Appendix S3

DETAILS ABOUT THE PROPERTIES OF DISSIMILARITY COEFFICIENTS

We describe four groups of properties and indicate the reason why we consider them relevant. The first two groups (i.e. from P1 to P9) contain the minimum requirements for assessing beta diversity. The remaining two groups (i.e. P10 to P14) are not necessarily required in all beta diversity assessments. Practitioners should determine whether the context of their analyses requires these latter properties or not. Similarity coefficients should be transformed into dissimilarities before assessing the following properties.

Property class 1: Basic necessary properties. — Properties P1 to P6 must be fulfilled by all resemblance coefficients used for beta diversity assessment. P1 and P2 are actually mathematical axioms that define a dissimilarity function. Thus, they are fulfilled by all coefficients considered in this paper. P3 (monotonicity) is a necessary condition for any coefficient used to study species assemblages. All coefficients investigated in the present study are monotonic. Properties P1 to P3 are therefore not shown in Table 1 of the paper.

P1 – Minimum of zero and positiveness. A dissimilarity value should never be negative and it should be zero when comparing a site to itself. When comparing two different sites, it can be zero or greater than zero, depending on the species abundance values and how the dissimilarity is defined. For example, with some coefficients, *D* is zero when comparing two site vectors whose abundance values are proportional to each other; that is the case with the profile, chi-square, chord, and Hellinger distances. Dissimilarities that violate this property by taking negative values are called nonmetric, by opposition to the metric and semimetric coefficients (see Legendre & Legendre 2012).

P2 – Symmetry. Consider two community abundance vectors, $\mathbf{x}_1$ and $\mathbf{x}_2$, whose dissimilarity is to be assessed. In symmetric indices, $D(\mathbf{x}_1,\mathbf{x}_2) = D(\mathbf{x}_2,\mathbf{x}_1)$. In the incidence-based counterparts of these coefficients (Table 1 in the main paper), the values *b* and *c* play exchangeable (symmetric) roles. When studying beta diversity, there is no reason to make a distinction between the two sampling units that are compared using a coefficient. Therefore, dissimilarity coefficients must be symmetric. The property of being *double-zero symmetrical*, referred to in P4, is different.

P3 – Monotonicity to changes in abundance. Increasing the difference in abundances of one of several species between two sites increases their dissimilarity. Property P3 was verified using ordered comparison case series (OCCAS), corresponding to linear changes in the abundances of two species along different types of simulated environmental gradients. The OCCAS method was proposed by Hajdu (1981) and used by Gower & Legendre (1986) to assess monotonicity in dissimilarity coefficients.

P4 – Double-zero asymmetry. Coefficients that have this property do not change when double-zeros are added to the data, but the dissimilarity decreases when double-X (where X > 0) values are added. The reasoning implies two conditions, derived from ecological niche theory.

1. Ecological statement: double-zeros in species abundance are not interpretable. — In his seminal paper, Whittaker (1972) published a figure (his Fig. 4, p. 228) showing simulated species represented by bell-shaped curves with different widths (species tolerances), succeeding one another along three ecological gradients. (1.1) For species *j* observed at a pair of sites, the presence of that species (in any abundance) at both sites indicates that the two sites are similar to some extent, i.e. they are close in positions along the gradient. Because the two sites are within the tolerance zone of species *j*, that species can be found at these two sites. (1.2) Ruling out sampling error, the presence at one site and absence at the other unambiguously indicates that the

sites occupy different positions along the gradient. (1.3) Ruling out sampling error again, double absence of species *j* is not interpretable, because it can result from the two sites being either at close positions along the gradient but outside the tolerance zone of species *j* (e.g. pH too high at both sites for that species), or at positions far away along the gradient, both sites being outside the tolerance zone of species *j* (e.g. pH too low at one site and too high at the other).

2. Lemma. — The presence of species *j* at two sites (point 1.1 above) is an indication of resemblance of these sites whatever the abundances observed. It follows that presence of species *j* with the same abundance X at two sites (a difference (X – X) = 0 for that species; point 1.1 above) has a different meaning from the double-absence (which also corresponds to a difference (0 – 0) = 0; point 1.3 above).

3. This reasoning leads to the following conclusions. (3.1) Double-zeros (0, 0) have a different meaning than double-presences (X, X), whatever the abundances. (3.2) Coefficients that produce the same effect (i.e. no change in dissimilarity) for double zeros as they do for double presences with identical abundances (i.e. (X,X), which we call double-X), where X > 0) are called *double-zero symmetrical* because they treat double zeros like any other pair of identical values. These coefficients are not admissible for the study of ecological differentiation of communities, i.e. for beta diversity studies. The Euclidean and Manhattan distances belong to that type: double-zeros and double-X produce no change in distance, whereas any other pair of non-identical abundances does produce a change in the distance. (3.3) Coefficients useful for beta diversity studies must be *double-zero asymmetrical* (term used in Legendre and Legendre 2012), meaning that their value does not change with the addition of double zeros, but it decreases when species with double-X abundances that are not double-zeros are added to the comparison of two sites.

The difference between double-zero and double-X is clearly recognized in binary coefficients. In double-zero symmetrical similarity coefficients such as the simple matching index, the numbers of double-presences and of double-absences (usually respectively represented by *a* and *d*) are used as indications of similarity. In contrast, double-zero asymmetrical similarity coefficients such as the Jaccard and Sørensen indices exclude double-absences (*d*) from both their numerator and denominator and use only the double-presences (*a*) in their formula in addition to *b* and *c*. Some binary dissimilarity coefficients can be expressed as function of *b* and *c* only (see, for example, the presence-absence formulas for the Euclidean and Manhattan distances in Table 1 of the main paper). These are also double-zero symmetrical, because neither double-presences (*a*) nor double-absences (*d*) can change the value of the coefficient. A double-zero asymmetrical binary dissimilarity coefficient must include double-presences (*a*) and exclude double-absences (*d*) from its formula.

P5 – Sites without species in common have the largest dissimilarity. Coefficients that violate this property are not suitable for beta diversity studies. For the coefficients that have a fixed upper bound (last column of Table 2), the largest dissimilarity is the upper bound. The following data sets were used to test the 16 coefficients analysed in this paper. In each example, the rows are sites S1-S4, the columns represent species.

Example data 1 –

	1	***2***	***3***	***4***	***5***	***6***	***7***	***8***	***9***	***10***	***11***	***12***	***13***	***14***	***15***	***16***	***17***	***18***	***19***	***20***
S1	3	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
S2	7	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
S3	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0

S4	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1

Example data 2 –

	1	***2***	***3***	***4***
S1	1	4	0	0
S2	4	1	0	0
S3	0	1	1	0
S4	1	0	0	1

Example data 3 –

	1	***2***	***3***	***4***
S1	1	2	0	0
S2	2	1	0	0
S3	0	1	40	0
S4	1	0	0	40

Some dissimilarity functions present the following paradox with one of more of these data sets: for sites 3 and 4 that have no species in common, these coefficients produce dissimilarities that are smaller than for sites that have some (e.g. sites 1 and 3) or all (sites 1 and 2) their species in common. These examples are extensions of the Orlóci (1978, p. 46) paradox data.

P6 – Dissimilarity does not decrease in series of nested species assemblages. For pairs of sites having any number of unique species, i.e. species that are not found at the other site, the dissimilarity should be the same or increase with the number of unique species. In particular, the dissimilarity should not decrease when the number of unique species in one or both sites increases. To assess this property, we carried out simulations where we added unique species to one of the sites (which corresponds to the data structure described as *nestedness* of species assemblages by Wright & Reeves 1992 and Baselga 2010) or to both sites (data structure described as monotonic by Jost *et al.* 2011). Violation of P6 leads to the paradox that the total sum of squares for a pair of sites, which is $D^2/2$ (eq. 8 in the main paper), tends to 0 as the number of unique species increases.

Property class 2: Comparability between data sets. — The following three properties are needed to appropriately compare beta diversity values calculated for different data tables, even if the sampling unit are the same size (e.g. quadrat size for vegetation) and the sampling effort is the same. Therefore, we consider they should also be required in beta diversity studies.

P7 – Species replication invariance. A community composition table with the columns in two or several copies should produce the same dissimilarities among sites as the original data table. Procedure: select a community composition data table, compute a dissimilarity index and obtain a dissimilarity matrix. Then, duplicate the data table, combining the two copies side by side, using for example function cbind() in R. Compute the same dissimilarity function and obtain a new dissimilarity matrix. Repeat the duplication step to include three copies of the data and compute the dissimilarity matrix again. The three dissimilarity matrices should be identical if the dissimilarity function has the property of *replication invariance*. For the abundance-based coefficients, the property is computed using the population formula, not the sample formula. This property was first described by Jost *et al.* (2011).

A special case of this property (not included as a separate property in the present study), for presence-absence data, is called *homogeneity*, for example by Janson and Vegelius (1981), Koleff *et al.* (2003) and Chao *et al.* (2006). The homogeneity property allows the comparison of beta values computed from data tables containing different total species richness. This property is verified on the binary form of the coefficients, by multiplying *a*, *b*, *c* and *d* by a constant factor and checking whether the resulting index value is changed.

P8 – Invariance to the measurement units. This property concerns abundance-based formulas only. It allows the comparison of beta values between data tables (e.g. regions) with different productivities (abundance or biomass), or where biomass has been measured using different units (e.g. in g and mg). To see whether a given quantitative coefficient is invariant to changes of measurement scale, we multiplied the abundance values by a constant factor and checked whether the resulting index was altered.

P9 – Existence of a fixed upper bound. The existence of an upper bound for a coefficient facilitates the interpretation and comparison of beta values because an upper bound in the dissimilarity index leads to an upper bound in the beta diversity value. The maximum beta value for a region is obtained when all site pairs have the maximum dissimilarity $D_{\max}$ permitted by the chosen coefficient. One can apply eq. 8 to that situation to compute the maximum sum of squares:

$$\mathrm{SS}_{\max} = \frac{1}{n}\left(\frac{n(n-1)}{2}D_{\max}^2\right) = \frac{n-1}{2}D_{\max}^2$$

then eq. 3 to obtain the maximum beta diversity value:

$$\mathrm{BD}_{\mathrm{max}} = \mathrm{SS}_{\mathrm{max}}/(n–1) = \left(\frac{n-1}{2}D^2_{\mathrm{max}}\right)\frac{1}{n-1} = \frac{1}{2}D^2_{\mathrm{max}}$$

The upper bound varies among dissimilarity coefficients (Table 2 of the pain paper, right-hand column). For coefficients with $D_{\mathrm{max}} = \sqrt{2}$, $\mathrm{BD}_{\mathrm{max}} = 1$; for those with $D_{\mathrm{max}} = 1$, $\mathrm{BD}_{\mathrm{max}} = 0.5$ (see section “Maximum value of BD” in the pain paper). For the chi-square distance, $D_{\mathrm{max}} = \sqrt{2y_{++}}$ and $\mathrm{BD}_{\mathrm{max}} = y_{++}$ which is the sum of the species abundances in **Y**. Although y_{++} varies from data table to data table, the chi-square distance is considered as belonging to the group of the coefficients that have a fixed upper bound because its sister index, the chi-square metric (not otherwise discussed in this paper), has an upper bound of $\sqrt{2}$ (see e.g. Legendre and Legendre 2012). The chi-square distance is the chi-square metric multiplied by $\sqrt{y_{++}}$, hence it has an upper bound of $\sqrt{2y_{++}}$. The chi-square distance is the one computed in software packages; this is why its properties are described here. The chi-square metric and distance have the same properties besides their different maximum values. Hence, for coefficients that have a fixed maximum (see section “The dissimilarity measures” in the main paper), we can compute a relative value of beta diversity, $\mathrm{BD}_{\mathrm{rel}}$, as follows:

$$\mathrm{BD}_{\mathrm{rel}} = \mathrm{BD}_{\mathrm{Total}}/\mathrm{BD}_{\mathrm{max}}$$

which is a value between 0 and 1. $\mathrm{BD}_{\mathrm{rel}}$ is useful to compare beta values computed using different coefficients.

Property class 3: Sampling issues. — This group of properties is mostly related to sampling issues. The fulfilment of properties P10 and P11 facilitates (but does not ensure) the comparability of beta values obtained from sampling units having different sizes or sampled using different efforts. Indeed, both the number of species and the total abundance may be strongly affected by changes in the size of sampling units or in sampling effort. On the other hand, if the size of sampling units and sampling effort are sufficiently homogeneous, ecologists may be interested in allowing differences in the numbers of species, and perhaps also in the total abundances between sites, to influence the dissimilarity and beta diversity assessments.

The last property deals with correction for undersampling (P12) of the community composition. This property is also related to sampling effort. It is related to sampling unit size as well because small sampling units can lead to undersampling the richness of the targeted community.

P10 – Invariance to the number of species in each sampling unit. This property analyses whether a double-zero asymmetrical binary coefficient changes its value depending on the number of species in each of two sampling units $\mathbf{x}_1$ and $\mathbf{x}_2$ that are compared. Does the dissimilarity value change if the two communities are species rich, compared to when the two communities are species poor or when one is rich and the other poor?

This property was verified algebraically on the binary form of the coefficients; it could not be checked for the chi-square metric which does not have a binary form. We start with the usual a, b, c notation for binary indices:

a = number of species shared between $\mathbf{x}_1$ and $\mathbf{x}_2$

b = number of unique species in $\mathbf{x}_1$ that do not appear in $\mathbf{x}_2$

c = number of unique species in $\mathbf{x}_2$ that do not appear in $\mathbf{x}_1$

We then define the number of species in $\mathbf{x}_1$ and $\mathbf{x}_2$ as $n_1 = a + b$ and $n_2 = a + c$, and the proportion of shared species with respect to each site as $p_1 = a / n_1 = a / (a + b)$ and $p_2 = a / n_2 = a / (a + c)$. After defining n_1, n_2, p_1 and p_2, one can reformulate the binary dissimilarity measures found in column *Incidence-based* of Table 1 of the main paper in terms of these four quantities, instead of using the notation a, b, c. The idea of the proof is to see whether a dissimilarity measure can be reformulated using a notation that uses n_1, n_2, p_1 and p_2, and then see if these terms n_1 and n_2 cancel out in the formula. In other words, we ask whether $D(a, b, c) = D(n_1, n_2, p_1, p_2)$ can be reduced to $D(p_1, p_2)$. The following equivalences are useful for reformulation:

$$a = (a + b) \times (a / (a + b)) = n_1 \times p_1 = (a + c) \times (a / (a + c)) = n_2 \times p_2$$

$$b = (a + b) \times (1 - (a / (a + b))) = n_1 \times (1 - p_1)$$

$$c = (a + c) \times (1 - (a / (a + c))) = n_2 \times (1 - p_2)$$

Using the presence-absence form of each dissimilarity measure (column *Incidence-based* in Table 1), property P10 is verified algebraically by trying to cancel n_1 and n_2 out of the formula.

P10 is a stricter property than *homogeneity* (see P7, second paragraph). It is easy to show that fulfilling P10 leads to a coefficient that is homogeneous (invariant to the total number of species of the data set), because if $D(a, b, c) = D(p_1, p_2)$, and knowing that $p_1 = a / (a + b) = k{\cdot}a / (k{\cdot}a + k{\cdot}b)$ and $p_2 = a / (a + c) = k{\cdot}a / (k{\cdot}a + k{\cdot}c)$, then we have $D(a, b, c) = D(ka, kb, kc)$, which proves *homogeneity*. However, the reverse does not follow. Indeed, a given coefficient may be homogeneous without satisfying P10. An example is the asymmetric binary similarity coefficient proposed by Kulczynski (1928), $S_{12} = a / (b + c)$ (S_{12} in Legendre & Legendre 2012), which is homogeneous but does not fulfil P10.

Proofs for individual coefficients –

(1) **Euclidean distance**. When this distance is calculated on presence-absence data it can be formulated as:

$$E = \sqrt{b+c} = \sqrt{n_1 \cdot (1 - p_1) + n_2 \cdot (1 - p_2)}$$

Because n_1 and n_2 cannot be cancelled out, the Euclidean distance does NOT satisfy P10.

(2) **Manhattan distance**. When this distance is calculated on presence-absence data it can be formulated as:

$$M = |b+c| = |n_1 \cdot (1-p_1) + n_2 \cdot (1-p_2)|$$

As before, the Manhattan distance does NOT satisfy P10.

(3) **Jaccard** similarity index (proof thanks to A. Chao).

$$J = \frac{a}{a+b+c} = \frac{a}{2a+b+c-a} = \frac{n_1 p_1}{(n_1 p_1 + n_2 p_2) + n_1(1-p_1) + n_2(1-p_2) - n_1 p_1}$$

$$= \frac{n_1 p_1}{n_1 + n_2 - n_1 p_1} = \frac{1}{(1/p_1) + (1/p_2) - 1}$$

Thus, the Jaccard index DOES satisfy P10. All resemblance coefficients that are equal to the Jaccard index for presence-absence data satisfy P10: the **modified mean character difference**, **coefficient of divergence**, **Canberra metric**, **Wishart coefficient** (1- Similarity ratio), and **abundance-based Jaccard**.

(4) **Sørensen** similarity index (proof thanks to A. Chao) The Sørensen similarity index is

$$S = \frac{2a}{2a+b+c} = \frac{2n_1 p_1}{n_1 p_1 + n_2 p_2 + n_1(1-p_1) + n_2(1-p_2)}$$

$$= \frac{2n_1 p_1}{n_1 + n_2} = \frac{2}{(1/p_1) + [n_2/(n_1 p_1)]}$$ and, since $n_1 p_1 = n_2 p_2$,

$$= \frac{2}{(1/p_1) + (1/p_2)}$$

Thus, the Sørensen index DOES satisfy P10. All resemblance coefficients that are equal to the Sørensen index for presence-absence data satisfy P10: the **percentage difference** and **abundance-based Sørensen**.

(5) **Ochiai** similarity index:

$$O = \frac{a}{\sqrt{(a+b)(a+c)}} = \sqrt{\frac{a}{(a+b)} \cdot \frac{a}{(a+c)}} = \sqrt{p_1 \cdot p_2}$$

Thus, the Ochiai index DOES satisfy P10. All resemblance coefficients that are equal to the Ochiai index for presence-absence data satisfy P10: the **Hellinger distance**, **chord distance**, and **abundance-based Ochiai**. In this coefficient, the similarity is the geometric mean of p_1 and p_2.

(6) **Species profile distance**. When this distance coefficient is calculated on presence-absence data, it can be written using the *a-b-c-d* notation as:

$$SP = \sqrt{\frac{b+c}{(a+b)(a+c)}}$$

After reformulating this index, we can reduce it to:

$$SP = \sqrt{\frac{n_1 \cdot (1-p_1) + n_2 \cdot (1-p_2)}{n_1 \cdot n_2}} = \sqrt{\frac{1-p_2}{n_1} + \frac{1-p_1}{n_2}}$$

Because n_1 and n_2 cannot be cancelled out, the species profile distance does NOT satisfy P10.

(7) **Whittaker's index of association**. When this distance coefficient is calculated on presence-absence data, it can be written using the *a-b-c-d* notation as:

$$W = \frac{1}{2}\left(\frac{b}{a+b} + \frac{c}{a+c} + \left|\frac{a}{a+b} - \frac{a}{a+c}\right|\right)$$

After reformulating this index, we can reduce it to:

$$W = \frac{1}{2}\left(\frac{n_1 \cdot (1-p_1)}{n_1} + \frac{n_2 \cdot (1-p_2)}{n_2} + \left|\frac{n_1 \cdot p_1}{n_1} - \frac{n_2 \cdot p_2}{n_2}\right|\right)$$

$$W = \frac{1}{2}\left((1-p_1) + (1-p_2) + |p_1 - p_2|\right) = \frac{1}{2}\left[2 - (p_1 + p_2) + |p_1 - p_2|\right]$$

Thus, the Whittaker index of association also DOES satisfy P10.

(8) **Kulczynski** similarity index. When this *similarity* is calculated on presence-absence data it can be straightforwardly formulated as:

$$K = \frac{1}{2}\left[\frac{a}{a+b} + \frac{a}{a+c}\right] = \frac{1}{2}\left[p_1 + p_2\right]$$

Thus, the Kulczynski similarity also DOES satisfy P10. In this coefficient, the similarity is the arithmetic mean of p_1 and p_2.

P11 – Invariance to the total abundance in each sampling unit. Except when researchers only count and identify a fixed number of individuals (which is often the case in plankton or palaeoecological studies), sampling units in the data table are likely to have different total abundances. Some abundance-based dissimilarity indices are only sensitive to relative abundances per site whereas others reflect differences in site total counts. This property was called “density invariance” by Jost *et al.* (2011). It is not the same as property P7 above. One can check property P11 by determining whether a coefficient is altered when the abundances are multiplied by a constant factor that is different for each sampling unit.

P12 – Coefficients with corrections for undersampling. With higher sampling effort, i.e. larger sampling units, rare species, and in particular those that are not found at the two sites under comparison, are more likely to be observed (Chao *et al.* 2006, Cardoso *et al.* 2009). For that reason, dissimilarity coefficients generally underestimate the dissimilarities among sites, the bias decreasing when sampling effort increases. For some binary similarity coefficients, Chao *et al.* (2006) and Jost *et al.* (2011) suggested abundance-based counterparts that incorporate corrections for undersampling bias.

Property class 4: Ordination-related properties. — The remaining properties are not related to the ecological interpretation of a coefficient or the comparability of beta diversity values. They are, however, useful for ordination and linear modelling of community composition data.

P13 – Euclidean property of D or D$^{(0.5)}$. A dissimilarity matrix **D** is *Euclidean* if it can be embedded in a Euclidean space of real axes such that the Euclidean distances among points are equal to the dissimilarity values in **D**. For coefficients that are Euclidean, principal coordinate analysis of **D** produces ordinations that are fully represented in Euclidean space (i.e. without negative eigenvalues). Several coefficients have the Euclidean property. Some coefficients that are not Euclidean for **D** become Euclidean after taking the square root of the dissimilarity values (Gower & Legendre 1986); the resulting matrix, which contains values [$D_{hi}^{0.5}$], is noted $\mathbf{D}^{(0.5)}$. Legendre & Legendre (2012, Tables 7.2 and 7.3) describe the Euclidean properties of 43 commonly-used similarity and dissimilarity coefficients, including several of the coefficients listed in Table 1.

P14 – Emulated by transformation of the raw frequency data followed by Euclidean distance. Legendre & Gallagher (2001) described how some distance coefficients can be obtained by computing the Euclidean distance (eq. 7 in the main paper) after transforming the raw data values in some appropriate way. Four such transformations are described in Appendix S1. Coefficients that can be obtained in that way are interesting because one can obtain $\mathrm{BD}_{\mathrm{Total}}$ by computing the transformation and then applying eqs 1-3. Moreover, transformed data allow the computation of the beta diversity contributions of individual species through eqs 4a and 4b (SCBD indices) and of sites through eqs 5a and 5b (LCBD indices). One can also use the transformed data directly in linear modelling of community composition data, e.g. by simple (PCA) or canonical (RDA) ordination, *K*-means partitioning, or multivariate regression tree analysis (MRT), because these methods implicitly preserve the Euclidean distance among sites.

In addition to the coefficients obtained by transformation followed by calculation of the Euclidean distance (Appendix S1), the Whittaker index can also be obtained by applying the Manhattan distance to profile transformed data; see section “The dissimilarity coefficients” in the main paper. This produces twice Whittaker’s index of association; for that reason, Whittaker’s index was dubbed “relativized Manhattan” by Faith *et al.* (1987). The Manhattan distance is, however, not the distance implicit in linear models, so that Whittaker’s index of association does not lend itself to linear modelling nor to the calculation of *Species Contributions to Beta Diversity* (SCDB indices) described in the main paper, eq. 4b.

ACKNOWLEDGEMENTS

The idea for example data set 1 used in paragraph P5 was provided by Prof. Anne Chao, Institute of Statistics, National Tsing Hua University, Hsin-Chu, Taiwan. Prof. Chao’s example was expanded for the present Appendix.

REFERENCES

Baselga, A. (2010). Partitioning the turnover and nestedness components of beta diversity. *Global Ecol. Biogeogr.*, 19, 134–143.

Faith, D.P., Minchin, P.R. & Belbin, L. (1987). Compositional dissimilarity as a robust measure of ecological distance. *Vegetatio*, 69, 57–68.

Gower, J.C. & Legendre, P. (1986). Metric and Euclidean properties of dissimilarity coefficients. *J. Classif.*, 3, 5–48.

Hajdu, L.J. (1981). Geographical comparison of resemblance measures in phytosociology. *Vegetatio*, 48, 47–59.

Jost, L., Chao, A. & Chazdon, R.L. (2011). Compositional similarity and beta diversity. In: *Biological diversity: frontiers in measurement and assessment* [eds Magurran, A. & McGill, B.]. Oxford University Press, Oxford, England, pp. 66–84.

Kulczynski, S. (1928). Die Pflanzenassoziationen der Pieninen. *Bull. Int. Acad. Pol. Sci. Lett. Cl. Sci. Math. Nat. Ser. B*, Suppl. II (1927), 57–203.

Legendre, P. & Gallagher, E.D. (2001). Ecologically meaningful transformations for ordination of species data. *Oecologia*, 129, 271–280.

Legendre, P. & Legendre, L. (2012). *Numerical ecology*. 3rd English edition. Elsevier Science BV, Amsterdam.

Orlóci, L. (1978). *Multivariate analysis in vegetation research.* 2nd edition. Dr. W. Junk B. V., The Hague, The Netherlands.

Whittaker, R.H. (1972). Evolution and measurement of species diversity. *Taxon*, 21, 213–251.

Wright, D.H. & Reeves, J.H. (1992). On the meaning and measurement of nestedness of species assemblages. *Oecologia*, 92, 416-428.