

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 $i, j = 1, 2, \dots, S$. 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 $k, m = 1, 2, \dots, N$. 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 $[d_{ij}]$, $i, j = 1, 2, \dots, S$. Let $p_{ik} \geq 0$ be the relative abundance of i th species in the k th assemblage, $i = 1, 2, \dots, S, k = 1, 2, \dots, N$. The assumption that the functional Hill number of order q for any pair of assemblages is a constant qD means

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

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

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

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

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

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

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

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

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} , $k, m = 1, 2, \dots, N$, 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} , $k, m = 1, 2, \dots, N$, 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 $k, m = 1, 2, \dots, N$ are identical, and the mean distances Q_{km} for $k, m = 1, 2, \dots, N$ 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 $k, m = 1, 2, \dots, N$ are identical, and the mean distances Q_{km} for $k, m = 1, 2, \dots, N$ 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 $\{z_{i+}/z_{++}; i = 1, 2, \dots, S\}$. Thus, it follows from Eq. 3 of the main text that the functional gamma Hill number of order q is:

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

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 $i = 1, 2, \dots, S, k = 1, 2, \dots, N$, let z_{ik} denote the species abundance of the i th species in the k th 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 $S \times N$ species abundance matrix \mathbf{Z} as

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

Let the k th column of \mathbf{Z} be denoted by $\mathbf{z}_k = (z_{1k}, z_{2k}, \dots, z_{Sk})^T$, (the super-script “ T ” denotes the “transpose” of a matrix) and this column denotes the species abundances of the k th assemblage. So the matrix \mathbf{Z} can be denoted by $(\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_N)$. The matrix total is $z_{++} = \sum_{i=1}^S \sum_{j=1}^N z_{ij}$. Let d_{ij} denote the functional distance between the i th and the j th species, $d_{ij} \geq 0$, and $d_{ij} = d_{ji}$. Denote the $S \times S$ 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 k th idealized assemblage, the abundance set can be expressed as an $A \times 1$ column vector $\mathbf{b}_k = (\bar{a}, \bar{a}, \dots, \bar{a})^T$, $k = 1, 2, \dots, N$. 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 $\mathbf{B} = (\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_N)$. A good feature of the idealized N -assemblage matrix \mathbf{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 \mathbf{Z} and \mathbf{B} are “equivalent” in the sense that any q -th power sum ($q \geq 0$) of the elements of the matrix $\mathbf{Z} = (\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_N)$ should give identical values when the same function is applied to the idealized reference matrix $\mathbf{B} = (\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_N)$ with $\mathbf{b}_j = (\bar{a}, \bar{a}, \dots, \bar{a})^T$ for all j . We consider the following two special functions:

- (1) The sum of all elements in \mathbf{Z} is thus equal to the corresponding sum in \mathbf{B} . This sum for $\mathbf{Z} = (\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_N)$ is z_{++} , whereas the sum for $\mathbf{B} = (\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_N)$ 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 \mathbf{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

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

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

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

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

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., $1 \leq {}^qD_\beta(Q) = {}^qMD_\beta(Q) \leq N$ for all $q \geq 0$. The functional beta diversity of order q is always in the range $[1, N^2]$, i.e., $1 \leq {}^qFD_\beta(Q) \leq N^2$ 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:

$${}^qD_\alpha(Q) \leq {}^qD_\gamma(Q) \leq N[{}^qD_\alpha(Q)], \quad (\text{B3})$$

or equivalently,

$$[{}^qD_\gamma(Q)]/N \leq {}^qD_\alpha(Q) \leq {}^qD_\gamma(Q). \quad (\text{B4})$$

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:

$${}^qFD_\alpha(Q) \leq {}^qFD_\gamma(Q) \leq N^2[{}^qFD_\alpha(Q)], \quad (\text{B5})$$

or equivalently,

$${}^qFD_\gamma(Q) / N^2 \leq {}^qFD_\alpha(Q) \leq {}^qFD_\gamma(Q). \quad (\text{B6})$$

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

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

and

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

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

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

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

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

For $q = 1$, note that we have

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

and

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

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 ${}^q D_\gamma(Q) \geq {}^q D_\alpha(Q)$, 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

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

yielding ${}^q D_\alpha(Q) \leq {}^q D_\gamma(Q)$ 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

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

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

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

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

Then we have

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

Equivalently,

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

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

Proposition S2.2 (A property of monotonicity used for Example 2 of the main text): Consider N assemblages indexed by $1, 2, \dots, N$. 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)

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

We can simplify the above formula as

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

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

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

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

$$\begin{aligned} & \left\{ (z_{a1} z_{b2})^q \left[\frac{1}{N^{2(1-q)}} \sum_{k,m=1}^N \sum_{i,j \neq (a,b)}^S d_{ij} (z_{ik} z_{jm})^q \right] - \frac{1}{N^{2(1-q)}} (z_{a1} z_{b2})^q \left[\sum_{i,j \neq (a,b)}^S d_{ij} (z_{i+} z_{j+})^q \right] \right\} / B^2 \\ & = \frac{1}{N^{2(1-q)}} (z_{a1} z_{b2})^q \left\{ \sum_{k,m=1}^N \sum_{i,j \neq (a,b)}^S d_{ij} (z_{ik} z_{jm})^q - \sum_{i,j \neq (a,b)}^S d_{ij} (z_{i+} z_{j+})^q \right\} / B^2 . \end{aligned} \quad (\text{B8})$$

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

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

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 ${}^qD_\beta(Q)$ ($={}^qMD_\beta(Q)$) always takes values between 1 and N (Proposition S2.1 in Appendix S2) and the functional beta diversity ${}^qFD_\beta(Q)$ 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 ${}^qD_\beta(Q)$ quantify *species-overlap* from different perspectives, whereas the similarity measures based on ${}^qFD_\beta(Q)$ quantify *distance-overlap* from different perspectives. In most applications, 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

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

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 ${}^0FD_\gamma(Q) = FAD_\gamma$ and ${}^0FD_\alpha(Q) = FAD_{pair} / N^2$, the measure $C_{0N}^*(Q)$ reduces to

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

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

the sum of $FADs$ 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

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

In this expression, the denominator is the alpha functional diversity of order 0, which is the average of $FADs$ 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

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

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 $h_{11} = \frac{z_{i1} z_{j1}}{z_{++} z_{++}}$ 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 $h_{12} = \frac{z_{i1} z_{j2}}{z_{++} z_{++}}$, $h_{21} = \frac{z_{i2} z_{j1}}{z_{++} z_{++}}$, $h_{22} = \frac{z_{i2} z_{j2}}{z_{++} z_{++}}$. Then the measure $C_{2N}^*(Q)$ for the special case of $N = 2$ reduces to

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

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

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

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

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

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

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

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:

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

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

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

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

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

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:

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

In the special case that all species are equally distinct, we have $FAD_\gamma = S^2 \times Q$ and $FAD_{pair} = (N\bar{S})^2 \times Q$, where S is the species richness in the pooled assemblage and \bar{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., $C_{0N} = \frac{(N\bar{S} - S)}{(N - 1)\bar{S}}$.

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 ${}^qD_\alpha(Q)$ equally abundant and functionally equally distinct species (all pairwise distances are identical), with exactly ${}^qD_\alpha(Q) \times p$ 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 ${}^qD_\gamma(Q)$. 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

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

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

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

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 ${}^qFD_\beta(Q)$ and ${}^qD_\beta(Q)$ 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

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

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 ${}^qD_\gamma(Q)$ equally abundant and equally distinct species, with exactly ${}^qD_\gamma(Q) \times u$ species shared by all N assemblages, and the remaining ${}^qD_\gamma(Q) \times (1-u)$ species evenly distributed in N assemblages. In each of the N idealized assemblages, the effective functional Hill number is ${}^qD_\alpha(Q)$. 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

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

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

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

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, $i, j = 1, 2, \dots, S$, and the indices k and m are used to identify assemblages, $k, m = 1, 2, \dots, N$.)

Order	Measure	Local species-overlap	Regional species-overlap
		$C_{qN}(Q) = \frac{N^{1-q} - [{}^q D_\beta(Q)]^{1-q}}{N^{1-q} - 1}$	$U_{qN}(Q) = \frac{[1/{}^q D_\beta(Q)]^{1-q} - (1/N)^{1-q}}{1 - (1/N)^{1-q}}$
$q = 0$	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 $1 - \frac{\log[{}^1 D_\gamma(Q)] - \log[{}^1 D_\alpha(Q)]}{\log N}$	
	Taxonomic	Classic Horn $\frac{H_{Sh,\alpha} - H_{Sh,\gamma} - \sum_{k=1}^N \frac{z_{+k}}{z_{++}} \log\left(\frac{z_{+k}}{z_{++}}\right)}{\log N}$; $1 - \frac{H_{Sh,\gamma} - H_{Sh,\alpha}}{\log N}$ (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 $1 - \frac{\sum_{i=1}^S \sum_{m>k} (z_{im} - z_{ik})^2}{(N-1)\sum_{i=1}^S \sum_{k=1}^N z_{ik}^2}$; $1 - \frac{H_{GS,\gamma} - H_{GS,\alpha}}{(1-1/N)(1-H_{GS,\alpha})}$ (if $z_{+k}=1, z_{++}=N$)	Classic regional-overlap $1 - \frac{\sum_{i=1}^S \sum_{m>k} (z_{im} - z_{ik})^2}{(N-1)\sum_{i=1}^S z_{i+}^2}$; $1 - \frac{H_{GS,\gamma} - H_{GS,\alpha}}{(N-1)(1-H_{GS,\gamma})}$ (if $z_{+k}=1, z_{++}=N$)

Notes:

- z_{ik} = the abundance of the i th species in the k th 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 $z_{+k}=1, z_{++}=N$.
- $q = 0$. FAD_γ = sum of the pairwise distances between species in the pooled assemblage; FAD_{pair} = sum of FAD s 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:

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

and

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

It follows from ${}^qFD_\gamma(Q) \leq N^2[{}^qFD_\alpha(Q)]$ (Proposition S2.1 of Appendix S2) that for $q > 1$, we have $[{}^qFD_\gamma(Q)]^{1-q} \geq N^{2(1-q)}[{}^qFD_\alpha(Q)]^{1-q}$. Then we obtain the inequality (D1). For $0 \leq q < 1$, 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 $0 \leq q < 1$) 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

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

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

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

and

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

Dividing $[{}^qFD_\gamma(Q)]^{1-q} - [{}^qFD_\alpha(Q)]^{1-q}$ (for $0 \leq q < 1$) 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

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

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

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

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

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 ${}^qFD_\alpha(Q) \leq {}^qFD_\gamma(Q) \leq N^2[{}^qFD_\alpha(Q)]$. This implies

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

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
$Q_\beta^* = \frac{Q_\gamma - Q_\alpha}{Q_\gamma}$	$q = 2$	0.003			0.134		
$Q_{e,\beta}^* = \frac{1 - 1/Q_{e,\beta}}{1 - 1/N}$	$q = 2$	0.006			0.035		
$Q_{e,\beta}^{**} = \frac{Q_{e,\beta} - 1}{N - 1}$	$q = 2$	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		
$Q_{e,\beta}^* = \frac{1 - 1/Q_{e,\beta}}{1 - 1/N}$	$q = 2$	0.022			0.325		
$Q_{e,\beta}^{**} = \frac{Q_{e,\beta} - 1}{N - 1}$	$q = 2$	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 ultrametric 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 $Q_\gamma = 0.161$, $Q_\alpha = 0.111$.

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

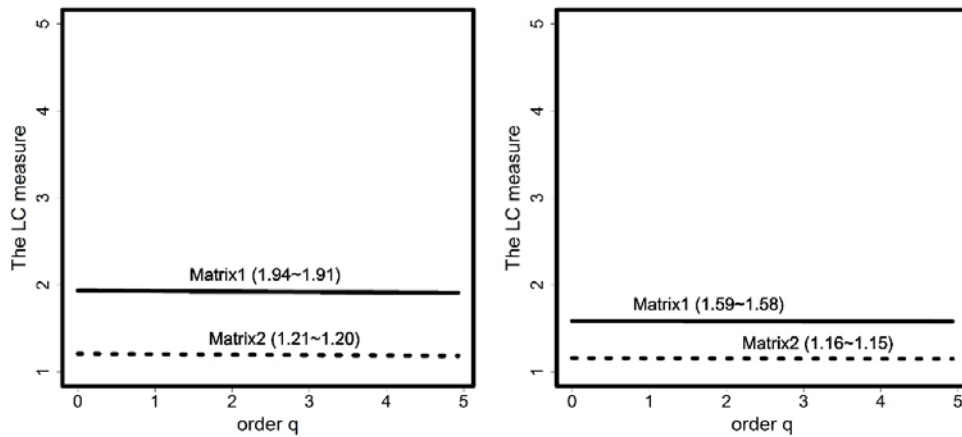
Measure	Order q	Ultrametric distance matrix (Case I)	Non-ultrametric distance matrix (Case II)
$1 - C_{qN}^*(Q)$	$q = 0$	0.273	0.448
	$q = 1$	0.364	0.517
	$q = 2$	0.455	0.586
$1 - U_{qN}^*(Q)$	$q = 0$	0.600	0.765
	$q = 1$	0.364	0.517
	$q = 2$	0.172	0.262
$Q_\beta^* = \frac{Q_\gamma - Q_\alpha}{Q_\gamma}$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

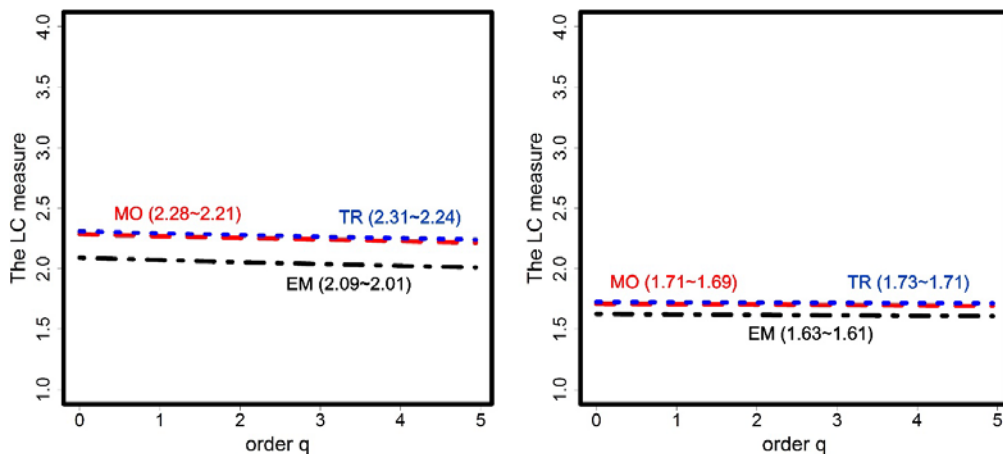


Figure S5.1. Diversity profiles as a function of the order q ($0 \leq q \leq 5$) of the Leinster and Cobbold (LC) measure for two examples. In the left panels, the similarity is defined as the one-complement of a distance. In the right panels, similarity is defined as $\exp(-d)$ of a distance d . The “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 ${}^q\bar{D}(T)$, is the effective number of equally abundant and equally phylogenetically distinct species with a constant branch length \bar{T} from the root node. Here \bar{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 \bar{T} . For an ultrametric tree with tree length T , then \bar{T} reduces to the tree length T , and the measure is simply denoted by ${}^q\bar{D}(T)$. Generally, if ${}^q\bar{D}(T) = z$, 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 \bar{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”), ${}^qPD(\bar{T}) = \bar{T} \times [{}^qD(\bar{T})]$. 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 ${}^qD(Q) = \nu$, then the functional Hill number of order q of the actual assemblage is the same as that of an idealized assemblage having ν 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 .

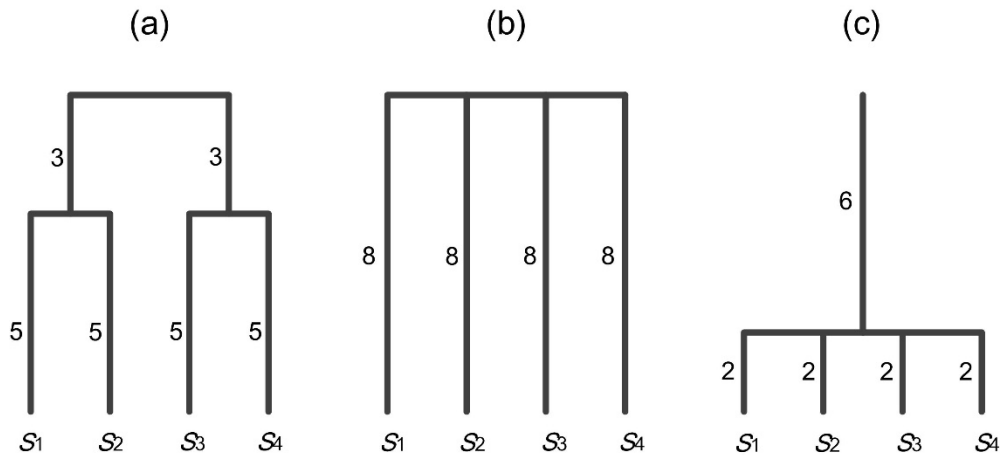
As with our phylogenetic Hill numbers, the functional Hill numbers ${}^qD(Q)$ (in units of “species equivalent”) are scale-free, so they need to be converted to our functional diversity ${}^qFD(Q)$ (the effective total functional distance between species), defined as ${}^qFD(Q) = Q \times [{}^qD(Q)]^2$. 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.



For the above three cladograms with $T = 8$ units, we give our phylogenetic Hill numbers ${}^q\bar{D}(T)$ (in units of “species equivalents”), phylogenetic diversity ${}^qPD(T)$ (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) phylogenetic diversity ${}^qPD(T)$	$q = 0$	26	32	14
	$q = 1$	24.78	32	11.31
	$q = 2$	23.27	32	9.85
Chao et al. (2010) phylogenetic Hill number ${}^q\bar{D}(T)$	$q = 0$	3.25	4	1.75
	$q = 1$	3.08	4	1.41
	$q = 2$	2.91	4	1.23
Scheiner (2012) phylogenetic diversity	$q = 0$	4	4	4
	$q = 1$	4	4	4
	$q = 2$	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 ${}^q\bar{D}(\bar{T})$ and ${}^qPD(\bar{T})$ 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 \bar{T} (or T in an ultrametric tree), or equivalently, all branch lengths must be equal to \bar{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 ${}^q\bar{D}(T)$ 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 $S^2 \times 0.1$. For Assemblage B, there are S species with a constant distance of 0.9 for all species pairs, and the functional diversity is $S^2 \times 0.9$. 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
<i>Ammophila arenaria</i>	0	0.118	0.006
<i>Anthemis maritima</i>	0.024	0.132	0.046
<i>Asparagus acutifolius</i>	0	0.003	0.003
<i>Bromus diandrus</i>	0	0.005	0.032
<i>Cakile maritima</i>	0.217	0.024	0.004
<i>Calystegia soldanella</i>	0.027	0.026	0.009
<i>Centaurea sphaerocephala</i>	0	0.008	0.015
<i>Chamaesyce peplis</i>	0.097	0.014	0.001
<i>Clematis flammula</i>	0	0.004	0.018
<i>Crucianella maritima</i>	0	0.022	0.080
<i>Cutandia maritima</i>	0.008	0.036	0.095
<i>Cyperus kalli</i>	0.003	0.047	0.037
<i>Daphne gnidium</i>	0	0	0.001
<i>Echinophora spinosa</i>	0.029	0.029	0.004
<i>Elymus farctus</i>	0.161	0.134	0.044
<i>Eryngium maritimum</i>	0.021	0.020	0.001
<i>Euphorbia terracina</i>	0	0.003	0.028
<i>Helicrisum stoechas</i>	0	0.004	0.029
<i>Juniperus oxycedrus</i>	0	0	0.008
<i>Lagurus ovatus</i>	0	0.003	0.022
<i>Lonicera implexa</i>	0	0.001	0.002
<i>Lophocloa pubescens</i>	0	0.005	0.006
<i>Lotus cytisoides</i>	0	0.005	0.062
<i>Medicago littoralis</i>	0	0.021	0.081
<i>Medicago marina</i>	0.003	0.047	0.012
<i>Ononis variegata</i>	0.005	0.070	0.025
<i>Otanthus maritimus</i>	0.043	0.005	0
<i>Pancratium maritimum</i>	0.016	0.042	0.039
<i>Phyllirea angustifolia</i>	0	0.002	0.005
<i>Pistacia lentiscus</i>	0	0.001	0.008
<i>Plantago coronopus</i>	0.003	0.013	0.017
<i>Polygonum maritimum</i>	0.038	0.003	0.001
<i>Prasium majus</i>	0	0.001	0.003
<i>Pseudorlaya pumila</i>	0	0.007	0.008
<i>Pycnocomon rutifolium</i>	0	0.003	0.042
<i>Quercus ilex</i>	0	0	0.003
<i>Rubia peregrina</i>	0	0	0.003
<i>Salsola kali</i>	0.193	0.025	0.002
<i>Silene canescens</i>	0	0.058	0.089
<i>Smilax aspera</i>	0	0.003	0.009
<i>Sonchus bulbosus</i>	0	0.006	0.005
<i>Sporobolus virginicus</i>	0.113	0.044	0.021
<i>Vulpia fasciculata</i>	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](https://arxiv.org/abs/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.