Appendix A. Proof and definitions

In our study, we compared indices that either used a matrix of dissimilarity among species directly or functional data transformed into a functional tree. We did not consider methods that use ordination approaches to position species into Euclidean space (Magnusson et al., 2009; Faith, 2015). Further studies on the need to use such approaches and how they compare to tree-based or distance-based approaches are still very much required. Connections exist among these alternative approaches: *Qb* can also be derived from an ordination method (Pavoine et al., 2005).

Definitions

For any real ($n \times n$) symmetric matrix $\mathbf{A}$, there exists a unitary matrix $\mathbf{U}$ with n rows (i.e. $\mathbf{U}^t\mathbf{U}=\mathbf{I}_n$, the $n \times n$ identity matrix) such that $\mathbf{U}^t\mathbf{A}\mathbf{U}=\mathbf{\Lambda}$, where $\mathbf{\Lambda}$ is a diagonal matrix of the eigenvalues of $\mathbf{A}$. Eigenvalues are real and eigenvectors, columns of $\mathbf{U}$, can be taken to be real (e.g. Seber 2008).

Proof that the first eigenvector of a dissimilarity matrix can be taken to be positive

Let $\mathbf{D}=(d_{ij})_{1\le i\le N, 1\le j\le N}$ be a matrix of pairwise dissimilarities with $d_{ij}=d_{ji}$ for any i,j, $d_{ij}>0$ for any $i\neq j$, and $d_{ii}=0$ for any i.

It is shown below that the first eigenvector of $\mathbf{D}$ has values of same sign, which can be taken to be positive.

According to Noutsos (2006, definition 2.2), a matrix $\mathbf{A} \in \mathbb{R}^{N,N}$ possesses the strong Perron-Frobenius property (in reference to the Perron-Frobenius theorem) if its dominant eigenvalue λ_1 is positive, the only one in the circle $|\lambda_1|$ ($\lambda_1 > |\lambda_i|$, $i = 2, 3, \ldots, N$) and the corresponding eigenvector $x^{(1)}$ is positive (which means that it has only positive values).

According to Noutsos (2006, definition 2.3), a matrix $\mathbf{A} \in \mathbb{R}^{N,N}$ is said to be *eventually positive* if there exists a positive integer k_0 such that $A^k > 0$ ($=A^k$ only has positive values) for all $k \ge k_0$. The demonstration that $\mathbf{D}$ is eventually positive stems from the fact that $d_{ij}>0$ for any $i\neq j$, and $d_{ii}=0$ for any i. With the latter conditions for the values of $\mathbf{D}$, $\mathbf{D}^k > 0$ for all $k \ge 2$.

According to Noutsos (2006, theorem 2.1), given that $\mathbf{D}$ is eventually positive and symmetric ($d_{ij}=d_{ji}$ for any i,j), then $\mathbf{D}$ possesses the strong Perron-Frobenius property.

□

Proof that the unique maximizing vector of R is the squared, first eigenvector of the matrix of dissimilarities

$R(\mathbf{p},\mathbf{D}) = \sum_{i=1}^{N}\sum_{j=1}^{N}\sqrt{p_i}\sqrt{p_j}d_{ij}$

Consider $u_i = \sqrt{p_i}$ and $\boldsymbol{u}=(u_1,...,u_N)$.

$R(\mathbf{u},\mathbf{D}) = \sum_{i=1}^{N}\sum_{j=1}^{N}u_i u_j d_{ij}$

It is well known that the maximum of **R**, given that $\sum_{i=1}^{N}u_i^2 = 1$ is the first eigenvalue of **D** that is reached when **u** is the first eigenvector of **D**. The fact that this eigenvector has positive values has been proved above so that the maximum of $\max_{\mathbf{p}}\left\{R(\mathbf{p},\mathbf{D})\right\}$ is the first eigenvalue of **D** that is reached for the squared, first eigenvector of **D**.

□

Details on Rao's Q

As above, we begin by considering a matrix $\mathbf{D}=(d_{ij})_{1\le i\le N,1\le j\le N}$ of dissimilarities among N species with N the number of species such that the dissimilarities are symmetric: $d_{ij}=d_{ji}$ for any i,j, $d_{ij}>0$ for any $i\neq j$, and $d_{ii}=0$ for each i. We also consider $\mathbf{p}=(p_1, ..., p_i, ..., p_N)$ a vector of species' relative abundance ($\sum_{i=1}^{N}p_i = 1$). The average dissimilarity among individuals from an assemblage is then an established measure of diversity, based on Rao's (1982) quadratic diversity or entropy (QE):

$$Q(\mathbf{p},\mathbf{D}) = \sum_{i}^{N} p_i \sum_{j}^{N} p_j d_{ij} .$$

If $d_{ij} = 1$ for all $i \neq j$, then the measure of diversity (Q) reduces to

$$T_2(\mathbf{p}) = 1 - \sum_{i}^{N} p_i^2 ,$$

which is the Gini-Simpson index and is a special case (when q=2) of a more generalized measure of entropy (the Tsallis entropy: T_q):

$$T_q(\mathbf{p}) = \frac{1-\sum_{i}^{N} p_i^q}{q-1},\ q>0,\ q\neq 1$$

When $q\to 1$, $T_q(\mathbf{p}) \underset{q\to 1}{\to} -\sum_{i=1}^{N} p_i \ln(p_i)$ and this is the Shannon entropy.

The use of quadratic entropy (Q) is then a generalization of the Tsallis entropy when q=2 that includes distances among species and can be used to measure functional and phylogenetic diversity. In the particular case where the d_{ij}'s are defined in the range [0,1], then $Q \leq T_2$. The maximum of T_2 is obtained when species' abundances are even (p_k=1/N for all i): $\max(T_2) = (N-1)/N$.

The Tsallis entropy is a function of Hill numbers (Hill 1973):

$${}^{q}D(\mathbf{p}) = \left[\sum_i^N (p_i)^q\right]^{\frac{1}{1-q}},\ q>0,\ q\neq 1$$

Indeed,

$$T_q(\mathbf{p}) = \frac{\left[{}^{q}D(\mathbf{p})\right]^{1-q}-1}{1-q},\ q>0,\ q\neq 1$$

$$\lim_{q\to 1}\left[T_q(\mathbf{p})\right] = \ln\left(\lim_{q\to 1}\left[{}^{q}D(\mathbf{p})\right]\right)$$

$1/\left[1-T_2(\mathbf{p})\right]$ thus is Hill number of order 2. It is an effective number of species: it is equal to the number of equally abundant species needed to obtain the same value of diversity. If the the d_{ij}'s are defined in the range [0,1], $1/\left[1-Q(\mathbf{p})\right]$ can then be interpreted as an effective number of evenly abundant and maximally dissimilar species (Ricotta & Szeidl 2009). Using $1/\left[1-Q(\mathbf{p})\right]$, abundant species are given high weight compared to rare species.

Pavoine et al. (2005) introduced the vector of species' abundances that maximizes Q, varying **p** and fixing **D**, as an index of species' originality (named QE-based index). However, this vector can contain null values (zeros) and several distinct vectors can simultaneously maximize Q for a given matrix **D**. To avoid this, Pavoine et al. (2005) proposed the restricted use of their index of species' originality to distances that have ultrametric properties (i.e. $d_{ij} \leq \max(d_{ik},d_{jk})$ for all i,j,k) including distances obtained from trees with constant tip-to-root distance (distances = half the sum of branch lengths on the shortest path that connects two tips in a tree).

Details on the new index of biodiversity, *R*

In the main text, we decreased the importance given to the most abundant species in Q as follows:

$$R(\mathbf{p},\mathbf{D}) = \sum_{i=1}^{N}\sum_{j=1}^{N}\sqrt{p_i}\sqrt{p_j}d_{ij}$$

The rest of the reasoning below follows exactly the same steps as for the QE-based index. For $d_{ij} = 1$ for all $i \neq j$, R reduces to

$$S_2(\mathbf{p}) = \left(\sum_i^N \sqrt{p_i}\right)^2 - 1,$$

S_2 is an established statistical metric (Behara & Chawla 1974) and physical quantity (Yamano 2014) as a special case of the generalized γ-entropy with γ=2:

$$S_\gamma(\mathbf{p}) = \frac{1 - \left[\sum_i^N (p_i)^{1/\gamma}\right]^\gamma}{1 - 2^{\gamma-1}},\ \gamma > 0,\ \gamma \neq 1$$

When $\gamma \to 1$, $S_\gamma(\mathbf{p}) \underset{\gamma \to 1}{\to} -\sum_{i=1}^N p_i \log_2(p_i)$. R is thus a generalization of the γ-entropy when γ=2 that includes distances among species and thus eventually measures functional and phylogenetic diversity. In the particular case where the d_{ij}'s are defined in the range [0,1], then $R \leq S_2$. The maximum of S_2 is obtained when species' abundances are even (p_k=1/N for all i): $\max(S_2) = N - 1$.

The generalized γ-entropy has been underexplored in ecology when analyzing biodiversity. As for Tsallis entropy, the generalized γ-entropy is a function of Hill numbers (Hill 1973). Indeed, considering q=1/γ,

$$S_\gamma(\mathbf{p}) = \frac{1 - \left[\sum_i^N (p_i)^q\right]^{1/q}}{1 - 2^{(1-q)/q}} = \frac{1 - \left[{}^qD(\mathbf{p})\right]^{(1-q)/q}}{1 - 2^{(1-q)/q}},\ q = 1/\gamma,\ \gamma \neq 1$$

$$\lim_{\gamma \to 1}\left[S_\gamma(\mathbf{p})\right] = \ln\left(\lim_{q \to 1}\left[{}^qD(\mathbf{p})\right]\right) / \ln(2)$$

As a particular case, $S_2(\mathbf{p}) = {}^{0.5}D(\mathbf{p}) - 1$. $S_2 + 1$ thus is Hill number of order 0.5. If the d_{ij}'s are defined in the range [0,1], R+1 can thus be interpreted as an equivalent number of evenly abundant and maximally dissimilar species. In R and R+1, rare species are given significant weight compared to Q and 1/[1-Q].

If the dissimilarities vary in the range [0,1] and if both the distances among species and the abundances of species vary, the maximum of R is N-1. In contrast, if the distances among species are fixed (with the only conditions that d_{ij}=d_{ji} for any i,j, d_{ij}>0 for any $i \neq j$, and d_{ii}=0 for any i) and only the abundances of the species vary, the maximum of R is the first eigenvalue of matrix **D** and the unique maximizing vector is its first eigenvector and has positive values (proof above). This maximizing vector was used by Champely and Chessel (2002) to rescale Q between 0 and 1.

The contribution of species to biodiversity

We provide in Fig. A.1 an illustration of the fact that increasing the abundance of original species compared to redundant species to a certain extent increases the diversity of the set of species.

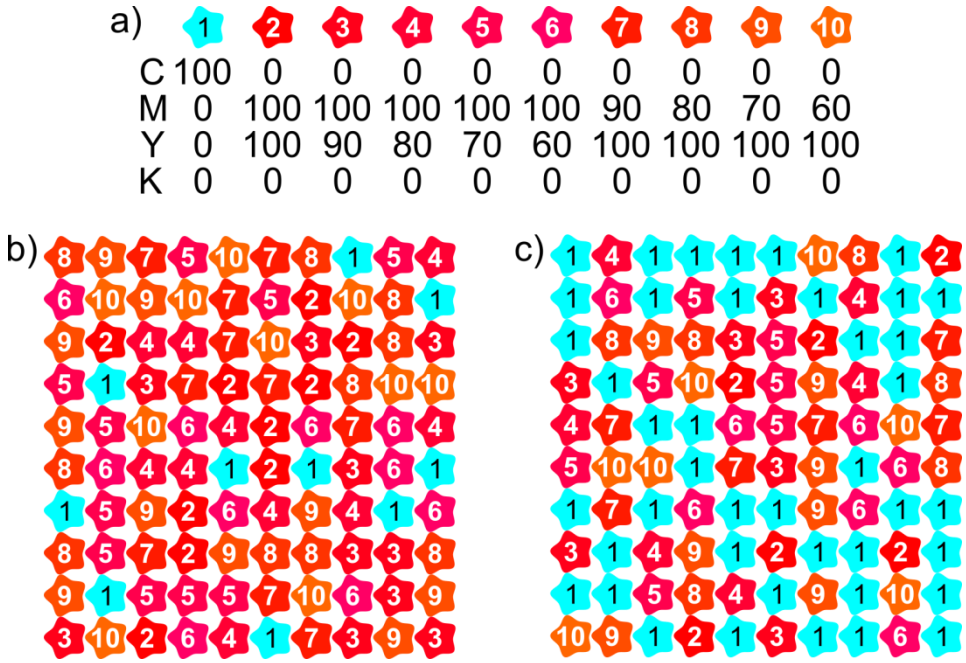

Fig. A.1. Theoretical example illustrating the increase in diversity when increasing the frequency of the most original species. Here species are symbolized by flowers and sets of species by fields of flowers. Each species numbered from 1 to 10 is associated with a color. We selected the CMYK color model used in color printing to define the color of the 10 flowers. The CMYK model is based on three elementary colors (C=cyan, M=magenta, Y=yellow), and on the black color (K). The percentage of each color used to define each flower is indicated in a). Species#1 is by far the most original, having only the cyan color, while others were created by mixing magenta and yellow. In b) each species is represented by 10 flowers. In c) the species with the most original color (species#1) is represented by 42 flowers, species#2 to #4 and species#7 and #8 by 6 flowers each and species#5, #6, #9, #10 by 7 flowers each. To obtain the number of flowers per species in c) we calculating Manhattan distance between the color profiles of species (given in a) and applied *Rb* to this inter-specific distance matrix.

References

Behara, M., Chawla, J.S., 1974. Generalized gamma-entropy. Selecta statistica canadiana 2, 15−38.

Champely, S., Chessel, D. 2002. Measuring biological diversity using Euclidean metrics. Environmental and Ecological Statistics, 9, 167−177.

Faith, D.P., 2015. The unimodal relationship between species' functional traits and habitat gradients provides a family of indices supporting the conservation of functional trait diversity. Plant Ecology 216, 725−740.

Hill, M.O., 1973. Diversity and evenness: a unifying notation and its consequences. Ecology 54, 427−432.
Magnuson-Ford, K., Ingram, T., Redding, D.W., Mooers, A.Ø., 2009. Rockfish (Sebastes) that are evolutionarily isolated are also large, morphologically distinctive and vulnerable to overfishing. Biological conservation 142, 1787–1796.
Noutsos, D. 2006. On Perron–Frobenius property of matrices having some negative entries. Linear Algebra and its Applications 412, 132–153.
Pavoine, S., Ollier, S., Pontier, D., 2005. Measuring diversity from dissimilarities with Rao's quadratic entropy: are any dissimilarities suitable? Theoretical population biology 67, 231–239.
Rao, C.R. 1982. Diversity and dissimilarity coefficients: a unified approach. Theoretical Population Biology 21, 24–43.
Ricotta, C., Szeidl, L. 2009. Diversity partitioning of Rao's quadratic entropy. Theoretical Population Biology 76, 299–302.
Seber, G.A.F. 2008. A matrix handbook for statisticians. Wiley, Hoboken, New Jersey.
Yamano, T., 2014. Thermodynamic meaning of the index of γ-entropy. Results in Physics 4, 50–51.