12}) \equiv FAD_{12}$, which represents an extension of Walker's FAD to a pair of assemblages. Thus FAD_{11} is identical to FAD for Assemblage I and FAD_{22} is identical to FAD for Assemblage II. Also, we have the following relationship:

$$^0MD(Q_{12}) = (FAD_{12} \times Q_{12})^{1/2}. \quad (5d)$$

Replication Principle. We generalize the concept of the replication principle to a functional version and show that the proposed functional Hill numbers and the mean functional diversity both satisfy the replication principle. Consequently, the product of these two measures (i.e., our proposed total functional diversity) satisfies a quadratic replication principle (i.e., the total functional diversity of the pooled assemblage is $\mathcal{N}^2$ times that of any individual assemblage.) A general proof of the replication principle for $\mathcal{N}$ completely distinct assemblages is given in Appendix S1. Throughout this paper, $\mathcal{N}$ assemblages are *completely distinct* if there are no shared species (and thus no shared species pairwise distances).

To simplify the concept, here we present the replication principle only for two assemblages. Assume that two equally large and completely distinct assemblages are pooled. Let Q_{11}, Q_{12}, Q_{21}, and Q_{22} denote respectively the mean distance between species of the four pairs of assemblages, (I, I), (I, II), (II, I) and (II, II). Assume that the functional Hill number of order q for all of the four pairs of assemblages is a constant ^qD. When the two assemblages are combined, the quadratic entropy in the pooled assemblage becomes $(1/4)(Q_{11} + Q_{12} + Q_{21} + Q_{22})$ and the functional Hill number of order q in the pooled assemblage is doubled. Consequently, if we further assume that the four mean distances (Q_{11}, Q_{12}, Q_{21} and Q_{22}) are identical, then the mean functional diversity in the pooled assemblage is also doubled, and the total functional diversity is quadrupled; see Appendix S1 for a general proof for $\mathcal{N}$ assemblage.

In Guiasu and Guiasu's work on the quadrupling property [74], they proved a weak version of the quadrupling property for their proposed weighted Gini-Simpson type index (Eq. 4c) when two equally large and completely distinct assemblages (I and II) are pooled. They assume that the joint probability matrices for the four pairs of assemblages, (I, I), (I, II), (II, I) and (II, II), are identical, and also assume that the species distance matrices for the four pairs of assemblages are also identical. The latter assumption implies the FAD for the four pairs is a constant (say, A), i.e., $FAD_{11} = FAD_{12} = FAD_{21} = FAD_{22} \equiv A$. This weak version can be directly used to understand why the functional diversity of order zero (i.e., FAD) satisfies a quadrupling property. In this simple case,

consider the distance matrix of the pooled assemblage when the two actual assemblages have no species shared. It is readily seen that the total distance between species in the pooled assemblage is quadrupled because the *FAD* in the pooled assemblage is $FAD_{11}+FAD_{12}+FAD_{21}+FAD_{22} = 4 \times A$. As shown in the proof (Appendix S1), our replication principle is a strong version in the sense that there are no restrictions on the joint probability matrices and on the distance matrices.

Partitioning Functional Diversity Measures

Assume that there are N assemblages defined by the investigator. The functional Hill number ${}^qD(Q)$ (Eq. 3), mean functional diversity ${}^qMD(Q)$ (Eq. 4a) and total functional diversity ${}^qFD(Q)$ (Eq. 4b) of the pooled assemblage can each be decomposed into independent alpha and beta components. Here we adopt the decomposition method developed by Chiu et al. [36]. We first briefly review Chiu et al.'s method for partitioning Hill numbers in order to provide necessary backgrounds. To calculate the gamma Hill numbers, species abundances are pooled over assemblages; the gamma diversity is Hill numbers computed from the species relative abundances in the pooled assemblage. However, as noted by de Bello et al. [39], how to define species relative abundances in the pooled assemblage depends on how we pool data over assemblages. The pooling scheme depends on the objective of the study. If the objective is to compare absolute abundances among assemblages (given the total abundance in the pooled assemblage), we should pool individuals over assemblages; if the objective is to compare relative abundances among assemblages, we should pool relative abundances of individual assemblages. These two kinds of pooling schemes lead to different relative abundances in the pooled assemblage. (The former is equivalent to an assemblage-size-weighted pooling scheme for relative abundances, whereas the latter naturally reduces to the equal-assemblage-weighted pooling scheme.) See Chao et al. [50, their Appendix 2] for a simple example for illustration.

Suppose in the pooled assemblage there are S species indexed by 1, 2, …, S. To incorporate both kinds of pooling schemes into our framework, we define z_{ik} as any measure of species "abundance" of the ith species in the kth assemblage, $i = 1, 2, …, S, k = 1, 2, …, N$. Some of the z_{ik} may be zero. The measure z_{ik} can be absolute abundances, relative abundances, incidence, biomasses, cover areas or any other species importance measure. Define $z_{+k} = \sum_{i=1}^{S} z_{ik}$ as the size of the kth assemblage. Let $z_{++} = \sum_{k=1}^{N} z_{+k}$ be the total abundance in pooled assemblage and $z_{i+} = \sum_{k=1}^{N} z_{ik}$ be the total abundances of the ith species in the pooled assemblage. Then the species relative abundance set in the pooled assemblage for both pooling schemes can be expressed as $\{z_{i+}/z_{++}; i = 1, 2, …, S\}$. Note that if we pool relative abundances over assemblages instead of absolute abundance, we have the special case that $z_{+k} = 1$ and $z_{++} = N$.

The abundance-based gamma diversity is Hill numbers computed from the species relative abundances $\{z_{i+}/z_{++}; i = 1, 2, …, S\}$ and is interpreted as the effective number of species in the pooled assemblage. The traditional definition of alpha diversity is "the mean of the diversities of individual assemblages". Routledge [55] and Jost [43] each derived a mathematical formula for alpha diversity based on this traditional definition and obtained the corresponding multiplicative beta component. As Chiu et al. [36] indicated, this traditional approach to alpha diversity based on Hill numbers leads to a beta that can only be used to produce differentiation measures to compare species *relative* abundances, but not *absolute* abundances. This is because in the framework of Hill numbers, diversity is a function of relative abundances only,

and thus "the mean of individual diversities" lose information about absolute abundances. Chiu et al. [36] expanded the conventional concept of alpha and proposed a modified definition for abundance-based alpha diversity: "alpha diversity is the effective number of species per assemblage" so that the resulting beta can be transformed to quantify the differentiation (or similarity) among N sets of vectors $(z_{1k}, z_{2k}, …, z_{Sk})$, $k = 1, 2, …, N$, for any measure of species importance z_{ik}, including absolute abundances. Based on this expanded definition, Chiu et al. derived a new formula for abundance-based alpha diversity.

When the data represent species relative abundances (i.e., equal-weight for assemblages), all three alpha formulas (Routledge, Jost and Chiu et al.) are identical. They differ, however, when the data represent species absolute abundances (i.e., assemblage-size as weight); Routledge's beta may exceed N and Jost's beta may be less than unity (i.e., gamma may be less than alpha) if $q \neq 1$. Also, for $q = 0$, Routledge's alpha and beta are not independent [49]. Chiu et al.'s new formula of the abundance-based alpha diversity has the following advantages: (1) it leads to a beta that can be applied to compare *any* types of data (z_{ik}), depending on the investigator's objective; (2) gamma is always greater than or equal to alpha for *all* orders $q \geq 0$; (3) beta is always between unity (when all assemblages are identical in species absolute abundances) and N (when the N assemblages have no species in common); and (4) alpha and beta components are independent for *all* orders $q \geq 0$.

Based on species abundance (z_{ik}) and a species pairwise distance matrix, we now extend Chiu et al. approach to decompose the functional diversity ${}^qFD(Q)$ of the pooled assemblage into within-assemblage component (functional alpha diversity) and between-assemblage component (functional beta diversity). As with the partition of Hill numbers, the functional gamma diversity of order q is based on the distance matrix $[d_{ij}]$ and the relative abundance $\bar{p}_{i+} = z_{i+}/z_{++}$ in the pooled assemblage. It then follows from Eq. 4b, the *functional gamma diversity* of order q can be written as

$$
\begin{aligned}
{}^qFD_\gamma(Q) &= \left[\sum_{i=1}^{S} \sum_{j=1}^{S} d_{ij} \left(\frac{\bar{p}_{i+} \bar{p}_{j+}}{Q} \right)^q \right]^{1/(1-q)} \\
&= \left[\sum_{i=1}^{S} \sum_{j=1}^{S} d_{ij} \left(\frac{z_{i+} z_{j+}}{Q z_{++}^2} \right)^q \right]^{1/(1-q)} \quad q \geq 0, q \neq 1,
\end{aligned}
\tag{6a}
$$

where $Q = \sum_i \sum_j d_{ij} \bar{p}_{i+} \bar{p}_{j+}$ is the quadratic entropy in the pooled assemblage. The limit when q approaches unity exists and is equal to

$$
\begin{aligned}
{}^1FD_\gamma(Q) &= \lim_{q \to 1} {}^qFD_\gamma(Q) \\
&= \exp\left[- \sum_{i=1}^{S} \sum_{j=1}^{S} d_{ij} \left(\frac{\bar{p}_{i+} \bar{p}_{j+}}{Q} \right) \log\left(\frac{\bar{p}_{i+} \bar{p}_{j+}}{Q} \right) \right].
\end{aligned}
\tag{6b}
$$

The functional gamma diversity is interpreted as the effective total distance between species in the pooled assemblage with a constant distance Q for all species pairs.

We follow Chiu et al.'s definition of alpha diversity to define the functional alpha diversity as the effective total distance between species of a pair of individual assemblages. Then we obtain (details of derivation are provided in Appendix S2) the *functional alpha diversity* of order q:

$$^{q}FD_{\alpha}(Q) = \frac{1}{N^{2}}\left[\sum_{k,m=1}^{N}\sum_{i,j=1}^{S}d_{ij}\left(\frac{z_{ik}z_{jm}}{Qz_{++}^{2}}\right)^{q}\right]^{1/(1-q)}, q \geq 0, q \neq 1. \quad (7a)$$

$$^{1}FD_{\alpha}(Q) = \lim_{q \to 1}{}^{q}FD_{\alpha}(Q)$$
$$= \frac{1}{N^{2}}\exp\left[-\sum_{k,m=1}^{N}\sum_{i,j=1}^{S}d_{ij}\left(\frac{z_{ik}z_{jm}}{Qz_{++}^{2}}\right)\log\left(\frac{z_{ik}z_{jm}}{Qz_{++}^{2}}\right)\right]. \quad (7b)$$

Note that in our alpha formula, Q still refers to the quadratic entropy for the *pooled* assemblage. When relative abundances are the relevant quantities for the investigative question, we simply define the measure z_{ik} as the ith species relative abundance in the kth assemblage. Then $z_{++} = N$ and thus z_{++} is replaced by N in all of the above formulas, Eqs. 6a, 6b, 7a and 7b.

As with ordinary Hill numbers [36,43], the complete partitioning of functional gamma diversity into independent within- and between-assemblage (alpha and beta) components is multiplicative. That is, the *functional beta diversity* is the ratio of functional gamma to functional alpha diversities:

$$^{q}FD_{\beta}(Q) = \frac{{}^{q}FD_{\gamma}(Q)}{{}^{q}FD_{\alpha}(Q)}. \quad (8)$$

The decomposition procedures for the other two measures are generally parallel and interpretations are similar. A summary of the decomposition of the three measures with interpretations is given in Table 2 and the formulas for decomposing functional Hill numbers are provided in Appendix S2. Note that for functional Hill number and mean functional diversity, their beta components are identical, i.e., $^{q}D_{\beta}(Q) = {}^{q}MD_{\beta}(Q)$. Also, we have $^{q}FD_{\beta}(Q) = [{}^{q}D_{\beta}(Q)]^{2}$.

For each of the three measures ($^{q}D(Q)$, $^{q}MD(Q)$ and $^{q}FD(Q)$), the gamma value is always greater than or equal to the corresponding alpha component for all orders $q \geq 0$ and all distance matrices; see Appendix S2 for a proof. When N assemblages are identical in species identities and abundance, the beta components of all three measures take their minimum value of unity. When all assemblages are completely distinct (no shared species and thus no shared pairwise distances), we have $^{q}D_{\beta}(Q) = {}^{q}MD_{\beta}(Q)$ and both attain the maximum value of N, and $^{q}FD_{\beta}(Q)$ attains the

maximum value of N^{2}. The functional beta Hill number, $^{q}D_{\beta}(Q)$, thus quantifies the effective number of *equally large and completely distinct assemblages*. The functional beta diversity, $^{q}FD_{\beta}(Q)$, quantifies the effective number of *equally large and completely distinct pairs of assemblages*. In Appendix S2, we show that $^{q}D_{\beta}(Q)$ is always between unity and N; and $^{q}FD_{\beta}(Q)$ is always between unity and N^{2}. Thus, the range of each beta component is independent of the corresponding alpha component, implying that the alpha and beta components based on the multiplicative partitioning for each of the three functional diversity measures ($^{q}D(Q), {}^{q}MD(Q), {}^{q}FD(Q)$) are unrelated (or independent).

We also note the following properties:

(1) When all species are equally distinct, the functional beta Hill numbers $^{q}D_{\beta}(Q)$ reduce to the beta diversity for ordinary Hill numbers, and the functional beta diversity $^{q}FD_{\beta}(Q)$ reduces to the squared beta diversity of ordinary Hill numbers [36].

(2) When $q = 0$, we have $^{0}D_{\gamma}(Q) = (FAD_{\gamma}/Q)^{1/2}$, $^{0}MD_{\gamma}(Q) = (FAD_{\gamma} \times Q)^{1/2}$ and $^{0}FD_{\gamma}(Q) = FAD_{\gamma}$, where FAD_{γ} denotes the sum of pairwise distances in the pooled assemblage. For the alpha components, we have $^{0}D_{\alpha}(Q) = (FAD_{pair}/Q)^{1/2}/N$, $^{0}MD_{\alpha}(Q) = (FAD_{pair} \times Q)^{1/2}/N$ and $^{0}FD_{\alpha}(Q) = FAD_{pair}/N^{2}$, where FAD_{pair} is the sum of FADs over all possible pairs of assemblages (there are N^{2} pairs of assemblages). So the functional alpha diversity for $q = 0$ is the average of FAD per pair of assemblages. Therefore, we have $^{0}D_{\beta}(Q) = {}^{0}MD_{\beta}(Q) = N(FAD_{\gamma}/FAD_{pair})^{1/2}$ and $^{0}FD_{\beta}(Q) = N^{2}FAD_{\gamma}/FAD_{pair}$.

(3) For the special case $q = 1$ and $q = 2$, we will present formulas for some special transformations of the functional beta diversity in the next section.

Four Classes of Normalized Functional Similarity Measures

Our functional beta components, $^{q}D_{\beta}(Q)(= {}^{q}MD_{\beta}(Q))$ and $^{q}FD_{\beta}(Q)$, all quantify pure functional differentiation among the N assemblages and their ranges depends only on N. Thus, each can be transformed to obtain the normalized similarity and differentiation measures in [0, 1] so that the dependence on N can be removed [36,43,49,79]. As stated in the preceding section and proved in Appendix S2, the range of $^{q}D_{\beta}(Q)$ is different from that of $^{q}FD_{\beta}(Q)$, so the transformations to normalized similarity measures in [0, 1] are thus different for these two beta

Table 2. Decomposition of the functional Hill number $^{q}D(Q)$ (Eq. 3), the mean functional diversity $^{q}MD(Q)$ (Eq. 4a) and the (total) functional diversity $^{q}FD(Q)$ (Eq. 4b) along with interpretations.

Measure	Functional Hill number $^{q}D(Q)$	Mean functional diversity $^{q}MD(Q)$	(Total) functional diversity $^{q}FD(Q)$
Gamma	$^{q}D_{\gamma}(Q)$	$^{q}MD_{\gamma}(Q) = Q \times [{}^{q}D_{\gamma}(Q)]$	$^{q}FD_{\gamma}(Q) = Q \times [{}^{q}D_{\gamma}(Q)]^{2}$
	The effective number of species in the pooled assemblage	The effective mean distance between species in the pooled assemblage	The effective total distance between species in the pooled assemblage (Eq. 6)
Alpha	$^{q}D_{\alpha}(Q)$	$^{q}MD_{\alpha}(Q) = Q \times [{}^{q}D_{\alpha}(Q)]$	$^{q}FD_{\alpha}(Q) = Q \times [{}^{q}D_{\alpha}(Q)]^{2}$
	The effective number of species in an individual assemblage	The effective mean distance between species in an individual assemblage	The effective total distance between species in a pair of local assemblage (Eq. 7)
Beta	$^{q}D_{\beta}(Q) = \frac{{}^{q}D_{\gamma}(Q)}{{}^{q}D_{\alpha}(Q)}$	$^{q}MD_{\beta}(Q) = \frac{{}^{q}MD_{\gamma}(Q)}{{}^{q}MD_{\alpha}(Q)} = {}^{q}D_{\beta}(Q)$	$^{q}FD_{\beta}(Q) = \frac{{}^{q}FD_{\gamma}(Q)}{{}^{q}FD_{\alpha}(Q)} = [{}^{q}D_{\beta}(Q)]^{2}$
	The effective number of equally large and completely distinct assemblages	The effective number of equally large and completely distinct assemblages	The effective number of equally large and completely distinct assemblage pairs

components. The similarity measures based on $^qD_\beta(Q)$ quantify *species-overlap* from different perspectives, whereas the similarity measures based on $^qFD_\beta(Q)$ quantify *distance-overlap* from different perspectives. In most applications of functional diversity, we suggest using the distance-overlap similarity measures and their corresponding differentiation measures. We now describe the two major classes of normalized distance-overlap measures based on $^qFD_\beta(Q)$; see Table 3 for all formulas and Appendix S3 for details.

(1) A class of local distance-overlap measures from the perspective of a pair of local assemblages

$$C^*_{qN}(Q) = \frac{N^{2(1-q)} - [^qFD_\beta(Q)]^{1-q}}{N^{2(1-q)} - 1}. \qquad (9a)$$

Here "local" refers to a property of a pair of individual assemblage. This measure gives the effective average proportion of the species pairwise distances in a pair of local assemblages that are shared with all other pairs of assemblages. See Appendix S3 for the interpretation of the "effective" average proportion.

We first give the special case of $q=0$ to intuitively explain its properties: since $^0FD_\gamma(Q) = FAD_\gamma$ and $^0FD_\alpha(Q) = FAD_{pair}/N^2$, the measure $C^*_{0N}(Q)$ reduces to

$$C^*_{0N}(Q) = \frac{(FAD_{pair} - FAD_\gamma)/(N^2 - 1)}{FAD_{pair}/N^2}, \qquad (9b)$$

where FAD_γ and FAD_{pair} are defined in the preceding section. In this expression, the denominator is the zero-order functional alpha diversity, which is the average of FADs over all N^2 assemblage pairs; the numerator is the average of all repeated pairwise distances in the $N^2 - 1$ pairs of assemblages (excluding the assemblage-pair in which a pairwise distance is first counted). The measure $C^*_{0N}(Q)$ thus quantifies the proportion of repeated distances in a pair of local assemblages. This interpretation is conceptually similar to the traditional Sørensen similarity index. The difference is that here we consider "assemblage-pairs" in functional distance-overlap measures rather than "individual assemblage" as in the traditional measure. Thus, this index can be regarded as an extension of the Sørensen index to functional similarity. Therefore, the measure $C^*_{0N}(Q)$ is referred to as "func-Sørensen" in Table 3. For $q=1$, this local distance-overlap measure is called "func-Horn" in Table 3 because when all distances are identical and (z_{ik}) represents species relative abundance within each assemblage, it reduces to the classical Horn measure [80]. For $q=2$, this measure is called "func-Morisita-Horn" in Table 3 because its interpretation is generally similar to the classic Morisita-Horn measure [81]. See Appendix S3 for more details.

(2) A class of regional distance-overlap measures in the pooled assemblage

$$U^*_{qN}(Q) = \frac{[1/^qFD_\beta(Q)]^{1-q} - (1/N)^{2(1-q)}}{1 - (1/N)^{2(1-q)}}. \qquad (10a)$$

Here "regional" refers to a property of the pooled assemblage. This class of measures differs from the local distance-overlap measures by taking a regional perspective. It quantifies the effective proportion of the species pairwise distances in the pooled assemblage that are shared with all pairs of local assemblages (Appendix S3).

For the special case of $q=0$, the measure $U^*_{0N}(Q)$ reduces to

$$U^*_{0N}(Q) = \frac{(FAD_{pair} - FAD_\gamma)/(N^2 - 1)}{FAD_\gamma}. \qquad (10b)$$

Again, the interpretation is similar to the measure $C^*_{0N}(Q)$ (in Eq. 9b) except that the denominator is replaced by the gamma FAD. Thus, our index can be regarded as an extension of the Jaccard

Table 3. Two major classes of distance-overlap (or similarity) measures and their special cases based on the functional beta diversity $^qFD_\beta(Q)$.

Order q	Local distance-overlap	Regional distance-overlap
	$C^*_{qN}(Q)$ $= \dfrac{N^{2(1-q)} - [^qFD_\beta(Q)]^{1-q}}{N^{2(1-q)} - 1}$	$U^*_{qN}(Q)$ $= \dfrac{[1/^qFD_\beta(Q)]^{1-q} - (1/N)^{2(1-q)}}{1 - (1/N)^{2(1-q)}}$
$q = 0$	Func-Sørensen $\dfrac{N^2 - N^2(FAD_\gamma/FAD_{pair})}{N^2 - 1}$	Func-Jaccard $\dfrac{FAD_{pair}/FAD_\gamma - 1}{N^2 - 1}$
$q = 1$	Func-Horn $1 - \dfrac{\log{^1FD_\gamma(Q)} - \log{^1FD_\alpha(Q)}}{2\log N}$	
$q = 2$	Func-Morisita-Horn $\dfrac{\sum_{i,j}^S d_{ij} \sum_{k,m}^N \left[\left(\frac{z_{i+}z_{j+}}{N}\right)^2 - (z_{ik}z_{jm})^2\right]}{(N^2-1)\sum_{i,j}^S \sum_{k,m}^N d_{ij}(z_{ik}z_{jm})^2}$	Func-regional-overlap $\dfrac{\sum_{i,j}^S d_{ij} \sum_{k,m}^N \left[\left(\frac{z_{i+}z_{j+}}{N}\right)^2 - (z_{jk}z_{jm})^2\right]}{(1-1/N^2)\sum_{i,j}^S d_{ij}(z_{i+}z_{j+})^2}$

The corresponding differentiation measures are the one-complements of the similarity measures. (The indices i and j are used to identify species, $i, j = 1, 2, \ldots, S$, and the indices k and m are used to identify assemblages, $k, m = 1, 2, \ldots, N$.)
Notation.
z_{ik} = the abundance of the ith species in the kth assemblage, $z_{+k} = \sum_{i=1}^S z_{ik}$, $z_{i+} = \sum_{k=1}^N z_{ik}$, and $z_{++} = \sum_{i=1}^S \sum_{k=1}^N z_{ik}$; see text for details. FAD_γ = sum of the pairwise distances between species in the pooled assemblage; FAD_{pair} = sum of FAD over all possible pairs of assemblages (there are N^2 pairs of assemblages).
S = species richness in the pooled assemblage. $\bar{S}$ = average species richness per assemblage.
doi:10.1371/journal.pone.0100014.t003

index to functional similarity. This is why the measure $U_{0N}^*(Q)$ is referred to as the N-assemblage "func-Jaccard" in Table 3. Since $U_{1N}^*(Q) = C_{1N}^*(Q)$, the measure $U_{qN}^*(Q)$ for $q = 1$ is also called "func-Horn" measure. For $q = 2$, this measure is referred to as "func-regional-overlap" measure; see Appendix S3.

As shown in Chiu et al. [36], we can also define two additional classes of functional distance-overlap measures. First, a class of functional distance-homogeneity measures which is a linear function of the inverse of the functional beta diversity:

$$S_{qN}^*(Q) = \frac{1/[^qFD_\beta(Q)] - 1/N^2}{1 - 1/N^2}. \quad (11)$$

Second, we have a class of measures which is a linear function of the functional beta diversity:

$$V_{qN}^*(Q) = \frac{N^2 - {}^qFD_\beta(Q)}{N^2 - 1}. \quad (12)$$

Its complement quantifies the functional distance-turnover rate.

All the four classes of similarity measures are continuous in $q \geq 0$, so a functional similarity or differentiation profile as a function of q can be made for any of them. We suggest using this method for conveying complete information about the functional similarity or differentiation of a set of assemblages. It is thus sufficient to focus on the two major classes ($C_{qN}^*(Q)$ and $U_{qN}^*(Q)$) because they include $S_{qN}^*(Q)$ and $V_{qN}^*(Q)$ as special cases for $q = 0$ and 2. See Example 3 for illustrative profiles.

Our decomposition presented above is based on the multiplicative scheme. We can also apply the additive decomposition to each of the three measures, ${}^qD(Q)$ (Eq. 3), ${}^qMD(Q)$ (Eq. 4a) and ${}^qFD_\beta(Q)$ (Eq. 4b). For example, we can define the "functional diversity excess" as ${}^qFD_\gamma(Q) - {}^qFD_\alpha(Q)$. The excess quantifies the effective total distances between species in the pooled assemblage not contained in a typical pair of local assemblages. As with ordinary Hill numbers [49], the functional diversity excess depends not only on the number of assemblages N, but also on the functional alpha diversity. Consequently, the excess measure cannot be directly applied to compare the similarity or differentiation among assemblages across multiple sets of assemblages even if the numbers of assemblages in these multiple regions are the same. We can eliminate these dependences by using appropriate normalizations [49]. In Appendix S4, we show that after proper normalizations, the multiplicative approach and additive approach both lead to the same four classes of normalized functional similarity and differentiation measures presented above. Thus, a consensus can be achieved on functional similarity and differentiation measures, including those measures given in Table 3.

Examples and Applications

To examine the performance of our functional diversity measures and to compare our proposed similarity and differentiation measures with previous indices, we use both artificial distance matrices (Examples 1 and 2) and real data (Example 3) for illustration. Although the distance matrices considered in our artificial examples are simple, they provide transparent answers so that we can clearly examine the performance of measures. Any meaningful differentiation measure should work properly for *all* matrices. If a functional diversity or similarity/differentiation measure cannot yield logical and sensible results for simple matrices, we would not expect it to work for complicated cases. The more complicated distance matrix calculated from real species traits is used in Example 3 for illustration.

In our comparisons, we consider various functional differentiation measures: (1) a differentiation measure (Eq. 2c) based on the traditional additive decomposition of quadratic entropy; (2) two differentiation measures (Eqs. 2e and 2f) based on the effective number of species with maximum distance; and (3) the proposed distance-based differentiation measures derived from our functional beta diversity (in Table 3 of this paper). Appendix S5 presents a simple example to show that the traditional measure based on the additive partitioning of the quadratic entropy (Eq. 2c) cannot work properly even for a very simple functional distance matrix; see Chiu et al. [36] for theoretic discussions and more examples.

Example 1: Effect of Functional Distances on Differentiation Measures. Consider two assemblages (I and II). Each assemblage contains 20 species, with 12 shared species and 8 non-shared species. There are 28 species in the pooled assemblage. For each assemblage, we first consider the equally abundant case in order to examine how differentiation measures vary with functional distances. (Two non-equally-abundant cases are given in Appendix S5.) The classical Sørensen-type dissimilarity index (the proportion of non-shared species in an individual assemblage) is $8/20 = 0.4$. (The abundance-based local differentiation measure based on Hill numbers is 0.4 for all $q \geq 0$; see [36].) The classical Jaccard-type dissimilarity index (the proportion of non-shared species in the pooled assemblage) is $1 - 12/28 = 0.571$; see Table 4 for abundance-based regional differentiation measure based on Hill numbers [36]. For functional differentiation measures, the quantifying target is shifted to the proportion of the total non-shared distances (incorporating abundances if $q > 0$) in an individual assemblage $(1 - C_{qN}^*(Q))$ or in the pooled assemblage $(1 - U_{qN}^*(Q))$.

We generated two contrasting types of distance matrices (Matrix I and Matrix II). Both matrices are displayed in Appendix S6. For easy presentation, species are indexed by 1, 2, ..., 28 in the pooled assemblage. Assemblage I includes Species 1–20, and Assemblage II includes Species 9–28 (Species 9–20 are shared). In Matrix I, the distances for two species within an assemblage follow the same distribution as those for species from the pooled assemblage so that the alpha quadratic entropy Q_α (the average distance between any two individuals within an assemblage) is close to the gamma quadratic entropy Q_γ (the average distance between any two individuals in the pooled assemblage). In this case, we expect that any meaningful functional differentiation measure is largely determined by species abundances. In Matrix II, the gamma quadratic entropy Q_γ is much higher than the alpha quadratic entropy Q_α, as described below. Consequently, we expect that functional distances should play an important role in characterizing functional differentiation.

(a) **Matrix I.** All the species pairwise distances in the 28×28 distance matrix of the pooled assemblage were generated from a beta (4, 4) distribution, which is a symmetric distribution with respect to 0.5. In this case, the alpha quadratic entropy ($Q_\alpha = 0.47$) is close to the gamma quadratic entropy ($Q_\gamma = 0.48$).

(b) **Matrix II.** We constructed the 28×28 distance matrix by generating substantially larger distances for pairs of "non-shared species" (s1, s2), where the first species s1 is a non-shared species in Assemblage I, and the second species s2 is a non-shared species in Assemblage II. The distances for such pairs of non-shared species were generated from a uniform (0.8, 1) distribution whereas the distances for other species pairs were generated from a uniform (0, 0.2) distribution. We have $Q_\alpha = 0.102$ and $Q_\gamma = 0.167$. There is large relative

Table 4. Comparison of various differentiation measures for Matrix I (with $Q_\gamma = 0.48$, $Q_\alpha = 0.47$) and Matrix II (with $Q_\gamma = 0.167$, $Q_\alpha = 0.102$) based on abundance and function (A&F), on function (F) only, and abundance (A) only.

Measure	Order	Matrix I			Matrix II		
		A&F	F	A[#]	A&F	F	A[#]
$1-C^*_{qN}(Q)$	q = 0	0.324	0.324	0.4	0.579	0.579	0.4
	q = 1	0.408	---	0.4	0.628	---	0.4
	q = 2	0.491	--	0.4	0.678	--	0.4
$1-U^*_{qN}(Q)$	q = 0	0.657	0.657	0.571	0.846	0.846	0.571
	q = 1	0.408	--	0.4	0.628	--	0.4
	q = 2	0.194	--	0.25	0.345	--	0.25
$Q^*_\beta = \dfrac{Q_\gamma - Q_\alpha}{Q_\gamma}$	q = 2	0.002			0.388		
$Q^*_{e,\beta} = \dfrac{1 - 1/Q_{e,\beta}}{1 - 1/N}$	q = 2	0.004			0.145		
$Q^{**}_{e,\beta} = \dfrac{Q_{e,\beta} - 1}{N - 1}$	q = 2	0.002			0.078		

[#] Differentiation measures are the abundance-based local differentiation measure $(1-C_{qN})$ and regional differentiation measure $(1-U_{qN})$ obtained from partitioning Hill numbers [36]; --- No measures for q = 1 and q = 2 because species abundances are not considered for measures based on function (F) only.
doi:10.1371/journal.pone.0100014.t004

difference between Q_α and Q_γ, as reflected by the high relative difference (with respect to the alpha) of 63.7%.

In Table 4, we first compare separately for Matrix I and Matrix II the differentiation measures incorporating both abundance and function (A&F), function (F) only, and abundance (A) only. The measures considering both (A&F) are based on our proposed measures $1-C^*_{qN}(Q)$ and $1-U^*_{qN}(Q)$ (with formulas in Table 3) derived from the functional beta diversity. The measure based only on function only (F) does not consider abundance, so it is identical to the zero-order of the measure considering A&F. The measures considering abundance only (A) refer to the abundance-based local differentiation measure $(1-C_{qN})$ and regional differentiation measure $(1-U_{qN})$ based on partitioning Hill numbers ([36], p. 31).

Comparing the column under A&F and the column under A within Matrix I, we find for each fixed order of $q = 0$ and $q = 2$ that there is appreciable difference between these two values (A&F and A) but the difference is limited to some extent (relatively to the corresponding difference for Matrix II); the difference is very little for $q = 1$. This is valid for both differentiation measures $1-C^*_{qN}(Q)$ and $1-U^*_{qN}(Q)$. Thus, for Matrix I (with similar distributional pattern of distances for all species pairs), functional differentiation is largely determined by species abundance pattern and function plays a minor factor.

In contrast, for Matrix II, the impact of function on our differentiation measures is clearly seen for both measures $1-C^*_{qN}(Q)$ and $1-U^*_{qN}(Q)$ by noting that our measure considering both (A&F) is much higher than the corresponding measure considering A for all orders $q = 0$, 1 and 2. This is because the functional distances for pairs of non-shared species are substantially larger than those of other species pairs, leading to a large increase in the proportion of non-shared distances in an assemblage (as reflected in our local distance-differentiation measure $1-C^*_{qN}(Q)$), and also in the pooled assemblage (as reflected in our regional distance-differentiation measure $1-U^*_{qN}(Q)$). In this case, function has profound effect on characterizing functional differentiation. Since the two measures (A&F and A) of $q = 1$ differ little for Matrix I whereas they differ substantially for Matrix II, their difference is a potentially useful indicator for the effect of function. All the above findings not only hold for equally abundant species as the example presented here but also are generally valid if species abundances are heterogeneous; see Appendix S5 for two heterogeneous cases.

For both matrices the proposed measures exhibit moderate differentiation between the two assemblages for Matrix I and moderate to high differentiation for Matrix II. For example, our proposed measure, $1-C^*_{qN}(Q)$, yields values 0.324 (for $q = 0$), 0.408 (for $q = 1$) and 0.491 (for $q = 2$) for Matrix I. The corresponding three values for Matrix II are 0.579 (for $q = 0$), 0.628 (for $q = 1$) and 0.678 (for $q = 2$). Table 4 reveals that the differentiation measure based on the additive partitioning of the quadratic entropy exhibits an unreasonably low differentiation value of 0.002 for Matrix I. As shown in reference [36], this measure does not properly quantify functional differentiation; also see the example in Appendix S5. The two measures based on the effective number of species with maximum distance (Eqs. 2e and 2f) for both matrices also show unreasonably low differentiation. For Matrix I, the measure in Eq. 2e gives a value of 0.004 and the measure in Eq. 2f gives a value of 0.002, implying that there is almost no differentiation among the two assemblages. These are counter-intuitive and unexpected values because function has almost no effect and thus all measures for Matrix I should yield

close results to those based on abundances only (the column under A in Table 4). This example also helps show that the measures in Eqs. 2e and 2f cannot be applied to non-ultrametric cases, as the two matrices are both non-ultrametric (Appendix S6). Similar findings about substantially low functional differentiation are also revealed in other papers [82,83]. For Matrix II, each of the two previously developed measures (Eqs. 2e and 2f) is also substantially lower than our proposed differentiation measure considering both ($A\&F$). More evidence from other perspectives is provided in Example 2 below.

Example 2: Ultrametric vs. Non-ultrametric Distance Matrices. In this example, we compare the performance of various differentiation measures when they are applied to an ultrametric matrix (Case I in Table 5) and a non-ultrametric matrix (Case II in Table 5). Each matrix represents a distance matrix for a pooled assemblage of four species. In each case, there are two completely distinct assemblages (no species shared). There are two equally common species (a, b) in the first assemblage, and two equally common species (c, d) in the other assemblage. We use this simple example to show that the effective approach based on the effective number of species with maximum distance (Eq. 2d) and the associated differentiation measures (Eqs. 2e and 2f) may lead to un-interpretable conclusions if they are applied to non-ultrametric distance matrices.

Comparing the two distance matrices, we see that the two matrices are identical except for the distances for the two pairs, (a, c) and (b, d). The distance between Species a and Species c is 0.2 in Case I but it is increased to 0.9 in Case II; the distance for Species b and Species d is 0.2 in Case I but it is increased to 0.8 in Case II. Thus, when the matrix is changed from Case I to Case II, the distance for any two species in different assemblages is either increased or kept as the same, whereas all the distances for species in the same assemblage are kept the same. By intuition and by theory for our measures (Proposition S2.2 in Appendix S2), any sensible differentiation measure should not decrease.

In Table 5, we compare various differentiation measures between the two assemblages separately for Case I and Case II. The measures based on Eqs. 2e and 2f both produce a maximum

differentiation of unity for Case I. This is intuitively understandable because the two assemblages are completely distinct and all distances for two species in different assemblages are higher than the distances for two species within an assemblage. In both Case 1 and Case II, the proposed differentiation measures, $1 - C_{qN}^*(Q)$ and $1 - U_{qN}^*(Q)$, attain the maximum differentiation of unity for all orders of q, showing the differentiation does not decrease from Case I to Case II. However, the two differentiation measures (Eqs. 2e and 2f) for Case II give unexpectedly lower differentiation than that of Case I. This example shows why application of Eq. 2d and the associated differentiation measures (Eqs. 2e and 2f) to non-ultrametric cases might be misleading. Although the measure based on additively partitioning quadratic entropy (Eq. 2c) yields higher differentiation for Case II, we have demonstrated its counter-intuitive behavior in Appendix S5 and in Example 1.

In this example, we specifically use the extreme case that two assemblages are completely distinct (no shared species) for illustrative purpose. A more general property of monotonicity is proved in Appendix S2 (Proposition S2.2): any differentiation measure based on our functional beta diversity is a non-decreasing function with respect to the distance of any non-shared species pair regardless of species abundance distributions. This property of monotonicity implies that the differentiation measure including $1 - C_{qN}^*(Q)$ and $1 - U_{qN}^*(Q)$ do not decrease if the distance for a non-shared species pair becomes larger even if the two assemblages are not completely distinct. In Appendix S5, we provide a supplementary example in which there are shared species between assemblages; our proposed measures yield the expected property of monotonicity, while the two previous differentiation measures (Eqs. 2e and 2f) do not.

Example 3: A Real Functional Distance Matrix for Dune Vegetation. We apply our proposed measures to the real data discussed by Ricotta et al. in [84]. The data contain a total of 43 vascular plant species collected from 272 random vegetation plots of 2×2 m in size during the period 2002–2009 in three successively less extreme fore dune habitats: embryo dunes (EM; 17 species in 70 plots), mobile dunes (MO; 39 species in 131 plots)

Table 5. Comparison of various differentiation measures between two assemblages for an ultramteric distance matrix (Case I below) and a non-ultrametric distance matrix (Case II below).

Measure	Ultrametric distance matrix (Case I)	Non-ultrametric distance matrix (Case II)
$1 - C_{qN}^*(Q)$	1 (for all $q \geq 0$)	1 (for all $q \geq 0$)
$1 - U_{qN}^*(Q)$	1 (for all $q \geq 0$)	1 (for all $q \geq 0$)
$Q_{\beta}^* = \dfrac{Q_{\gamma} - Q_{\alpha}}{Q_{\gamma}}$	0.6	0.826
$Q_{e,\beta}^* = \dfrac{1 - 1/Q_{e,\beta}}{1 - 1/N}$	1	0.559
$Q_{e,\beta}^{**} = \dfrac{Q_{e,\beta} - 1}{N - 1}$	1	0.388

Assume the two assemblages are completely distinct. There are two equally common species (a, b) in the fisrt assemblage, and two equally common species (c, d) in the second assemblage. In the pooled assemblage, there are four species (a, b, c, d) with relative abundances (0.25, 0.25, 0.25, 0.25). As explained in the text, we expect that the differentiation for Case II should not be lower than that for Case I. See Appendix S5 for a non-completely-distinct case.

Case I: An ultrametric distance matrix for four species (a, b, c, d) with $Q_{\gamma} = 0.125$, $Q_{\alpha} = 0.05$. $\begin{bmatrix} 0 & 0.1 & 0.2 & 0.2 \\ 0.1 & 0 & 0.2 & 0.2 \\ 0.2 & 0.2 & 0 & 0.1 \\ 0.2 & 0.2 & 0.1 & 0 \end{bmatrix}$

Case II: A non-ultrametric distance matrix for four species (a, b, c, d) with $Q_{\gamma} = 0.288$, $Q_{\alpha} = 0.05$. $\begin{bmatrix} 0 & 0.1 & 0.9 & 0.2 \\ 0.1 & 0 & 0.2 & 0.8 \\ 0.9 & 0.2 & 0 & 0.1 \\ 0.2 & 0.8 & 0.1 & 0 \end{bmatrix}$
doi:10.1371/journal.pone.0100014.t005

and transition dunes (TR; 42 species in 71 plots) along the Tyrrhenian coast, where EM is closest to the sea, MO is between EM and TR, and TR is farthest from the sea; see [85,86,87] for details. There are 17 shared species (out of a total of 39 species) between EM and MO, 16 shared species (out of a total of 43 species) between EM and TR, and 38 shared species (out of a total of 43 species) between MO and TR. In each habitat, we pooled species abundance data over plots and applied various diversity and differentiation measures based on the species relative abundances (Table S5.4 in Appendix S5) in the three type habitats.

All species were described by a set of sixteen functional traits which include seven quantitative variables: plant height, leaf size, leaf thickness, seed mass, seed shape, leaf dry mass and specific leaf area, together with nine categorical variables: life form, growth form, leaf texture, dispersal mode, leaf persistence, plant life span, pollination system, clonality and flowering phenology. Based on these sixteen traits, the species distance matrix in the pooled assemblage was calculated by a Gower mixed-variables coefficient of distance with equal weights for all traits [71]. The Gower species pairwise distance matrix of the pooled assemblage is provided in Appendix S6. The matrix and the three sub-matrices (corresponding to those of three habitats) are all non-ultrametric. The two idealized examples (Example 1 and Example 2) just given showed that previously-proposed functional differential measures led to unexpected conclusions when applied to non-ultrametric matrices. This real example shows how such mathematical problems can lead to misinterpretation of important ecological patterns.

For each of the three habitats, we present four diversity measures: ordinary Hill numbers ^qD (Eq. 1a), our functional Hill number $^qD(Q)$ (Eq. 3), mean functional diversity $^qMD(Q)$ (Eq. 4a) and functional diversity $^qFD(Q)$ (Eq. 4b). The diversity profiles for the four diversity measures as a function of order q are shown in Fig. 1. A consistent pattern is revealed in Fig. 1: EM has the lowest diversity, MO is intermediate, and TR has the highest diversity. This pattern is valid for all orders of q, and is expected from ecologists' perspectives [84]. The EM is closest to the sea, and hence exposed to wind disturbance, flooding, salt spray, and other harsh environmental factors. Therefore, the assemblage in the EM is mainly composed of a few specialized pioneer species with similar functional traits (as reflected by the value of quadratic entropy, which is respectively 0.513, 0.556, and 0.561 in EM, MO and TR) to adapt the extreme environment, leading to the lowest functional diversity in this habitat. The vegetation of the MO is less affected by harsh environment factors, so the vegetation presents more diverse species composition, resulting in larger functional distances and thus higher functional diversity. The species richness and evenness in the TR are the highest among the three habitats and the vegetation of TR is even more weakly constrained by these environmental factors, supporting an even higher functional diversity. The diversity pattern for Hill numbers is similar to those based on functional diversity measures, as will be discussed later. In each of the three functional diversity profiles (the two middle panels and the right panel of Fig. 1), the initial value (i.e., the value for $q = 0$) represents the diversity when only function is considered.

The formula in Eq. 2d produces much lower values of species equivalents: 2.94 (EM), 3.39 (MO) and 2.95 (TR), substantially lower than the corresponding functional Hill numbers ($q = 2$): 7.72 (EM), 15.27 (MO), 19.42 (TR); see the second panel of Fig. 1. Moreover, the number of species equivalents from Eq. 2d give a diversity ordering MO>TR≈EM, which does not conform to ecologists' expectation.

In Fig. 2, we show the differentiation profiles of the two proposed measures $1 - C^*_{qN}(Q)$ and $1 - U^*_{qN}(Q)$ as a function of order q for q between 0 and 5. In Table 6, we compare various differentiation measures between any two habitats (EM vs. MO, EM vs. TR and MO vs. TR). In the same table, as we did in Table 4, we also show the differentiation values incorporating both abundance and function ($A\&F$), function (F) only, and abundance (A) only. Table 6 reveals that in any pair of assemblages, we have a pattern similar to that in Table 4 for Matrix I. That is, our differentiation measures considering both ($A\&F$) yield comparable results to those considering abundance only (A) for $q = 0$ and for $q = 2$, and yield very close results for $q = 1$. As with Example 1, this may be explained by the fact that the gamma quadratic entropy in each pair of assemblage is only slightly higher than the alpha quadratic entropy. The relative differences between gamma and alpha quadratic entropies is respectively 2.8%, 4.5% and 2.7% for EM vs. MO, EM vs. TR and MO vs. TR. Therefore, abundance is the major factor that determines the differentiation between any two habitats, implying that the four measures incorporating abundances with or without considering function exhibit very similar patterns in Fig. 1.

Our proposed differentiation measures, $1 - C^*_{qN}(Q)$ and $1 - U^*_{qN}(Q)$ (Table 6 and Fig. 2) implies that EM vs. TR has the highest functional differentiation, MO vs. TR has the lowest differentiation, and EM vs. MO is somewhat in between for any fixed order q between 0 and 5. This pattern is anticipated. As discussed above, the vegetation within EM is composed by few specialized plants with similar ecological functions to adapt the extreme environmental stress. However, these traits are unique to species in EM when compared with species in the other two habitats. There are also fewer shared species between EM and TR (also EM and MO). In contrast, the vegetation in MO and TR is similarly diverse and most species in these two habitats are shared. These explain why MO vs. TR exhibits the lowest functional differentiation, whereas EM vs. TR (also EM vs. MO) exhibit higher functional differentiation.

Table 6 and Fig. 2 further reveal that the two measures $1 - C^*_{qN}(Q)$ and $1 - U^*_{qN}(Q)$ for the three pairs of habitats give moderate to high differentiation. For example, for $q = 2$, our differentiation measure $1 - C^*_{2N}(Q)$ for the three pairs (EM vs. MO, EM vs. TR and MO vs. TR) is respectively 0.658, 0.885 and 0.539, and the corresponding differentiation measure $1 - U^*_{2N}(Q)$ is respectively 0.324, 0.659 and 0.226. In sharp contrast, the three previous measures based on the quadratic entropy (Eqs. 2c, 2e and 2f) show substantially lower differentiation. For these data, the differentiation measure based on the additive decomposition of quadratic entropy (Eq. 2c) for EM vs. MO, EM vs. TR and MO vs. TR is respectively 0.028, 0.042 and 0.026. This wrongly implies substantially low differentiation between any two habitats. For the differentiation measure based on Eq. 2f are also low (0.034, 0.054 and 0.035). These values also give an unexpected ordering in that EM vs. MO exhibits the lowest functional differentiation, which is counter-intuitive. Similarly, the measure given in Eq. 2e gives a wrong ordering. All three examples demonstrate that our functional diversity measures and their associated differentiation measures yield the expected results and ecologically sensible interpretations.

Conclusion and Discussion

We have extended ordinary Hill numbers to the distance-based functional Hill number $^qD(Q)$ to take into account the pairwise functional distance between species (see Eq. 3, in units of effective number of equally abundant and equally distinct species). Here Q
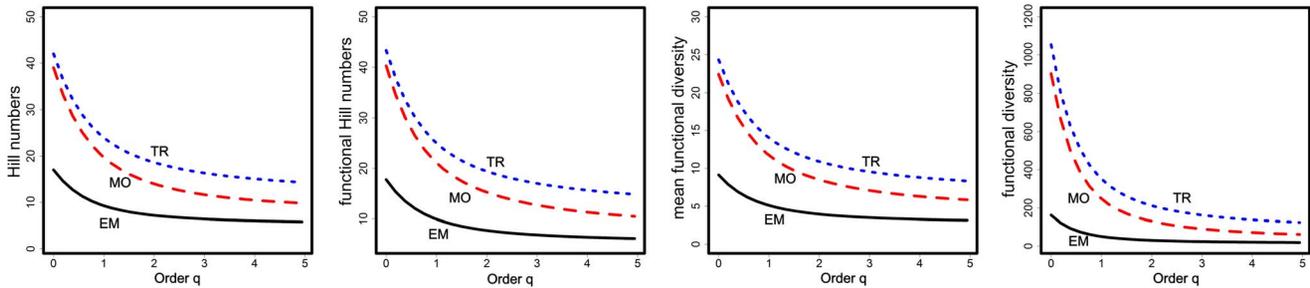
Figure 1. Diversity profiles as a function of order q for ordinary Hill numbers ^{q}D (left panel), functional Hill numbers $^{q}D(Q)$ (the second panel from the left), mean functional diversity $^{q}MD(Q)$ (the third panel from the left) and (total) functional diversity $^{q}FD(Q)$ (right panel) for three habitats (TR, MO, and EM). All the profiles show a consistent diversity pattern about the ordering of the three habitats: TR>MO>EM.

(Rao's quadratic entropy) plays an important indirect role, even though the measure Q itself cannot be directly used to measure functional diversity as noted by several authors [26,35,36,39]. We have also proposed a class of mean functional diversity $^{q}MD(Q) = {}^{q}D(Q) \times Q$; see Eq. 4a. The product of the functional Hill number and the mean functional diversity quantifies the (total) functional diversity $^{q}FD(Q) = [{}^{q}D(Q)]^{2} \times Q$, i.e., the effective total distance between species of the assemblage. See Fig. 1 for an example comparing ordinary Hill numbers and the three functional diversity measures. The three proposed measures quantify different aspects of species trait space. Our approach is valid not only for any symmetric distance matrices in ecology, but also for all types of symmetric matrices in other disciplines.

Since the pioneering work by MacArthur [40] and Hill [41], the replication principle has been identified as an essential property for characterizing abundance-based species diversity. As we reviewed in this paper, Hill numbers obey the replication principle. Hill numbers have been extended to phylogenetic Hill numbers (in units of "species equivalent") and related branch diversity (in units of "branch length"); both satisfy a phylogenetic generalization of the replication principle [36,45]. In this paper, we have proved that the functional Hill numbers (in units of "species equivalent") and the mean functional diversity (in units of "functional distance") both satisfy a functional version of the replication principle, and also proved that the functional diversity (in units of "functional distance") satisfies a quadratic replication principle. Therefore, we think replication principle is an essential property for measures in units of species equivalents, but for other related measures this property may be valid (e.g., branch diversity, mean functional diversity), or may be replaced by a quadratic (or a power function of $\mathcal{N}$) property; see [48] for a different perspective.

Recently, Chao et al. [50] integrated species diversity, phylogenetic diversity and functional diversity into a unified framework of "attribute diversity" based on Hill numbers. Both Leinster and Cobbold [75] and Scheiner [63] derived their integrative metrics under a framework of Hill numbers and their metrics are also in units of "species equivalents". In Appendix S5, we provide detailed comparison to distinguish these two previous approaches from ours. Generally, we find that Leinster & Cobbold's measure may not be sensitive to species abundances when species similarity matrix is computed from species traits in functional analysis. If species similarity matrix deviates greatly from a naïve identity matrix, then their measure typically yields very low diversity values especially for assemblages with many species; this causes problems for the interpretation of "species equivalents" in their approach. Reeve et al. [88] recently proposed a diversity partition based on Leinster & Cobbold's measure. In
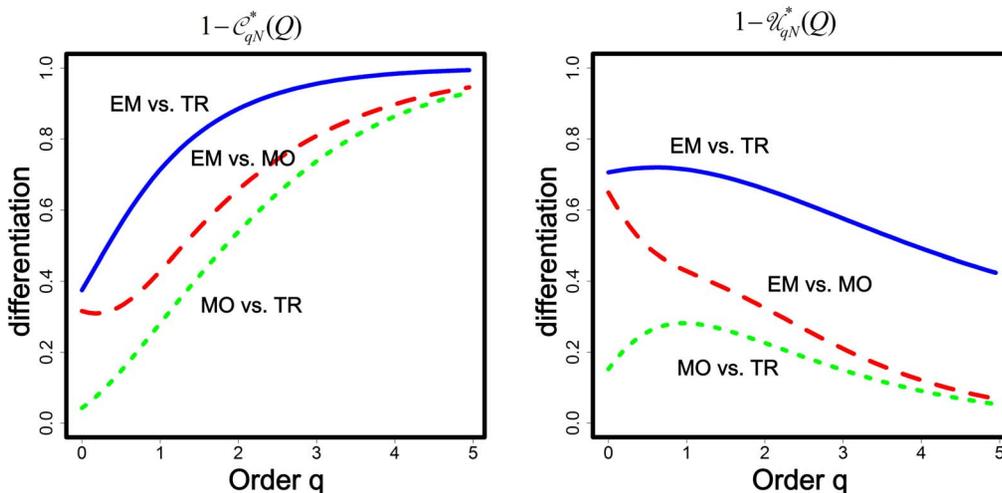

Figure 2. Differentiation profiles for the functional differentiation measures $1 - C^{*}_{qN}(Q)$ (left panel) and $1 - U^{*}_{qN}(Q)$ (right panel) as a function of order q for three pairs of habitats (EM vs. MO, EM vs. TR and MO vs. TR.)

Table 6. Comparison of various differentiation measures for three pairs of habitats in the real data analysis based on abundance and function (A&F), on function (F) only, and abundance (A) only.

Measure	Order	EM vs. MO			EM vs. TR			MO vs. TR		
		A&F	F	A#	A&F	F	A#	A&F	F	A#
$1 - C^*_{qN}(Q)$	q = 0	0.316	0.316	0.392	0.375	0.375	0.457	0.043	0.043	0.062
	q = 1	0.428	---	0.427	0.714	---	0.721	0.282	---	0.278
	q = 2	0.658	---	0.573	0.885	---	0.854	0.539	---	0.457
$1 - U^*_{qN}(Q)$	q = 0	0.649	0.649	0.564	0.706	0.706	0.628	0.152	0.152	0.118
	q = 1	0.428	---	0.427	0.714	---	0.721	0.282	---	0.278
	q = 2	0.324	---	0.401	0.659	---	0.746	0.226	---	0.296
$Q^*_\beta = \dfrac{Q_\gamma - Q_\alpha}{Q_\gamma}$	q = 2	0.028			0.042			0.026		
$Q^*_{e,\beta} = \dfrac{1 - 1/Q_{e,\beta}}{1 - 1/N}$	q = 2	0.066			0.102			0.067		
$Q^{**}_{e,\beta} = \dfrac{Q_{e,\beta} - 1}{N - 1}$	q = 2	0.034			0.054			0.035		

$Q_\gamma = 0.550$ and $Q_\alpha = 0.535$ for the pair (EM, MO); $Q_\gamma = 0.561$, $Q_\alpha = 0.537$ for the pair (EM, TR); $Q_\gamma = 0.574$, $Q_\alpha = 0.559$ for the pair (MO, TR).
#Differentiation measures are the abundance-based local differentiation measure $(1 - C_{qN})$ and regional differentiation measure $(1 - U_{qN})$ obtained from partitioning Hill numbers [36];
--- No measures for $q = 1$ and $q = 2$ because species abundances are not considered for measures based on function (F) only.
doi:10.1371/journal.pone.0100014.t006

the same Appendix, we show by an example that their gamma diversity may be less than their proposed alpha diversity even in equal weight case. Scheiner's approach and our measures have different meanings of "species equivalents" and thus quantify different aspects and properties of ecosystems. Scheiner's measure cannot be directly linked to most of the previous commonly used phylogenetic diversity (e.g., Faith's *PD*) and functional measures (e.g., *FAD*).

Except for Rao's quadratic entropy, the decomposition of other functional diversity measures is rarely discussed in the literature. In this paper, we have developed the decomposition of the proposed three functional diversity measures of any order q; see Table 2 for a summary. In the decomposition of each of the three measures, the alpha and beta components are unrelated (or independent), and thus each beta component measures pure functional differentiation among assemblages and can be transformed onto the range [0,1] to obtain the normalized distance-overlap measures (from the beta component based on the functional diversity) or species-overlap measures (from the beta components based on decomposing the functional Hill numbers and the mean functional diversity). In most applications, we recommend applying the distance-overlap measures (given in Table 3 for the two major classes of similarity measures). An important advantage of using the framework of Hill numbers is that there is a direct link between functional diversity measures and functional similarity (or differentiation) among assemblages. To convey the information about functional differentiation among multiple assemblages, we suggest plotting the differentiation profiles for two differentiation measures, $1 - C^*_{qN}(Q)$ (from the perspective of a pair of local assemblages) and $1 - U^*_{qN}(Q)$ (from the perspective of the pooled assemblage) with respect to q; see Fig. 2 for examples. For the special case of $q = 0$, the measure $C^*_{0N}(Q)$ (Eq. 9b and Table 3) in terms of *FAD* represents the functional generalization of the Sørensen similarity index, and the measure $U^*_{0N}(Q)$ (Eq. 10b and Table 3) represents the functional generalization of the classic Jaccard similarity index. Conceptually different approach to diversity decomposition was proposed by Kosman [48].

Villéger et al. [20] proposed that trait-based functional diversity should include three aspects: functional richness, functional evenness and functional divergence. In ordinary Hill numbers, zero-order diversity represents species richness, and Hill numbers (or their transformations) of different orders can be used to construct various functional evenness measures as those proposed by Jost [89]. Jost used partitioning theory to derive Hill's useful class of evenness measures, the ratios of Hill numbers ${}^qD$ and species richness, ${}^qD/S$ for $q > 0$, and he showed that the ratio of the logarithms of Hill numbers and logarithm of richness, $\log({}^qD)/\log(S)$, expresses the corresponding relative evenness. Applying this idea to our framework, we can construct measures of functional richness and functional evenness based on any of our three functional diversity measures. For example, we can regard the effective total distance between species of order $q = 0$ as a measure of functional richness and use ${}^qFD(Q)/[{}^0FD(Q)]$ and $\log[{}^qFD(Q)]/\log[{}^0FD(Q)]$ as measures of functional evenness.

Finally, we mention a potential application of our proposed measures. In genetics, the nucleotide diversity [90] is based on a nucleotide distance matrix. The elements of this distance matrix are obtained as the nucleotide differences between any two DNA sequences. Since our approach can be applied to any type of symmetric distance matrix, we expect our proposed measures would be useful in genetics. The nucleotide diversity in genetics is equivalent to Rao's quadratic entropy, and the corresponding differentiation measure N_{ST} is identical to the traditional

differentiation measure in Eq. 2c. We have shown in this paper by hypothetical and real examples that the measure in Eq. 2c does not provide legitimate measure of differentiation; see [36] for theoretical discussion. We have proposed unified and rigorous distance-overlap measures and their corresponding differentiation measures (in Table 3), and thus those measures merit investigation for applications to genetics.

Supporting Information

Appendix S1 Some properties of the proposed functional diversity measures.
(PDF)

Appendix S2 Decomposition of the proposed functional diversity measures.
(PDF)

Appendix S3 Four classes of functional similarity/differentiation measures.
(PDF)

Appendix S4 Functional beta diversity and functional diversity excess lead to the same classes of similarity and differentiation measures.
(PDF)

Appendix S5 Supplementary examples and comparisons.
(PDF)

Appendix S6 Distance matrices used in Example 1 and Example 3.
(XLSX)

Author Contributions

Conceived and designed the experiments: CHC AC. Performed the experiments: CHC AC. Analyzed the data: CHC AC. Contributed reagents/materials/analysis tools: CHC AC. Wrote the paper: CHC AC. Figures: CHC.

References

1. Tilman D, Knops J, Wedin D, Reich P, Ritchie M, et al. (1997) The influence of functional diversity and composition on ecosystem processes. Science 277: 1300–1302.
2. Tilman D (2001) Functional diversity. In: Levin AS, editor. Encyclopedia of Biodiversity, Vol. 3. San Diego: Academic Press. pp. 109–120.
3. Diaz S, Cabido M (2001) Vive la différence: plant functional diversity matters to ecosystem processes. Trends Ecol Evol 16: 646–655.
4. Petchey OL, Gaston KJ (2002) Functional diversity (FD), species richness and community composition. Ecol Lett 5: 402–411.
5. Ricotta C (2005) A note on functional diversity measures. Basic Appl Ecol 6: 479–486.
6. Swenson NG (2011) Phylogenetic beta metrics, trait evolution and inferring the functional beta diversities of communities. Plos One 6: e21264.
7. Swenson NG, Erickson DL, Mi X, Bourg NA, Forero-Montaña J, et al. (2012) Phylogenetic and functional alpha and beta diversity in temperate and tropical tree communities. Ecology 93: 112–125.
8. Norberg J, Swaney DP, Dushoff J, Lin J, Casagrandi R, et al. (2001) Phenotypic diversity and ecosystem functioning in changing environments: a theoretical framework. Proc Natl Acad Sci USA 98: 11376–11381.
9. Suding KN, Lavorel S, Chapin FS, Cornelissen JH, Diaz S, et al. (2008) Scaling environmental change through the community-level: a trait-based response-and-effect framework for plants. Glob Change Biol 14: 1125–1140.
10. Cadotte MW, Cavender-Bares J, Tilman D, Oakley TH (2009) Using phylogenetic, functional and trait diversity to understand patterns of plant community productivity. PLoS One 4: e5695.
11. Flynn DF, Mirotchnick N, Jain M, Palmer MI, Naeem S (2011) Functional and phylogenetic diversity as predictors of biodiversity-ecosystem-function relationships. Ecology 92: 1573–1581.
12. Dukes JS (2001) Productivity and complementarity in grassland microcosms of varying diversity. Oikos 94: 468–480.
13. Folke C, Carpenter S, Walker B, Scheffer M, Elmqvist T, et al. (2004) Regime shifts, resilience, and biodiversity in ecosystem management. Annu Rev Ecol Evol Syst: 557–581.
14. Cardinale BJ, Duffy JE, Gonzalez A, Hooper DU, Perrings C, et al. (2012) Biodiversity loss and its impact on humanity. Nature 486: 59–67.
15. Mouillot D, Graham NA, Villéger S, Mason NW, Bellwood DR (2013) A functional approach reveals community responses to disturbances. Trends Ecol Evol 28: 167–177.
16. Petchey OL, Gaston KJ (2009) Dendrograms and measures of functional diversity: a second installment. Oikos 118: 1118–1120.
17. Walker B, Kinzig A, Langridge J (1999) Plant attribute diversity, resilience, and ecosystem function: The nature and significance of dominant and minor species. Ecosystems 2: 95–113.
18. Schmera D, Erös T, Podani J (2009) A measure for assessing functional diversity in ecological communities. Aquat Ecol 43: 157–167.
19. Díaz S, Lavorel S, de Bello F, Quétier F, Grigulis K, et al. (2007) Incorporating plant functional diversity effects in ecosystem service assessments. Proc Natl Acad Sci USA 104: 20684–20689.
20. Villéger S, Mason NW, Mouillot D (2008) New multidimensional functional diversity indices for a multifaceted framework in functional ecology. Ecology 89: 2290–2301.
21. Poos MS, Walker SC, Jackson DA (2009) Functional-diversity indices can be driven by methodological choices and species richness. Ecology 90: 341–347.
22. Mouchet MA, Villéger S, Mason NWH, Mouillot D (2010) Functional diversity measures: an overview of their redundancy and their ability to discriminate community assembly rules. Funct Ecol 24: 867–876.
23. Cadotte MW, Carscadden K, Mirotchnick N (2011) Beyond species: functional diversity and the maintenance of ecological processes and services. J Appl Ecol 48: 1079–1087.
24. Weiher E (2011) A primer of trait and functional diversity. In: Magurran AE, McGill BA, editors. Biological diversity: frontiers in measurement and assessment: pp.175–193.
25. Pla L, Casanoves F, Di Rienzo J. (2012) Functional diversity indices. In: Quantifying Functional Biodiversity, London: Springer. pp. 27–51.
26. Mouillot D, Graham NAJ, Villéger S, Mason NWH, Bellwood DR (2013) Functional approach reveals community responses to disturbances. Trends Ecol Evol 28: 167–177.
27. Mason NW, MacGillivray K, Steel JB, Wilson JB (2003) An index of functional diversity. J Veg Sci 14: 571–578.
28. Petchey OL, Gaston KJ (2006) Functional diversity: back to basics and looking forward. Ecol Lett 9: 741–758.
29. Petchey OL, Gaston KJ (2007) Dendrograms and measuring functional diversity. Oikos 116: 1422–1426.
30. Podani J, Schmera D (2006) On dendrogram-based measures of functional diversity. Oikos 115: 179–185.
31. Podani J, Schmera D (2007) How should a dendrogram-based measure of functional diversity function? A rejoinder to Petchey and Gaston. Oikos 116: 1427–1430.
32. Rao CR (1982) Diversity and dissimilarity coefficients: A unified approach. Theor Popul Biol 21: 24–43.
33. Botta-Dukát Z (2005) Rao's quadratic entropy as a measure of functional diversity based on multiple traits. J Veg Sci 16: 533–540.
34. Ricotta C, Szeidl L (2006) Towards a unifying approach to diversity measures: bridging the gap between the Shannon entropy and Rao's quadratic index. Theor Popul Biol 70: 237–243.
35. Ricotta C, Szeidl L (2009) Diversity partitioning of Rao's quadratic entropy. Theor Popul Biol 76: 299–302.
36. Chiu C-H, Jost L, Chao A (2014) Phylogenetic beta diversity, similarity, and differentiation measures based on Hill numbers. Ecol Monogr 84: 21–44.
37. Hardy OJ, Jost L (2008) Interpreting and estimating measures of community phylogenetic structuring. J Ecol 96: 849–852.
38. Guiasu RC, Guiasu S (2011) The weighted quadratic index of biodiversity for pairs of species: a generalization of Rao's index. Nat Sci 3: 795–801.
39. de Bello F, Lavergne S, Meynard CN, Lepš J, Thuiller W (2010) The partitioning of diversity: showing Theseus a way out of the labyrinth. J Veg Sci 21: 992–1000.
40. MacArthur RH (1965) Patterns of species diversity. Biol Rev 40: 510–533.

41. Hill M (1973) Diversity and evenness: A unifying notation and its consequences. Ecology 54: 427–432.
42. Jost L (2006) Entropy and diversity. Oikos 113: 363–375.
43. Jost L (2007) Partitioning diversity into independent alpha and beta components. Ecology 88: 2427–2439.
44. Ellison AM (2010) Partitioning diversity. Ecology 91: 1962–1963.
45. Chao A, Chiu C-H, Jost L (2010) Phylogenetic diversity measures based on Hill numbers. Phil Trans R Soc B 365: 3599–3609.
46. Kosman E (1996) Difference and diversity of plant pathogen populations: a new approach for measuring. Phytopathology 86: 1152–1155.
47. Kosman E, Leonard KJ (2007) Conceptual analysis of methods applied to assessment of diversity within and distance between populations with asexual or mixed mode of reproduction. New Phytol 174: 683–696.
48. Kosman E (2014) Measuring diversity: from individuals to populations. Eur J Plant Pathol 138: 467–486.
49. Chao A, Chiu C-H, Hsieh TC (2012) Proposing a resolution to debates on diversity partitioning. Ecology 93: 2037–2051.
50. Chao A, Chiu C-H, Jost L (2014) Unifying species diversity, phylogenetic diversity, functional diversity and related similarity/differentiation measures through Hill numbers. Annu Rev Ecol Evol Syst (in press)
51. Chao A, Gotelli NJ, Hsieh TC, Sander E, Ma KH, et al. (2014) Rarefaction and extrapolation with Hill numbers: a framework for sampling and estimation in species diversity studies. Ecol Monogr 84: 45–67.
52. Tóthmérész B (1995) Comparison of different methods for diversity ordering. J Veg Sci 6: 283–290.
53. MacArthur RH (1972) Geographical Ecology: patterns in the distribution of species. Princeton: Princeton University Press. 288 p.
54. Whittaker RH (1972) Evolution and measurement of species diversity. Taxon 12: 213–251.
55. Routledge R (1979) Diversity indices: which ones are admissible? J Theor Biol 76: 503–515.
56. Jost L, DeVries P, Walla T, Greeney H, Chao A, et al. (2010) Partitioning diversity for conservation analyses. Divers Distrib 16: 65–76.
57. Grime J (1998) Benefits of plant diversity to ecosystems: immediate, filter and founder effects. J Ecol 86: 902–910.
58. Vile D, Shipley B, Garnier E (2006) Ecosystem productivity can be predicted from potential relative growth rate and species abundance. Ecol Lett 9: 1061–1067.
59. Stuart-Smith RD, Bates AE, Lefcheck JS, Duffy JE, Baker SC, et al. (2013) Integrating abundance and functional traits reveals new global hotspots of fish diversity. Nature 501: 539–542.
60. Devictor V, Mouillot D, Meynard C, Jiguet F, Thuiller W, et al. (2010) Spatial mismatch and congruence between taxonomic, phylogenetic and functional diversity: the need for integrative conservation strategies in a changing world. Ecol Lett 13: 1030–1040.
61. Stegen JC, Hurlbert AH (2011) Inferring ecological processes from taxonomic, phylogenetic and functional trait β-diversity. PloS one 6: e20906.
62. Swenson NG (2011) Phylogenetic beta diversity metrics, trait evolution and inferring the functional beta diversity of communities. PloS one 6: e21264.
63. Scheiner SM (2012) A metric of biodiversity that integrates abundance, phylogeny, and function. Oikos 121: 1191–1202.
64. Mason NW, Mouillot D, Lee WG, Wilson JB (2005) Functional richness, functional evenness and functional divergence: the primary components of functional diversity. Oikos 111: 112–118.
65. Laliberté E, Legendre P (2010) A distance-based framework for measuring functional diversity from multiple traits. Ecology 91: 299–305.
66. Jost L (2008) GST and its relatives do not measure differentiation. Mol Ecol 17: 4015–4026.
67. Champely S, Chessel D (2002) Measuring biological diversity using Euclidean metrics. Environ Ecol Stat 9: 167–177.
68. Pavoine S, Dufour AB, Chessel D (2004) From dissimilarities among species to dissimilarities among communities: a double principal coordinate analysis. J Theor Biol 228: 523–537.
69. Pavoine S, Ollier S, Pontier D (2005) Measuring diversity from dissimilarities with Rao's quadratic entropy: Are any dissimilarities suitable? Theor Popul Biol 67: 231–239.
70. Hardy OJ, Senterre B (2007) Characterizing the phylogenetic structure of communities by an additive partitioning of phylogenetic diversity. J Ecol 95: 493–506.
71. Pavoine S, Vallet J, Dufour AB, Gachet S, Daniel H (2009) On the challenge of treating various types of variables: application for improving the measurement of functional diversity. Oikos 118: 391–402.
72. Villéger S, Miranda JR, Hernandez DF, Mouillot D (2012) Low functional β-diversity despite high taxonomic β-diversity among tropical estuarine fish communities. PloS One 7: e40679.
73. Escalas A, Bouvier T, Mouchet MA, Leprieur F, Bouvier C, et al. (2013) A unifying quantitative framework for exploring the multiple facets of microbial biodiversity across diverse scales. Environ. Microbiol. 15: 2642–2657.
74. Guiasu RC, Guiasu S (2012) The weighted Gini-Simpson index: revitalizing an old index of biodiversity. International J Ecol 2012, Article ID 478728, 10 pages.
75. Leinster T, Cobbold CA (2012) Measuring diversity: the importance of species similarity. Ecology 93: 477–489.
76. de Bello F, Lavorel S, Albert CH, Thuiller W, Grigulis K, et al. (2011) Quantifying the relevance of intraspecific trait variability for functional diversity. Methods Ecol Evol 2: 163–174.
77. Marshall A, Olkin I (1979) Inequalities: Theory of Majorization and its Applications. New York: Academic Press.
78. Weikard HP, Punt M, Wessler J (2006) Diversity measurement combining relative abundances and taxonomic distinctiveness of species. Divers Distrib 12: 215–217.
79. Chao A, Jost L, Chiang SC, Jiang Y-H, Chazdon RL (2008) A two-stage probabilistic approach to multiple-community similarity indices. Biometrics 64: 1178–1186.
80. Horn HS (1966) Measurement of "overlap" in comparative ecological studies. Am Nat: 419–424.
81. Morisita M (1959) Measuring of interspecific association and similarity between communities. Memoires of the Faculty of Science, Kyushu University, Series E (Biology) 3: 65–80.
82. Bernard-Verdier M, Flores O, Navas ML, Garnier E (2013) Partitioning phylogenetic and functional diversity into alpha and beta components along an environmental gradient in a Mediterranean rangeland. J Veg Sci 24: 877–889.
83. Villéger S, Miranda JR, Hernandez DF, Mouillot D (2012) Low functional β-diversity despite high taxonomic β-diversity among tropical estuarine fish communities. PloS One 7: e40679.
84. Ricotta C, Pavoine S, Bacaro G, Acosta ATR (2012) Functional rarefaction for species abundance data. Methods Ecol Evol 3: 519–525.
85. Carboni M, Santoro R, Acosta A (2010) Are some communities of the coastal dune zonation more susceptible to alien plant invasion? J Plant Ecol 3: 139–147.
86. Carboni M, Santoro R, Acosta AT (2011) Dealing with scarce data to understand how environmental gradients and propagule pressure shape fine-scale alien distribution patterns on coastal dunes. J Veg Sci 22: 751–765.
87. Carboni M, Acosta AT, Ricotta C (2013) Are differences in functional diversity among plant communities on Mediterranean costal dunes driven by phylogenetic history? J Veg Sci 24: 932–941.
88. Reeve R, Matthews L, Cobbold CA, Leinster T, Thompson J, Brummitt N (2014) How to partition diversity. Cornell University Library, arXiv:1404.6520, assessed on May 28, 2014
89. Jost L (2010) The relation between evenness and diversity. Diversity 2: 207–232.
90. Nei M, Li W-H (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc Natl Acad Sci USA 76: 5269–5273.

Supporting Information

Distance-based functional diversity measures and their decomposition: a framework based on Hill numbers

Chun-Huo Chiu and Anne Chao
Institute of Statistics, National Tsing Hua University, Hsin-Chu, Taiwan 30043

Appendix S1: Some properties of the proposed functional diversity measures

In this Appendix, we summarize some basic properties of the three classes of functional diversity measures: (1) functional Hill number $^qD(Q)$ which quantifies the effective number of equally abundant and equally distinct species in an assemblage with a constant species pairwise distance Q, where Q denotes Rao's quadratic entropy; (2) mean functional diversity $^qMD(Q)$ $=[^qD(Q)]\times Q$ which quantifies the effective sum of functional distances between a fixed species to all other species; and (3) (total) functional diversity $^qFD(Q)$ ($=\ ^qD(Q)\times\ ^qMD(Q)$), which quantifies the effective total distance between species of an assemblage. Since the three measures are closely related, all the properties are shared.

Without loss of generality, we assume that all distances d_{ij} are integers for intuitively understanding our derivation. This is because the measure $^qD(Q)$ is independent of the scale of d_{ij} and the other two measures are proportional to $^qD(Q)$ and thus properties discussed in this Appendix are not affected. As discussed by Chao et al. [1], we could conceptually think of all species pairwise distances as forming a single assemblage of pairwise distances. In this assemblage of "distances", there are d_{ij} units of "distance" for $i, j = 1, 2, \ldots, S$. Each of the d_{ij} units has relative abundance $p_i p_j / Q$. From the definition of Rao's quadratic entropy, the sum of all these relative abundances is unity, i.e.,

$$\sum_{i=1}^{S}\sum_{j=1}^{S}d_{ij}\left(\frac{p_i p_j}{Q}\right)=1.$$

Then we can apply the concept of Hill numbers to this assemblage of "distances". Our proposed (total) functional diversity of order q, $^qFD(Q)$, is the Hill number of order q for this assemblage of "distances":

$$\left[\sum_{i=1}^{S}\sum_{j=1}^{S}d_{ij}\left(\frac{p_i p_j}{Q}\right)^q\right]^{1/(1-q)}.$$

The measure $^qFD(Q)$ quantifies the effective total distances between species in the assemblage. From this perspective, the following properties are direct consequences of the theoretical properties of Hill numbers; see the appendices of [1] for details.

Proposition S1.1: The three measures ($^qD(Q)$, $^qMD(Q)$ and $^qFD(Q)$) are Schur-concave with respect to the product of relative abundances $p_i p_j$, and all three measures are non-increasing function of the order q.

Proposition S1.2: (Weak monotonicity) If a rarest new species is added to an assemblage, then the measure $^qFD(Q)$ does not decrease. Also, if a rarest new species is added to an assemblage such that the quadratic entropy remains unchanged, then all three measures do not decrease regardless of distance matrices.

Replication principle

Proposition S1.3: (Replication principle for functional Hill numbers) Suppose we have N equally large and completely distinct assemblages (no shares species). Assume that the functional Hill number $^qD(Q_{km})$ between Assemblage k and Assemblage m is a constant ^qD, for $k, m = 1, 2, …, N$. If these N assemblages are pooled, then the functional Hill number of order q in the pooled assemblage is $N \times {}^qD$.

Proof: Assume and there are S species in the pooled assemblage and the $S \times S$ symmetric species pairwise distance matrix is denoted by $[d_{ij}]$, $i, j = 1, 2, …, S$. Let $p_{ik} \geq 0$ be the relative abundance of ith species in the kth assemblage, $i = 1, 2, …, S$, $k = 1, 2, …, N$. The assumption that the functional Hill number of order q for any pair of assemblages is a constant ^qD means

$$\sum_{i=1}^{S}\sum_{j=1}^{S}\frac{d_{ij}}{Q_{km}}(p_{ik}p_{jm})^q = ({}^qD)^{2(1-q)}, \quad k, m = 1, 2, …, N, \tag{A1}$$

where $Q_{km} = \sum_{i=1}^{S}\sum_{j=1}^{S}d_{ij}p_{ik}p_{jm}$ denotes the (abundance-weighted) mean distance between species of Assemblage k and Assemblage m. Eq. (A1) implies

$$\sum_{i=1}^{S}\sum_{j=1}^{S}d_{ij}(p_{ik}p_{jm})^q = Q_{km}\times({}^qD)^{2(1-q)}, \quad k, m = 1, 2, …, N. \tag{A2}$$

When these N equally large assemblages are combined, the relative abundance of the ith species in the pooled assemblage is $\overline{p}_{i+} = \sum_{k=1}^{N}p_{ik}/N$, $i = 1, 2,…, S$. Then the quadratic entropy Q in the pooled assemblage is

$$Q = \sum_{i,j=1}^{S}d_{ij}\overline{p}_{i+}\overline{p}_{j+} = \sum_{i,j=1}^{S}d_{ij}\left(\sum_{k=1}^{N}\frac{p_{ik}}{N}\right)\left(\sum_{m=1}^{N}\frac{p_{jm}}{N}\right) = \frac{1}{N^2}\sum_{k=1}^{N}\sum_{m=1}^{N}\sum_{i,j=1}^{S}d_{ij}p_{ik}p_{jm} = \frac{1}{N^2}\sum_{k=1}^{N}\sum_{m=1}^{N}Q_{km}. \tag{A3}$$

Therefore, the functional Hill number of order q in the pooled assemblage using Eq. 3 of the main text can be expressed as

$$^qD_{\text{pooled}}(Q) = \left[\sum_{i=1}^{S}\sum_{j=1}^{S}\frac{d_{ij}}{Q}(\overline{p}_{i+}\overline{p}_{j+})^q\right]^{\frac{1}{2(1-q)}} = \left[\sum_{i=1}^{S}\sum_{j=1}^{S}\frac{d_{ij}}{Q}\left(\sum_{k=1}^{N}\frac{p_{ik}}{N}\right)^q\left(\sum_{m=1}^{N}\frac{p_{jm}}{N}\right)^q\right]^{\frac{1}{2(1-q)}}$$

$$= \left[\frac{1}{N^{2q}} \sum_{k,m=1}^{N} \sum_{i,j=1}^{S} \frac{d_{ij}}{Q} (p_{ik} p_{jm})^q \right]^{\frac{1}{2(1-q)}}.$$

It then follows from Eqs. (A2) and (A3) that we have

$$^q D_{\text{pooled}}(Q) = \left[\frac{1}{N^{2q}} \sum_{k=1}^{N} \sum_{m=1}^{N} \frac{Q_{km}}{Q} (^q D)^{2(1-q)} \right]^{\frac{1}{2(1-q)}} = \left[\frac{1}{N^{2q}} N^2 (^q D)^{2(1-q)} \right]^{\frac{1}{2(1-q)}} = N \times {}^q D .$$

Thus, the functional Hill number of the same order q in the pooled assemblage is N times that of each individual assemblage. Notice that in our proof of this replication principle, the mean distances Q_{km}, k, $m = 1, 2, …, N$, are not required to be a constant. Also, the species abundance distributions are allowed to be different across assemblages, and the result is valid for any symmetric matrices. This is a strong version of the replication principle.

If we further assume that all the mean distances Q_{km}, k, $m = 1, 2, …, N$, are the same, then the mean functional diversity of the pooled assemblage will be N times that of each assemblage. However, the total functional diversity (as a product of functional Hill number and the mean functional diversity) satisfies a "*quadratic* replication principle", i.e., the total functional diversity of the pooled assemblage will be N^2 times that of each assemblage. We only state the propositions below and the proof follows directly from Proposition S1.3.

Proposition S1.4: (Replication principle for the mean functional diversity $^q MD(Q)$) Suppose we have N equally large and completely distinct assemblages (no shares species). Assume that for all pairs of assemblages (k, m) the mean functional diversity $^q MD(Q_{km})$ for k, $m = 1, 2, …, N$ are identical, and the mean distances Q_{km} for k, $m = 1, 2, …, N$ are also identical. When these N equally large assemblages are pooled, the mean functional diversity of the pooled assemblage is N times that of an individual assemblage.

Proposition S1.5: (Quadratic replication principle for the total functional diversity $^q FD(Q)$) Suppose we have N equally large and completely distinct assemblages (no shares species). Assume that for all pairs of assemblages (k, m) the total functional diversity $^q FD(Q_{km})$ for k, $m = 1, 2, …, N$ are identical, and the mean distances Q_{km} for k, $m = 1, 2, …, N$ are also identical. When these assemblages are pooled, the total functional diversity of the pooled assemblage is N^2 times that of an individual assemblage.

Reference

1. Chao A, Chiu C-H, Jost L (2010) Phylogenetic diversity measures based on Hill numbers. Phil Trans R Soc B 365: 3599–3609.

Supporting Information

Distance-based functional diversity measures and their decomposition: a framework based on Hill numbers

Chun-Huo Chiu and Anne Chao
Institute of Statistics, National Tsing Hua University, Hsin-Chu, Taiwan 30043

Appendix S2: Decomposition of the proposed functional diversity measures

When there are multiple assemblages defined by the investigator, as indicated in the main text, the functional Hill number $^qD(Q)$ (Eq. 3 of the main text), mean functional diversity $^qMD(Q)$ (Eq. 4a) and total functional diversity $^qFD(Q)$ (Eq. 4b) of the pooled assemblage can each be decomposed into independent alpha and beta components using a derivation similar to that developed by Chiu et al. [1] for ordinary Hill numbers. The decomposition procedures of all three measures are generally parallel and interpretations are similar. A summary of the decomposition of all three measures along with their interpretations are given in Table 2 of the main text.

Since the decomposition procedures for the mean functional diversity and total functional diversity are better understood via partitioning the functional Hill numbers, we first present the details of decomposing the functional Hill numbers $^qD(Q)$ and then apply the results to the total functional diversity $^qFD(Q)$. In this Appendix, we mainly focus on the decomposition of the functional Hill numbers and the total functional diversity.

The functional gamma Hill number is defined as the effective number of equally abundant and equally distinct species in the pooled assemblage, where species abundances are pooled over assemblages. As discussed in the main text, the species relative abundance set in the pooled assemblage can be expressed as $\{ z_{i+} / z_{++}; i = 1, 2,\ldots, S\}$. Thus, it follows from Eq. 3 of the main text that the functional gamma Hill number of order q is:

$$^qD_\gamma(Q) = \left[\sum_{i=1}^{S} \sum_{j=1}^{S} \frac{d_{ij}}{Q} \left(\frac{z_{i+}}{z_{++}} \frac{z_{j+}}{z_{++}} \right)^q \right]^{1/2(1-q)} \qquad q \geq 0, \ q \neq 1. \qquad (B1)$$

$$^1D_\gamma(Q) = \lim_{q \to 1} {}^qD_\gamma(Q) = \exp\left[-\frac{1}{2} \sum_{i=1}^{S} \sum_{j=1}^{S} \frac{d_{ij}}{Q} \left(\frac{z_{i+}}{z_{++}} \frac{z_{j+}}{z_{++}} \right) \log\left(\frac{z_{i+}}{z_{++}} \frac{z_{j+}}{z_{++}} \right) \right].$$

For a single assemblage, the Hill number of order q, or the "effective number of species", is defined as the number of equally abundant species that would be needed to give the same value of the diversity measure. We now extend this concept to functional alpha Hill number. Assume that there are N assemblages and there are S species in the pooled assemblage. For $i = 1, 2, \ldots, S, k = 1, 2, \ldots, N$, let z_{ik} denote the species abundance of the ith species in the kth assemblage. As discussed in the main text, the variable z_{ik} could be absolute abundance, relative abundances, incidences,

biomasses, or cover areas. We can formulate the S x N species abundance matrix $\boldsymbol{Z}$ as

$$\boldsymbol{Z} = \begin{bmatrix} z_{11} & z_{12} & \cdots & z_{1N} \\ z_{21} & . & \cdots & . \\ . & . & \cdots & . \\ z_{S1} & z_{S2} & \cdots & z_{SN} \end{bmatrix}.$$

Let the kth column of $\boldsymbol{Z}$ be denoted by $z_k = (z_{1k}, z_{2k}, ..., z_{Sk})^T$, (the super-script "$T$" denotes the "transpose"of a matrix) and this column denotes the species abundances of the kth assemblage. So the matrix $\boldsymbol{Z}$ can be denoted by $(z_1, z_2, ..., z_N)$. The matrix total is $z_{++} = \sum_{i=1}^{S} \sum_{j=1}^{N} z_{ij}$. Let d_{ij} denote the functional distance between the ith and the jth species, $d_{ij} \geq 0$, and $d_{ij} = d_{ji}$. Denote the S x S symmetric pairwise distance matrix by $\Delta = [d_{ij}]$.

Following Chiu, we define functional alpha Hill number as the effective number of equally abundant and equally distinct species per assemblage. Then the following derivation leads to a formula. Assume that each of the N assemblages is equivalent to an idealized reference assemblage which contains A species and all species are equally common with a constant abundance $\bar{a}$. That is, in the kth idealized assemblage, the abundance set can be expressed as an A x1 column vector $\boldsymbol{b}_k = (\bar{a}, \bar{a}, ..., \bar{a})^T$, $k = 1, 2, ..., N$. Whether there are shared species among these N idealized assemblages is not relevant because functional alpha Hill number is independent of shared information (i.e., the value is independent of the between-assemblage information). Without loss of generality, we assume the N idealized assemblages form an abundance matrix $\boldsymbol{B} = (\boldsymbol{b}_1, \boldsymbol{b}_2, ..., \boldsymbol{b}_N)$. A good feature of the idealized N-assemblage matrix $\boldsymbol{B}$ is that each assemblage has A equally abundant species and assemblage sizes are all equal.

Let Q be the quadratic entropy of the pooled assemblage of the actual assemblages, i.e., $Q = \sum_i \sum_j d_{ij} z_{i+} z_{j+} / z_{++}^2$. In the idealized assemblage, we have a constant distance Q for all species pairs; see Table 1 of the main text for illustration. The idealized distance matrix is denoted by $\Delta(Q)$.

The two matrices $\boldsymbol{Z}$ and $\boldsymbol{B}$ are "equivalent" in the sense that any q-th power sum ($q \geq 0$) of the elements of the matrix $\boldsymbol{Z} = (z_1, z_2, ..., z_N)$ should give identical values when the same function is applied to the idealized reference matrix $\boldsymbol{B} = (\boldsymbol{b}_1, \boldsymbol{b}_2, ..., \boldsymbol{b}_N)$ with $\boldsymbol{b}_j = (\bar{a}, \bar{a}, ..., \bar{a})^T$ for all j. We consider the following two special functions:

(1) The sum of all elements in $\boldsymbol{Z}$ is thus equal to the corresponding sum in $\boldsymbol{B}$. This sum for $\boldsymbol{Z} = (z_1, z_2, ..., z_N)$ is z_{++}, whereas the sum for $\boldsymbol{B} = (\boldsymbol{b}_1, \boldsymbol{b}_2, ..., \boldsymbol{b}_N)$ is $AN\bar{a}$ (since each column has A species and there are N columns with all elements being equal to $\bar{a}$). Thus we obtain $\bar{a} = z_{++} / (AN)$.

(2) Choose all possible combinations of any two columns of the matrix $\boldsymbol{Z}$ and form the weighted (by distance) q-th power sum:

$$h(\mathbf{Z}, \Delta) = \sum_{k,m=1}^{N} \sum_{i,j=1}^{S} d_{ij} (z_{ik} z_{jm})^q .$$

Then the same function is applied to the simple reference assemblage to obtain

$$h(\mathbf{B}, \Delta(Q)) = \sum_{k,m=1}^{N} \sum_{i,j=1}^{A} Q(\overline{a}\overline{a})^q = \sum_{k,m=1}^{N} \sum_{i,j=1}^{A} Q\left(\frac{z_{++}}{AN}\right)^{2q} = QA^{2(1-q)} N^{2(1-q)} (z_{++})^{2q} .$$

Equating these two functions shows that A (the proposed functional alpha Hill number) has the following form:

$$A = {}^q D_\alpha(Q) = \frac{1}{N} \left[\sum_{k,m=1}^{N} \sum_{i,j=1}^{S} \frac{d_{ij}}{Q} \left(\frac{z_{ik}}{z_{++}} \frac{z_{jm}}{z_{++}} \right)^q \right]^{\frac{1}{2(1-q)}} \quad q \geq 0, \quad q \neq 1 ; \qquad \text{(B2)}$$

$$ {}^1 D_\alpha(Q) = \lim_{q \to 1} {}^q D_\alpha(Q) = \frac{1}{N} \exp\left[-\frac{1}{2} \sum_{k,m=1}^{N} \sum_{i,j=1}^{S} \frac{d_{ij}}{Q} \left(\frac{z_{ik}}{z_{++}} \frac{z_{jm}}{z_{++}} \right) \log\left(\frac{z_{ik}}{z_{++}} \frac{z_{jm}}{z_{++}} \right) \right], \quad q = 1.$$

This new alpha Hill number is interpreted as the "effective number of equally abundant and equally distinct species per assemblage". Note here in the above alpha formula, Q refers to the quadratic entropy of the pooled assemblage. Then based on a multiplicative partitioning, we define the functional beta Hill number of order q as the ratio:

$$ {}^q D_\beta(Q) = {}^q D_\gamma(Q) / {}^q D_\alpha(Q) .$$

This functional beta Hill number is interpreted as the effective number of equally large and completely distinct assemblages.

The above theory can be directly applied to decompose the (total) functional diversity ${}^q FD(Q)$. The functional gamma diversity is the effective total distance between species in the pooled assemblage with a constant distance Q for all species pairs, where Q denotes the quadratic entropy of the pooled assemblage. It follows from the derivation of the functional gamma Hill numbers that we have ${}^q FD_\gamma(Q) = Q \times [{}^q D_\gamma(Q)]^2$, and formulas are given in Eqs. 6a and 6b in the main text. The functional alpha diversity is defined as the effective total distance between species of a pair of assemblages. Our above derivation implies that ${}^q FD_\alpha(Q) = Q \times [{}^q D_\alpha(Q)]^2$. Substituting the functional alpha Hill number, we then obtain the formulas of the functional alpha diversity as shown in Eqs. 7a and 7b in the main text. The functional beta diversity based on a multiplicative rule is:

$$ {}^q FD_\beta(Q) = \frac{{}^q FD_\gamma(Q)}{{}^q FD_\alpha(Q)} = [{}^q D_\beta(Q)]^2 .$$

Similar gamma and alpha components for the mean functional diversity can be also derived respectively as ${}^q MD_\gamma(Q) = Q \times [{}^q D_\gamma(Q)]$ and ${}^q MD_\alpha(Q) = Q \times [{}^q D_\alpha(Q)]$; the corresponding beta component is

$$ {}^q MD_\beta(Q) = \frac{{}^q MD_\gamma(Q)}{{}^q MD_\alpha(Q)} = {}^q D_\beta(Q) .$$

See Table 2 of the main text for a summary.

In the following proposition, we prove for all $q \geq 0$ that the functional beta Hill number of order q is always between 1 and N regardless of the functional alpha Hill numbers. Also, the functional beta diversity of order q is always between 1 and N^2 regardless of the functional alpha diversity. Then the alpha and beta components obtained from decomposing each measure are unrelated (or independent).

For any arbitrary symmetric matrices and all orders of $q \geq 0$, when the number of assemblages, N, is fixed, the functional beta Hill number of order q is always in the range $[1, N]$, i.e., $1 \leq {}^{q}D_{\beta}(Q) = {}^{q}MD_{\beta}(Q) \leq N$ for all $q \geq 0$. The functional beta diversity of order q is always in the range $[1, N^2]$, i.e., $1 \leq {}^{q}FD_{\beta}(Q) \leq N^2$ for all $q \geq 0$. These conclusions are based on the following proposition.

Proposition S2.1:

(a) For *all* $q \geq 0$, the functional alpha and gamma Hill numbers satisfy the following inequality:

$$ {}^{q}D_{\alpha}(Q) \leq {}^{q}D_{\gamma}(Q) \leq N[{}^{q}D_{\alpha}(Q)], \tag{B3} $$

or equivalently,

$$ [{}^{q}D_{\gamma}(Q)]/N \leq {}^{q}D_{\alpha}(Q) \leq {}^{q}D_{\gamma}(Q). \tag{B4} $$

Similar inequalities are also valid for the mean functional alpha and gamma diversities.

(b) For *all* $q \geq 0$, the functional alpha and gamma diversities satisfy the following inequality:

$$ {}^{q}FD_{\alpha}(Q) \leq {}^{q}FD_{\gamma}(Q) \leq N^2[{}^{q}FD_{\alpha}(Q)], \tag{B5} $$

or equivalently,

$$ {}^{q}FD_{\gamma}(Q)/N^2 \leq {}^{q}FD_{\alpha}(Q) \leq {}^{q}FD_{\gamma}(Q). \tag{B6} $$

Proof: It is sufficient to prove (B3) and (B4) because (B5) and (B6) follow directly from (B3) and (B4). From Eqs. (B1) and (B2), the functional gamma and alpha Hill number for $q \neq 1$ is respectively

$$ {}^{q}D_{\gamma}(Q) = \left[\sum_{i,j=1}^{S} \frac{d_{ij}}{Q} \left(\frac{z_{i+}}{z_{++}} \frac{z_{j+}}{z_{++}} \right)^{q} \right]^{\frac{1}{2(1-q)}} = \left[\sum_{i,j=1}^{S} \frac{d_{ij}}{Q} \left(\sum_{k=1}^{N} \frac{z_{ik}}{z_{++}} \sum_{m=1}^{N} \frac{z_{jm}}{z_{++}} \right)^{q} \right]^{\frac{1}{2(1-q)}}, $$

and

$$ {}^{q}D_{\alpha}(Q) = \frac{1}{N} \left[\sum_{k,m=1}^{N} \sum_{i,j=1}^{S} \frac{d_{ij}}{Q} \left(\frac{z_{ik}}{z_{++}} \frac{z_{jm}}{z_{++}} \right)^{q} \right]^{\frac{1}{2(1-q)}}. $$

We first prove ${}^{q}D_{\gamma}(Q) \leq N \times [{}^{q}D_{\alpha}(Q)]$. For $q > 1$, this conclusion follows directly from the following inequality:

$$[{}^qD_\gamma(Q)]^{2(1-q)} = \sum_{i,j=1}^{S} \frac{d_{ij}}{Q} \left(\sum_{k=1}^{N} \frac{z_{ik}}{z_{++}} \sum_{m=1}^{N} \frac{z_{jm}}{z_{++}} \right)^q \geq \sum_{i,j=1}^{S} \frac{d_{ij}}{Q} \sum_{k,m=1}^{N} \left(\frac{z_{ik}}{z_{++}} \frac{z_{jm}}{z_{++}} \right)^q = [N \times {}^qD_\alpha(T)]^{2(1-q)}.$$

Similarly, for $0 \leq q \leq 1$, the same conclusion follows from the following inequality:

$$[{}^qD_\gamma(Q)]^{2(1-q)} = \sum_{i,j=1}^{S} \frac{d_{ij}}{Q} \left(\sum_{k=1}^{N} \frac{z_{ik}}{z_{++}} \sum_{m=1}^{N} \frac{z_{jm}}{z_{++}} \right)^q \leq \sum_{i,j=1}^{S} \frac{d_{ij}}{Q} \sum_{k,m=1}^{N} \left(\frac{z_{ik}}{z_{++}} \frac{z_{jm}}{z_{++}} \right)^q = [N \times {}^qD_\alpha(Q)]^{2(1-q)}.$$

For $q = 1$, note that we have

$$\log {}^1D_\gamma(Q) = -\frac{1}{2} \sum_{i=1}^{S} \sum_{j=1}^{S} \frac{d_{ij}}{Q} \left(\sum_{k=1}^{N} \frac{z_{ik}}{z_{++}} \sum_{m=1}^{N} \frac{z_{jm}}{z_{++}} \right) \log \left(\sum_{k=1}^{N} \frac{z_{ik}}{z_{++}} \sum_{m=1}^{N} \frac{z_{jm}}{z_{++}} \right),$$

and

$$\log[N \times {}^1D_\alpha(Q)] = -\frac{1}{2} \sum_{k,m=1}^{N} \sum_{i,j=1}^{S} \frac{d_{ij}}{Q} \frac{z_{ik}}{z_{++}} \frac{z_{jm}}{z_{++}} \log \left(\frac{z_{ik}}{z_{++}} \frac{z_{jm}}{z_{++}} \right).$$

Since $\log \left(\sum_{k=1}^{N} \frac{z_{ik}}{z_{++}} \sum_{m=1}^{N} \frac{z_{jm}}{z_{++}} \right) \geq \log \left(\frac{z_{ik}}{z_{++}} \frac{z_{jm}}{z_{++}} \right)$, the proof is readily seen for $q = 1$.

To prove ${}^qD_\gamma(Q) \geq {}^qD_\alpha(Q)$, note that for $q > 1$, $f(x) = x^q$ is a convex function, so the Jensen inequality implies that for any species pair (i, j), we have

$$\left(\frac{1}{N^2} \sum_{k=1}^{N} \frac{z_{ik}}{z_{++}} \sum_{m=1}^{N} \frac{z_{jm}}{z_{++}} \right)^q \leq \frac{1}{N^2} \sum_{k=1}^{N} \sum_{m=1}^{N} \left(\frac{z_{ik}}{z_{++}} \frac{z_{jm}}{z_{++}} \right)^q.$$

That is,

$$\frac{1}{N^{2q}} \sum_{i,j=1}^{S} \frac{d_{ij}}{Q} \left(\sum_{k=1}^{N} \frac{z_{ik}}{z_{++}} \sum_{l=1}^{N} \frac{z_{jm}}{z_{++}} \right)^q \leq \frac{1}{N^2} \sum_{i,j=1}^{S} \frac{d_{ij}}{Q} \sum_{k=1}^{N} \sum_{m=1}^{N} \left(\frac{z_{ik}}{z_{++}} \frac{z_{jm}}{z_{++}} \right)^q,$$

which is equivalent to

$$N^{-2q}[{}^qD_\gamma(Q)]^{2(1-q)} \leq N^{-2}[N \times {}^qD_\alpha(Q)]^{2(1-q)},$$

yielding ${}^qD_\alpha(Q) \leq {}^qD_\gamma(Q)$ for $q > 1$.

For $0 \leq q < 1$, $f(x) = x^q$ is a concave function, so the Jensen inequality implies that for species pair (i, j), we have

$$\left(\frac{1}{N^2} \sum_{k=1}^{N} \frac{z_{ik}}{z_{++}} \sum_{m=1}^{N} \frac{z_{jm}}{z_{++}} \right)^q \geq \frac{1}{N^2} \sum_{k=1}^{N} \sum_{m=1}^{N} \left(\frac{z_{ik}}{z_{++}} \frac{z_{jm}}{z_{++}} \right)^q,$$

and thus

$$\frac{1}{N^{2q}} \sum_{i,j=1}^{S} \frac{d_{ij}}{Q} \left(\sum_{k=1}^{N} \frac{z_{ik}}{z_{++}} \sum_{l=1}^{N} \frac{z_{jm}}{z_{++}} \right)^q \geq \frac{1}{N^2} \sum_{i,j=1}^{S} \frac{d_{ij}}{Q} \sum_{k=1}^{N} \sum_{m=1}^{N} \left(\frac{z_{ik}}{z_{++}} \frac{z_{jm}}{z_{++}} \right)^q.$$

The above inequality is equivalent to

$$N^{-2q}[{}^qD_\gamma(Q)]^{2(1-q)} \ge N^{-2}[N\times{}^qD_\alpha(Q)]^{2(1-q)},$$

which implies $\ {}^qD_\gamma(Q)\le{}^qD_\alpha(Q)\ $ for $\ 0\le q<1$.

To prove $\ {}^qD_\gamma(Q)\ge{}^qD_\alpha(Q)\ $ for $q = 1$, note that $\ f(x) = -x\log x\ $ is a concave function. The Jensen inequality leads to

$$-\left(\frac{1}{N^2}\sum_{k,m=1}^{N}\frac{z_{ik}}{z_{++}}\frac{z_{jm}}{z_{++}}\right)\log\left(\frac{1}{N^2}\sum_{k,m=1}^{N}\frac{z_{ik}}{z_{++}}\frac{z_{jm}}{z_{++}}\right) \ge -\frac{1}{N^2}\sum_{k,m=1}^{N}\left(\frac{z_{ik}}{z_{++}}\frac{z_{jm}}{z_{++}}\right)\log\left(\frac{z_{ik}}{z_{++}}\frac{z_{jm}}{z_{++}}\right).$$

Then we have

$$-\sum_{i,j=1}^{S}\frac{d_{ij}}{Q}\left(\sum_{k,m=1}^{N}\frac{z_{ik}}{z_{++}}\frac{z_{jm}}{z_{++}}\right)\log\left(\frac{1}{N^2}\sum_{k,m=1}^{N}\frac{z_{ik}}{z_{++}}\frac{z_{jm}}{z_{++}}\right) \ge -\sum_{i,j=1}^{S}\frac{d_{ij}}{Q}\sum_{k,m=1}^{N}\left(\frac{z_{ik}}{z_{++}}\frac{z_{jm}}{z_{++}}\right)\log\left(\frac{z_{ik}}{z_{++}}\frac{z_{jm}}{z_{++}}\right).$$

Equivalently,
$$2\log N + 2\log{}^1D_\gamma(Q) \ge 2\log N + 2\log{}^1D_\alpha(Q).$$

Then the inequality $\ {}^1D_\gamma(Q)\ge{}^1D_\alpha(Q)$ is obtained.

Proposition S2.2 (A property of monotonicity used for Example 2 of the main text): Consider N assemblages indexed by 1, 2, …, N. Assume that species a is a non-shared species in an assemblage (say, Assemblage 1), i.e., species a does not exist in any other assemblages. Assume that species b is a non-shared species in a different assemblage (say, Assemblage 2). Let the functional distance between species a and species b be denoted by d_{ab}. Then the functional beta Hill number of order q is always a non-decreasing function with respect to d_{ab}. This implies that any differentiation measure based on the functional beta Hill number is also a non-decreasing function with respect to d_{ab}.

Proof: From the formulas of the functional alpha and gamma diversities, the functional beta diversity of order q is expressed as (see the main text for notation)

$$^qFD_\beta(Q) = \frac{{}^qFD_\gamma(Q)}{{}^qFD_\alpha(Q)} = \left[\frac{\sum_{i,j=1}^{S}d_{ij}\left(\frac{z_{i+}z_{j+}}{Qz_{++}^2}\right)^q}{\frac{1}{N^{2(1-q)}}\sum_{k,m=1}^{N}\sum_{i,j=1}^{S}d_{ij}\left(\frac{z_{i+}z_{j+}}{Qz_{++}^2}\right)^q}\right]^{\frac{1}{(1-q)}}.$$

We can simplify the above formula as

$$[{}^qFD_\beta(Q)]^{1-q} = \frac{\sum_{i,j=1}^{S}d_{ij}(z_{i+}z_{j+})^q}{\frac{1}{N^{2(1-q)}}\sum_{k,m=1}^{N}\sum_{i,j=1}^{S}d_{ij}(z_{ik}z_{jm})^q} \equiv \frac{A}{B}.$$

We first prove the proposition for $q > 1$. For the non-shared species pair (a, b) with distance d_{ab} between these two species, we have

$$\frac{A}{B} = \frac{\sum\limits_{i,j \neq (a,b)}^{S} d_{ij}(z_{i+}z_{j+})^q + d_{ab}(z_{a1}z_{b2})^q}{\frac{1}{N^{2(1-q)}} \sum\limits_{k,m=1}^{N} \sum\limits_{i,j \neq (a,b)}^{S} d_{ij}(z_{ik}z_{jm})^q + \frac{1}{N^{2(1-q)}} d_{ab}(z_{a1}z_{b2})^q} \quad .$$ (B7)

The derivative of Eq. (B7) with respective to d_{ab} is

$$\left\{ (z_{a1}z_{b2})^q \left[\frac{1}{N^{2(1-q)}} \sum\limits_{k,m=1}^{N} \sum\limits_{i,j \neq (a,b)}^{S} d_{ij}(z_{ik}z_{jm})^q \right] - \frac{1}{N^{2(1-q)}} (z_{a1}z_{b2})^q \left[\sum\limits_{i,j \neq (a,b)}^{S} d_{ij}(z_{i+}z_{j+})^q \right] \right\} / B^2$$

$$= \frac{1}{N^{2(1-q)}} (z_{a1}z_{b2})^q \left\{ \sum\limits_{k,m=1}^{N} \sum\limits_{i,j \neq (a,b)}^{S} d_{ij}(z_{ik}z_{jm})^q - \sum\limits_{i,j \neq (a,b)}^{S} d_{ij}(z_{i+}z_{j+})^q \right\} / B^2.$$ (B8)

Since for $q > 1$ and for any distance between species i and j, we have

$$d_{ij}\left[\sum\limits_{k,m=1}^{N} (z_{ik}z_{jm})^q \right] \leq d_{ij}\left[\left(\sum\limits_{k=1}^{N} z_{ik} \right)\left(\sum\limits_{m=1}^{N} z_{jm} \right) \right]^q = d_{ij}(z_{i+}z_{j+})^q.$$

Thus, the derivative in Eq. (B8) must be non-positive. Therefore, $[^q FD_\beta(Q)]^{1-q}$ is non-increasing with d_{ab} for $q > 1$, implying that $^q FD_\beta(Q)$ is non-decreasing with d_{ab}. For $q < 1$, parallel steps also lead the same conclusion. The proof for $q = 1$ is direct.

Reference

1. Chiu C-H, Jost L, Chao A (2014) Phylogenetic beta diversity, similarity, and differentiation measures based on Hill numbers. Ecol Monogr 84: 21–44.

Supporting Information

Distance-based functional diversity measures and their decomposition: a framework based on Hill numbers

Chun-Huo Chiu and Anne Chao
Institute of Statistics, National Tsing Hua University, Hsin-Chu, Taiwan 30043

Appendix S3: Four classes of functional similarity/differentiation measures

As proved in Appendix S2, our functional beta Hill number $^{q}D_{\beta}(Q)$ $(= ^{q}MD_{\beta}(Q))$ always takes values between 1 and N (Proposition S2.1 in Appendix S2) and the functional beta diversity $^{q}FD_{\beta}(Q)$ always takes values between 1 and N^{2}. Thus both can be transformed onto [0, 1] to obtain normalized similarity and differentiation measures among assemblages. The similarity measures based on $^{q}D_{\beta}(Q)$ quantify *species-overlap* from different perspectives, whereas the similarity measures based on $^{q}FD_{\beta}(Q)$ quantify *distance-overlap* from different perspectives. In most applications, we suggest using the distance-overlap (or similarity) measures and their corresponding differentiation measures. In the main text, we have briefly introduced the two major classes of normalized distance-overlap measures; see Table 3 of the main text. Here we provide more details along with interpretations for all four classes of distance-overlap measures.

(A) Similarity/differentiation measures based on the functional beta diversity (summarized in Table 3 of the main text)

(1) A class of local (functional) distance-overlap measures from the perspective of a pair of local assemblages

$$C_{qN}^{*}(Q) = \frac{N^{2(1-q)} - [^{q}FD_{\beta}(Q)]^{1-q}}{N^{2(1-q)} - 1}. \tag{C1}$$

This measure gives the effective average proportion of the species pairwise distances in a pair of local assemblages that are shared with all other pairs of assemblages.

(1a) $q = 0$

We first give the special case of $q = 0$ to intuitively explain its properties: since $^{0}FD_{\gamma}(Q) = FAD_{\gamma}$ and $^{0}FD_{\alpha}(Q) = FAD_{pair} / N^{2}$, the measure $C_{0N}^{*}(Q)$ reduces to

$$C_{0N}^{*}(Q) = \frac{(FAD_{pair} - FAD_{\gamma})/(N^{2} - 1)}{FAD_{pair} / N^{2}}, \tag{C2}$$

where FAD_{γ} denotes the sum of pairwise distances in the pooled assemblage, and FAD_{pair} is

the sum of *FAD*s over all possible pairs of assemblages (there are N^2 pairs of assemblages). As discussed in the text, the measure $C^*_{0N}(Q)$ thus quantifies the proportion of repeated distances in a pair of local assemblages. Its interpretation is conceptually analogous to the classic Sørensen similarity index and can be regarded as an extension of the classic Sørensen index to functional similarity. It is referred to as "func-Sørensen" in Table 3 of the main text. For the special case of $N = 2$, let the index set Ω_1 denote the species in Assemblage I, Ω_2 denote the species in Assemblage II, and Ω_{12} denote the shared species (species in both assemblages). Then we can rewrite Eq. (C2) as

$$C^*_{02}(Q) = \frac{\left(2\sum_{i\in\Omega_1}\sum_{j\in\Omega_{12}}d_{ij} + 2\sum_{i\in\Omega_2}\sum_{j\in\Omega_{12}}d_{ij} - \sum_{i\in\Omega_{12}}\sum_{j\in\Omega_{12}}d_{ij}\right)/3}{\left(\sum_{i\in\Omega_1}\sum_{j\in\Omega_1}d_{ij} + 2\sum_{i\in\Omega_1}\sum_{j\in\Omega_2}d_{ij} + \sum_{i\in\Omega_2}\sum_{j\in\Omega_2}d_{ij}\right)/4} . \tag{C3}$$

In this expression, the denominator is the alpha functional diversity of order 0, which is the average of *FAD*s over all four assemblage pairs (I, I), (I, II), (II, I) and (II, II); the numerator is the average of all repeated distances in the $N^2 - 1 = 3$ pairs of assemblages (excluding the assemblage-pair in which a distance is first counted).

(1b) $q = 1$

Letting q in Eq. (C1) tend to 1, we have

$$C^*_{1N}(Q) = 1 - \frac{\log^1 FD_\gamma(Q) - \log^1 FD_\alpha(Q)}{2\log N}. \tag{C4}$$

We refer to it as the N-assemblage "func-Horn" distance-overlap measure because when all distances are identical and (z_{ik}) represents species relative abundance within each assemblage, it reduces to the classical Horn measure. [1].

(1c) $q = 2$

The general formula for $C^*_{2N}(Q)$ shown in Table 3 of the main text is:

$$C^*_{2N}(Q) = \frac{\sum_{i,j}^S d_{ij}\sum_{k,m}^N\left[\left(\frac{z_{i+}z_{j+}}{N}\right)^2 - (z_{ik}z_{jm})^2\right]}{(N^2-1)\sum_{i,j}^S\sum_{k,m}^N d_{ij}(z_{ik}z_{jm})^2} .$$

Here we give the formula for $N = 2$ to interpret the measure. Let $h_{11} = \frac{z_{i1}}{z_{++}}\frac{z_{j1}}{z_{++}}$ be the product of relative abundances of species pair (i, j) in the assemblage-pair (I, I), and a similar interpretation pertains to the notation $h_{12} = \frac{z_{i1}}{z_{++}}\frac{z_{j2}}{z_{++}}$, $h_{21} = \frac{z_{i2}}{z_{++}}\frac{z_{j1}}{z_{++}}$, $h_{22} = \frac{z_{i2}}{z_{++}}\frac{z_{j2}}{z_{++}}$. Then the measure $C^*_{2N}(Q)$ for the special case of $N = 2$ reduces to

$$C_{22}^*(Q) = \frac{\sum_{i,j}^{S} d_{ij} 2[h_{11}h_{12} + h_{11}h_{21} + h_{11}h_{22} + h_{12}h_{21} + h_{12}h_{22} + h_{21}h_{22}]/12}{\sum_{i,j}^{S} d_{ij}[h_{11}h_{11} + h_{12}h_{12} + h_{21}h_{21} + h_{22}h_{22}]/4} \quad . \tag{C5}$$

It is called "func-Morisita-Horn" similarity measure because the concept is generally similar to the classic Morisita-Horn measure [2]. The denominator in Eq. (C5) represents the weighted (by abundances) average of distances for any two identical pairs of assemblages (there are four identical pairs, i.e., (I, I) vs. (I, I), (I, II) vs. (I, II), (II, I) vs. (II, I) and (II, II) vs. (II, II)).The numerator in Eq. (C5) represents the weighted (by abundances) average of *shared distances* for any two non-identical pairs of assemblages (there are 12 pairs, i.e., (I, I) vs. (I, II) (same as (I, II) vs. (I, I)), (I, I) vs. (II, I) (same as (II, I) vs. (I, I)), …, (II, I) vs. (II, II) (same as (II, II) vs. (II, I).) This interpretation extends the formulation of Chao et al. [2] for the classic Morisita-Horn measure to its functional version. The difference is that here we consider "assemblage-pairs" rather than "individual assemblage" as in the classical measure.

(1d) A general order of $q \geq 0$

Here the sense of "effective" is similar to that for the abundance-based local overlap measure C_{qN} [1,2], but "N assemblages" should be replaced with "N^2 pairs of assemblages", and "species" should be replaced by the unit of "distance". The meaning of "effective" average proportion for a value of $C_{qN}^*(Q) = p$ is described as follows. We can construct a set of N^2 idealized assemblage-pairs: the total species pairwise distance of the assemblage-pair is $^qFD_\alpha(Q)$ units of "distance", with exactly $^qFD_\alpha(Q) \times p$ units of "distance" shared by all of them, and the remaining units of "distance" of each assemblage-pair not shared by any other assemblage-pairs. In the pooled assemblage, the gamma functional diversity is $^qFD_\gamma(Q)$ units of "distance". Then the $C_{qN}^*(Q)$ measure of the actual set of N^2 assemblage-pairs is the same as that of this set of idealized assemblage-pairs. This gives a simple set of idealized assemblage-pairs to intuitively understand the interpretation of the measure $C_{qN}^*(Q)$.

(2) A class of regional (functional) distance-overlap measures in the pooled assemblage

$$U_{qN}^*(Q) = \frac{[1/^qFD_\beta(Q)]^{1-q} - (1/N)^{2(1-q)}}{1 - (1/N)^{2(1-q)}} \quad . \tag{C6}$$

This measure quantifies the effective proportion of the species pairwise distances in the pooled assemblage that are shared with all pairs of local assemblages.

(2a) $q = 0$

For the special case of $q = 0$, the measure $U_{0N}^*(Q)$ reduces to

$$U^*_{0N}(Q) = \frac{(FAD_{pair} - FAD_\gamma)/(N^2 - 1)}{FAD_\gamma}.$$ (C7)

The interpretation is similar to the measure $C^*_{0N}(Q)$ except that it is normalized by the gamma *FAD*. Thus, our index can be regarded as an extension of the Jaccard index to functional similarity. So this measure $U^*_{0N}(Q)$ is referred to as the *N*-assemblage "func-Jaccard" in Table 3 of the main text. For the special case of $N = 2$, we have

$$U^*_{02}(Q) = \frac{\left(2\sum\limits_{i\in\Omega_1}\sum\limits_{j\in\Omega_{12}} d_{ij} + 2\sum\limits_{i\in\Omega_2}\sum\limits_{j\in\Omega_{12}} d_{ij} - \sum\limits_{i\in\Omega_{12}}\sum\limits_{j\in\Omega_{12}} d_{ij} \right)/3}{\sum\limits_{i\in\Omega}\sum\limits_{j\in\Omega} d_{ij}},$$ (C8)

Where Ω_1, Ω_2 and Ω_{12} are defined earlier and Ω denotes the species in the pooled assemblage. Comparing this expression with $C^*_{02}(Q)$ in Eq. (C3), we see the numerators of the two measures are identical, but here it is normalized by FAD_γ in the denominator (i.e., from a regional view).

(2b) $q = 1$

Letting q in Eq. (C6) tend to 1, we obtain the same formula as the *N*-assemblage local distance-overlap measure $C^*_{1N}(Q)$. That is, $U^*_{1N}(Q) = C^*_{1N}(Q)$. Hence the measure $U^*_{1N}(Q)$ is also called "func-Horn" measure in Table 3 of the main text.

(2c) $q = 2$

The general formula for $U^*_{2N}(Q)$ shown in Table 3 of the main text is:

$$U^*_{2N}(Q) = \frac{\sum\limits_{i,j}^{S} d_{ij} \sum\limits_{k,m}^{N}\left[\left(\frac{z_{i+}z_{j+}}{N} \right)^2 - (z_{ik}z_{jm})^2 \right]}{(1 - 1/N^2)\sum\limits_{i,j}^{S} d_{ij}(z_{i+}z_{j+})^2}.$$

Here we give the formula for $N = 2$ to interpret the measure:

$$U^*_{22}(Q) = \frac{\sum\limits_{i,j}^{S} d_{ij}\, 2[h_{11}h_{12} + h_{11}h_{21} + h_{11}h_{22} + h_{12}h_{21} + h_{12}h_{22} + h_{21}h_{22}]/12}{\sum\limits_{i,j}^{S} d_{ij}(\bar{p}_{i+}\bar{p}_{j+})^2}.$$ (C9)

The numerator is the same as that in Eq. (C5), but here it is from a regional perspective. So it is referred to as "func-regional-overlap" measure in Table 3 of the main text.

(2d) A general order of $q \geq 0$

Here the sense of "effective" is similar to that for the abundance-based measure U_{qN} [1], but "N assemblages" should be replaced with "N^2 pairs of assemblages", and "species" should be replaced by the unit of "distance". The meaning of "effective" proportion for the measure $U_{qN}^*(Q)$ is slightly different from that for the measure $C_{qN}^*(Q)$. When $U_{qN}^*(Q) = u$, consider the following set of N idealized assemblages: in the pooled assemblage, the total species pairwise distance is $^qFD_\gamma(Q)$ units of "distance", with exactly $^qFD_\gamma(Q) \times u$ units of "distance" shared by all N^2 assemblage-pairs, and the remaining $^qFD_\gamma(Q) \times (1-u)$ units of "distance" evenly distributed in N^2 assemblage-pairs. In each of the N idealized assemblages, the functional diversity is $^qFD_\alpha(Q)$ units of "distance". Then the $U_{qN}^*(Q)$ measure of the set of actual assemblages is the same as that of this set of idealized assemblage-pairs. Thus, we can intuitively understand the interpretation of the measure $U_{qN}^*(Q)$.

(3) A class of functional distance-homogeneity measures

$$S_{qN}^*(Q) = \frac{1/[^qFD_\beta(Q)] - 1/N^2}{1 - 1/N^2}.$$

For $q = 0$, it is the N-assemblage "func-Jaccard" measure $U_{0N}^*(Q)$. For $q = 2$, this measure is identical to $C_{2N}^*(Q)$, the "func-Morisita-Horn" measure. Thus, we have $S_{0N}^*(Q) = U_{0N}^*(Q)$ and $S_{2N}^*(Q) = C_{2N}^*(Q)$. However, for $q = 1$, this measure does not reduce to the "func-Horn" measure.

(4) A class of the complement of distance-turnover rate

$$V_{qN}^*(Q) = \frac{N^2 - {^qFD_\beta(Q)}}{N^2 - 1}.$$

The corresponding differentiation measure $[^qFD_\beta(Q) - 1]/(N^2 - 1)$ quantifies distance-turnover rate per assemblage pair. When $q = 0$, the measure $V_{0N}^*(Q)$ is identical to the "func-Sørensen" measure. For $q = 2$, this measure is identical to $U_{2N}^*(Q)$, the "func-regional-overlap" measure. That is, we have $V_{0N}^*(Q) = C_{0N}^*(Q)$ and $V_{2N}^*(Q) = U_{2N}^*(Q)$. However, for $q = 1$, this measure does not reduce to the "func-Horn" measure.

(B) Similarity/differentiation measures based on the functional beta Hill numbers (Table S3.1)

(1) A class of local (functional) species-overlap measures

$$C_{qN}(Q) = \frac{N^{1-q} - [^qD_\beta(Q)]^{1-q}}{N^{1-q} - 1}. \tag{C10}$$

This measure is interpreted as the effective average proportion of the species in an individual

assemblage that are shared with all other assemblages. The corresponding differentiation measure $1 - C_{qN}(Q)$ quantifies the effective average proportion of the total non-shared species in an individual assemblage.

(1a) $q = 0$

We first write the formula for the special case of $q = 0$ to intuitively understand the measure:

$$C_{0N}(Q) = \frac{N - N(FAD_\gamma / FAD_{pair})^{1/2}}{N - 1} = \frac{(FAD_{pair})^{1/2} - (FAD_\gamma)^{1/2}}{(N-1)(FAD_{pair})^{1/2} / N}. \qquad (C11)$$

In the special case that all species are equally distinct, we have $FAD_\gamma = S^2 \times Q$ and $FAD_{pair} = (N\overline{S})^2 \times Q$, where S is the species richness in the pooled assemblage and $\overline{S}$ is the average species richness per assemblage. Consequently, the measure $C_{0N}(Q)$ reduces to the classic measure C_{0N} (N-assemblage Sørensen similarity measure [2,3]), i.e., $C_{0N} = \frac{(N\overline{S} - S)}{(N-1)\overline{S}}$.

From this view, the measure $C_{0N}(Q)$ is referred to as the "func-Sørensen (species-overlap)" measure in Table S3.1. We add "species-overlap" in order to distinguish it from "func-Sørensen" (distance-overlap) based on the functional beta diversity in Table 3 of the main text.

(1b) $q = 1$

For $q = 1$, we have the following formula:

$$C_{1N}(Q) = 1 - \frac{\log[{}^1D_\gamma(Q)] - \log[{}^1D_\alpha(Q)]}{\log N}.$$

This measure is referred to as the "func-Horn" measure because $C_{1N}^*(Q) = C_{1N}(Q)$ and $C_{1N}^*(Q)$ is called "func-Horn" in Table 3 of the main text. This indicates that for $q = 1$, the local distance- and species-overlap measures are identical.

(1c) $q = 2$

For $q = 2$, the general formula of $C_{2N}(Q)$ can be expressed as

$$C_{2N}(Q) = \frac{\left(\sum_{i,j} d_{ij}(z_{i+} z_{j+})^2\right)^{1/2} - \left(\sum_{k,m} \sum_{i,j} d_{ij}(z_{ik} z_{jm})^2\right)^{1/2}}{(N-1)\left(\sum_{k,m} \sum_{i,j} d_{ij}(z_{ik} z_{jm})^2\right)^{1/2}}.$$

This measure is called as the "func-Morisita-Horn (species-overlap)" measure because when all species are equally distinct and the data (z_{ik}) represents species relative abundance within each assemblage, it reduces to the classic N-assemblage Morisita-Horn overlap measure [2,3].

(1d) A general order of $q \geq 0$

The measure $C_{qN}(Q)$ is interpreted as the effective average proportion of the species in an individual assemblage that are shared with all other assemblages. The interpretation of "effective" for this measure is a direct extension of the classic measure C_{qN} discussed in Chao et al. [2,3] and Chiu et al. [1]. When $C_{qN}(Q) = p$, We can construct a set of N idealized assemblages: each of the N assemblages has ${}^{q}D_{\alpha}(Q)$ equally abundant and functionally equally distinct species (all pairwise distances are identical), with exactly ${}^{q}D_{\alpha}(Q) \times p$ species shared by all of them, and the remaining species of each assemblage not shared by any other assemblages. In the pooled assemblage, the gamma functional Hill number is ${}^{q}D_{\gamma}(Q)$. Then the $C_{qN}(Q)$ measure of the actual set of N assemblages is the same as that of this set of idealized assemblages. This gives a simple set of idealized assemblages to intuitively understand the interpretation of the measure $C_{qN}(Q)$.

(2) A class of regional (functional) species-overlap measures

$$U_{qN}(Q) = \frac{[1/{}^{q}D_{\beta}(Q)]^{1-q} - (1/N)^{1-q}}{1 - (1/N)^{1-q}}. \tag{C12}$$

This class of measures differs from the local functional species-overlap measures by taking a regional perspective. They give the effective proportion of the species in the pooled assemblage that are shared with all other assemblages. The corresponding differentiation measure $1 - U_{qN}(Q)$ quantifies the effective proportion of the total non-shared species in the pooled assemblage.

(2a) $q = 0$

For $q = 0$, the measure reduces to

$$U_{0N}(Q) = \frac{(FAD_{pair} / FAD_{\gamma})^{1/2} - 1}{N - 1} = \frac{(FAD_{pair})^{1/2} - (FAD_{\gamma})^{1/2}}{(N-1)(FAD_{\gamma})^{1/2}}. \tag{C13}$$

In the special case that all species are equally distinct, $U_{0N}(Q)$ reduces to the classic measure U_{0N} (N-assemblage Jaccard similarity measure [1]), i.e., $U_{0N} = \frac{(N\bar{S} - S)}{(N-1)S}$; see Table S3.1 at the end of this appendix. So the measure $U_{0N}(Q)$ is referred to as the N-assemblage "func-Jaccard (species-overlap)" measure in Table S3.1. We add "species-overlap" in order to distinguish it from "func-Jaccard" (distance-overlap) measure based on the functional beta diversity in Table 3 of the main text.

(2b) For $q = 1$, this measure is identical to the N-assemblage "func-Horn" overlap measure $C_{1N}(Q)$. So this measure has unique feature that both ${}^{q}FD_{\beta}(Q)$ and ${}^{q}D_{\beta}(Q)$ lead to the

same local and regional overlap measures, i.e., $C_{1N}^{*}(Q) = C_{1N}(Q) = U_{1N}^{*}(Q) = U_{1N}(Q)$.

(2c) For $q = 2$, the general formula can be expressed as

$$U_{2N}(Q) = \frac{\left(\sum\limits_{i,j} d_{ij}(z_{i+}z_{j+})^2\right)^{1/2} - \left(\sum\limits_{k,m}\sum\limits_{i,j} d_{ij}(z_{ik}z_{jm})^2\right)^{1/2}}{(1-1/N)\left(\sum\limits_{i,j} d_{ij}(z_{i+}z_{j+})^2\right)^{1/2}}.$$

The numerator is the same as that of the measure $C_{2N}(Q)$, but the denominator is from a regional view. So it is referred to as "func-regional-overlap (species-overlap)" in Table S3.1.

(2d) A general order of $q \geq 0$

The measure $U_{qN}(Q)$ is interpreted as the effective average proportion of the species in the pooled assemblage that are shared with all other assemblages. The meaning of "effective" for this class is slightly different from that for $C_{qN}(Q)$. When $U_{qN}(Q) = u$, consider the following set of idealized assemblages: in the pooled assemblage, we have ${}^q D_\gamma(Q)$ equally abundant and equally distinct species, with exactly ${}^q D_\gamma(Q) \times u$ species shared by all N assemblages, and the remaining ${}^q D_\gamma(Q) \times (1-u)$ species evenly distributed in N assemblages. In each of the N idealized assemblages, the effective functional Hill number is ${}^q D_\alpha(Q)$. Then the measure $U_{qN}(Q)$ of the set of actual assemblages is the same as that of this set of idealized assemblages. This gives a simple set of idealized assemblages to intuitively understand the interpretation of the measure $U_{qN}(Q)$.

(3) A class of functional species-homogeneity measures

$$S_{qN}(Q) = \frac{1/{}^q D_\beta(Q) - 1/N}{1 - 1/N}.$$

For $q = 0$, this measure is the "func-Jaccard (species-overlap)" measure $U_{0N}(Q)$. For $q = 2$, this measure is identical to $C_{2N}(Q)$, the "func-Morisita-Horn (species-overlap)" similarity measure. Thus, we have $S_{0N}(Q) = U_{0N}(Q)$ and $S_{2N}(Q) = C_{2N}(Q)$. However, for $q = 1$, this measure does not reduce to the "func-Horn" overlap measure.

(4) A class of the complement of functional species-turnover rate

$$V_{qN}(Q) = \frac{N - {}^q D_\beta(Q)}{N - 1}.$$

The corresponding differentiation measure $1 - V_{qN}(Q)$ is the normalized functional species-turnover rate per assemblage. When $q = 0$, the measure is identical to the "func-Sørensen

(species-overlap)" measure. For $q = 2$, this measure is identical to $U_{2N}(Q)$, the "func-regional-overlap (species-overlap)" measure. That is, we have $V_{0N}(Q) = C_{0N}(Q)$ and $V_{2N}(Q) = U_{2N}(Q)$. However, for $q = 1$, this measure does not reduce to the "func-Horn" overlap measure.

Table S3.1. Two classes of functional species-overlap (or similarity) measures and their special cases. The corresponding differentiation measures are the one-complements of the overlap measures. In the second column, "taxonomic" measures refer to the special case that all species are equally distinct as in the classic abundance-based measures. (The indices i and j are used to identify species, $i, j = 1, 2, …, S$, and the indices k and m are used to identify assemblages, $k, m = 1, 2, …, N$.)

Order	Measure	Local species-overlap $$C_{qN}(Q) = \frac{N^{1-q} - [^qD_\beta(Q)]^{1-q}}{N^{1-q} - 1}$$	Regional species-overlap $$U_{qN}(Q) = \frac{[1/^qD_\beta(Q)]^{1-q} - (1/N)^{1-q}}{1 - (1/N)^{1-q}}$$
$q = 0$	Functional	Func-Sørensen (species-overlap) $$\frac{N - N(FAD_\gamma)^{1/2}/(FAD_{pair})^{1/2}}{N - 1}$$	Func-Jaccard (species-overlap) $$\frac{(FAD_{pair})^{1/2}/(FAD_\gamma)^{1/2} - 1}{N - 1}$$
	Taxonomic	Classic Sørensen $$\frac{N - S/\bar{S}}{N - 1}$$	Classic Jaccard $$\frac{N\bar{S}/S - 1}{N - 1}$$
$q = 1$	Functional	Func-Horn $$1 - \frac{\log[^1D_\gamma(Q)] - \log[^1D_\alpha(Q)]}{\log N}$$	
	Taxonomic	Classic Horn $$\frac{H_{Sh,\alpha} - H_{Sh,\gamma} - \sum_{k=1}^{N} \frac{z_{+k}}{z_{++}}\log\left(\frac{z_{+k}}{z_{++}}\right)}{\log N} \quad ; \quad 1 - \frac{H_{Sh,\gamma} - H_{Sh,\alpha}}{\log N} \text{ (if } z_{+k}=1, z_{++}=N)$$	
$q = 2$	Functional	Func-Morisita-Horn (species-overlap) $$\frac{\left(\sum_{i,j} d_{ij}(z_{i+}z_{j+})^2\right)^{1/2} - \left(\sum_{k,m}\sum_{i,j} d_{ij}(z_{ik}z_{jm})^2\right)^{1/2}}{(N-1)\left(\sum_{k,m}\sum_{i,j} d_{ij}(z_{ik}z_{jm})^2\right)^{1/2}}$$	Func-regional-overlap (species-overlap) $$\frac{\left(\sum_{i,j} d_{ij}(z_{i+}z_{j+})^2\right)^{1/2} - \left(\sum_{k,m}\sum_{i,j} d_{ij}(z_{ik}z_{jm})^2\right)^{1/2}}{(1-1/N)\left(\sum_{i,j} d_{ij}(z_{i+}z_{j+})^2\right)^{1/2}}$$
	Taxonomic	Classic Morisita-Horn $$1 - \frac{\sum_{i=1}^{S}\sum_{m>k}(z_{im} - z_{ik})^2}{(N-1)\sum_{i=1}^{S}\sum_{k=1}^{N} z_{ik}^2} ;$$ $$1 - \frac{H_{GS,\gamma} - H_{GS,\alpha}}{(1-1/N)(1 - H_{GS,\alpha})} \text{ (if } z_{+k}=1, z_{++}=N)$$	Classic regional-overlap $$1 - \frac{\sum_{i=1}^{S}\sum_{m>k}(z_{im} - z_{ik})^2}{(N-1)\sum_{i=1}^{S} z_{i+}^2} ;$$ $$1 - \frac{H_{GS,\gamma} - H_{GS,\alpha}}{(N-1)(1 - H_{GS,\gamma})} \text{ (if } z_{+k}=1, z_{++}=N)$$

Notes:

(1) z_{ik} = the abundance of the ith species in the kth assemblage, $z_{+k} = \sum_{i=1}^{S} z_{ik}$, $z_{i+} = \sum_{k=1}^{N} z_{ik}$, and $z_{++} = \sum_{i=1}^{S}\sum_{k=1}^{N} z_{ik}$; see the main text for details. If z_{ik} represents species relative abundance, then in all formulas we have $z_{+k}=1$, $z_{++} = N$.

(2) $q = 0$. FAD_γ = sum of the pairwise distances between species in the pooled assemblage; FAD_{pair} = sum of FADs over all possible pairs of assemblages (there are N^2 pairs of assemblages). S = species richness in the pooled assemblage. $\bar{S}$ = average species richness per

assemblage.

(3) $q = 1$. $H_{Sh,\gamma}$, $H_{Sh,\alpha}$ = gamma and alpha Shannon entropy.

(4) $q = 2$. $H_{GS,\gamma}$, $H_{GS,\alpha}$ = gamma and alpha Gini-Simpson index.

References

1. Chiu C-H, Jost L, Chao A (2014) Phylogenetic beta diversity, similarity, and differentiation measures based on Hill numbers. Ecol Monogr 84: 21–44.
2. Chao A, Jost L, Chiang SC, Jiang Y-H, Chazdon RL (2008) A two-stage probabilistic approach to multiple-community similarity indices. Biometrics 64: 1178–1186.
3. Chao A, Chiu C-H, Hsieh TC (2012) Proposing a resolution to debates on diversity partitioning. Ecology 93: 2037–2051.

Supporting Information

Distance-based functional diversity measures and their decomposition: a framework based on Hill numbers

Chun-Huo Chiu and Anne Chao
Institute of Statistics, National Tsing Hua University, Hsin-Chu, Taiwan 30043

Appendix S4: Functional beta diversity and functional diversity excess lead to the same classes of similarity and differentiation measures

As indicated in the main text, we can apply the additive decomposition to the three measures, $^qD(Q)$ (Eq. 3), $^qMD(Q)$ (Eq. 4a) and $^qFD(Q)$ (Eq. 4b); here the equations numbers refer to those in the main text. For example, we define the "functional diversity excess" as $^qFD_\gamma(Q) - {}^qFD_\alpha(Q)$. This is also an interpretable measure. However, this excess measure cannot be directly applied to compare the functional similarity or differentiation among assemblages across multiple sets of assemblages because it depends not only on the number of assemblages, but also on the corresponding functional alpha (equivalently, gamma) diversity. Following Chao et al. [1] and Chiu et al. [2], we can readily eliminate these dependences by using an appropriate normalization. For easy demonstration, we prove the theorem for differentiation measures instead of similarity measures.

(1) *A class of functional distance-differentiation measures from a local perspective*: $1 - C_{qN}^*(Q)$

We first prove the following two inequalities:

$$0 \le [^qFD_\alpha(Q)]^{1-q} - [^qFD_\gamma(Q)]^{1-q} \le (1 - N^{2(1-q)})[^qFD_\alpha(Q)]^{1-q}, \quad q > 1, \qquad (D1)$$

and

$$0 \le [^qFD_\gamma(Q)]^{1-q} - [^qFD_\alpha(Q)]^{1-q} \le (N^{2(1-q)} - 1)[^qFD_\alpha(Q)]^{1-q}, \quad 0 \le q < 1. \quad (D2)$$

It follows from $^qFD_\gamma(Q) \le N^2[^qFD_\alpha(Q)]$ (Proposition S2.1 of Appendix S2) that for $q > 1$, we have $[^qFD_\gamma(Q)]^{1-q} \ge N^{2(1-q)}[^qFD_\alpha(Q)]^{1-q}$. Then we obtain the inequality (D1). For $0 \le q < 1$, the inequality (D2) also follows directly from the same inequality. Dividing $[^qFD_\alpha(Q)]^{1-q} - [^qFD_\gamma(Q)]^{1-q}$ (for $q > 1$) and $[^qFD_\gamma(Q)]^{1-q} - [^qFD_\alpha(Q)]^{1-q}$ (for $0 \le q < 1$) by their respective maximum possible value gives the measure $1 - C_{qN}^*(Q)$. That is,

$$\frac{[^qFD_\gamma(Q)]^{1-q} - [^qFD_\alpha(Q)]^{1-q}}{(N^{2(1-q)} - 1)[^qFD_\alpha(Q)]^{1-q}} = \frac{[^qFD_\gamma(Q)/^qFD_\alpha(Q)]^{1-q} - 1}{N^{2(1-q)} - 1} = \frac{[^qFD_\beta(Q)]^{1-q} - 1}{N^{2(1-q)} - 1} = 1 - C_{qN}^*(Q).$$

Also, for $q = 1$, we have

$$\lim_{q\to 1}\frac{[{}^qFD_\gamma(Q)]^{1-q}-[{}^qFD_\alpha(Q)]^{1-q}}{(N^{2(1-q)}-1)[{}^qFD_\alpha(Q)]^{1-q}}=\frac{\log{}^1FD_\gamma(Q)-\log{}^1FD_\alpha(Q)}{2\log N}=1-C_{1N}^*(Q).$$

(2) *A class of functional distance-differentiation measures from a regional perspective*:
$1-U_{qN}^*(Q)$

The inequality in Proposition S2.1 of Appendix S2, ${}^qFD_\gamma(Q)/N^2\le {}^qFD_\alpha(Q)\le {}^qFD_\gamma(Q)$,

implies $(1/N^{2(1-q)})[{}^qFD_\gamma(Q)]^{1-q}\le[{}^qFD_\alpha(Q)]^{1-q}\le[{}^qFD_\gamma(Q)]^{1-q}$ for $0\le q<1$. Thus, we obtain

$$0\le[{}^qFD_\gamma(Q)]^{1-q}-[{}^qFD_\alpha(Q)]^{1-q}\le(1-1/N^{2(1-q)})[{}^qFD_\gamma(Q)]^{1-q},\quad 0\le q<1;$$

and

$$0\le[{}^qFD_\alpha(Q)]^{1-q}-[{}^qFD_\gamma(Q)]^{1-q}\le(1/N^{2(1-q)}-1)[{}^qFD_\gamma(Q)]^{1-q},\quad q>1.$$

Dividing $[{}^qFD_\gamma(Q)]^{1-q}-[{}^qFD_\alpha(Q)]^{1-q}$ (for $0\le q<1$) and $[{}^qFD_\alpha(Q)]^{1-q}-[{}^qFD_\gamma(Q)]^{1-q}$ (for $q>1$) by their respective maximum possible value gives the measure, $1-U_{qN}^*(Q)$. That is,

$$\frac{[{}^qFD_\gamma(Q)]^{1-q}-[{}^qFD_\alpha(Q)]^{1-q}}{(1-1/N^{2(1-q)})[{}^qFD_\gamma(Q)]^{1-q}}=\frac{1-[{}^qFD_\alpha(Q)/{}^qFD_\gamma(Q)]^{1-q}}{1-1/N^{2(1-q)}}=\frac{1-[1/{}^qFD_\beta(Q)]^{1-q}}{1-1/N^{2(1-q)}}=1-U_{qN}^*(Q).$$

Moreover, we have

$$\lim_{q\to 1}\frac{[{}^qFD_\gamma(Q)]^{1-q}-[{}^qFD_\alpha(Q)]^{1-q}}{(1-1/N^{2(1-q)})[{}^qFD_\gamma(Q)]^{1-q}}=\frac{\log{}^1FD_\gamma(Q)-\log{}^1FD_\alpha(Q)}{2\log N}=1-U_{1N}^*(Q)=1-C_{1N}^*(Q).$$

(3) *A class of measures of functional distance-heterogeneity*: $1-S_{qN}^*(Q)$

As we proved in Appendix S2, $[{}^qFD_\gamma(Q)]/N^2\le {}^qFD_\alpha(Q)\le {}^qFD_\gamma(Q)$. We then have

$$0\le {}^qFD_\gamma(Q)]-{}^qFD_\alpha(Q)\le(1-1/N^2)[{}^qFD_\gamma(Q)],$$

which shows the functional diversity excess depends on functional gamma diversity and *N*. The normalized measure turns out to be $1-S_{qN}^*(Q)$:

$$[{}^qFD_\gamma(Q)-{}^qFD_\alpha(Q)]/\{(1-1/N^2)[{}^qFD_\gamma(Q)]\}=[1-1/{}^qFD_\beta(Q)]/(1-1/N^2)=1-S_{qN}^*(Q).$$

(4) *A class of functional distance-turnover rate*: $1-V_{qN}^*(Q)$

As proved in Appendix S2, we have ${}^qFD_\alpha(Q)\le {}^qFD_\gamma(Q)\le N^2[{}^qFD_\alpha(Q)]$. This implies

$$0\le{}^qFD_\gamma(Q)-{}^qFD_\alpha(Q)\le(N^2-1)[{}^qFD_\alpha(Q)],$$

which shows the functional diversity excess depends on the functional alpha diversity and *N*. The dependence can be removed in this case by normalization and the resulting measure is
$1-V_{qN}^*(Q)$:

$$[^qFD_\gamma(Q) - \; ^qFD_\alpha(Q)]/\{(N^2-1)[^qFD_\alpha(Q)]\} = [^qFD_\beta(Q) - 1]/(N^2-1) = 1 - V^*_{qN}(Q).$$

Thus, the functional diversity excess leads to the same four classes of normalized similarity measures ($C^*_{qN}(Q)$, $U^*_{qN}(Q)$, $V^*_{qN}(Q)$ and $S^*_{qN}(Q)$) derived from the functional beta diversity. Although normalization is not a general cure for problems of dependence on alpha or gamma functional diversity, all the resulting normalized measures discussed above are in terms of functional beta diversity only. Thus, dependence on the alpha (or gamma) functional diversity can be removed in our cases as our functional beta diversity is independent of the functional alpha diversity and also independent of the functional gamma diversity, based on a similar argument in [1].

References

1. Chao A, Chiu C-H, Hsieh TC (2012) Proposing a resolution to debates on diversity partitioning. Ecology 93: 2037–2051.
2. Chiu C-H, Jost L, Chao A (2014) Phylogenetic beta diversity, similarity, and differentiation measures based on Hill numbers. Ecol Monogr 84: 21–44.

Supporting Information

Distance-based functional diversity measures and their decomposition: a framework based on Hill numbers

Chun-Huo Chiu and Anne Chao
Institute of Statistics, National Tsing Hua University, Hsin-Chu, Taiwan 30043

Appendix S5: Supplementary examples and comparisons

(All equation numbers refer to those in the main text.)

Example: A simple functional distance matrix to show the counter-intuitive behavior of the traditional differentiation measure based on an additive partitioning of the quadratic entropy

Consider two hypothetical completely distinct assemblages (i.e. no shared species) each with S equally common species. Assume that the functional distance is a constant d for distinct-species pairs and 0 for same-species pairs. We first apply Rao's quadratic entropy to this simple distance matrix. The quadratic entropy Q of each assemblage is $d(1-1/S)$. This would be the alpha quadratic entropy Q_α of the assemblages. For any fixed d, the alpha value will be high (i.e., tends to the maximum possible value of d) if and only if species richness S is high. In the pooled assemblage, there are $2S$ equally common species with a constant functional distance d for distinct-species pairs and 0 for same-species pairs, so the gamma quadratic entropy of the pooled assemblage is $Q_\gamma = d[1-1/(2S)]$.

The traditional differentiation measure based on the additive partitioning is $Q_\beta^* = (Q_\gamma - Q_\alpha)/Q_\gamma$; see Eq. 2c. Thus, the differentiation measure is $[d/(2S)]/[d - d/(2S)] = 1/(2S-1)$, a number that approaches zero (wrongly indicating that there is almost no differentiation) when alpha is high (equivalently, when S is high), even though the two assemblages are completely distinct. Chiu et al. [1] proved that when the alpha quadratic entropy is high, the additive differentiation measure based on the quadratic entropy always tends to zero for any assemblages, not only for simple completely distinct assemblages, but also for more complicated assemblages and real data; see Example 3 of the main text.

The resolution in Eq. 2d of the main text works because this simple distance matrix is ultrametric. The effective number of species for the alpha diversity is $1/(1-Q_\alpha/d) = S$ and the effective number of species for the gamma diversity is $1/(1-Q_\gamma/d) = 2S$. Thus the multiplicative beta based on these effective measures is 2 and the two transformed differentiation measures in Eqs. 2e and 2f are both unity, which correctly indicates the differentiation attains the maximum. Our proposed normalized functional differentiation measures $1-C_{qN}^*(Q)$ and $1-U_{qN}^*(Q)$ (Table 3 of the main text) are always 1 for any q, any richness S, and any functional distance d. For this example, the correct answer is unequivocal: the two completely distinct assemblages in this simple case should attain the maximum differentiation of unity. This example shows that the traditional

measures based on the additive partitioning of the quadratic entropy cannot work properly even for this simple example.

Two supplementary cases for Example 1 of the main text

In Example 1 of the main text, we specifically consider the special case that all species in the two assemblages are equally abundant for illustrative purposes. Here we append two more cases that species abundances are heterogeneous in both assemblages.

Case 1: Shared species are abundant species

In Assemblage I, species abundances are 0.01 (for Species 1–Species 5), 0.02 (for Species 6–Species 10), and 0.085 (for Species 11–Species 20);
In Assemblage II, species abundances are 0.05 (for Species 9, 10, 11–18), 0.075 (for Species 19, 20, 21–23), and 0.025 (for Species 24 –Species 28);
There are 12 shared species (Species 9, 10, 11– 20), and a total of 28 species in the pooled assemblage.

Table S5.1. Comparison of various differentiation measures for Matrix I (with $Q_\gamma = 0.463$, $Q_\alpha = 0.462$) and Matrix II (with $Q_\gamma = 0.118$, $Q_\alpha = 0.102$) based on abundance and function (*A&F*), on function (*F*) only, and abundance (*A*) only.

Measure	Order	Matrix I			Matrix II		
		A&F	*F*	*A*[#]	*A&F*	*F*	*A*[#]
$1 - C_{qN}^*(Q)$	$q = 0$	0.324	0.324	0.4	0.579	0.579	0.4
	$q = 1$	0.271	---	0.267	0.365	---	0.267
	$q = 2$	0.322	---	0.256	0.342	---	0.256
$1 - U_{qN}^*(Q)$	$q = 0$	0.658	0.658	0.571	0.846	0.846	0.571
	$q = 1$	0.271	---	0.267	0.365	---	0.267
	$q = 2$	0.106	---	0.147	0.115	---	0.147
$Q_\beta^* = \dfrac{Q_\gamma - Q_\alpha}{Q_\gamma}$	$q = 2$	0.003			0.134		
$Q_{e,\beta}^* = \dfrac{1 - 1/Q_{e,\beta}}{1 - 1/N}$	$q = 2$	0.006			0.035		
$Q_{e,\beta}^{**} = \dfrac{Q_{e,\beta} - 1}{N - 1}$	$q = 2$	0.003			0.018		

[#] Differentiation measures are the abundance-based local differentiation measure ($1 - C_{qN}$) and regional differentiation measure ($1 - U_{qN}$) obtained from partitioning Hill numbers [1];
--- No measures for $q = 1$ and $q = 2$ because species abundances are not considered for measures based on function (*F*) only.

Case 2: Shared species are rare species

In Assemblage I, species abundances are 0.085 (for Species 1–Species 10), 0.01 (for Species 11–Species 15), and 0.002 (for Species 16–Species 20);
In Assemblage II, species abundances are 0.025 (for Species 9, 10, 11– 13), 0.05 (for Species 14, 15, 16 –23), and 0.075 (for Species 24 –species 28);
There are 12 shared species (Species 9, 10, 11– 20), and a total of 28 species in the pooled assemblage.

Table S5.2. Comparison of various differentiation measures for Matrix I (with $Q_\gamma = 0.480$, $Q_\alpha = 0.475$) and Matrix II (with $Q_\gamma = 0.244$, $Q_\alpha = 0.098$) based on abundance and function ($A\&F$), on function (F) only, and abundance (A) only.

Measure	Order	Matrix I			Matrix II		
		$A\&F$	F	$A^{\#}$	$A\&F$	F	$A^{\#}$
$1-C_{qN}^{*}(Q)$	$q = 0^{\#}$	0.324	0.324	0.4	0.579	0.579	0.4
	$q = 1$	0.681	---	0.680	0.864	---	0.680
	$q = 2$	0.879	---	0.832	0.957	---	0.832
$1-U_{qN}^{*}(Q)$	$q = 0$	0.658	0.658	0.571	0.846	0.846	0.571
	$q = 1$	0.681	---	0.680	0.864	---	0.680
	$q = 2$	0.646	---	0.712	0.848	---	0.712
$Q_{\beta}^{*} = 1 - Q_\alpha / Q_\gamma$	$q = 2$	0.010			0.598		
$Q_{e,\beta}^{*} = \dfrac{1 - 1/Q_{e,\beta}}{1 - 1/N}$	$q = 2$	0.022			0.325		
$Q_{e,\beta}^{**} = \dfrac{Q_{e,\beta} - 1}{N - 1}$	$q = 2$	0.011			0.194		

[#] Differentiation measures are the abundance-based local differentiation measure ($1-C_{qN}$) and regional differentiation measure ($1-U_{qN}$) obtained from partitioning Hill numbers [1];
--- No measures for $q = 1$ and $q = 2$ because species abundances are not considered for measures based on function (F) only.

A supplementary case for Example 2 of the main text

In Example 2 of the main text, we consider the special case that the two focal assemblages are completely distinct (no species shared, and thus no pairwise distances shared). Here we present similar results for a case that there are shared species between the two assemblages.

Table S5.3. Comparison of various differentiation measures between two assemblages for an ultramteric distance matrix (Case I below) and a non-ultrametric distance matrix (Case II below). Assume the two assemblages are not completely distinct. There are three equally common species (*a*, *b, c*) in the first assemblage, and three equally common species (*b*, *c*, *d*) in the second assemblage. In the pooled assemblage, there are four species (*a*, *b*, *c*, *d*) with relative abundances (0.167, 0.333, 0.333, 0.167). As explained in the text, we expect that the differentiation for Case II should not be lower than that for Case I.

Case I: An ultrametric distance matrix for four species (*a*, *b*, *c*, *d*) with $Q_\gamma = 0.122$, $Q_\alpha = 0.111$.

$$\begin{bmatrix} 0 & 0.1 & 0.2 & 0.2 \\ 0.1 & 0 & 0.2 & 0.2 \\ 0.2 & 0.2 & 0 & 0.1 \\ 0.2 & 0.2 & 0.1 & 0 \end{bmatrix}$$

Case II: A non-ultrametric distance matrix for four species (*a*, *b*, *c*, *d*) with $Q_\gamma = 0.161$, $Q_\alpha = 0.111$.

$$\begin{bmatrix} 0 & 0.1 & 0.2 & 0.9 \\ 0.1 & 0 & 0.2 & 0.2 \\ 0.2 & 0.2 & 0 & 0.1 \\ 0.9 & 0.2 & 0.1 & 0 \end{bmatrix}$$

Measure	Order q	Ultrametric distance matrix (Case I)	Non-ultrametric distance matrix (Case II)
$1 - C^*_{qN}(Q)$	$q = 0$	0.273	0.448
	$q = 1$	0.364	0.517
	$q = 2$	0.455	0.586
$1 - U^*_{qN}(Q)$	$q = 0$	0.600	0.765
	$q = 1$	0.364	0.517
	$q = 2$	0.172	0.262
$Q^*_\beta = \dfrac{Q_\gamma - Q_\alpha}{Q_\gamma}$	$q = 2$	0.091	0.310
$Q^*_{e,\beta} = \dfrac{1 - 1/Q_{e,\beta}}{1 - 1/N}$	$q = 2$	0.250	0.127
$Q^{**}_{e,\beta} = \dfrac{Q_{e,\beta} - 1}{N - 1}$	$q = 2$	0.143	0.068

Comparison of our framework with Leinster & Cobbold (2012) approach

Leinster & Cobbold [2] derived a parametric class of measures sensitive to species similarity based on a framework of Hill numbers. The similarity may be based on phylogeny, ecosystem function, or any other species character. We find that their measure (referred to as the LC measure hereafter) may not be sensitive to species abundances when species similarity matrix is computed from species traits in functional analysis. When species similarity matrix deviates greatly from a naïve identity matrix, their measure typically yields very low diversity values especially for assemblages with many species; this causes problems for the interpretation of "species equivalents" in their approach. Note that in the bottom right panel of Fig. 3 of [2], as q varies between 0 and 5, their measure for a non-naive similarity matrix decreases from 1.27 to 1.25 for Case "TS1" with ~250 species and decreases from 1.25 to 1.22 for Case "TS3" with ~ 200 species, as shown in the bottom left panel of their Fig. 3. This reveals that the LC measure hardly varies with the order q for the two cases considered in their Fig. 3. We thus computed several other real examples to see whether the LC measure generally exhibits a similar pattern. We describe two typical examples to show our findings.

We applied the LC measure to the artificial data (Example 2 of the main text) and the real data (Example 3 of the main text), so that readers can make comparisons. In Example 2, there are 20 equally abundant species in each of the two focal assemblages, and 12 species are shared. Two simulated distance matrices with all distances between 0 and 1 (Matrix I and Matrix II, displayed in Appendix S6) are considered; see the main text. In Example 3, the full data contain a total of 43 vascular plant species collected three fore-dune habitats: embryo dunes (EM; 17 species), mobile dunes (MO; 39 species) and transition dunes (TR; 42 species). The species relative abundances are provided in Table S5.1 of this appendix). The distance matrix for 43 species is displayed in Appendix S6. All species distances (between 0 and 1) are obtained from the Gower mixed-variables coefficient of distance. Since the LC measure is based on species similarity matrix, we considered two types of similarity metrics computed from the distance matrices: (i) the one-complement of each distance; (ii) the transformation exp(–d) of each distance d. The plot of the LC measure with respect to the order q for the two types of similarity metrics is given in the following figure for the two examples.

(a) The Leinster & Cobbold (LC) measure for two matrices based on artificial data in Example 2 of the main text
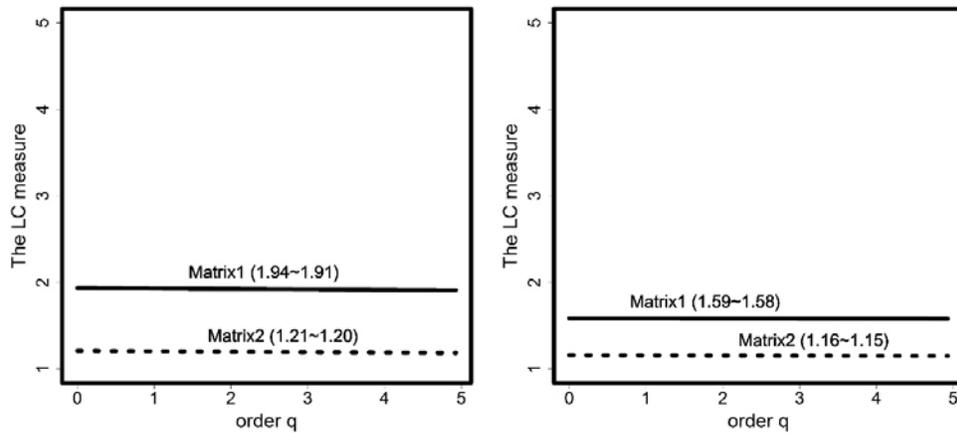
(b) The Leinster & Cobbold (LC) measure for three habitats (EM, MO and TR) based on real data in Example 3 of the main text
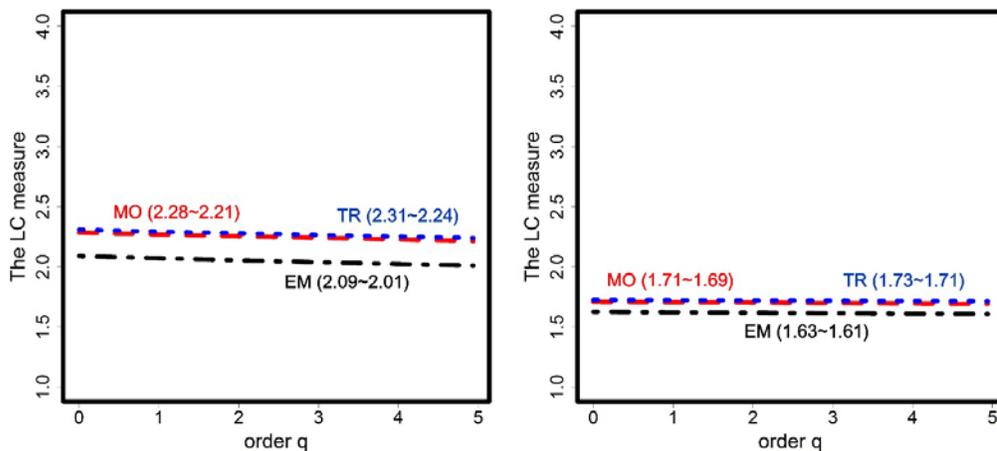

Figure S5.1. Diversity profiles as a function of the order q ($0 \leq q \leq 5$) of the Leinster and Cobbold (LC) measure for two examples. In the left panels, the similarity is defined as the one-complement of a distance. In the right panels, similarity is defined as $\exp(-d)$ of a distance d. The "range" in the plots and in the following captions refers to the range of the LC measure when q is increased from 0 to 5.

(Upper left panel) Matrix I (black line, in range: 1.94~ 1.91) and Matrix II (dotted line, in range: 1.21~1.20).

(Upper right panel) Matrix I (black line, in range: 1.59~ 1.58) and Matrix II (dotted line, in range: 1.16~1.15).

(Lower left panel) EM Habitat (black line, in range: 2.09~ 2.01), MO Habitat (red line, in range: 2.28~ 2.21) and TR Habitat (blue line, in range: 2.31~ 2.24).

(Lower right panel) EM Habitat (black line, in range: 1.63~ 1.61), MO Habitat (red line, in range: 1.71~ 1.69) and TR Habitat (blue line, in range: 1.73~ 1.71).

The above plots reveal that the LC measure takes values in a very narrow range and this range hardly changes for the two different types of similarity matrices. Since the order q controls the measure's sensitivity to species relative abundances and a larger value of q place progressively more weight on common species, these plots in Figure S5.1 demonstrate that the LC measure may not be sensitive to the species abundances. We have found similar patterns for many other data sets. A related question is thus how to interpret the magnitude and "effective numbers" of the LC measure.

Recently, Reeve et al. [3] proposed formulas for the alpha, beta and gamma diversities based on the LC measure and on the decomposition framework of ordinary Hill numbers. Here we consider the simplest equal weight case and apply Reeve et al. formulas to a simple similarity matrix. Consider two communities each with four species (1, 2, 3, 4). The relative abundances of the four species in community I are (0.98, 0.003, 0.001, 0.016). The relative abundances of the same four species for community II are (0.90, 0.009, 0.082, 0.009). Assume that the pairwise similarity for the four species is given by this matrix:

$$\begin{bmatrix} 1 & 0.9 & 0.9 & 0 \\ 0.9 & 1 & 0 & 0 \\ 0.9 & 0 & 1 & 0.9 \\ 0 & 0 & 0.9 & 1 \end{bmatrix}$$

Then we obtain the following alpha and gamma diversity for four values of q:

Order	Gamma	Alpha
$q = 0$	1.2605	1.5375
$q = 0.5$	1.1120	1.1570
$q = 1$	1.0611	1.0679
$q = 2$	1.0343	1.0345

For this case, the gamma LC measure is less than the alpha LC measure not only for the four specific values of q in the above table but also for all values of $q \geq 0$. This situation violates the necessary condition that alpha must always be less than or equal to gamma.

Leinster & Cobbold ([2], p. 478) indicated that their metric has close connections with the phylogenetic indices of Faith [4] and Chao et al. [5]. This may be a misleading statement. In their Appendix, Leinster & Cobbold demonstrated that their formula could include Faith's *PD* and Chao et al. [5] phylogenetic Hill number only for a *particular* constructed similarity matrix (possibly non-symmetric) and a special set of species abundances. Note that their particular similarity matrix for *species* depends on species relative abundance. Thus, when two communities have the same set of species with different sets of species abundances, the corresponding particular similarity matrices are then different. Even within a single community, if two samples result in different

species abundances, then LC's particular similarity matrices are different. Thus, the "connection" between their metric and Chao et al. measure is only based on an uninterpretable similarity matrix. A useful "connection" between two measures should be based on any *given* matrix in a broad class, not just for a single particular constructed matrix.

For any given ultrametric tree, we can divide each species pairwise phylogenetic distance by the tree depth so that all distances are scaled to be in the range [0, 1]. When the similarity between any two species is defined as the one-complement of the scaled distance, the LC measure for $q = 2$ reduces to the Chao et al.'s phylogenetic Hill number of the same order. This is the only general connection that we have found between the LC measure and Chao et al. [5] phylogenetic Hill numbers if the similarity matrix is not a naive identity matrix.

Comparison of our framework with Scheiner (2012) approach

Scheiner [6] proposed a metric that integrates abundance, phylogeny and function based on a framework of Hill numbers. Our framework (Chao et al. [5] for phylogenetic diversity, and this paper for functional diversity) is also based on Hill numbers. However, the two approaches are completely different. In this section, we describe our fundamental concept and discuss the differences between our framework and Scheiner's approach.

The basic difference

The major difference lies in the interpretation of Hill numbers. In Scheiner's approach, the ordinary Hill numbers are interpreted as the *variability* in relative abundances among species. Based on this approach, Scheiner's phylogenetic diversity quantifies the variability of proportional phylogenetic divergences of species, and his functional diversity quantifies the variability of proportional functional distinctiveness.

Our interpretation of Hill numbers is different. The fundamental concept in our approach is based on the fact that there is a *unique* idealized assemblage with equally abundant species so that the actual assemblage and this idealized assemblage have the same diversity of order *q*. Thus, our extension to phylogenetic diversity and functional diversity leads to completely different measures as briefly described as follows.

(i) Phylogenetic diversity measures

Our phylogenetic Hill number (or mean phylogenetic diversity) of order *q*, denoted by $^{q}\overline{D}(T)$, is the effective number of equally abundant and equally phylogenetically distinct species with a constant branch length $\overline{T}$ from the root node. Here $\overline{T}$ denotes the abundance-weighted mean distance from a tip node to the root node; see Fig. 1 of Chao et al. [5] for the definition of $\overline{T}$. For an ultrametric tree with tree length *T*, then $\overline{T}$ reduces to the tree length *T*, and the measure is simply denoted by $^{q}\overline{D}(T)$. Generally, if $^{q}\overline{D}(T) = z$, then the phylogenetic Hill number of the assemblage is the same as the diversity of an idealized assemblage consisting of *z* equally abundant and (phylogenetically) equally distinct lineages all with branch length $\overline{T}$ from the root node. The basic concept is that there exists a *unique* idealized assemblage with equally abundant and equally

distinct species so that the actual assemblage and this idealized assemblage have the same diversity of order q.

The phylogenetic Hill number (in units of "species equivalent") does not incorporate information about the actual length of the phylogenetic tree because it is independent of the scale of lineage lengths. To incorporate the units of lineage length, we also proposed the phylogenetic diversity $^qPD(\bar{T})$ (in units of "lineage length"), $^qPD(\bar{T}) = \bar{T} \times [^q\bar{D}(\bar{T})]$. Thus, we not only have a measure in units of "species equivalents" but also a measure in units of "lineage length". This is more useful biologically since it expresses the amount of evolutionary history in the tree (with branches weighted by the size of their contribution to the present-day assemblage), and is also fruitful mathematically because we then can link our measures to Faith PD (for $q = 0$), phylogenetic entropy [7] (for $q = 1$), and Rao's quadratic entropy (for $q = 2$). Scheiner's measure cannot be linked to the phylogenetic entropy, nor to Rao's quadratic entropy.

(ii) Functional diversity measures

Our functional Hill number denoted by $^qD(Q)$ (see Table 1 of the main text) is interpreted as "the effective number of equally abundant and (functionally) equally distinct species". Thus if $^qD(Q) = v$, then the functional Hill number of order q of the actual assemblage is the same as that of an idealized assemblage having v equally abundant and equally distinct species with a constant distance Q for all species pairs. Our concept for functional diversity is based on the fact that there exists a *unique* idealized assemblage with equally abundant and equally distinct species so that the actual assemblage and this idealized assemblage have the same diversity of order q.
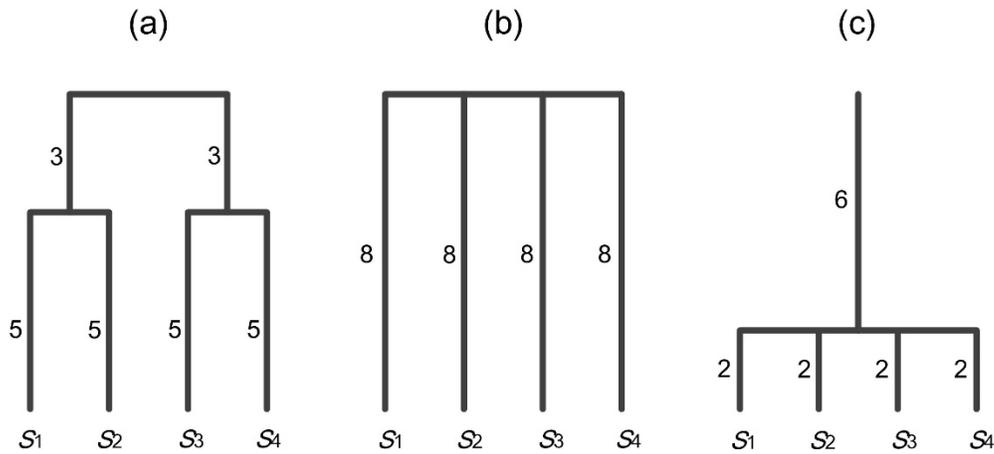
As with our phylogenetic Hill numbers, the functional Hill numbers $^qD(Q)$ (in units of "species equivalent") are scale-free, so they need to be converted to our functional diversity $^qFD(Q)$ (the effective total functional distance between species), defined as $^qFD(Q) = Q \times [^qD(Q)]^2$. Thus, we can link our measures to FAD (for $q = 0$), and to the weighted Gini-Simpson index (for $q = 2$) defined by Guiasu & Guiasu [8,9]; see the main text for details. To our knowledge, Scheiner's metric cannot be linked to these two previous measures.

Different meanings of "species equivalents"

Scheiner's integrated metric and our phylogenetic (and functional) Hill number are both in units of "effective number of species" or "species equivalents", and the "species equivalent" in both approaches is interpreted as the equally abundant and equally distinct species. However, the definition of "equally distinct" diverges between Scheiner's approach and ours. We use a simple example to illustrate the two different meanings.

(i) Phylogenetic diversity measures

Consider the following three assemblages with ultrametric cladograms. Each assemblage includes four species and the tree depth is $T = 8$ units. The number along each branch segment denotes the length of that branch. For each assemblage, we assume that all four species are equally abundant.

<center>(a)   (b)   (c)</center>

For the above three cladograms with $T = 8$ units, we give our phylogenetic Hill numbers $^{q}\overline{D}(T)$ (in units of "species equivalents"), phylogenetic diversity $^{q}PD(T)$ (in units of "lineage length") and Scheiner's phylogenetic diversity in the following table.

Measure	Order	Assemblage/Cladogram		
		(a)	(b)	(c)
Chao et al. (2010)	$q = 0$	26	32	14
phylogenetic	$q = 1$	24.78	32	11.31
diversity $^{q}PD(T)$	$q = 2$	23.27	32	9.85
Chao et al. (2010)	$q = 0$	3.25	4	1.75
phylogenetic Hill	$q = 1$	3.08	4	1.41
number $^{q}\overline{D}(T)$	$q = 2$	2.91	4	1.23
Scheiner (2012)	$q = 0$	4	4	4
phylogenetic	$q = 1$	4	4	4
diversity	$q = 2$	4	4	4

All the above three cladograms have the same proportional divergences as defined by Scheiner. For any $q \geq 0$, his phylogenetic diversity which quantifies the variability of proportional phylogenetic divergences of species thus yields four equally distinct species for all three assemblages. When his measure takes a maximum value of four, the assemblage may correspond to the four equally abundant species in cladograms (a), (b), (c) or any other symmetric or balanced cladograms. This explains why Scheiner ([6], p. 1195) indicated that his metric is a measure of tree "symmetry" or "balance". His measure cannot distinguish the difference among the three assemblages, and thus "species equivalents" does not have a unique reference assemblage.

Our phylogenetic measures $^{q}\overline{D}(\overline{T})$ and $^{q}PD(\overline{T})$ both satisfy the "weak monotonicity" property [5]. This property requires that if a newly added rarest species is maximally distinct from all other

species in the assemblage, then a phylogenetic measure should not decrease. However, Scheiner's phylogenetic diversity measure does not satisfy this property. Note that for $q > 0$, if such a species is added to the assemblage (a) or (c) in the above figure, that tree becomes non-symmetric, implying a possible decline in a measure of symmetry.

Note that in Scheiner's measure, "equally distinct" means species are equally divergent from the age of the root node. Our definition of "equally distinct" implies that *any two species must have a constant phylogenetic distance of* $\overline{T}$ (*or T* in an ultrametric tree), or equivalently, *all branch lengths must be equal to* $\overline{T}$ (*or T*) as cladogram (b) given above. For cladogram (a), the distance between Species 1 and Species 2 is 5 units whereas the distance between Species 1 and Species 3 is 8 units, so the species in cladogram (a) are not "equally distinct" in our perspective. Similarly, the four species in cladogram (c) are not "equally distinct" with branch lengths of 8 units either. Only cladogram (b) is the unique idealized assemblages with all species being "equally distinct" with all branch lengths of 8 units. Thus, for the "effective number of species" in our phylogenetic diversity measure, there exists a unique reference assemblage so that the actual assemblage and this idealized assemblage have the same diversity of order q. For example, in the special case of $q = 0$, the phylogenetic Hill number of cladogram (c) is 1.75. Then this means the zero-order diversity of the assemblage is the same as an idealized assemblage with 1.75 equally abundant species with a constant branch length of 8 units, i.e., the cladogram of the idealized reference assemblage is like cladogram (b) but only with 1.75 species.

The above table reveals that when diversity is based on our phylogenetic Hill number $^q\overline{D}(T)$ and phylogenetic diversity $^qPD(T)$, the three assemblages for any q have consistent ordering: (b) > (a) > (c), whereas Scheiner's measure shows (a) = (b) = (c). Ecologists may use this example to choose the measure to be used in their analysis.

(ii) Functional diversity measures

We use a simple example to compare the difference between our functional diversity measures and Scheiner's approach. Consider the following example: In Assemblage A, all species are equally distinct with species pairwise distance $d_{ij} = 0.1$ units; In Assemblage B, all species are equally distinct with $d_{ij} = 0.9$ units. Scheiner's functional diversity quantifying the variability of functional distinctiveness will give the same functional diversity for these two assemblages. Yet, from our approach, there are S species with a constant distance of 0.1 for all species pairs in Assemblage A, and our functional diversity (i.e., effective total distance) between species is S^2 x 0.1. For Assemblage B, there are S species with a constant distance of 0.9 for all species pairs, and the functional diversity is S^2 x 0.9. The effective numbers of species are the same for the two assemblages, but the total distance between species for the two assemblages are different. Thus, Scheiner's measure loses the information about the magnitude of species pairwise distances, which we think is important to characterize distance-based traits diversity.

Table S5.4. Species identities and relative abundances in three dune habitats: upper beach and embryo dunes (EM, 17 species), mobile dunes (MO, 39 species) and transition dunes (TR, 42 species). A relative abundance of 0 for a species in a habitat means that the species does not exist in that habitat. See the main text for data description and details.

Species	EM	MO	TR
Ammophila arenaria	0	0.118	0.006
Anthemis maritima	0.024	0.132	0.046
Asparagus acutifolius	0	0.003	0.003
Bromus diandrus	0	0.005	0.032
Cakile maritima	0.217	0.024	0.004
Calystegia soldanella	0.027	0.026	0.009
Centaurea sphaerocephala	0	0.008	0.015
Chamaesyce peplis	0.097	0.014	0.001
Clematis flammula	0	0.004	0.018
Crucianella maritima	0	0.022	0.080
Cutandia maritima	0.008	0.036	0.095
Cyperus kalli	0.003	0.047	0.037
Daphne gnidium	0	0	0.001
Echinophora spinosa	0.029	0.029	0.004
Elymus farctus	0.161	0.134	0.044
Eryngium maritimum	0.021	0.020	0.001
Euphorbia terracina	0	0.003	0.028
Helicrisum stoechas	0	0.004	0.029
Juniperus oxycedrus	0	0	0.008
Lagurus ovatus	0	0.003	0.022
Lonicera implexa	0	0.001	0.002
Lophocloa pubescens	0	0.005	0.006
Lotus cytisoides	0	0.005	0.062
Medicago littoralis	0	0.021	0.081
Medicago marina	0.003	0.047	0.012
Ononis variegata	0.005	0.070	0.025
Otanthus maritimus	0.043	0.005	0
Pancratium maritimum	0.016	0.042	0.039
Phillirea angustifolia	0	0.002	0.005
Pistacia lentiscus	0	0.001	0.008
Plantago coronopus	0.003	0.013	0.017
Polygonum maritimum	0.038	0.003	0.001
Prasium majus	0	0.001	0.003
Pseudorlaya pumila	0	0.007	0.008
Pycnocomon rutifolium	0	0.003	0.042
Quercus ilex	0	0	0.003
Rubia peregrina	0	0	0.003
Salsola kali	0.193	0.025	0.002
Silene canescens	0	0.058	0.089
Smilax aspera	0	0.003	0.009
Sonchus bulbosus	0	0.006	0.005
Sporobolus virginicus	0.113	0.044	0.021
Vulpia fasciculata	0	0.009	0.073

References

1. Chiu C-H, Jost L, Chao A (2014) Phylogenetic beta diversity, similarity, and differentiation measures based on Hill numbers. Ecol Monogr 84: 21–44.
2. Leinster T, Cobbold CA (2012) Measuring diversity: the importance of species similarity. Ecology 93: 477–489.
3. Reeve R, Matthews L, Cobbold CA, Leinster T, Thompson J, Brummitt N. (2014) How to partition diversity. Cornell University Library, arXiv:1404.6520, assessed on May 28, 2014
4. Faith DP (1992) Conservation evaluation and phylogenetic diversity. Biol Conserv 61: 1–10.
5. Chao A, Chiu C-H, Jost L (2010) Phylogenetic diversity measures based on Hill numbers. Phil Trans R Soc B 365: 3599–3609.
6. Scheiner SM (2012) A metric of biodiversity that integrates abundance, phylogeny, and function. Oikos 121: 1191–1202.
7. Allen B, Kon M, Bar-Yam Y (2009) A new phylogenetic diversity measure generalizing the Shannon index and its application to phyllostomid bats. Am Nat 174: 236–243.
8. Guiasu RC, Guiasu S (2011) The weighted quadratic index of biodiversity for pairs of species: a generalization of Rao's index. Nat Sci 3:795–801.
9. Guiasu RC, Guiasu S (2012) The weighted Gini-Simpson index: revitalizing an old index of biodiversity. International J Ecol 2012, Article ID 478728, 10 pages.