Global Ecology and Biogeography
A Journal of Macroecology

Interpreting the replacement and richness difference components of beta diversity

Journal:	*Global Ecology and Biogeography*
Manuscript ID:	GEB-2013-0328.R2
Manuscript Type:	Macroecological Methods
Date Submitted by the Author:	n/a
Complete List of Authors:	Legendre, Pierre; Université de Montréal, Département de sciences biologiques
Keywords:	Beta diversity, community composition data, community ecology, dissimilarity coefficients, local contributions to beta diversity, replacement, richness difference

Interpreting the replacement and richness difference components of beta diversity

Pierre Legendre*

Département de sciences biologiques, Université de Montréal

E-mail addresses: pierre.legendre@umontreal.ca

Keywords

Beta diversity, community composition data, community ecology, dissimilarity coefficients, local contributions to beta diversity, replacement, richness difference.

Short running title: Replacement and richness difference components

*Correspondence

Pierre Legendre
Département de sciences biologiques
Université de Montréal
C.P. 6128, succursale Centre-ville
Montréal, Québec, Canada H3C 3J7

Phone: 514-343-7591
Fax: 514-343-2293
E-mail: Pierre.Legendre@umontreal.ca

Words in abstract: 296 # Words in main text (Introduction to Biosketch): 7237

References: 33 # Figures: 4 # Tables: 0

Article type: Macroecological methods

ABSTRACT

Aim The variation in species composition among sites, or beta diversity, can be decomposed into replacement and richness difference. A debate is ongoing in the literature concerning the best ways of computing and interpreting these indices. This paper first reviews the historical development of the formulas for decomposing dissimilarities into replacement, richness difference and nestedness indices. These formulas are presented using a unified algebraic framework, for species presence-absence and abundance. The indices decomposing beta play different roles in ecological analysis than beta diversity indices.

Innovation Replacement and richness difference indices can be interpreted and related to ecosystem processes. The pairwise index values can be summed across all pairs of sites; these sums form a valid decomposition of total beta diversity into total replacement and total richness difference components. Different communities and study areas can be compared: some may be dominated by replacement, others by richness/abundance difference processes. Within a region, differences among sites measured by these indices can then be analysed and interpreted using explanatory variables or experimental factors. The paper also shows that local contributions of replacement and richness difference to total beta diversity can be computed and mapped. A case study is presented involving fish communities along a river.

Main conclusions The different forms of indices are based upon the same functional numerators. These indices are complementary; they can help researchers understand different aspects of ecosystem functioning. The methods of analysis used in this paper apply to any of the indices recently proposed. Further work, based on ecological theory and numerical simulations, is required to clarify the precise meaning and domain of application of the different forms. The

forms available for presence-absence and quantitative data are both useful because these different data types allow researchers to answer different types of ecological or biogeographic questions.

INTRODUCTION

Since Whittaker (1972), ecological dissimilarities have been used to measure beta diversity between sampling units. Koleff et al. (2002) reviewed 24 beta diversity indices proposed in the literature whereas Legendre & De Cáceres (2013) described 14 properties of 11 dissimilarity indices that are appropriate for beta diversity studies.

Harrison *et al.* (1992), Williams (1996) and Lennon *et al.* (2001) pioneered the idea that dissimilarities among communities result from two different processes: species replacement (also called turnover), and richness difference or nestedness (species gain and loss). Following these foundation publications, a series of papers (see Methods) appeared, proposing more refined or alternative ways of partitioning dissimilarity indices in beta diversity studies.

All authors use the same quantitative components, found when comparing community compositions at two sites (a, b, c, $(b + c)$, $|b - c|$, $\min(b,c)$, Fig. 1), to construct replacement, richness difference and nestedness indices from species presence-absence data. Figures similar to Fig. 1 are found in several papers: Williams (1996), Baselga (2010, 2012), Podani & Schmera (2011) and Carvalho *et al.* (2013).

Species replacement refers to the well-known fact that species tend to replace each other along ecological gradients that are sufficiently long (e.g. Whittaker, 1952); the replacement rate is also a function of the ecological tolerance, or niche breadth, of the species. Species replacement is also called turnover when analyzed along spatial or environmental gradients. It implies the simultaneous gain and loss of species due to environmental filtering and historical

events (Leprieur *et al.*, 2011). The replacement component of dissimilarity measures (Appendix S1) may thus reflect the influence on community structure of the variables controlling ecological gradients.

Richness difference refers to the fact that a community may include a larger number of species than another. It may reflect the diversity of niches available at different locations along the sampling axis or throughout the study area. Differences in richness may be due to species thinning causing nestedness, or to other ecological processes.

Nestedness is a type of richness difference pattern characterized by the species at a site being a strict subset of the species at a richer site (Atmar & Patterson, 1993; Baselga, 2012). This is also the concept that underlies the nestedness indices of Podani & Schmera (2011). How to translate the concept of nestedness into an index is, however, a subject open to discussion as there is no unequivocal way to do it.

In the comparison of two sites, richness difference can be interpreted as nestedness *sensu stricto* only if the sites have 'a' species in common, with $a > 0$, and they differ in other species, one site being richer than the other. When $a = 0$, the richness difference between two sampling units cannot be interpreted as nestedness, which is logically 0 in that situation (Podani & Schmera, 2011; Carvalho *et al.*, 2013). In that case, the difference in species composition measured by dissimilarity indices is equal to species replacement plus richness difference, without reference to ecological processes producing nestedness. For real ecological data, replacement and richness difference (or nestedness) contribute jointly to the differentiation of the sites, and it is interesting to partition a dissimilarity matrix into these two components, which correspond to different ecological processes, provided that the components add up to the dissimilarity value (Fig. 1).

87 The first contribution of this paper is to review the historical development of the formulas
88 proposed in the literature for decomposing dissimilarity indices into replacement, richness
89 difference and nestedness indices. These formulas are casted into a unified algebraic framework,
90 for species presence-absence and abundance data (Appendix S1).

91 The main contributions of this article are the following: (1) The paper demonstrates that the
92 sums of the replacement and richness difference indices corresponding to all pairwise
93 comparisons over a study area form a proper decomposition of the total beta diversity over that
94 area. (2) It describes how the differences among sites measured by these indices can be analysed
95 and interpreted using descriptive or experimental factors, or explanatory environmental variables.
96 They can also be used to produce ordinations. (3) Finally, it shows how to compute local
97 contributions of replacement and richness difference to total beta diversity. These contributions
98 can be mapped to facilitate interpretation. The calculations are illustrated using an ecological case
99 study. R software is presented in appendices.

100 **METHODS**

101 Many papers appeared, mostly during the past four years, describing or discussing
102 replacement, richness (or abundance) difference, and nestedness indices (Williams 1996; Lennon
103 *et al.* 2001; Cardoso *et al.*, 2009; Baselga, 2010, 2012, 2013a; Podani & Schmera 2011; Schmera
104 & Podani, 2011; Carvalho *et al.*, 2012, 2013; Podani *et al.,* 2013); index notations differed
105 among papers. Because the present paper is limited in page space, a brief history of the
106 development of these indices is presented in Appendix S1 together with a summary of the
107 criticisms formulated against suggested indices.

That discussion shows that all authors agree on the dissimilarity coefficients that seem to be the most useful bases for decomposition: firstly, the Jaccard (D_J) and Sørensen (D_S) indices for presence-absence data, which are usually expressed in terms of the following quantities found in a contingency table crossing two vectors of presence-absence data: a = number of species present at both sites, b = number of species present at site 1 but not at site 2, c = number of species present at site 2 but not at site 1; secondly, the Ružička (D_R) and percentage difference ($D_{\%diff}$) dissimilarity coefficients for quantitative data. They also agree that the operational portion of the equations estimating replacement is the quantity $\min(b,c)$ or $2\min(b,c)$ used as the numerator of indices for presence-absence data and the quantity $\min(B,C)$ or $2 \times \min(B,C)$ used as the replacement numerator for quantitative data (meanings of B and C: see Appendix S1). For richness difference, the operational portion of the equations is $|b - c|$ for presence-absence data; to estimate abundance difference for quantitative data, it is $|B - C|$.

Replacement, richness difference and nestedness indices

Appendix S1 describes the development of indices that decompose dissimilarity coefficients into replacement, richness/abundance difference and nestedness components. These indices belong to families that were developed by two groups of authors; for simplicity, they are called the Podani and Baselga families in this paper. Each family contains indices for species presence-absence and for abundance data. Both families were completed by new indices described in Appendix S1. The indices are designated as follows in this paper (formulas shown in Appendix S1, Table S1.1):

• Podani family, presence-absence data – Replacement indices: $Repl_J$ and $Repl_S$, which are components of the Jaccard (D_J) and Sørensen (D_S) dissimilarities, respectively. Richness difference indices: $RichDiff_J$ and $RichDiff_S$.

• Podani family, abundance data – Replacement indices: $Repl_R$ and $Repl_{\%diff}$, which are components of the Ružička and percentage difference dissimilarities, respectively. Abundance difference indices: $AbDiff_R$ and $AbDiff_{\%diff}$.

• Baselga family, presence-absence data – Replacement indices: $Repl_{BJ}$ and $Repl_{BS}$, which are components of the Jaccard and Sørensen dissimilarities, respectively. Nestedness indices: Nes_{BJ} and Nes_{BS}.

• Baselga family, abundance data – Replacement indices: $Repl_{BR}$ and $Repl_{B\%diff}$, which are components of the Ružička and percentage difference dissimilarities, respectively. Nestedness indices: Nes_{BR} and $Nes_{B\%diff}$.

From the exchanges that appeared in the recent literature, the following points are discussed in Appendix S1. This section only presents the conclusions reached in the appendix, where details are given.

• It is the numerators of the proposed indices that estimate replacement and richness difference. One can then scale the indices to values between 0 and 1 with denominators of one's choice, depending on the purpose of the study. The denominators of the Jaccard, Sørensen, Ružička and percentage difference dissimilarities, or those used by Baselga (2010, 2012) in his replacement (turnover) indices, can all be used. Ecologists should understand, however, that the chosen denominators might create distortions in the positioning of sites in an ordination, compared to using the numerator values only. None of the denominators proposed up to now have all the optimal qualities. The discussions about over- or under-estimation of species replacement by indices of the two families of indices are, actually, discussions about the choice of a denominator.

• The indices in the Podani family correspond to the concepts of replacement and richness/abundance difference. Those in the Baselga family are replacement (or turnover) and nestedness indices. Richness difference is not the same as nestedness. Podani & Schmera (2011) proposed an index of nestedness (N_{rel}) that differs from their index of richness difference; they explained that the latter only represents a portion of nestedness. Hence the Baselga nestedness indices (Nes_{BJ} and Nes_{BS}) should be compared to Podani & Schmera's relativized nestedness index, not to the richness difference indices of the Podani family ($RichDiff_J$ and $RichDiff_S$).

• In the two families, the replacement and richness difference (Podani family) or replacement and nestedness indices (Baselga family) sum to dissimilarity measures (D_J, D_S, D_R, $D_{\%diff}$). These four dissimilarities are appropriate for beta diversity assessment, following the criteria of Legendre & De Cáceres (2013). An important point is that the replacement, richness difference and nestedness indices are not themselves indices of beta diversity; they decompose dissimilarity coefficients that can be used to estimate beta diversity.

• Replacement and richness difference or nestedness indices should have an ecological interpretation. In that respect, indices in the Podani family are easy to interpret due to the logic of their construction. Likewise, interpretation of Baselga's replacement indices is clear, whereas that of his nestedness indices is more intricate, albeit logical.

• When matrices of indices are to be used to produce ordinations of the sites, the Podani-family richness/abundance difference indices ($RichDiff_S$ and $AbDiff_{\%diff}$) that decompose the Sørensen and percentage difference dissimilarities present clear advantages for ordination because the $\mathbf{RichDiff}_S$ and $\mathbf{AbDiff}_{\%diff}$ matrices are Euclidean, meaning that the data points can be fully represented in Euclidean space by principal coordinate analysis (PCoA) without production of negative eigenvalues and complex ordination axes.

• Claims have been made that the *Repl* indices in the Podani family were correlated to species richness differences between the sampling units whereas indices in the Baselga family were not. Actually, all indices described in Appendix S1 do not depend directly (and linearly) on site richness since they can all be expressed without recourse to the richness of the compared sites, r_1 and r_2. Section 6 of the appendix shows that they can all be expressed by equations containing only p_1 and p_2, where p_1 is the proportion of shared species in the first sampling unit, $p_1 = a/r_1$, and p_2 is the proportion of shared species in the second sampling unit, $p_2 = a/r_2$; a is the number of species in common between the two sites. Section 5 of the appendix shows an example where some replacement indices vary as an inverse function of richness difference and another example where they do not. Hence this criticism does not apply to any of the indices described in this paper.

Partitioning total beta diversity

Pelissier *et al.* (2003), Legendre *et al.* (2005) and Anderson *et al.* (2006) showed that the total variance of a community composition table is an appropriate measure of its variation in species composition, or beta diversity. Total beta can be directly computed from the community data table or from a dissimilarity matrix derived from it using an appropriate dissimilarity coefficient (Legendre & De Cáceres, 2013). Beta diversity can thus be computed as $BD_{Total} = \sum_{h=1}^{n-1}\sum_{i=h+1}^{n} D_{hi}^2/(n(n-1))$. In that equation, $\sum_{h=1}^{n-1}\sum_{i=h+1}^{n} D_{hi}^2/n$ is the total sum of squares (SS_{Total}); further division by $(n-1)$ produces the total variance, or total beta diversity (BD_{Total}).

For the D_S (Sørensen), D_J (Jaccard), $D_{\%diff}$ and D_R dissimilarity coefficients, which are non-Euclidean, taking the square root of the dissimilarities makes the resulting matrices $\mathbf{D}^{(0.5)}$ =

$\left[D_{hi}^{0.5}\right]$ Euclidean (meaning described in the previous subsection; Legendre & Legendre, 2012, Tables 7.2 and 7.3). For that reason, in the variance approach (previous paragraph), Legendre & De Cáceres (2013) recommended, for these four coefficients, to take the square root of the distances before computing total beta diversity, BD_{Total}:

$$BD_{Total} = \sum_{h=1}^{n-1}\sum_{i=h+1}^{n}\sqrt{D_{hi}}^{2}/(n(n-1)) = \sum_{h=1}^{n-1}\sum_{i=h+1}^{n}D_{hi}/(n(n-1)) \quad (1)$$

as shown by Whittaker (1972; Whittaker's formula gives twice the value of BD_{Total} produced by eq. 1) and Legendre & De Cáceres (2013), where D_{hi} is the dissimilarity between sites h and i.

For any two sites h and i, in the Podani family of indices, replacement ($Repl_{hi}$) plus richness difference ($RichDiff_{hi}$) is equal to D_{hi}. Likewise, in the Baselga family, replacement ($Repl_{hi}$) plus nestedness (Nes_{hi}) is equal to D_{hi}. Hence, the sum of the $Repl_{hi}$ plus the sum of the $RichDiff_{hi}$ values is equal to the sum of D_{hi} values, and this relationship can be written:

$$BD_{Total} = Repl_{Total} + RichDiff_{Total} \quad (2)$$

where

$$Repl_{Total} = \sum_{h=1}^{n-1}\sum_{i=h+1}^{n}Repl_{hi}/(n(n-1)) \text{ and } RichDiff_{Total} = \sum_{h=1}^{n-1}\sum_{i=h+1}^{n}RichDiff_{hi}/(n(n-1)).$$

Because these equations are similar to eq. 1, the quantities $Repl_{Total}$ and $RichDiff_{Total}$ form a true partition of BD_{Total} for any community data matrix analysed using one of the four dissimilarity functions considered in this paper. This allows ecologists to calculate the proportion of BD_{Total} accounted for by the replacement and richness (or abundance) difference fractions as

$$Repl_{Prop} = Repl_{Total}/BD_{Total} \quad \text{and} \quad RichDiff_{Prop} = RichDiff_{Total}/BD_{Total} \quad (3)$$

The sum of $Repl_{Prop}$ and $RichDiff_{Prop}$ is 1. The denominators of the terms in these ratios, $n(n-1)$, cancel out and the proportions can be computed as

$$Repl_{Prop} = \sum_{h=1}^{n-1}\sum_{i=h+1}^{n} Repl_{hi} \Big/ \sum_{h=1}^{n-1}\sum_{i=h+1}^{n} D_{hi} \quad (4a)$$

$$\text{and } Rich_{Prop} = \sum_{h=1}^{n-1}\sum_{i=h+1}^{n} Rich_{hi} \Big/ \sum_{h=1}^{n-1}\sum_{i=h+1}^{n} D_{hi} \quad (4b)$$

For a single pair of sampling units h and i, $n = 2$, so that BD_{Total} for that pair is $D_{hi}/2$, which can be partitioned into $Repl_{Total} = Repl_{hi}/2$ and $RichDiff_{Total} = RichDiff_{hi}/2$.

While the replacement and richness difference components are required for detailed gradient analysis, the $Repl_{Total}$ and $RichDiff_{Total}$ indices are useful to determine which of the two processes dominates among the sampling sites under study.

Similar relationships can be computed for the replacement ($Repl_{BJ}$, $Repl_{BS}$, $Repl_{BR}$, $Repl_{B\%diff}$) and nestedness (Nes_{BJ}, Nes_{BS}, Nes_{BR}, $Nes_{B\%diff}$) components in the Baselga family of indices. These global indices offer simple alternatives to the multiple-site indices proposed by Baselga (2010, 2013b). For ecological interpretation, the interest of analysing pairwise replacement and nestedness matrices, instead of multiple-site indices, will be shown in the following subsections and in the Case study.

An R function is provided in Appendix S3 to compute all distance, replacement, richness difference and nestedness matrices described in this paper, as well as the partitioning of total beta diversity described in the present section.

Explaining *Repl* and *RichDiff* variation

When replacement and richness difference have been estimated for all pairs of sites in a study, the next step is to test hypotheses of explanations for their variation. For the dissimilarity matrices $\mathbf{D}_J$, $\mathbf{D}_S$, $\mathbf{D}_R$ and $\mathbf{D}_{\%diff}$, which represent beta variation, the distance-based redundancy analysis (db-RDA) method of Legendre & Anderson (1999) can be used to explain the variation of the species data by canonical analysis. That method consists of the following steps: first, a principal coordinate analysis (PCoA) of the dissimilarity matrix is computed; the principal coordinates are then used as the response data in redundancy analysis (RDA) against a matrix of explanatory variables. That method works best when the dissimilarity matrix is Euclidean; in that case, all principal coordinate axes are real and they account together for the entire variation represented by the dissimilarity matrix. As mentioned in the previous section, $\mathbf{D}_J$, $\mathbf{D}_S$, $\mathbf{D}_R$ and $\mathbf{D}_{\%diff}$ matrices are non-Euclidean, but taking the square root of the dissimilarities makes the resulting matrices $\mathbf{D}^{(0.5)} = [\, D_{hi}^{0.5}]$ Euclidean (meaning described above). Using the complete set of principal coordinates computed from $\mathbf{D}^{(0.5)}$ in redundancy analysis allows ecologists to select the most interesting explanatory variables in a stepwise manner or test the effect of a given explanatory variable or factor on the among-site variation in community composition.

In most instances, the replacement (**Repl**), richness/abundance difference and nestedness matrices are not fully Euclidean even after taking the square root of the coefficients, although square-rooting reduces in important ways the non-Euclidean component of these matrices that shows up as negative eigenvalues and complex ordination axes (in Appendix S1, Table S1.4 shows that only the $\mathbf{RichDiff}_S$ and $\mathbf{AbDiff}_{\%diff}$ matrices of the Podani family are fully Euclidean). McArdle & Anderson (2001) described a way of correctly testing the significance of the canonical relationship between a dissimilarity matrix **D** and a set of explanatory variables when

D is non-Euclidean. The method, which is of course also valid for Euclidean matrices, consists in computing the projector (or *'hat'*) matrix

$$\mathbf{H} = \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}' \tag{5}$$

where **X** is the matrix of explanatory variables. Matrix **H** is also used in multiple regression and RDA. Then compute the Gower-centred matrix **G**, which is the first step of principal coordinate analysis (PCoA, Gower, 1966) before eigenvalue decomposition:

$$\mathbf{G} = \left(\mathbf{I} - \frac{\mathbf{1}\mathbf{1}'}{n}\right)\ \mathbf{A}\ \left(\mathbf{I} - \frac{\mathbf{1}\mathbf{1}'}{n}\right) \tag{6}$$

where $\mathbf{A} = [a_{hi}] = [-0.5D_{hi}^2]$. Use these matrices to construct the F-statistic of RDA as follows:

$$F = \frac{\mathrm{tr}(\mathbf{HGH})/m}{\mathrm{tr}\left[(\mathbf{I}-\mathbf{H})\mathbf{G}(\mathbf{I}-\mathbf{H})\right]/(n-m-1)} \tag{7}$$

where tr() is the trace operator which sums the diagonal values of a square matrix, **I** is an identity matrix of the same size as **H**, m is the rank of **X** centred by columns, and n is the number of observations. This statistic can be tested by permutation, as done by the demonstration function in Appendix S4. This method improves upon the test originally proposed by Legendre & Anderson (1999), which is approximate in the case of non-Euclidean dissimilarity matrices. It is recommended to square-root the values in the **Repl**, **RichDiff**, **AbDiff** and **Nes** matrices before this analysis in order to reduce their non-Euclideanarity.

Local contributions

Local contributions to beta diversity (LCBD indices) are comparative indicators of the ecological uniqueness of the sites for their contributions to beta diversity (Legendre & De Cáceres, 2013). In an ordination diagram by principal coordinate analysis, the sites with high

277 LCBD values are those found far from the multivariate centroid of the graphs; LCBD is actually
278 the squared distance of a site to the data centroid, which is the point of origin of the multivariate
279 graph.

280 Large LCBD values indicate sites that have strongly different species compositions
281 compared to a mean site. For conservation biology, large LCBD values may indicate sites that
282 have unusual species combinations and high conservation value, or degraded and species-poor
283 sites that are good candidates for ecological restoration. They may also correspond to special
284 ecological conditions or result from the effect of invasive species on communities. LCBD values
285 can be mapped to facilitate interpretation, as shown in the Legendre & De Cáceres (2013) paper.

286 For readers interested in computational details, LCBD indices are the diagonal values of the
287 Gower-centred matrix **G** (eq. 6) computed during principal coordinate analysis of the
288 dissimilarity matrix; each value is then divided by the total sum of squares of the data, which is
289 the sum of the diagonal values of **G**. Because of that division, the sum of LCBD values computed
290 over all sites is 1. Several dissimilarity functions, for presence-absence or abundance data, are
291 available to compute total beta diversity as well as LCBD indices (Legendre & De Cáceres,
292 2013).

293 Here the calculation of LCBD indices is extended to replacement and richness difference.
294 These new LCDB indices, $Repl_{LCBD}$ and $RichDiff_{LCBD}$, measure how exceptional each site is,
295 when compared to the other sites, in terms of replacement or richness (or abundance) difference.
296 In an ordination of the replacement or richness difference indices, the sites with high values of
297 $Repl_{LCBD}$ and $RichDiff_{LCBD}$ are those that are far from the multivariate centroid of the graph. Like
298 total beta LCBD, $Repl_{LCBD}$ and $RichDiff_{LCBD}$ are the diagonal values of the Gower-centred
299 dissimilarity matrix computed during principal coordinate analysis of the **Repl** and **RichDiff**

matrices, divided by the total sum of squares of the matrix. An R function is provided in Appendix S5 to compute LCBD from a dissimilarity, replacement, or richness difference matrix. Note that $Repl_{LCBD}$ and $RichDiff_{LCBD}$ computed from *D* do not add up to LCBD indices computed from the dissimilarity matrix. That is because in each case, the LCBD indices measure the squared distance of the sites to the multivariate centroid in an ordination diagram, and the three ordinations have unrelated axes, despite the fact that the **Repl** and **RichDiff** matrices add up to the corresponding **D** matrix.

Simplex analysis

Podani & Schmera (2011) and Podani *et al.* (2013) noted that *Repl* + *RichDiff* = *D* and that the similarity $S = (1 - D)$, which imply that *S* + *Repl* + *RichDiff* = 1. These authors proposed to represent the triplets of values {*S*, *Repl*, *RichDiff*} corresponding to a point in a triangular graph that they called an SDR-simplex; examples are shown in the Case study section. The sides of the triangle bear scales of similarity (bottom edge, with zero on the left), richness difference (left edge, with 0 at the top) and replacement (right edge, with 0 at the bottom). Each triplet of values represents the similarity $S = 1 - D$ as well as the two components of *D*. If one of the three components is zero, the point is found on the edge where that component has the value 0; for example, if replacement is 0, the point is found along the bottom edge. If two components are equal, the point is found on a median that originates on one of the vertices and ends in the centre of the opposite edge; for example, if replacement and richness difference are equal (values between 0 and 0.5), the points are on the median that originates in the lower-right corner and ends in the centre of the richness difference edge. Podani and Schmera (2011) provided a series of examples that help familiarize oneself with the interpretation of SDR simplices.

CASE STUDY

Freshwater fish were collected by Verneaux (1973) in the Doubs River, a tributary of the Saône that runs near the France-Switzerland border in the Jura Mountains in eastern France. In his paper, Verneaux proposed to use fish communities to characterize ecological zones along European rivers and streams. The data include fish community composition at 30 sites along the 453 km course of the river, the site geographic coordinates, and environmental data (source: http://adn.biol.umontreal.ca/~numericalecology/numecolR/). Twenty-seven species were captured and identified. No fish were caught at site 8, hence that site was excluded from the reanalyses made by Borcard *et al.* (2011) as well as here.

This data set exhibits a strongly nested structure along the first five sites in the headwaters of the river. For example, a single species is present at site 1, then 3, 4, 8 and 11 at sites 2 to 5 (Fig. 2c). Using the binary forms of the coefficients, D_S or D_J, this strong ecological gradient produced replacement values of 0 among these 5 sites because min(b,c) was always 0; all differences were captured by the richness difference part of the dissimilarity D. Looking at the 29 sites, the fish community is dominated by richness difference, which accounts for 72% of total beta diversity measured through D_S or D_J ($RichDiff_{Prop}$), compared to 28% for species replacement ($Repl_{Prop}$). In the **Repl** and **RichDiff** matrices, 27 pairs of sites had richness difference values of 0 compared to 150 pairs that had replacement values of 0. With presence-absence data, richness difference of 0 occurs when all differences between sites are attributed to species replacement, so that $b = c$ and hence $|b - c| = 0$.

Let us examine the graphs of the indices comparing sites 1–29 with site 30 located downstream, where richness is high (Fig. 2a). Site 30 was selected as the reference because it corresponds to the historical point of entry of the species into the river. Reading the graph from

345 right to left, richness differences (triangles) are null or very small up to site 17, except for sites 25
346 to 23, discussed a few sentences down. Thereafter, the values increase markedly towards site 1
347 upstream. Replacement values (squares) are also very small until about site 19, where they
348 increase slowly up to site 15, after which they drop along the upper course of the river. The
349 replacement (squares) and richness difference (triangles) indices sum to the Jaccard distances
350 (circles, often hidden by the red triangles in the right-hand portion of the graph). The same
351 observations can be made for this data set from the indices computed using the Ružička index,
352 which is the quantitative form of the Jaccard dissimilarity (Fig. 2b). Sites 25 to 23 stand out as
353 different (shaded areas in Fig. 2); these sites suffer from agricultural pollution and have
354 impoverished fish communities compared to the sites down- or upriver. These three sites have
355 high phosphate, nitrate and ammonium concentrations, low dissolved oxygen, and high biological
356 oxygen demand.

357 Ward hierarchical clustering of the sites based upon the matrix of environmental variables
358 showed a first division of the sites between the upper (sites 1 to 22) and lower courses of the river
359 (sites 23 to 30). The lower course is characterized by lower altitude and higher water discharge,
360 higher hardness and biological oxygen demand, and lower oxygen concentrations. A test of
361 significance of the variation of the $\mathbf{D}_J$, $\mathbf{Repl}_J$ and $\mathbf{RichDiff}_J$ matrices against a factor representing
362 the upper and lower groups of sites was carried out to illustrate the test of significance (eq. 7) of
363 the variation in these matrices by a factor, using the function in Appendix S4. Note that the
364 indices derived from the community composition data subjected to the test are independent of the
365 environmental data used in the cluster analysis that produced the factor. The binary factor
366 significantly explained the variation between the two groups for $\mathbf{D}_J$ and $\mathbf{Repl}_J$ (p-values of 0.005
367 and 0.001 respectively after 999 random permutations), but not for $\mathbf{RichDiff}_J$ (p-value of 0.943).

368 This result was expected from the ordinations shown in Appendix S6, where sites 1-22 are clearly
369 separated from sites 23-30 in the replacement ordinations, but not in the richness difference
370 ordinations (Figs. S6.1 and S6.2). Tests based upon the quantitative forms of the indices
371 produced similar results.

372 Following that, the variation among sites in the upper course of the river only (sites 1 to
373 22) was studied in more detail using the environmental variables. In order to identify the
374 environmental variables that best explained each index, a principal coordinate analysis of each of
375 the Jaccard-based matrices (with indices square-rooted) was carried out, with Lingoes correction
376 for the negative eigenvalues in the case of the $\mathbf{Repl}_J$ and $\mathbf{RichDiff}_J$ matrices; matrix $\mathbf{D}_J^{(0.5)}$ was
377 Euclidean and did not produce negative eigenvalues. The principal coordinates were used as
378 input response variables into a forward selection procedure in RDA (function forward.sel() of R
379 package packfor, Dray *et al.*, 2012). The selected explanatory variables were {slope, hardness,
380 nitrate and O_2} for $\mathbf{D}_J$, {O_2} for $\mathbf{Repl}_J$, and {slope, hardness and nitrate} for $\mathbf{RichDiff}_J$. The tests
381 of significance using principal coordinates corrected for negative eigenvalues were only
382 approximate (McArdle & Anderson, 2001). An exact test of significance was then computed for
383 each matrix and its set of explanatory variables, using the function in Appendix S4, which
384 implements the *F*-test described in eq. 7 in a permutation testing procedure. The variation in the
385 $\mathbf{D}_J$, $\mathbf{Repl}_J$ and $\mathbf{RichDiff}_J$ matrices explained by the selected environmental variables (those listed
386 above) was highly significant in all cases (p-values = 0.001) with high values of R^2_{adj} of 0.36,
387 0.44 and 0.47 respectively. Thus the *Repl* and *RichDiff* components of the species variation were
388 significantly related to different environmental variables.

389 LCBD indices were computed for the replacement and richness difference indices
390 decomposing Jaccard dissimilarities. These indices are shown on schematic maps of the river in

Fig. 3. Results obtained using decomposition of the quantitative Ružička distances are nearly identical. Note that the *Replacement LCBD* and *Richness difference LCBD* indices show the most exceptional sites for each type of index separately. They do not allow researchers to determine the relative importance of the replacement and richness difference processes in the study; that information will be provided by triangular graphs (below).

• In a principal coordinate ordination of the $Repl_J$ matrix (Appendix S6, Fig. S6.1a), sites 11-15 and 23-25 were the farthest from the centroid of the ordination diagram when considering all ordination dimensions. As a consequence, high LCBD values for replacement ($Repl_{LCBD}$ index) were found at sites 11 to 15, which is a transition zone (sites located at intermediate altitudes, with strong slope, high oxygen concentration like the upper section, and water discharge like the next lower section) between the head section of the river and the more quiet sections downstream, and at the polluted sites 23-25 (Fig. 3a) which are exceptional because of their low species richness reflecting extreme ecological conditions.

• In a principal coordinate ordination of the $RichDiff_J$ matrix (Appendix S6, Fig. S6.1b), sites 1-3 and 23 were the farthest from the centroid of the ordination diagram. Hence, high LCBD values for richness difference ($RichDiff_{LCBD}$ index) were found at sites 1-3 and 23 (Fig. 3b). These sites had very small species richness (1 to 4 species) and were thus very different in richness from the other sites (Fig. 2c). Among the remaining sites on the map, site 29 has the highest LCBD (albeit smaller than sites 1-3 and 23) because it has the highest richness.

Triangular graphs were used to represent the pairwise indices in the **S** = (1 – **D**), **Repl**, **RichDiff** and **AbDiff** matrices using all algebraic forms described in this paper. Graphs of this type were recommended by Podani & Schmera (2011) and Podani *et al.* (2013) for interpretation of the *Repl* and *RichDiff* indices. The results for the Jaccard and Sørensen indices on the one hand

414 (Fig. 4a,b) and for the Ružička and percentage difference indices on the other hand (Fig. 4c,d),
415 are very similar. The graphs show that the among-site variation is dominated by richness
416 difference (mean points [large dots] along the *RichDiff* or *AbDiff axis* near 0.5 in Figs 4a and 4c,
417 and near 0.4 in 4b and 4d, compared to lower mean values along the *Repl* axis), confirming the
418 values found above for $RichDiff_{Prop}$ = 72% of total beta diversity versus 28% for $Repl_{Prop}$.
419 Remember that these two statistics use the sum of the *D* values as their denominator (eq. 4a,b),
420 and that sum is $n(n–1)/2$ minus the sum of the *S* values used in the triangular plots. That is why
421 the $RichDiff_{Prop}$ and $Repl_{Prop}$ statistics are larger than the means of *RichDiff* and *Repl* shown by
422 large dots along the graph margins. Note also the many triplets (points in the triangular graphs)
423 along the *Similarity* axis that have *Repl* values of 0.

424 **DISCUSSION**

425 Several forms of replacement and richness difference indices have been described in the
426 literature, and arguments have been presented about the way these indices should be computed
427 and about their interpretation. The key papers are listed in the introduction of Appendix S1.
428 Further work, based on ecological theory and numerical simulations, is required to clarify the
429 precise meaning and the domain of application of the different forms of these indices, which are
430 most likely complementary and should help researchers understand different aspects of
431 ecosystem functioning.

432 The present paper focussed on two families of indices and used a unified algebraic
433 framework for their computation. These indices have interesting properties: the numerators of the
434 indices are chosen to estimate the ecological phenomena of interest (replacement and
435 richness/abundance difference), and the denominators are added afterwards to normalize the

436 indices. The sum of the replacement and richness difference components is always equal to the
437 corresponding dissimilarity, binary (Jaccard or Sørensen) or quantitative (Ružička or percentage
438 difference). The indices react monotonically to species composition (presence-absence) or
439 abundance gradients, as shown by numerical simulations. The R function provided with this
440 paper allows readers to check empirically that when the quantitative indices are computed on
441 presence-absence data, they produce the same results as the binary versions of the indices; that
442 fact can also be checked algebraically.

443 Podani & Schmera (2011) found a logical advantage in the Jaccard-based indices where the
444 denominator is the total richness, (a+b+c), or its quantitative equivalent, but it is too early to rule
445 out the Sørensen-based indices that have a different denominator. In the analysis of real data sets,
446 little differences were found between the two sets of results and the correlation between the
447 Jaccard-based and Sørensen-based indices was very high. This result was expected because the
448 numerators of the indices are the same (Table S1.2).

449 In the Podani family, the indices decomposing D_S present an advantage for ordination
450 because $RichDiff_S$ is Euclidean. For triangular plots on the contrary, interpretation of the indices
451 decomposing D_J is algebraically marginally simpler because the denominator is the total number
452 of species in each pair of sites under study. By opposition, in indices decomposing D_S, the
453 denominator gives double weight to the a fraction, and this moves the points in the simplex
454 towards the lower-right corner of the plot.

455 Computing indices is a first step; interpreting their variation is the next necessary step. This
456 paper described statistical tools and methods for the interpretation of these new indices. That is
457 its main contribution. The paper showed that the replacement and richness difference indices can
458 be summed over the study sites in an area, and the sums can be divided by the sum of the

459 corresponding distances. These ratios estimate the relative contributions of the two ecological
460 processes to beta diversity over the study area. This is because, for the Jaccard, Sørensen,
461 Ružička and percentage difference dissimilarities, which are non-Euclidean, the sum of all
462 pairwise dissimilarities over a study area, divided by $n(n-1)$, estimates the beta diversity of the
463 area. Different communities in the same area, as well as a community of interest in different
464 study areas, may be compared using these ratios. Some communities may be dominated by
465 replacement processes, others by richness/abundance difference processes, and this gives
466 ecologists insights into the influence of environmental variation on communities and the
467 importance of species interactions.

468 The paper then showed that the portions of beta diversity represented by the **Repl** and
469 **RichDiff** or **AbDiff** matrices could be analysed and interpreted with respect to explanatory
470 factors or sets of environmental variables, using a *F*-test of significance designed for the analysis
471 of non-Euclidean matrices in redundancy analysis. In the same spirit, Dobrovolski *et al.* (2012)
472 conducted a macroecological study relating the ratios Nes_{BS}/D_S, averaged over the eight
473 neighbours of each map grid cell of the New World (in the manner of Lennon *et al.* 2001), to cell
474 age (time since glaciation), for different groups of vertebrates: amphibians, birds and mammals.
475 Their results supported the hypothesis that the nestedness component of beta diversity is more
476 important in areas affected by glaciations until recent times.

477 The case study illustrates how the calculations are done. The last methodological section
478 showed that local contributions can be computed for the replacement and richness difference
479 components of beta diversity, and that these contributions can be plotted on a map of the study
480 sites. Again, the case study illustrates that point.

Should one use the presence-absence or the quantitative forms of the indices? That same question must always be answered as part of the process of choosing dissimilarity indices. Experience shows that binary dissimilarity coefficients produce interesting results when communities differ in the species complements composing them and partly harbour different species. Quantitative indices are clearly appropriate when the species are largely the same and the communities differ mostly by the abundances of their species (species composition). The former situation is expected from communities that are geographically far apart, hence the binary forms of the beta diversity component indices are likely to produce interesting results in studies carried out at broad spatial scales, whereas the abundance forms of the indices are likely to be preferable in studies carried out within small spatial extents, where species composition should differ mostly in the abundances of the species. For the comparison of communities that are so geographically distant that they do not share species, analyses based on groups of species with similar traits, e.g. species guilds, may bring about useful comparative information (e.g. Villéger *et al.*, 2013). The indices can also be modified to incorporate phylogenetic information about species relatedness (Leprieur *et al.*, 2012; Cardoso *et al.*, 2014).

Methodological issues should also be factored into the decision. Quantitative indices can be used only when the abundance assessments are based on appropriate sampling procedures and are comparable among sites. Biomass, or diameters at breast height (dbh) for trees, can be used instead of abundance data to compute quantitative dissimilarity indices and partition them into replacement and richness difference indices. Binary coefficients, although likely to produce coarser results, are the only choice when the data come from different sources, or have been collected by different researchers over long periods, or come from indirect sources like governmental reports, grey literature or museum collections. Palaeolimnologists analysing

sediment cores prefer in most studies to rely on species presence-absence data; their confidence in the numbers of individuals or pollen grains observed for the various species is low because these numbers may not reflect the quantitative structure of the ancient communities they are studying. Presence-absence data are also often the most reliable basis for comparison of distant areas in macroecological studies.

For species abundance data with lognormal distributions, researchers often log-transform the data before computing the percentage difference dissimilarity. For abundance data with less extreme distributions, the square root transformation is often used prior to ordination analysis. The same can be done before computing the Ružička dissimilarity. These data transformations shrink the differences caused by high abundance values and change the computed dissimilarities (Legendre & Legendre, 2012, section 7.7). A word of caution is in order about the effect of these variance-reducing transformations on the *Repl* and *AbDiff* indices. For abundance data, these indices are computed on an individual basis. Log-transformed data produce values of *Repl* and *AbDiff* indices that are intermediate between the results obtained with raw abundance and with binary data. Hence, the results are not equivalent to those obtained from raw abundance data and their ecological interpretation may differ. Further research, including simulation studies, is required about the effect of transformations on *Repl* and *AbDiff* indices.

ACKNOWLEDGEMENTS

I am thankful to Yu-Yun Chen, I-Fang Sun and Thomas Lamy who experimented with the indices and computed them on real ecological data during the development phase of this paper, and to Andrés Baselga, Daniel Borcard, Miquel De Cáceres, José C. Carvalho, Fabien Leprieur and János Podani, as well as an anonymous reviewer, who sent useful comments that greatly

526 helped in improving the manuscript. This research was funded by a Natural Sciences and
527 Engineering Research Council of Canada (NSERC) research grant to P. Legendre.

528 **REFERENCES**

529 Anderson, M.J. , Ellingsen, K.E. & McArdle, B.H. (2006) Multivariate dispersion as a measure
530 of beta diversity. *Ecology Letters*, **9**, 683–693.

531 Atmar, W. & Patterson, B.D. (1993) The measure of order and disorder in the distribution of
532 species in fragmented habitat. *Oecologia*, **96**, 373–382.

533 Baselga, A. (2010) Partitioning the turnover and nestedness components of beta diversity. *Global*
534 *Ecology and Biogeography*, **19**, 134–143.

535 Baselga, A. (2012) The relationship between species replacement, dissimilarity derived from
536 nestedness, and nestedness. *Global Ecology and Biogeography*, **21**, 1223–1232.

537 Baselga, A. (2013a) Separating the two components of abundance-based dissimilarity: balanced
538 changes in abundance vs. abundance gradients. *Methods in Ecology and Evolution*, **4**, 552–
539 557.

540 Baselga, A. (2013b) Multiple site dissimilarity quantifies compositional heterogeneity among
541 several sites, while average pairwise dissimilarity may be misleading. *Ecography*, **36**, 124–
542 128.

543 Borcard, D., Gillet, F. & Legendre, P. (2011) *Numerical ecology with R.* Use R! series, Springer
544 Science, New York.

545 Cardoso, P., Borges, P.A.V. & Veech, J.A. (2009) Testing the performance of beta diversity
546 measures based on incidence data: the robustness to undersampling. *Diversity and*
547 *Distributions*, **15**, 1081–1090.

548 Cardoso, P., Rigal, F., Carvalho, J.C., Fortelius, M., Borges, P.A.V., Podani, J. & Schmera, D.
549 (2014) Partitioning taxon, phylogenetic and functional beta diversity into replacement and
550 richness difference components. *Journal of Biogeography*, **41**, (first published online: 4 Dec
551 2013, DOI: 10.1111/jbi.12239).

552 Carvalho, J.C., Cardoso, P. and Gomes, P. (2012) Determining the relative roles of species
553 replacement and species richness differences in generating beta-diversity patterns. *Global*
554 *Ecology and Biogeography*, **21**, 760–771.

555 Carvalho, J.C., Cardoso, P., Borges, P.A.V., Schmera, D. & Podani, J. (2013) Measuring
556 fractions of beta diversity and their relationships to nestedness: a theoretical and empirical
557 comparison of novel approaches. *Oikos*, **122**, 825–834.

558 Dobrovolski, R., Melo, A.S., Cassemiro, F.A.S. & Diniz-Filho, J.A.F. (2012) Climatic history
559 and dispersal ability explain the relative importance of turnover and nestedness components of
560 beta diversity. *Global Ecology and Biogeography*, **21**: 191–197.

561 Dray, S. [with contributions of Legendre, P. & Blanchet, G.]. (2012) packfor: Forward selection
562 with permutation (Canoco p. 46). R package version 0.0-8. https://r-forge.r-
563 project.org/R/?group_id=195.

564 Gower, J.C. (1966) Some distance properties of latent root and vector methods used in
565 multivariate analysis. *Biometrika*, **53**, 325–338.

566 Harrison, S., Ross, S.J. & Lawton, J.H. (1992) Beta-diversity on geographic gradients in Britain.
567 *Journal of Animal Ecology*, **61**, 151–158.

568 Koleff, P., Gaston, K.J. & Lennon, J.J. (2003). Measuring beta diversity for presence-absence
569 data. *Journal of Animal Ecology*, **72**, 367–382.

570 Legendre, P. & Anderson, M.J. (1999) Distance-based redundancy analysis: testing multispecies
571 responses in multifactorial ecological experiments. *Ecological Monographs*, **69**, 1–24.

572 Legendre, P., Borcard, D. & Peres-Neto, P.R. (2005) Analyzing beta diversity: partitioning the
573 spatial variation of community composition data. *Ecological Monographs*, **75**, 435–450.

574 Legendre, P. & De Cáceres, M. (2013) Beta diversity as the variance of community data:
575 dissimilarity coefficients and partitioning. *Ecology Letters*, **16**, 951–963.

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

578 Lennon, J.J., Koleff, P., Greenwood, J.J.D. & Gaston, K.J. (2001) The geographical structure of
579 British bird distributions: diversity, spatial turnover and scale. *Journal of Animal Ecology,* **70**,
580 966–979.

581 Leprieur, F., Albouy, C., De Bortoli, J., Cowman, P.F., Bellwood, D.R. & Mouillot, D. (2012)
582 Quantifying phylogenetic beta diversity: distinguishing between 'true' turnover of lineages
583 and phylogenetic diversity gradients. *PLOS One*, **7**(8), 7(8): e42760.
584 doi:10.1371/journal.pone.0042760.

585 Leprieur, F., Tedesco, P.A., Hugueny, B., Beauchard, O., Dürr, H.H., Brosse, S. & Oberdorff, T.
586 (2011) Partitioning global patterns of freshwater fish beta diversity reveals contrasting
587 signatures of past climate changes. *Ecology Letters*, **14**: 325–334.

588 McArdle, B.H. & Anderson, M.J. (2001) Fitting multivariate models to community data: a
589 comment on distance-based redundancy analysis. *Ecology*, **82**, 290–297.

590 Pelissier, R., Couteron, P., Dray, S. & Sabatier, D. (2003) Consistency between ordination
591 techniques and diversity measurements: two strategies for species occurrence data. *Ecology*,
592 **84**, 242–251.

593 Podani, J., Ricotta C. & Schmera D. (2013) A general framework for analyzing beta diversity,
594 nestedness and related community-level phenomena based on abundance data. *Ecological*
595 *Complexity*, **15**, 52–61.

596 Podani, J. & Schmera, D. (2011) A new conceptual and methodological framework for exploring
597 and explaining pattern in presence-absence data. *Oikos*, **120**, 1625–1638.

598 Schmera, D. & Podani, J. (2011) Comments on separating components of beta diversity.
599 *Community Ecology*, **12**, 153–160.

600 Verneaux, J. (1973) Cours d'eau de Franche-Comté (Massif du Jura). Recherches écologiques sur
601 le réseau hydrographique du Doubs – Essai de biotypologie. *Annales Scientifiques de*
602 *l'Université de Franche-Comté, Biologie Animale,* **3**, 1–260.

603 Villéger, S, Grenouillet, G. & Brosse, S. (2013) Decomposing functional β-diversity reveals that
604 low functional β-diversity is driven by low functional turnover in European fish assemblages.
605 *Global Ecology and Biogeography*, **22**, 671–681.

606 Whittaker, R.H. (1952) A study of summer foliage insect communities in the Great Smoky
607 Mountains. *Ecological Monographs*, **22**, 1–44.

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

609 Williams, P.H. (1996) Mapping variations in the strength and breadth of biogeographic transition
610 zones using species turnover. *Proceedings of the Royal Society of London B: Biological*
611 *Sciences*, **263**, 579-588.

612 **BIOSKETCH**

613 Pierre Legendre is a specialist of numerical ecology, Fellow of the Royal Society of Canada
614 (Academy of Science), and ISI Highly Cited Researcher in Ecology/Environment. He is
615 especially interested in the ecological and biogeographic processes that organize biodiversity
616 spatially and temporally. As part of his research in fundamental ecology, he develops quantitative
617 methods for ecological data analysis. Web page: http://numericalecology.com.
618

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article at the publisher's web-site.

Appendix S1 Description and discussion of the replacement, richness difference and nestedness indices.

Appendix S2 Simulations showing that the replacement (*Repl*) and richness/abundance difference (*RichDiff* and *AbDiff*) indices respond monotonically to species gradients.

Appendix S3 R function to compute the Podani- and Baselga-family decompositions of the Jaccard or Sørensen groups into replacement and richness difference (or nestedness) components, for species presence-absence or abundance data.

Appendix S4 R function to compute the dbRDA *F*-test of significance between response data represented by a Euclidean or non-Euclidean dissimilarity matrix and a matrix of explanatory variables, following McArdle and Anderson (2001).

Appendix S5 R function to compute LCBD indices from a dissimilarity matrix (**D**) or from beta diversity component matrices (**Repl**, **RichDiff**, **AbDiff** or **Nes**).

Appendix S6 Principal coordinate ordinations of the replacement and richness difference indices for the fish case study data.

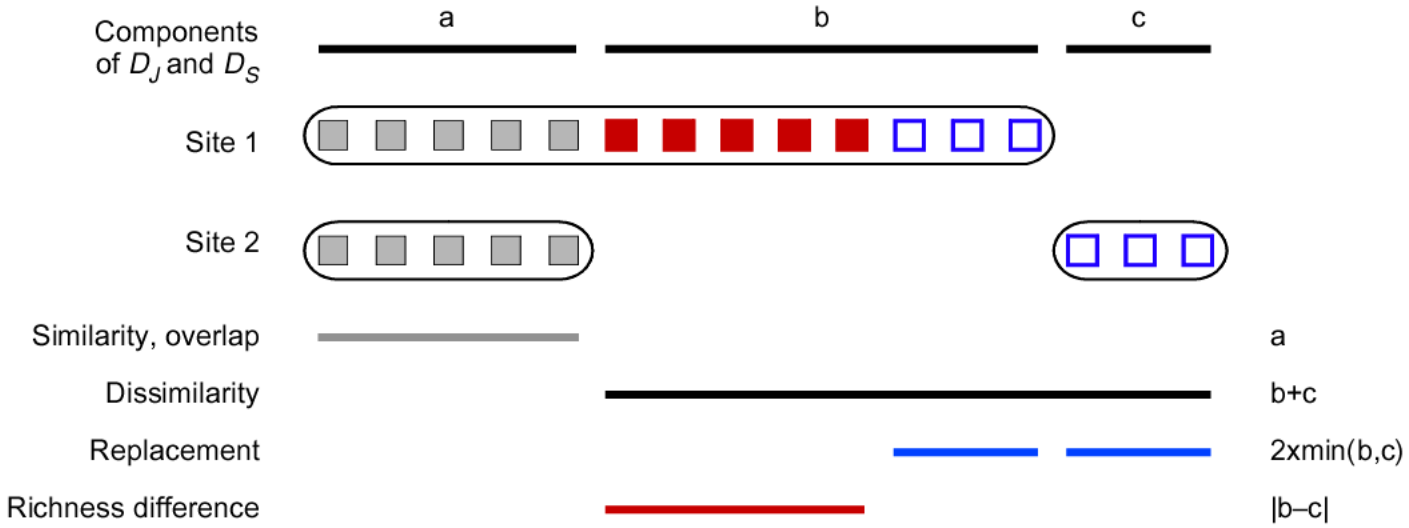
635 **FIGURE LEGENDS**

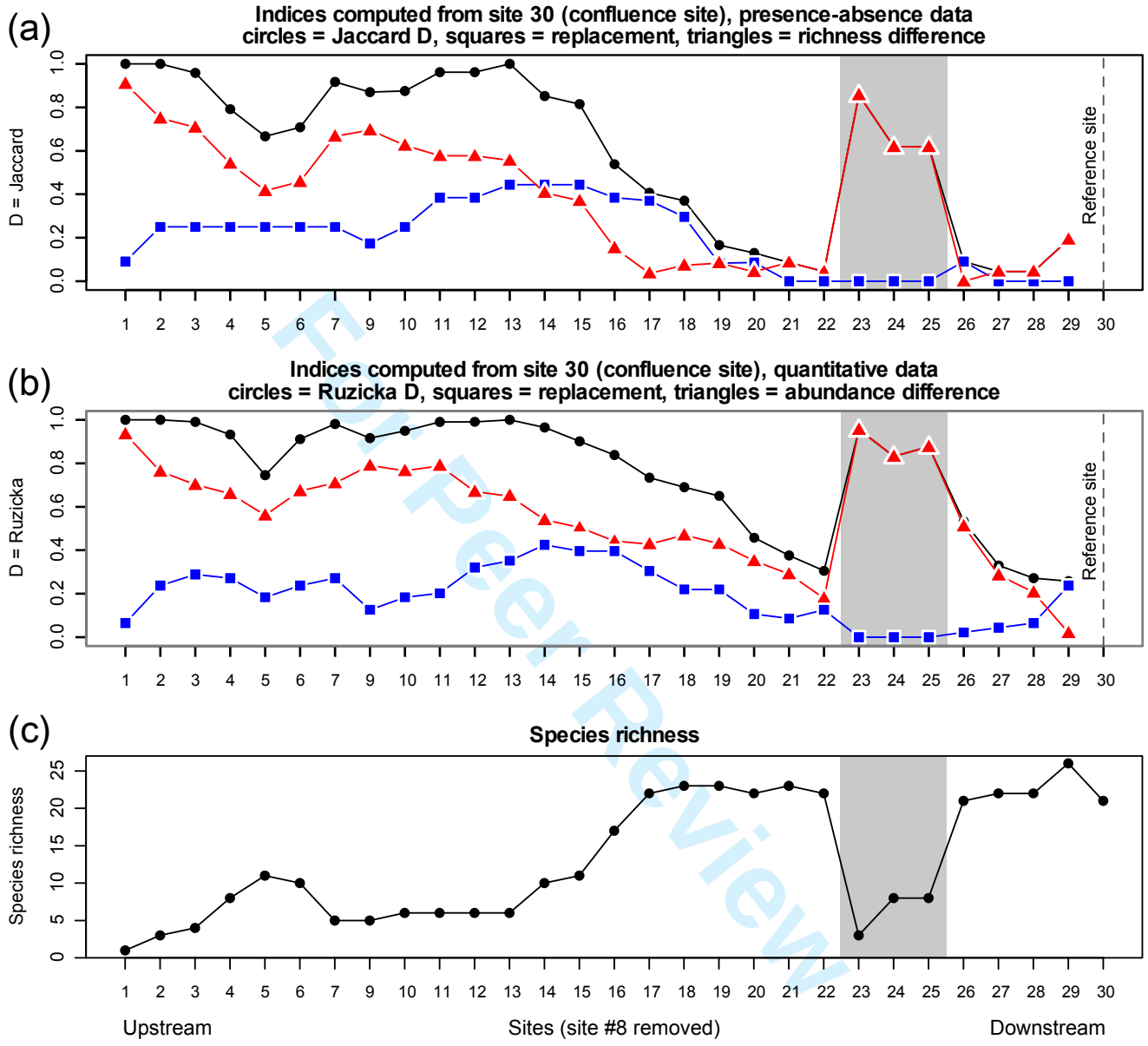
636 Figure 1 Comparison of species composition (16 species in total, squares) at two sites showing
637 the components (*a*, *b*, *c*) of the Jaccard (*J*) and Sørensen (*S*) dissimilarity coefficients for
638 presence-absence data, and how these components are used to assess replacement (species with
639 open squares) and richness difference (species with filled squares). Inspired from Williams
640 (1996), Podani & Schmera (2011) and Carvalho *et al.* (2013). (Online version in colour.)

641 Figure 2 (a) *Jaccard dissimilarity* (*D*, circles), *Replacement* (squares) and *Richness difference*
642 (triangles) indices for presence-absence data, comparing sites 1–29 with site 30. (b) Same, using
643 species abundance data and the Ružička dissimilarity index. (c) Species richness at the study
644 sites. The grey rectangles highlight sites 23 to 25 influenced by agricultural pollution. (Online
645 version in colour.)

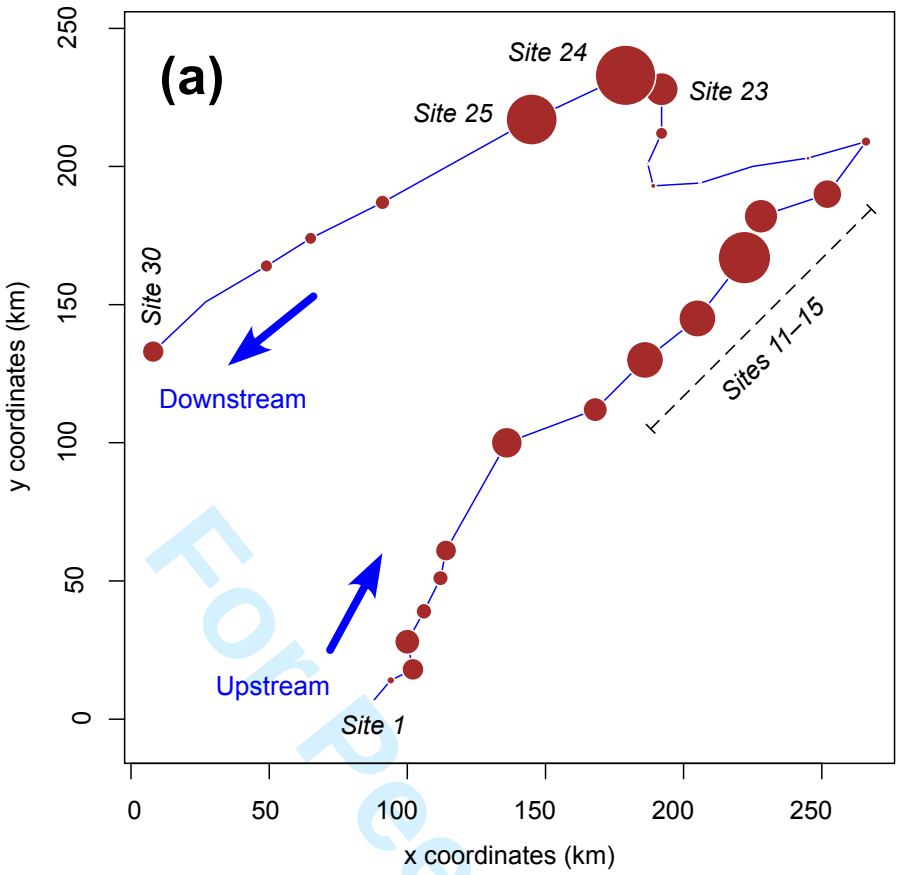
646 Figure 3 Schematic maps of Doubs River (line) showing the Podani-family (a) *Replacement*
647 *LCBD* ($Repl_{LCBD}$) and (b) *Richness difference LCBD* ($RichDiff_{LCBD}$) of the binary fish assemblage
648 data at the 29 study sites decomposing the Jaccard dissimilarity. Circle sizes are proportional to
649 LCBD values. The arrows indicate water flow direction. (Online version in colour.)

650 Figure 4 Triangular plots (simplices) of the relationships among the 406 pairs of sites for the
651 Doubs river fish data. Each point (black dot) represents a pair of sites. Its position is determined
652 by a triplet of values from the **S** = (1 – **D**) (similarity), **Repl** (replacement), **RichDiff** and **AbDiff**
653 (richness and abundance difference) matrices; each triplet sums to 1. Graphs are shown for all
654 algebraic forms of the Podani-family indices: (a) Jaccard, (b) Sørensen, (c) Ružička and
655 (d) percentage difference. The large central dot in each graph (blue in the online version) is the
656 centroid of the points; the smaller dots (blue in the online version) represent the mean values of
657 the *S*, *Repl* and *Rich* components.

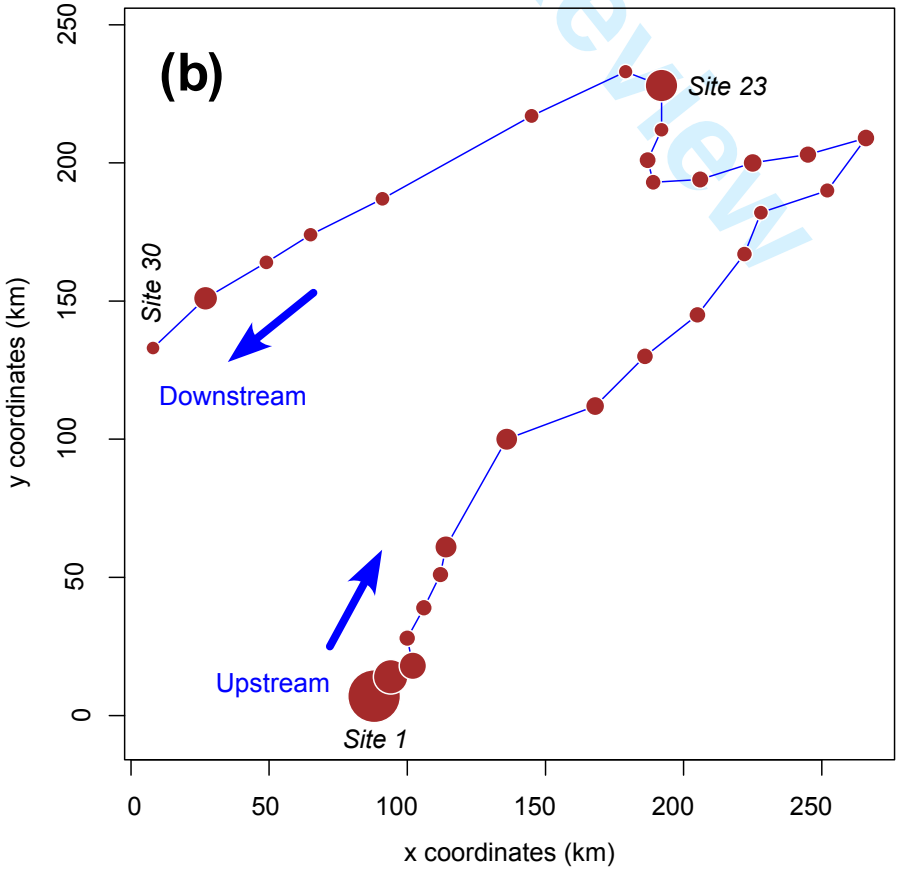

Components of D_J and D_S
a
b
c
Site 1
Site 2
Similarity, overlap
a
Dissimilarity
b+c
Replacement
2xmin(b,c)
Richness difference
|b–c|

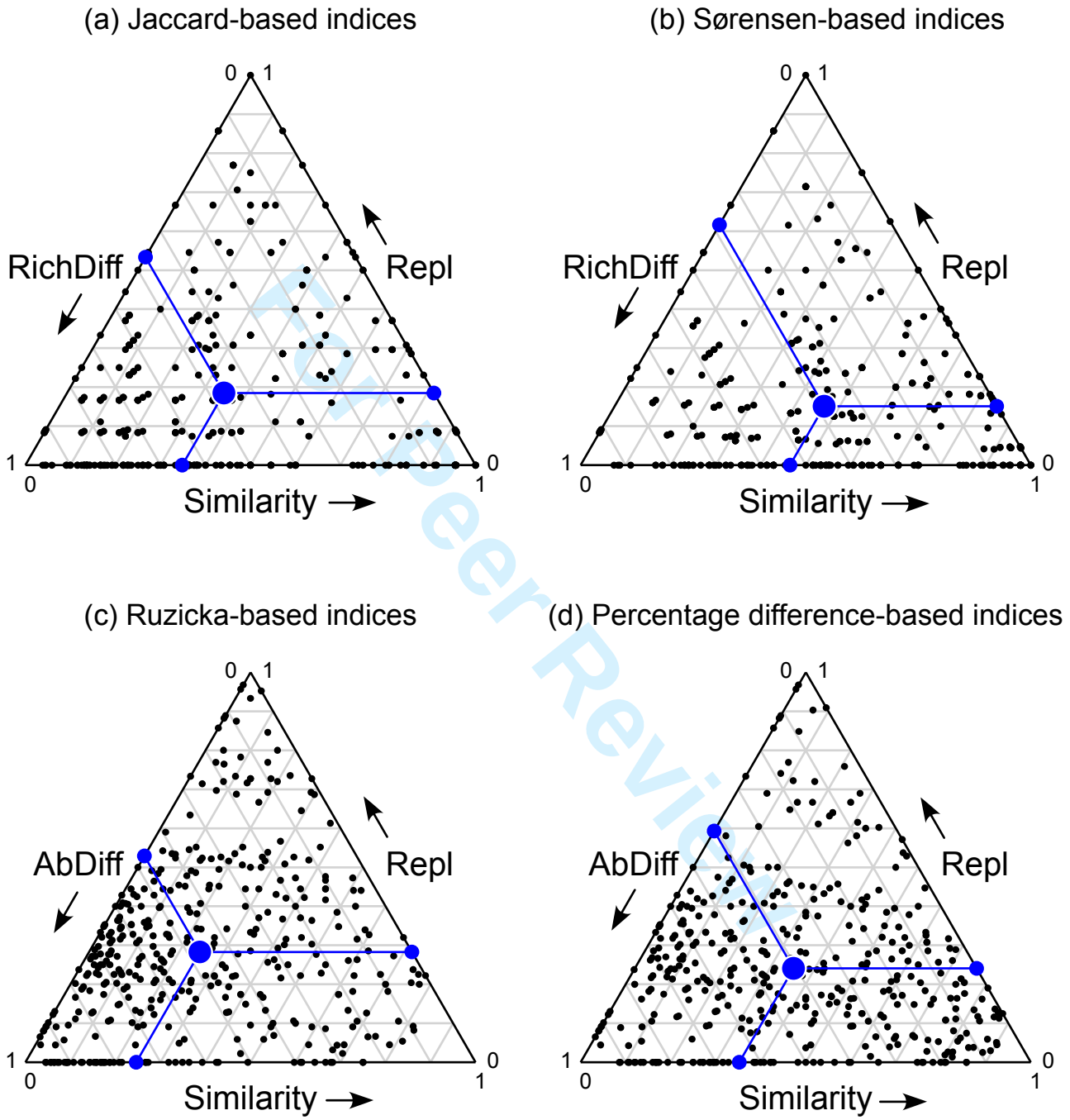


Map of *Replacement LCBD*



Map of *Richness difference LCBD*





Appendix to:

Legendre, P. (2014) Interpreting the replacement and richness difference components of beta diversity. *Global Ecology and Biogeography*, **23**, xxx–xxx.

Appendix S1

Replacement, richness difference and nestedness indices

1. Introduction

Whittaker (1960, 1972) described the alpha, beta and gamma diversity levels of natural communities. Alpha is local diversity, beta is spatial differentiation (or variation in species composition among sites), and gamma is regional diversity. Koleff *et al.* (2003) reviewed 24 beta diversity measures for species presence-absence data, reexpressed them in terms of the traditional *a*, *b* and *c* values used for the comparison of pairs of sites, and classified them according to domains of application. This appendix focuses on some of these indices (Table S1.1) that are now used to estimate species replacement and richness difference, and are the subject of heated comparisons and discussions in the literature.

For species presence-absence data, two sites are compared using a 2×2 contingency table crossing the observed communities. The table contains the following frequencies: a = number of species present at both sites, b = number of species present at site 1 but not at site 2, and c = number of species present at site 2 but not at site 1. The number of species present at neither site, d, is also found in the contingency table, but it is not used in the calculation of the indices.

The two most widely used dissimilarity coefficients for presence-absence data are the Jaccard (1908; $D_J = (b+c)/(a+b+c)$) and Sørensen (1948; $D_S = (b+c)/(2a+b+c)$) indices. These coefficients have different mathematical properties: D_J is a metric whereas D_S is a semimetric (Legendre & Legendre, 2012). These dissimilarities can be partitioned into two components, the portions due to species replacement (*Repl* indices) and to either richness (*RichDiff*) or abundance difference (*AbDiff*), or nestedness (*Nes*); the corresponding concepts are discussed in the introduction of the main paper. Different forms of these indices have been proposed and can be computed from the values a, b and c. Several papers appeared on the subject during the past years: Williams (1996), Lennon *et al.* (2001), Cardoso *et al.* (2009), Baselga (2010, 2012, 2013), Podani and Schmera (2011), Schmera & Podani (2011) and Carvalho *et al.* (2012, 2013) and Podani *et al.* (2013); index notations have also diverged among papers.

This brief historical account will focus on the contributions and indices that I consider the most important for users to understand how these indices are interrelated. The literature review is admittedly incomplete. It reflects personal choices from among the abundant literature published during the past few years on the subject.

Early indices

A first spatial turnover (or replacement) index was proposed by Williams (1996) who called it beta-3, meaning that it was the third modified form of Whittaker's (1972) famous β

diversity index, $\beta = \gamma/\alpha$, where γ is regional richness and α is the mean of the richness at individual sites. Williams applied the index to measure species turnover in small neighbourhoods of sampling units, e.g. quadrats. Koleff *et al.* (2003) labelled that index β_{-3} in their review paper on measures of beta diversity. Cardoso *et al.* (2009) modified the formula for β_{-3}, proposing to multiply the value by 2 to make it vary between 0 and 1; this is the current form of the β_{-3} index (Table S1.1).

A second replacement index was proposed by Lennon *et al.* (2001) who called it β_{sim}; "sim" refers to palaeontologist G. G. Simpson, although β_{sim} differs from the asymmetric index used by Simpson to compare paleontological faunas in his 1943 paper. In similarity form, this index was applied by Lennon *et al.* (2001) to compare the community at a site to its eight first neighbours in a regular checkerboard grid of sampling units and compute the mean of these comparisons. Baselga (2010) based his first decomposition of D_S on β_{sim} (section 2.2).

Sections 2 and 3 describe the *Repl, RichDiff* or *AbDiff*, and *Nes* indices available in the recent literature for, respectively, species presence-absence and abundance data. The main paper shows how these indices can be interpreted and related to ecosystem functioning. *Repl*, *RichDiff*, *AbDiff* and *Nes* are new abbreviations introduced here to make the relationships among the indices easier to understand.

2. Replacement and richness difference indices: presence-absence data

2.1. Podani family

The indices proposed by Podani & Schmera (2011) and Carvalho *et al.* (2012, 2013), that expand upon Williams' (1996) beta-3 index, are described first. For simplicity, these indices are referred to in this appendix as the Podani family. These authors suggested to estimate replacement ($Repl_J$) as $2\times\min(b,c)$ and richness difference ($Rich_J$) as $|b - c|$. These equations are justified by the test case illustrated in Fig. 1 of the main paper: three species of site 1 (represented by squares with positive-slope stripes, blue in the online version of the paper) are replaced by three species of site 2, whereas the richness difference (5 species represented by squares with negative-slope stripes, red in the online version, $|b - c|$) is the remainder of the dissimilarity ($b + c$) between the two sites. Because $2\times\min(b,c) = (b + c) - |b - c|$, the sum of the replacement and richness difference values is equal to ($b + c$), which is the numerator of the Jaccard dissimilarity coefficient (D_J, introduction section). Hence, if the replacement and richness difference indices are divided by the denominator of the Jaccard dissimilarity, ($a + b + c$), the resulting scaled indices have a minimum value of 0 and a maximum of 1 and they sum to the Jaccard dissimilarity. The calculations are broken down into steps in Table S1.2. The resulting indices were called *relativized species replacement* (R_{rel}) and *relativized richness difference* by Podani & Schmera (2011), and β_{-3} and β_{rich} by Carvalho *et al.* (2012, 2013).

Podani & Schmera (2011) also defined an index of nestedness: $N = a + |b - c|$ if $a > 0$ and $N = 0$ if $a = 0$. N can be scaled to N_{rel} by division by ($a + b + c$), which produces values in the range [0,1]. They clearly stated that nestedness is not the same as richness difference ($|b - c|$).

Note that (b+c) is also the numerator of the Sørensen dissimilarity coefficient (D_S, introduction section). Carvalho *et al.* (2013) indicated that the replacement and richness difference indices could be divided by the denominator of the Sørensen dissimilarity, ($2a$+b+c), although they did not find the resulting indices meaningful. Podani & Schmera's (2011) view was

that one meaningful way of making the *Repl* and *RichDiff* indices independent of total species richness was to divide them by (a+b+c); see sections 5 and 6 (below) for a discussion of dependence/independence of species richness. With the (2a+b+c) denominator, the resulting standardized $Repl_S$ and $Rich_S$ indices now sum to the Sørensen dissimilarity (Table S1.2). This is a second, mathematically valid way of constructing indices. It satisfies Carvalho *et al.*'s (2013) first criterion for an ecologically meaningful way of partitioning beta diversity (Appendix S2).

Consider the following binary data example, which contains both a replacement and a richness difference component:

	Species 1	Species 2	Species 3	Species 4	Species 5
Site 1	1	1	1	0	1
Site 2	1	1	0	1	0

For these data, $a = 2$, $b = 2$ and $c = 1$; $2\times\min(b,c) = 2$ (i.e. a species of Site 1 replaces a species of Site 2 and vice-versa); $|b–c| = 1$ (i.e. after elimination of the two species involved in replacement, say species 3 and 4, species 5 is the only one showing richness difference); (a+b+c) = 5, and (2a+b+c) = 7. Hence,

• the Jaccard-based indices (J) are

$$Repl_J = 2\times\min(b,c)/(a+b+c) \quad (R_{rel} \text{ in Podani \& Schmera 2011, } \beta_{-3} \text{ in Cardoso } et\ al. \text{ 2009}) \quad (1)$$

$$Rich_J = |b–c|/(a+b+c) \quad (D_{rel} \text{ in Podani \& Schmera 2011}) \quad (2)$$

For the example data, $Repl_J = 0.4$ and $Rich_J = 0.2$; these values sum to $D_J = (b+c)/(a+b+c) = 0.6$.

• the Sørensen-based indices (S) are

$$Repl_S = 2\times\min(b,c)/(2a+b+c) \quad (3)$$

$$Rich_S = |b–c|/(2a+b+c) \quad (4)$$

For the example data, $Repl_S = 0.2857$ and $Rich_S = 0.1429$; these values sum to $D_S = (b + c)/(2a + b + c) = 0.4286$ (rounded values). So in each case, *Repl* and *RichDiff* add up to D.

Construction of indices for presence-absence data in the Podani family is summarized in Table S1.2.

2.2. Baselga family

Baselga (2010, 2012) described replacement and nestedness indices for species presence-absence data, calling the latter the *nestedness component* of dissimilarity indices. Baselga (2010) chose the β_{sim} index of Lennon *et al.* (2001) to estimate replacement because, according to these authors, that index clearly accounts for replacement (spatial turnover) and is only weakly correlated to the difference in species richness of the sites (but see section 6). He used β_{sim} in a more general way than Lennon *et al.* (2001) (see Introduction section), computing matrices of β_{sim} indices among all pairs of sites with the following formula:

$$Repl_{BS} = \beta_{sim} = \frac{\min(b,c)}{a + \min(b,c)} \quad (5)$$

($Repl_{BS}$ in Table S1.3). Because the Sørensen dissimilarity D_S accounts for both replacement and nestedness, Baselga (2010) proposed to subtract β_{sim} from the Sørensen dissimilarity D_S to estimate the nestedness-resultant dissimilarity, or amount of dissimilarity derived from the nestedness component:

$$Nes_{BS} = \beta_{nes} = D_S - \beta_{sim} = \frac{|b-c|}{2a+b+c} \times \frac{a}{a+\min(b,c)} \quad (6)$$

(Nes_{BS} in Table S1.3). Baselga (2010) offered the following interpretation for the algebraic form of that coefficient, whose meaning is not immediately clear: it is a measure of richness difference, $|b - c|$, divided by the denominator of D_S; this part is also the Podani-family index $Rich_S$. The ratio is then multiplied (scaled) by the Simpson similarity ($1 - \beta_{sim}$); if there are no species in common (i.e., if $a = 0$), then β_{nes} is zero. This is a way of incorporating the constraint that nestedness is larger than 0 only if $a > 0$. That constraint is also found, in another algebraic form, in Podani & Schmera's (2011) nestedness index described above.

For the example data, $Repl_{BS}$ = 0.3333 and Nes_{BS} = 0.0952; these values sum to D_S = $(b+c)/(2a+b+c)$ = 0.4286 (rounded values). The product extracts the difference in richness caused by nestedness from other types of richness differences (Baselga, 2010). Baselga did not consider his nestedness index to be a measure of richness difference.

Baselga (2012) proposed an alternative pair of indices designed to sum to the Jaccard dissimilarity D_J. He called the new turnover index β_{jtu} ($Repl_{BJ}$ in Table S1.3) and the new nestedness index β_{jne} (Nes_{BJ} in Table S1.3):

$$Repl_{BJ} = \beta_{jtu} = \frac{2\min(b,c)}{a+2\min(b,c)} \quad (7)$$

$$Nes_{BJ} = \beta_{jne} = D_J - \beta_{jtu} = \frac{|b-c|}{a+b+c} \times \frac{a}{a+2\min(b,c)} \quad (8)$$

For the example data, $Repl_{BJ}$ = 0.5 and Nes_{BJ} = 0.1; the sum of these values is D_J = $(b + c)/(a + b + c)$ = 0.6.

Construction of indices for presence-absence data in the Baselga family is summarized in Table S1.3.

3. Replacement and abundance difference indices: quantitative data

For presence-absence data, two sites that have the exact same species complement but differ in species abundances have replacement and richness difference values of 0 for coefficients D_S or D_J because $b = c = 0$. These 0 values are not informative of the quantitative differences that exist between the sites. The quantitative forms of the indices, described in this section, provide finer, more interpretable results when species abundance data are trustworthy.

Podani *et al.* (2013) and Baselga (2013) described ways of estimating replacement and richness difference based on species abundances, related to quantitative dissimilarities that are extensions of the Jaccard and Sørensen indices. Consider the following example:

	Species 1	Species 2	Species 3	Species 4	Species 5
Site 1	7	3	5	0	6
Site 2	2	4	0	3	0

Tamas *et al.* (2011) proposed the following notation to construct indices based on abundance data: A designates the sum of the minimum abundances of the various species, each minimum being the abundance at the site where the species is the rarest (this quantity is called W in Legendre & Legendre 2012, Chapter 7); B is the sum of abundances at site 1 minus A; and C is the sum of abundances at site 2 minus A. The symbols have meanings similar to the lower-case letters a, b and c used in the description of presence-absence indices: A is the sum of intersections (or the minima) of the abundances of the species at the two sites under comparison, B and C are the site-specific abundance complements. For the example data, $A = 5$, $B = 16$ and $C = 4$.

3.1. Podani family

In the context of quantitative community data, computation of the indices is individual-based instead of species-based. *Replacement* is easier to compute than to explain. For the site with the smallest total abundance (site 2 in the example), consider only the species (#2 and #4) that have larger abundances at that site than at site 1. Replacement refers to the fact that the individuals (4) that exceed the number of individuals of these species at site 1 are replaced, at site 1, by the same number (4) of individuals of different species. Replacement is computed as $\min(B,C)$; this value is multiplied by 2 because the *replaced* and *replacing* individuals are both counted. Hence, replacement difference is $2\times\min(B,C)$, which is $2\times4 = 8$ for the example; computing this index for binary data produces the binary form of replacement, $2\times\min(b,c) = 2$.

Abundance difference (abbreviated *AbDiff*) refers to the individuals that are not involved in replacement, that is, for the example, $21 - 4 = 17$ individuals at site 1 and $9 - 4 = 5$ individuals at site 2. The absolute value difference of these two values, 12, is the abundance difference. *AbDiff* can also be computed as the absolute value of the difference between B and C, which is $|B - C| = 12$. Again, computing this index for binary data produces the binary form of richness difference, $|b - c| = 1$.

What should we choose for denominator? The quantitative dissimilarity index of Ružička (1958) is one of the quantitative forms of the Jaccard index. Its formula is

$$D_R = (B + C)/(A + B + C) \tag{9a}$$

This coefficient can also be written as:

$$D_R = 1 - \left(\sum_{j=1}^{p} \min(y_{1j}, y_{2j}) \Big/ \sum_{j=1}^{p} \max(y_{1j}, y_{2j}) \right) \tag{9b}$$

D_R computed for presence-absence data produces the Jaccard dissimilarity. As shown in Table S1.2, $2\times\min(B,C)$ (replacement) plus $|B - C|$ (abundance difference) is equal to $(B + C)$. *Note* – There are other quantitative dissimilarity functions that correspond to the Jaccard index when applied to presence-absence data. Four are listed in Legendre & De Cáceres (2013, Table 1). However, none of those has the quantity $(B + C)$ as its numerator, quantity that is the

sum of the replacement and abundance difference components; hence the choice of the Ružička index for the decomposition proposed by Podani *et al.* (2013).

Now if we apply the denominator of the D_R coefficient, $(A + B + C)$ (eq. 9a), to the replacement and abundance difference numerators, we obtain $Repl_R$ and $Rich_R$ indices (Table S1.2):

$$Repl_R = 2 \times \min(B,C)/(A + B + C) \tag{10}$$

$$\text{and } AbDiff_R = |B - C|/(A + B + C) \tag{11}$$

Their sum is D_R (Table S1.2). For the example data, $Repl_R = 0.32$ and $AbDiff_R = 0.48$, which sum to $D_R = 0.80$. Podani *et al.* (2013) described $Repl_R$ as the *relativized abundance replacement index*, ${}^aR_{\mathrm{rel}(jk)}$, and $AbDiff_R$ as the *relativized abundance difference*, ${}^aD_{\mathrm{rel}(jk)}$. For the example data, $Repl_R = 0.32$ and $AbDiff_R = 0.48$, which sum to $D_R = 0.80$.

The quantitative dissimilarity that corresponds to the Sørensen index is the percentage difference ($D_{\%diff}$, Odum, 1950), incorrectly referred to by some authors as the Bray-Curtis index; see Legendre & De Cáceres (2013), footnote of their Table 1, about that story. The formula of that well-known dissimilarity index is

$$D_{\%diff} = (B + C)/(2A + B + C) \tag{12a}$$

Note that $D_{\%diff}$ has the same numerator as D_R. The formula can also be written as (Odum, 1950):

$$D_{\%diff} = \sum_{j=1}^{p} |y_{1j} - y_{2j}| \Big/ \sum_{j=1}^{p} (y_{1j} + y_{2j}) = \sum_{j=1}^{p} |y_{1j} - y_{2j}| \Big/ (2A + B + C) \tag{12b}$$

where $\mathbf{y}_1$ and $\mathbf{y}_2$ are the two site vectors under comparison and p is the number of species in the data matrix (Legendre & Legendre, 2012, eq. 7.58). As mentioned above, the numerators of replacement and abundance difference (eqs, 10 and 11) sum to $(B+C)$, which is also the numerator of the $D_{\%diff}$ index. Hence we can apply the denominator of the $D_{\%diff}$ coefficient to these numerators to obtain $Repl_{\%diff}$ and $AbDiff_{\%diff}$ indices that sum to $D_{\%diff}$ (Table S1.2):

$$Repl_{\%diff} = 2 \times \min(B,C)/(2A + B + C) \tag{13}$$

$$\text{and } AbDiff_{\%diff} = |B - C|/(2A + B + C) \tag{14}$$

The $Repl_{\%diff}$ and $AbDiff_{\%diff}$ indices do not seem to have been described in the literature yet and are thus new, although they stem from the same logic as the Podani *et al.* (2013) indices for quantitative data (eqs. 10 and 11). For the example data, $Repl_{\%diff} = 0.2667$ and $AbDiff_{\%diff} = 0.4000$, which sum to $D_{\%diff} = 0.6667$.

Construction of indices for abundance data in the Podani family is summarized in Table S1.2.

3.2. Baselga family

Baselga (2013) described replacement and nestedness indices for species abundance data that sum to the percentage difference and correspond to the presence-absence indices described in his 2010 paper. He called the former *balanced variation component* and the latter *abundance gradient component* of dissimilarity indices. They are obtained by replacing the *a*, *b* and *c*

components of the indices for presence-absence data (eqs. 5 and 6) by the quantities A, B and C defined above, which produces the following equations:

$$Repl_{B\%diff} = \frac{\min(B,C)}{A + \min(B,C)} \tag{15}$$

$$\text{and } Nes_{B\%diff} = \frac{|B - C|}{2A + B + C} \times \frac{A}{A + \min(B,C)} \tag{16}$$

For the example data, $Repl_{B\%diff} = 0.4444$ and $Nes_{B\%diff} = 0.2222$, which sum to $D_{\%diff} = 0.6667$.

Replication and nestedness indices that sum to the Ružička dissimilarity can now be described. The $Repl_{BR}$ and Nes_{BR} indices were not described by Baselga and are thus new.

$$Repl_{BR} = \frac{2\min(B,C)}{A + 2\min(B,C)} \tag{17}$$

$$\text{and } Nes_{BR} = \frac{|B - C|}{A + B + C} \times \frac{A}{A + 2\min(B,C)} \tag{18}$$

For the example data, $Repl_{BR} = 0.6154$ and $Nes_{BR} = 0.1846$, which sum to $D_R = 0.8000$.

Construction of indices for abundance data in the Baselga family is summarized in Table S1.3.

All indices decomposing dissimilarities in the Podani and Baselga families, described above, have the property that when the quantitative indices are computed on presence-absence data, they produce the same results as the binary versions.

4. Comparison of indices: which set of indices should one use?

During the past few years, arguments have been presented in favour or against each family of indices. Here is a short and possibly incomplete review of these exchanges.

4.1. Richness difference versus nestedness

After the publication by Baselga (2010) of a first pair of indices (β_{sim} and β_{nes}) decomposing D_S, Schmera & Podani (2011) argued that β_{nes} was simply the arithmetic difference between D_S and β_{sim}, without any connection to an ecological measure of species replacement or nestedness. They then summarised the partitioning of D_J into R_{rel} ($Repl_J$) and D_{rel} ($Rich_J$) described in Podani & Schmera (2011), where each component has clear ecological meaning, and they illustrated the differences between their indices and Baselga's using simulated and real ecological data. Baselga had explained, however, that β_{nes} was a *nestedness* index, whereas D_{rel} was defined by Podani & Schmera (2011) as a *richness difference* index. Podani & Schmera had recognized the difference between the two types of indices and defined a *nestedness* index (N) associated with their R_{rel} ($Repl_J$) and D_{rel} ($Rich_J$) indices in their 2011 paper; see section 2.1

Baselga explained as follows the intricate formula of the nestedness index β_{nes}, which is the product of two components: "This product is needed to separate differences in richness caused by nestedness from other differences in richness" (Baselga 2010, p. 138). The difference between richness difference and nestedness is emphasized again in subsection 4.3 (below).

The Podani and Baselga families of decompositions have different objectives. In the Baselga family, dissimilarities are decomposed into turnover (or replacement) and nestedness components; the latter indices do not measure richness difference. In the Podani family, dissimilarities are decomposed into replacement and richness difference components; nestedness indices have also been proposed, but they are not an additive component of the dissimilarity indices. The message here is that these families of decompositions are not intended for the same purposes. They may be used to fulfil different objectives, and they may bring out complementary information when applied to the same data.

Among other criticisms that were voiced, Almeida-Neto (2012) criticized Baselga's β_{nes} on the ground that it did not behave like a true nestedness index because it did not satisfy the following conditions for selected simulated data: (1) it did not always increase when nestedness increased, (2) it should not have varied when nestedness remained constant, and (3) it should not have yielded positive values when there was no nestedness in the data.

4.2. Over- or under-estimation of species replacement

The β_{sim} index of replacement ($Repl_{BS}$) of Baselga was criticized by Carvalho *et al.* (2012) on the ground that it overestimated species replacement. Conversely, Baselga (2012) criticized Podani's β_{-3} index of replacement ($Repl_J$) on the ground that it underestimated species replacement. *The difference in point of view is due to the denominator that serves as the reference in each index*; see formulas in Table S1.1. Podani & Schmera (2011) defined species replacement as $R = 2\times\min(b,c)$ without a denominator, indicating that it is the numerators of the proposed functions ($Repl_J$, $Repl_S$, $Repl_{BJ}$, $Repl_{BS}$, all having the same numerator) that estimate replacement. One can then use the denominator of one's choice, depending on the purpose of the study. In the two families of indices, the replacement and richness difference (or nestedness) indices sum to dissimilarity measures (D_J, D_S, D_R, $D_{\%diff}$) that are all appropriate for beta diversity assessment, following the criteria of Legendre & De Cáceres (2013).

4.3. Monotonicity

Carvalho *et al.* (2013) suggested that the replacement and richness difference indices should increase monotonically with increases in the amounts of replacement and richness difference in data. They carried out simulations using species presence-absence data. They found that the indices of Podani & Schmera (2011) that decompose D_J increased monotonically. That was not the case for Baselga's (2010, 2012) β_{jne} (Nes_{BJ}) and β_{nes} (Nes_{BS}), which are nestedness indices; results of simulations similar to those of Carvalho *et al.* (2013) are shown in Appendix S2, Fig. S2.2. I also computed the relativized nestedness index of Podani & Schmera (2011) for the simulated data; values of that index *decreased* monotonically with the amounts of replacement and richness difference. In any case, this is not a problem because these indices measure nestedness, not richness difference. According to Podani & Schmera (2011), the quantity $|b - c|$, which measures richness difference, does not fully reflect nestedness but only a contribution to it. Hence there is no constraining reason why nestedness indices should increase monotonically with increases in replacement and richness difference in data series.

I carried out further simulations to complete the picture. First, I verified that the indices of the Podani family that decompose D_S obeyed the monotonicity condition of Carvalho *et al.* (2013) (Appendix S2, Fig. S2.1). Then I carried out simulations using quantitative data to check that the quantitative indices developed by Podani *et al.* (2013) obeyed the monotonicity

condition. The results showed that the *Repl* and *AbDiff* forms of the Ružička and percentage difference indices reacted monotonically to gradual changes in quantitative community composition (Appendix S2, Fig. S2.3).

4.4. Metric and Euclidean properties

Table S1.4 shows the metric and Euclidean properties of the dissimilarity coefficients (D_J, D_R, D_S, $D_{\%diff}$) and the corresponding *Repl* and *RichDiff / AbDiff / Nes* indices of the Podani (Table S1.4a) and Baselga (Table S1.4b) families. These properties are useful to select a coefficient to produce an ordination. The Jaccard and Ružička dissimilarities are metric whereas the Sørensen and percentage difference dissimilarities are semimetric. In all four cases, the dissimilarity matrix **D** is not Euclidean but $\mathbf{D}^{(0.5)} = [D_{hi}^{0.5}]$ is Euclidean (Legendre & Legendre, 2012, Table 7.2), meaning that the data points can be fully represented in Euclidean space by principal coordinate analysis (PCoA) without production of negative eigenvalues and complex ordination axes. In simple terms, they are fully appropriate for ordination by PCoA.

The Podani-family *RichDiff* indices decomposing the Sørensen ($RichDiff_S$) and percentage difference ($AbDiff_{\%diff}$) indices present clear advantages for ordination over the $RichDiff_J$, $AbDiff_R$ and *Nes* indices because the $\mathbf{RichDiff}_S$, $\mathbf{AbDiff}_{\%diff}$, $\mathbf{RichDiff}_S^{(0.5)}$ and $\mathbf{AbDiff}_{\%diff}^{(0.5)}$ matrices are metric and Euclidean. All other forms of indices are not metric nor Euclidean, so they should be subjected to corrections for negative eigenvalues when used to produce ordinations by principal coordinate analysis (Legendre & Legendre, 2012, section 9.3.4).

4.5. Property P5 of Legendre & De Cáceres

Leprieur & Oikonomou (2014) argued that the Podani & Schmera (2011) measure of replacement β_{-3} lacked an important property that a dissimilarity measure should have when it is used in beta diversity studies, namely property P5 of Legendre & De Cáceres (2013), which states that *sites without species in common should have the largest dissimilarity*. For Legendre & De Cáceres (2013), that was indeed one of the important properties that should be fulfilled by dissimilarity coefficients used for beta diversity assessment. When we consider the pair $Repl_J$ and $Rich_J$, or $Repl_S$ and $Rich_S$, however, these indices are not used to measure beta diversity as a whole; they decompose the D_J and D_S dissimilarities that can be used to estimate beta diversity. It is D_J and D_S that should have property P5, not their components, and indeed they do. Replacement and richness difference indices are not used to carry out the same types of data analyses as dissimilarity indices; they have different purposes.

5. Artificial numerical examples

Leprieur & Oikonomou (2014) created an artificial data set to illustrate the fact that β_{-3} ($Repl_J$) was correlated with differences in species richness between sites, whereas $Repl_{BJ}$, (β_{jtu}) and $Repl_{BS}$ (β_{sim}) were not. That example is revisited here (Table S1.5). Only the first five sites of their example are used here. The data comparing the community of site A to those of sites B-E are presented in the form of the counts *a*, *b* and *c* used in the coefficient formulas (Table S1.1).

The replacement coefficients proposed by Baselga (2010, 2012), $Repl_{BJ}$ and $Repl_{BS}$, are not correlated to richness difference in this example but $Repl_J$ and $Repl_S$ are negatively correlated to unscaled richness difference, *Rich.diff*.

Table S1.5. Coefficients and indices computed for the first artificial example. *Rich.diff*: unscaled richness difference computed as $|b - c|$ or $|r_1 - r_2|$. r_1, r_2, p_1, p_2: see text. Two dissimilarities: D_J (Jaccard) and D_S (Sørensen). Four replacement indices: $Repl_{BJ}$, $Repl_{BS}$, $Repl_J$ and $Repl_S$; four richness difference indices: Nes_{BJ}, Nes_{BS}, $RichDiff_J$ and $RichDiff_S$.

(a) Coefficients

Site pair	a	b	c	r_1	p_1	r_2	p_2	*Rich.diff*
A-B	10	10	10	20	1/2	20	1/2	0
A-C	10	10	20	20	1/2	30	1/3	10
A-D	10	10	30	20	1/2	40	1/4	20
A-E	10	10	40	20	1/2	50	1/5	30

(b) Dissimilarity, replacement and richness difference indices

Site pair	D_J	D_S	$Repl_{BJ}$	$Repl_{BS}$	$Repl_J$	$Repl_S$	Nes_{BJ}	Nes_{BS}	$RichDiff_J$	$RichDiff_S$
A-B	0.667	0.500	0.667	0.500	0.667	0.500	0.000	0.000	0.000	0.000
A-C	0.750	0.600	0.667	0.500	0.500	0.400	0.083	0.100	0.250	0.200
A-D	0.800	0.667	0.667	0.500	0.400	0.333	0.133	0.167	0.400	0.333
A-E	0.833	0.714	0.667	0.500	0.333	0.286	0.167	0.214	0.500	0.429

Legendre & De Cáceres (2013, Appendix S3, property P10) showed, however, that D_J and D_S do not depend on the species richness, r_1 and r_2, in the two sampling units that are compared; $r_1 = a + b$ and $r_2 = a + c$. For proof, they showed that these dissimilarity coefficients can be expressed by formulas that do not contain r_1 and r_2 but only p_1 and p_2, where p_1 is the proportion of shared species in the first sampling unit, $p_1 = a/r_1$; likewise, p_2 is the proportion of shared species in the second sampling unit, $p_2 = a/r_2$. They considered this property P10 important for dissimilarity coefficients used for studying beta diversity; this property facilitates (although it does not ensure) the comparability of beta diversity values obtained for sampling units having different sizes or sampled using different efforts. That property can be demonstrated for all coefficients computed in Table S1.5 (see section 6 of the present Appendix): these coefficients are all invariant to the number of species in each sampling unit since they can be expressed by equations that only contain p_1 and p_2.

The example in Table S1.5 had been tailored to support the claim of dependence of $Repl_J$ and $Repl_S$ on species richness r_2; it does not show what happens when p_1 and p_2 remain constant across site comparisons. When they do, the ten coefficients produce constant values (i.e. they do

not change their values) when one changes r_1 and r_2 to any positive integer value under the constraint that $r_1p_1 = r_2p_2$.

To illustrate that property, a second example was generated using four pairs of unrelated sites (Table S1.6). In this example, the dissimilarity and replacement indices do not vary with changes in unscaled richness difference (*Rich.diff* = $|b - c|$ or $|r_1 - r_2|$, last column in section (a) of the table), provided that $r_1p_1 = r_2p_2 = a$.

Table S1.6. Coefficients and indices computed for the second artificial example.

(a) Coefficients

Site pair	*a*	*b*	*c*	r_1	p_1	r_2	p_2	*Rich.diff*
A-B	3	9	3	12	0.25	6	0.5	6
C-D	5	15	5	20	0.25	10	0.5	10
E-F	10	30	10	40	0.25	20	0.5	20
G-H	25	75	25	100	0.25	50	0.5	50

(b) Dissimilarity, replacement and richness difference indices

Site pair	D_J	D_S	$Repl_{BJ}$	$Repl_{BS}$	$Repl_J$	$Repl_S$	Nes_{BJ}	Nes_{BS}	$RichDiff_J$	$RichDiff_S$
A-B	0.800	0.667	0.667	0.500	0.400	0.333	0.133	0.167	0.400	0.333
C-D	0.800	0.667	0.667	0.500	0.400	0.333	0.133	0.167	0.400	0.333
E-F	0.800	0.667	0.667	0.500	0.400	0.333	0.133	0.167	0.400	0.333
G-H	0.800	0.667	0.667	0.500	0.400	0.333	0.133	0.167	0.400	0.333

In Table S1.6, the ten indices have identical values for all pairs of sites and they are totally unrelated to the unscaled richness difference between sites (*Rich.diff* column) since they are invariant. When the values of p_1 and p_2 do not change, the values of all indices do not change. The indices are related in highly nonlinear ways to the values of p_1 and p_2, as shown by the equations in section 6. Data where the condition ($r_1p_1 = r_2p_2 = a$) is violated may, of course, show correlations of *Repl* indices with richness difference on a case-by-case basis.

6. Proofs of property P10 for individual indices

Property P10 of Legendre & De Cáceres (2013) states the following:

Indices for binary (species presence-absence) data that have this property do not directly depend on the number of species in the compared sampling units.

Does the index 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? Indices having property P10 do not change their values with changes in numbers of species r_1 and r_2, in two sampling units $\mathbf{x}_1$ and $\mathbf{x}_2$ that are compared, provided that $r_1p_1 = r_2p_2$.

The following equalities and notations are used in the demonstrations:

• the total number of species (richness) in $\mathbf{x}_1$ and $\mathbf{x}_2$ are noted $r_1 = a + b$ and $r_2 = a + c$;
• the proportion of shared species with respect to the total richness at each site is $p_1 = a/r_1 = a/(a + b)$ and $p_2 = a/r_2 = a/(a + c)$;
• hence $a = r_1p_1 = r_2p_2$, $b = r_1(1 – p_1)$ and $c = r_2(1 – p_2)$.

The minimum of b and c, which we will need in the demonstrations, can be computed as $\min(b,c) = ((b + c) – \text{abs}(b – c))/2 = (r_1(1 – p_1) + r_2(1 – p_2) – | r_1(1 – p_1) – r_2(1 – p_2)|)/2$

Also, because $r_1p_1 = r_2p_2 = a$, we have

• if $r_1 > r_2$ then $p_1 < p_2$; else if $r_1 < r_2$ then $p_1 > p_2$;

• $\min(b,c) = (r_1 + r_2 – 2r_1p_1 – |r_1 – r_2|)/2 = (r_1 + r_2 – 2r_2p_2 – |r_1 – r_2|)/2$.

For the Jaccard and Sørensen similarity indices, demonstrations are provided in Appendix S3 of Legendre & De Cáceres (2013). The equations are reported again here, transformed into dissimilarities D_J and D_S.

Indices that have property P10 may still be empirically correlated (linearly or not) with differences in species richness for particular data sets.

(1) Jaccard dissimilarity

$$D_J = 1 - \frac{1}{(1/p_1) + (1/p_2) - 1} = \frac{p_1 + p_2 - 2p_1p_2}{p_1 + p_2 - p_1p_2}$$

(2) Sørensen dissimilarity

$$D_S = 1 - \frac{2}{(1/p_1) + (1/p_2)} = \frac{p_1 + p_2 - 2p_1p_2}{p_1 + p_2}$$

These two dissimilarity coefficients are thus independent of r_1 and r_2. The dissimilarity functions recommended in that paper for beta diversity studies all have property P10, but some other dissimilarity coefficients do not.

Following are the demonstrations for the replacement and richness difference indices, which all have property P10.

(3) β_{-3} or $Repl_J$ (Williams 1996, Podani & Schmera 2011)

$$\beta_{-3} = \frac{2\min(b,c)}{a + b + c} = \frac{r_1 + r_2 - 2r_1p_1 - |r_1 - r_2|}{r_1p_1 + r_1(1 - p_1) + r_2(1 - p_2)} = \frac{r_1 + r_2 - |r_1 - r_2| - 2r_1p_1}{r_1 + r_2 - r_2p_2}$$

If $r_1 > r_2$, and considering that $r_1 p_1 = r_2 p_2$,

$$\beta_{-3} = \frac{r_1 + r_2 - r_1 + r_2 - 2r_2 p_2}{r_1 + r_2 - r_2 p_2} = \frac{2r_2(1 - p_2)}{r_1 + r_2 - r_2 p_2}$$

Since $r_1 = r_2\, p_2/p_1$,

$$\beta_{-3} = \frac{2r_2(1 - p_2)}{\frac{r_2 p_2}{p_1} + r_2 - r_2 p_2} = \frac{2r_2(1 - p_2)}{\frac{r_2 p_2}{p_1} + \frac{r_2 p_1}{p_1} - \frac{r_2 p_1 p_2}{p_1}} = \frac{2p_1(1 - p_2)}{p_1 + p_2 - p_1 p_2}$$

If $r_1 < r_2$,

$$\beta_{-3} = \frac{r_1 + r_2 + r_1 - r_2 - 2r_1 p_1}{r_1 + r_2 - r_2 p_2} = \frac{2r_1(1 - p_1)}{r_1 + r_2 - r_2 p_2}$$

Since $r_2 = r_1 p_1/p_2$,

$$\beta_{-3} = \frac{2r_1(1 - p_1)}{r_1 + \frac{r_1 p_1}{p_2} - \frac{r_1 p_1 p_2}{p_2}} = \frac{2r_1(1 - p_1)}{\frac{r_1 p_2}{p_2} + \frac{r_1 p_1}{p_2} - \frac{r_2 p_1 p_2}{p_2}} = \frac{2p_2(1 - p_1)}{p_1 + p_2 - p_1 p_2}$$

Hence, considering that if $r_1 > r_2$ then $p_1 < p_2$, or if $r_1 < r_2$ then $p_1 > p_2$,

$\beta_{-3} = \dfrac{2\min(p_1, p_2)(1 - \max(p_1, p_2)}{p_1 + p_2 - p_1 p_2}$, which is independent of r_1 and r_2.

(4) *Repl*S (Carvalho *et al.* 2013)

$$Repl_S = \frac{2\min(b,c)}{2a + b + c} = \frac{r_1 + r_2 - 2r_1 p_1 - |r_1 - r_2|}{2r_1 p_1 + r_1(1 - p_1) + r_2(1 - p_2)} = \frac{r_1 + r_2 - |r_1 - r_2| - 2r_1 p_1}{r_1 + r_2 + r_1 p_1 - r_2 p_2}$$

If $r_1 > r_2$, and considering that $r_1 p_1 = r_2 p_2$,

$$Repl_S = \frac{r_1 + r_2 - r_1 + r_2 - 2r_2 p_2}{r_1 + r_2} = \frac{2r_2(1 - p_2)}{r_1 + r_2}$$

Since $r_1 = r_2p_2/p_1$,

$$Repl_S = \frac{2r_2(1-p_2)}{\frac{r_2p_2}{p_1}+r_2} = \frac{2r_2(1-p_2)}{\frac{r_2p_2}{p_1}+\frac{r_2p_1}{p_1}} = \frac{2p_1(1-p_2)}{p_1+p_2}$$

If $r_1 < r_2$, and considering that $r_1p_1 = r_2p_2$,

$$Repl_S = \frac{r_1+r_2+r_1-r_2-2r_1p_1}{r_1+r_2} = \frac{2r_1(1-p_1)}{r_1+r_2}$$

Since $r_2 = r_1p_1/p_2$,

$$Repl_S = \frac{2r_1(1-p_1)}{r_1+\frac{r_1p_1}{p_2}} = \frac{2r_1(1-p_1)}{\frac{r_1p_2}{p_2}+\frac{r_1p_1}{p_2}} = \frac{2p_2(1-p_1)}{p_1+p_2}$$

Hence, considering that if $r_1 > r_2$ then $p_1 < p_2$, or if $r_1 < r_2$ then $p_1 > p_2$,

$Repl_S = \frac{2\min(p_1,p_2)(1-\max(p_1,p_2)}{p_1+p_2}$, which is independent of r_1 and r_2.

(5) β_{jtu} (Baselga 2012) (*$Repl_{B,J}$* in this Appendix)

$$\beta_{jtu} = \frac{2\min(b,c)}{a+2\min(b,c)} = \frac{r_1+r_2-2r_1p_1-|r_1-r_2|}{r_1p_1+r_1+r_2-2r_1p_1-|r_1-r_2|} = \frac{r_1+r_2-|r_1-r_2|-2r_1p_1}{r_1+r_2-|r_1-r_2|-r_1p_1}$$

If $r_1 > r_2$, and considering that $r_1p_1 = r_2p_2$,

$$\beta_{jtu} = \frac{r_1+r_2-r_1+r_2-2r_2p_2}{r_1+r_2-r_1+r_2-r_2p_2} = \frac{2r_2(1-p_2)}{2r_2(1-0.5p_2)} = \frac{1-p_2}{1-0.5p_2}$$

If $r_1 < r_2$,

$$\beta_{jtu} = \frac{r_1+r_2+r_1-r_2-2r_1p_1}{r_1+r_2+r_1-r_2-r_1p_1} = \frac{2r_1(1-p_1)}{2r_1(1-0.5p_1)} = \frac{1-p_1}{1-0.5p_1}$$

Hence, considering that if $r_1 > r_2$ then $p_1 < p_2$, or if $r_1 < r_2$ then $p_1 > p_2$,

$\beta_{jtu} = \frac{1-\max(p_1,p_2)}{1-\max(p_1,p_2)/2}$, which is independent of r_1 and r_2.

(6) β_{sim} (Baselga 2010) (*Repl*$_{BS}$ in this Appendix)

$$\beta_{\text{sim}} = \frac{\min(b,c)}{a + \min(b,c)} = \frac{(r_1 + r_2 - 2r_1p_1 - |r_1 - r_2|)/2}{r_1p_1 + (r_1 + r_2 - 2r_1p_1 - |r_1 - r_2|)/2} = \frac{r_1 + r_2 - |r_1 - r_2| - 2r_1p_1}{r_1 + r_2 - |r_1 - r_2|}$$

If $r_1 > r_2$, and considering that $r_1p_1 = r_2p_2$,

$$\beta_{\text{sim}} = \frac{r_1 + r_2 - r_1 + r_2 - 2r_2p_2}{r_1 + r_2 - r_1 + r_2} = \frac{2r_2(1 - p_2)}{2r_2} = 1 - p_2$$

If $r_1 < r_2$,

$$\beta_{\text{sim}} = \frac{r_1 + r_2 + r_1 - r_2 - 2r_1p_1}{r_1 + r_2 + r_1 - r_2} = \frac{2r_1(1 - p_1)}{2r_1} = 1 - p_1$$

Hence, considering that if $r_1 > r_2$ then $p_1 < p_2$, or if $r_1 < r_2$ then $p_1 > p_2$,

$\beta_{\text{sim}} = 1 - \max(p_1, p_2)$, which is independent of r_1 and r_2.

(7) Richness difference and nestedness indices

Because the D_J and D_S dissimilarities and the four replacement indices obey P10, the corresponding richness difference indices must also have property P10. Indeed, the richness difference and nestedness indices are computed as a dissimilarity (either D_J and D_S) minus a replacement index. From the equations above, the following relationships can easily be derived:

$$Rich_J = D_J - \beta_{-3} = \frac{\text{abs}(p_1 - p_2)}{p_1 + p_2 - p_1p_2}$$

$$Rich_S = D_S - Repl_S = \frac{\text{abs}(p_1 - p_2)}{p_1 + p_2}$$

$$\beta_{jne} = D_J - \beta_{jtu} = \frac{\max(p_1,p_2)\,\text{abs}(p_1 - p_2)}{2(p_1 + p_2) - 3p_1p_2 - \max(p_1^2,p_2^2)(1 - \min(p_1,p_2))} \quad (Nes_{BJ} \text{ in this Appendix})$$

$$\beta_{nes} = D_S - \beta_{sim} = \frac{\max(p_1,p_2)\,\text{abs}(p_1 - p_2)}{p_1 + p_2} \quad (Nes_{BS} \text{ in this Appendix})$$

These equations show that richness difference and nestedness are independent of the values of r_1 and r_2. They depend only on p_1 and p_2.

7. Conclusion and summary

This appendix attempted to describe in an orderly fashion the development of indices that decompose dissimilarity coefficients into replacement and richness/abundance difference (or nestedness) components. In the Podani and Baselga families, each one contains indices for presence-absence data that decompose the Jaccard and Sørensen dissimilarities, as well as indices

for species abundance data that decompose the Ružička and percentage difference dissimilarities. Both families were completed by new indices described in the present paper.

From the discussions that appeared in the recent literature, I chose to discuss the following points:

• As shown by Podani & Schmera (2011) and following papers, it is the numerators of the proposed indices that estimate replacement and richness difference. One can then scale the indices to values between 0 and 1 with denominators of one's choice, depending on the purpose of the study. The denominators of the Jaccard, Sørensen, Ružička and percentage difference dissimilarities, or those used by Baselga (2010, 2012) in his replacement (turnover) indices, can all be used. Ecologists should understand, however, that the chosen denominators may create distortions in the positioning of sites in an ordination, compared to using the numerator values only. None of the denominators proposed up to now have all the optimal qualities. The discussions about over- or under-estimation of species replacement by indices of the two families of indices are, actually, discussions about the choice of a denominator.

• The indices in the Podani family correspond to the concepts of replacement and richness/abundance difference. Those in the Baselga family are replacement (or turnover) and nestedness indices. Richness difference is not the same as nestedness. Podani & Schmera (2011) proposed an index of nestedness (N_{rel}) that differs from their index of richness difference; they explained that the latter only represents a portion of nestedness; see the Introduction of the main paper. Hence the Baselga nestedness indices (Nes_{BJ} and Nes_{BS}) should be compared to Podani & Schmera's relativized nestedness index, not to the richness difference indices of the Podani family ($Rich_J$ and $Rich_S$).

• In the two families, the replacement and richness difference (Podani family) or replacement and nestedness indices (Baselga family) sum to dissimilarity measures (D_J, D_S, D_R, $D_{\%diff}$) that are appropriate for beta diversity assessment, following the criteria of Legendre & De Cáceres (2013). An important point is that these indices are not themselves indices of beta diversity; they decompose dissimilarity coefficients that can be used as estimates of beta diversity.

• Replacement and richness difference or nestedness indices should have an ecological interpretation. In that respect, indices in the Podani family are easy to interpret due to the logic of their construction. Likewise, interpretation of Baselga's replacement indices is clear, whereas that of his nestedness indices is more intricate, albeit logical.

• When matrices of indices are to be used to produce ordinations of the sites, the Podani-family *RichDiff/AbDiff* indices ($RichDiff_S$ and $AbDiff_{\%diff}$) that decompose the Sørensen and percentage difference dissimilarities present clear advantages for ordination because the $\mathbf{RichDiff}_S$ and $\mathbf{AbDiff}_{\%diff}$ matrices are Euclidean.

• Claims have been made that the *Repl* indices in the Podani family were correlated to species richness differences between the sampling units whereas indices in the Baselga family are not. Actually, all indices described in this appendix do not depend directly on site richness since they can all be expressed without recourse to the species richness of the sites that are compared, r_1 and r_2. Section 6 of this appendix has shown that they can all be expressed by equations containing only p_1 and p_2, where p_1 is the proportion of shared species in the first sampling unit, $p_1 = a/r_1$, and p_2 is the proportion of shared species in the second sampling unit, $p_2 = a/r_2$; a is the number

of species in common between the two sites. So this criticism does not apply to any of the indices described in this appendix.

References

Almeida-Neto, M., Frensel, D.M.B. & Ulrich, W. (2012) Rethinking the relationship between nestedness and beta diversity: a comment on Baselga (2010). *Global Ecology and Biogeography*, **21**, 772–777.

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

Baselga, A. (2012) The relationship between species replacement, dissimilarity derived from nestedness, and nestedness. *Global Ecology and Biogeography*, **21**, 1223–1232.

Baselga, A. (2013) Separating the two components of abundance-based dissimilarity: balanced changes in abundance vs. abundance gradients. *Methods in Ecology and Evolution*, **4**, 552–557.

Cardoso, P., Borges, P.A.V. & Veech, J.A. (2009) Testing the performance of beta diversity measures based on incidence data: the robustness to undersampling. *Diversity and Distributions*, **15**, 1081–1090.

Carvalho, J.C., Cardoso, P. & Gomes, P. (2012) Determining the relative roles of species replacement and species richness differences in generating beta-diversity patterns. *Global Ecology and Biogeography*, **21**, 760–771.

Carvalho, J.C., Cardoso, P., Borges, P.A.V., Schmera, D. & Podani, J. (2013) Measuring fractions of beta diversity and their relationships to nestedness: a theoretical and empirical comparison of novel approaches. *Oikos*, **122**, 825–834.

Jaccard, P. (1908) Nouvelles recherches sur la distribution florale. *Bulletin de la Société Vaudoise des Sciences naturelles* **44**, 223–270.

Koleff, P., Gaston, K.J. & Lennon, J.J. (2003). Measuring beta diversity for presence-absence data. *Journal of Animal Ecology*, **72**, 367–382.

Legendre, P. & De Cáceres, M. (2013) Beta diversity as the variance of community data: dissimilarity coefficients and partitioning. *Ecology Letters*, **16**, 951–963.

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

Lennon, J.J., Koleff, P., Greenwood, J.J.D. & Gaston, K.J. (2001) The geographical structure of British bird distributions: diversity, spatial turnover and scale. *Journal of Animal Ecology,* **70**, 966–979.

Leprieur, F. & Oikonomou, A. (2014) The need for richness-independent measures of turnover when delineating biogeographical regions. *Journal of Biogeography* **41**, 417-420.

Odum, E.P. (1950) Bird populations of the Highlands (North Carolina) plateau in relation to plant succession and avian invasion. *Ecology*, **31**, 587–605.

Podani, J., Ricotta C. & Schmera D. (2013) A general framework for analyzing beta diversity, nestedness and related community-level phenomena based on abundance data. *Ecological Complexity*, **15**, 52–61.

Podani, J. & Schmera, D. (2011) A new conceptual and methodological framework for exploring and explaining pattern in presence-absence data. *Oikos*, **120**, 1625–1638.

Ružička, M. (1958) Anwendung mathematisch-statisticher Methoden in der Geobotanik (synthetische Bearbeitung von Aufnahmen). *Biologia, Bratislava*, **13**, 647–661.

Schmera, D. & Podani, J. (2011) Comments on separating components of beta diversity. *Community Ecology*, **12**, 153–160.

Simpson, G.G. (1943) Mammals and the nature of continents. *American Journal of Science*, **241**, 1–31.

Sørensen, T. (1948) A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analysis of the vegetation on Danish commons. *Biologiske Skrifter*, **5**, 1–34.

Tamas, J., Podani, J. & Csontos, P. (2001) An extension of presence/absence coefficients to abundance data: a new look at absence. *Journal of Vegetation Science*, **12**, 401–410.

Whittaker, R.H. (1960). Vegetation of the Siskiyou mountains, Oregon and California. *Ecological Monographs*, **30**, 279–338.

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

Williams, P.H. (1996) Mapping variations in the strength and breadth of biogeographic transition zones using species turnover. *Proceedings of the Royal Society of London B: Biological Sciences*, **263**, 579-588.

Table S1.1. Measures of dissimilarity, replacement and richness difference for presence-absence data with names used by different authors. Modified and expanded from Carvalho *et al.* (2013).

Indices	Equation	Reference
Dissimilarity indices		
D_J or D_{Jaccard} or β_{CC} or β_{jac}	$(b+c) / (a+b+c)$	Jaccard (1908)
D_S or $D_{\text{Sørensen}}$ or β_{sor}	$(b+c) / (2a+b+c)$	Sørensen (1948)
Podani family, Jaccard		
β_{-3} or R_{rel} or $Repl_J$	$\frac{2\min(b,c)}{a+b+c}$	Williams (1996) Cardoso *et al.* (2009) Podani & Schmera (2011)
β_{rich} or D_{rel} or $RichDiff_J$	$\frac{\lvert b-c\rvert}{a+b+c}$	Podani & Schmera (2011) Carvalho *et al.* (2012, 2013)
Podani family, Sørensen		
$Repl_S$	$\frac{2\min(b,c)}{2a+b+c}$	Carvalho *et al.* (2013)
$RichDiff_S$	$\frac{\lvert b-c\rvert}{2a+b+c}$	Carvalho *et al.* (2013)
Baselga family, Jaccard		
β_{jtu} ($Repl_{BJ}$)	$\frac{2\min(b,c)}{a+2\min(b,c)}$	Baselga (2012)
β_{jne} (Nes_{BJ})	$D_J - \beta_{jtu} = \frac{\lvert b-c\rvert}{a+b+c} \times \frac{a}{a+2\min(b,c)}$	Baselga (2012)
Baselga family, Sørensen		
β_{sim} ($Repl_{BS}$)	$\frac{2\min(b,c)}{2a+2\min(b,c)} = \frac{\min(b,c)}{a+\min(b,c)}$	Lennon *et al.* (2001), Baselga (2010)
β_{nes} (Nes_{BS})	$D_S - \beta_{sim} = \frac{\lvert b-c\rvert}{2a+b+c} \times \frac{a}{a+\min(b,c)}$	Baselga (2010)

Table S1.2. Podani-family indices: construction of the replacement (*Repl*), richness difference (or abundance difference for quantitative data) (*RichDiff or AbDiff*), and dissimilarity (*D*) indices, for presence-absence and abundance data. References for the indices, presence-absence data, Jaccard group: Podani & Schmera (2011); Sørensen group: Carvalho *et al.* (2013). Abundance indices, Jaccard group: Podani *et al.* (2013); Sørensen group: new in this paper.

	Presence-absence data	**Species abundance data**
***Numerators* (num.)**		
Replacement num.	$2\times\min(b,c) = (b+c) - \lvert b-c\rvert$	$2\times\min(B,C)$
Richness or abund. difference num.	$\lvert b-c\rvert = \max(b,c) - \min(b,c)$	$\lvert B - C\rvert$
Dissimilarity num.	$(b+c) = 2\times\min(b,c) + \lvert b-c\rvert$	$(B+C) = 2\times\min(B,C) + \lvert B - C\rvert$
***Jaccard group* (*J*)**		
Denominator	$(a+b+c)$	$(A+B+C)$
D	Jaccard dissimilarity: $D_J = (b+c) / (a+b+c)$	Ružička dissimilarity: $D_R = (B+C) / (A+B+C)$
Repl	$Repl_J = 2\times\min(b,c) / (a+b+c)$	$Repl_R = 2\times\min(B,C) / (A+B+C)$
RichDiff/AbDiff	$RichDiff_J = \lvert b-c\rvert / (a+b+c)$	$AbDiff_R = \lvert B - C\rvert / (A+B+C)$
***Sørensen group* (*S*)**		
Denominator	$(2a+b+c)$	$(2A+B+C)$
D	Sørensen dissimilarity: $D_S = (b+c) / (2a+b+c)$	Percentage difference dissimilarity: $D_{\%diff} = (B+C) / (2A+B+C)$
Repl	$Repl_S = 2\times\min(b,c) / (2a+b+c)$	$Repl_{\%diff} = 2\times\min(B,C) / (2A+B+C)$
RichDiff/AbDiff	$RichDiff_S = \lvert b-c\rvert / (2a+b+c)$	$AbDiff_{\%diff} = \lvert B - C\rvert / (2A+B+C)$

Table S1.3. Baselga-family indices: construction of the replacement (*Repl*), nestedness (or abundance gradient component for quantitative data, *Nes*), and dissimilarity (*D*) indices, for presence-absence and abundance data. References for the indices, presence-absence data, Jaccard group: Baselga (2012); Sørensen group: Baselga (2010). Abundance indices, Jaccard group: new in this paper; Sørensen group: Baselga (2013).

	Presence-absence data	**Species abundance data**
***Jaccard group* (*J*)**		
D	Jaccard dissimilarity: $D_J = (b+c) / (a+b+c)$	Ružička dissimilarity: $D_R = (B+C) / (A+B+C)$
Repl	$Repl_{BJ} = \frac{2\min(b,c)}{a+2\min(b,c)}$	$Repl_{BR} = \frac{2\min(B,C)}{A+2\min(B,C)}$
Nes	$Nes_{BJ} = \frac{\lvert b-c\rvert}{a+b+c} \times \frac{a}{a+2\min(b,c)}$	$Nes_{BR} = \frac{\lvert B-C\rvert}{A+B+C} \times \frac{A}{A+2\min(B,C)}$
***Sørensen group* (*S*)**		
D	Sørensen dissimilarity: $D_S = (b+c) / (2a+b+c)$	Percentage difference dissimilarity: $D_{\%diff} = (B+C) / (2A+B+C)$
Repl	$Repl_{BS} = \frac{\min(b,c)}{a+\min(b,c)}$	$Repl_{B\%diff} = \frac{\min(B,C)}{A+\min(B,C)}$
Nes	$Nes_{BS} = \frac{\lvert b-c\rvert}{2a+b+c} \times \frac{a}{a+\min(b,c)}$	$Nes_{B\%diff} = \frac{\lvert B-C\rvert}{2A+B+C} \times \frac{A}{A+\min(B,C)}$

Table S1.4. Metric and Euclidean properties of the two families of indices: (a) Podani family, (b) Baselga family. Yes: the index has the property; No: the index does not have the property. Abbreviations: *D* = dissimilarity, *Repl* = replacement, *RichDiff/AbDiff* = richness/abundance difference, *Nes* = nestedness, *J* = Jaccard, *R* = Ružička, *S* = Sørensen, *%diff* = percentage difference.

(a) Podani family	Metric property		Euclidean property	
	index	$\sqrt{index}$	*index*	$\sqrt{index}$
Jaccard group				
D_J	Yes	Yes	No	Yes
D_R	Yes	Yes	No	Yes
$Repl_J$	No	No	No	No
$Repl_R$	No	No	No	No
$RichDiff_J$	No	No	No	No
$AbDiff_R$	No	No	No	No
Sørensen group				
D_S	No	Yes	No	Yes
$D_{\%diff}$	No	Yes	No	Yes
$Repl_S$	No	No	No	No
$Repl_{\%diff}$	No	No	No	No
$RichDiff_S$	Yes	Yes	Yes	Yes
$AbDiff_{\%diff}$	Yes	Yes	Yes	Yes

(b) Baselga family	Metric property		Euclidean property	
	index	$\sqrt{index}$	*index*	$\sqrt{index}$
Jaccard group				
D_J	Yes	Yes	No	Yes
D_R	Yes	Yes	No	Yes
$Repl_{BJ}$	No	No	No	No
$Repl_{BR}$	No	No	No	No
Nes_{BJ}	No	No	No	No
Nes_{BR}	No	No	No	No
Sørensen group				
D_S	No	Yes	No	Yes
$D_{\%diff}$	No	Yes	No	Yes
$Repl_{BS}$	No	No	No	No
$Repl_{B\%diff}$	No	No	No	No
Nes_{BS}	No	No	No	No
$Nes_{B\%diff}$	No	No	No	No

Appendix to:

Legendre, P. (2014) Interpreting the replacement and richness difference components of beta diversity. *Global Ecology and Biogeography*, **23**, xxx–xxx.

Appendix S2

Simulations for monotonicity

Carvalho *et al.* (2013, pp. 830-831) carried out a simulation study to determine if the replacement and richness difference indices increased monotonically along ecological gradients showing increases in the amounts of replacement and richness difference.

• The replacement gradient was generated as follows: starting with a community containing p species (p = 105 in our study), an increasing number of species were replaced at each time step during 25 steps. Each replacement involved one species that was present at time 1 and was lost at time 2, and one new species that appeared at time 2. At each time step, one more species was replaced than at the previous step.

• Three scenarios were used to simulate gradients of richness difference, with one, two, or three species lost at each step. In scenario 1, replacement dominates the gradient because two new species are involved in replacement at each step compared to one for richness difference. The two processes are of equal importance in scenario 2, and richness difference dominates in scenario 3.

There would, of course, be other ways of generating these gradients.

First, I reproduced the study made by these authors for the Podani & Schmera (2011) indices decomposing the Jaccard dissimilarity D_J. The results were similar to those obtained by Carvalho *et al.* (2013) (Fig. S2.1, upper row).

The study was repeated for the indices derived from the Sørensen dissimilarity D_S, which had not been considered by Carvalho *et al.* (2013) in their simulation study. The indices derived from D_S were monotonic to replacement and richness gradients (Fig. S2.1, lower row).

Next, I carried out simulations involving the relativized nestedness index of Podani & Schmera (2011) and Baselga's nestedness indices decomposing D_J and D_S. (Fig. S2.2). The results showed that the relativized nestedness index (N_{rel}) decreased monotonically when replacement and richness difference increased (Fig. S2.2, upper row), whereas Baselga's nestedness-resultant D indices increased, then decreased (red lines in the middle and lower rows).

Finally, simulations were carried out to verify if the quantitative indices of the Podani-family obeyed the monotonicity condition of Carvalho *et al.* (2013).

• At each step in the simulations, which involved 100 species, selected species were replaced by new species, but without loss of total individuals, forming a gradient of increasing species replacement.

• Other species were lost, contributing to a smooth gradient of monotonic change in the abundance difference. The remaining initial species lost a constant number of individuals at each

step, following the *Ordered Comparison Case Series* simulation method proposed by Hajdu (1981) and used by Gower & Legendre (1986) to assess the properties of dissimilarity coefficients; these changes contributed to forming a smooth gradient of individual-based abundance differences along the series.

The *Repl* and *AbDiff* forms of the Ružička and percentage difference indices reacted monotonically to these quantitative changes in community composition (Fig. S2.3). The R code used to simulate the data is shown below. The indices were computed by function beta.div.comp (Appendix S3).

References

Carvalho, J.C., Cardoso, P., Borges, P.A.V., Schmera, D. & Podani, J. (2013) Measuring fractions of beta diversity and their relationships to nestedness: a theoretical and empirical comparison of novel approaches. *Oikos*, **122**, 825–834.

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

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

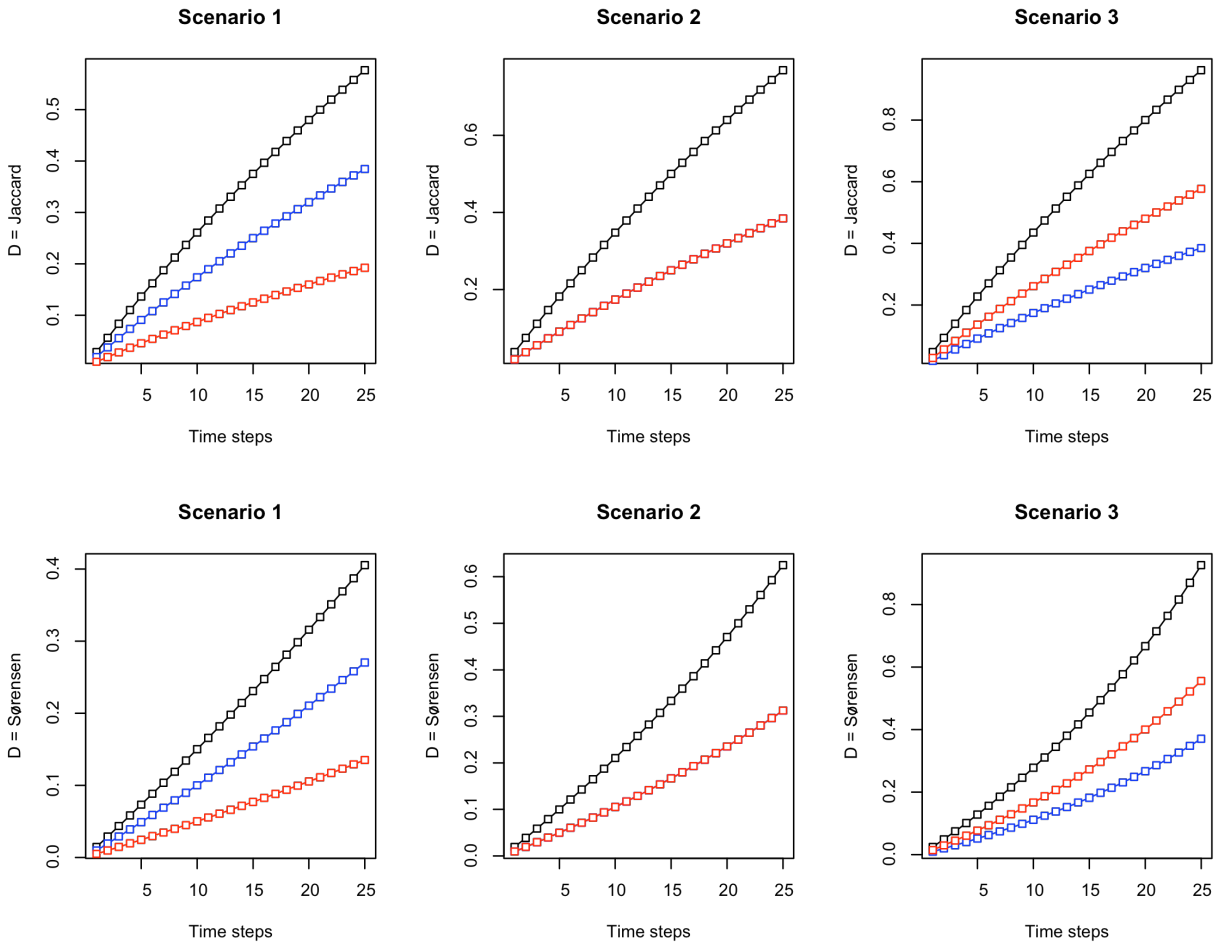


Figure S2.1 Simulation results for binary data. Podani & Schmera (2011) indices decomposing the Jaccard (D_J, upper row) and the Sørensen (D_S, lower row) dissimilarities. In each graph, the *initial* species data are compared to those at the given step along the simulated gradient (e.g. time). Black: dissimilarity values; blue: *Repl* index; red: *RichDiff* index. Under scenario 2, the *Repl* and *RichDiff* values are identical; the red squares are masking the blue symbols.

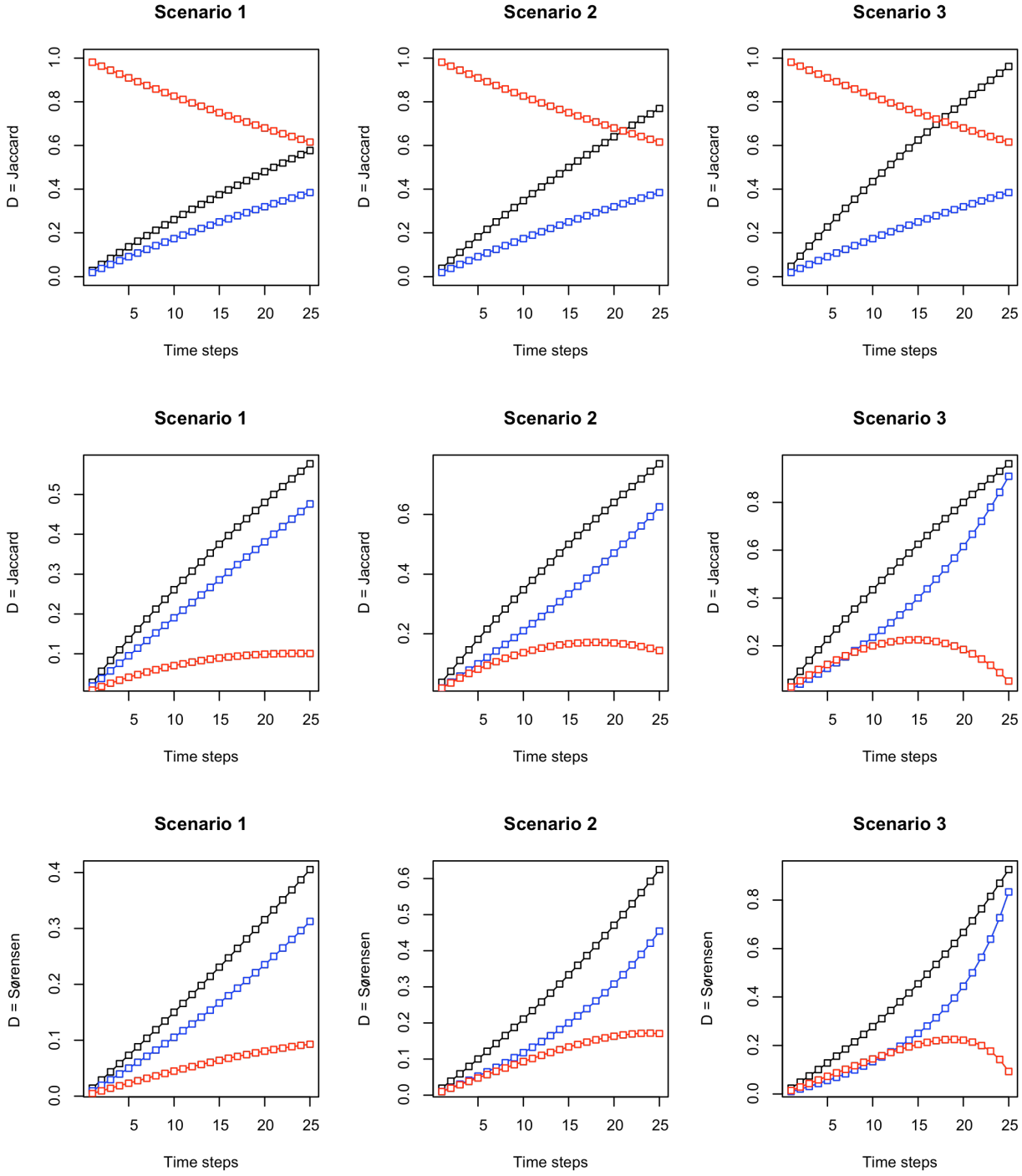


Figure S2.2 Simulation results for binary data. Upper row: Podani & Schmera (2011) decomposition of D_J into *Repl* and nestedness (N_{rel}); middle row: Baselga-family decomposing D_J; lower row: Baselga-family decomposing D_S. Black: dissimilarity values; blue: *Repl* indices; red: *nestedness* indices.

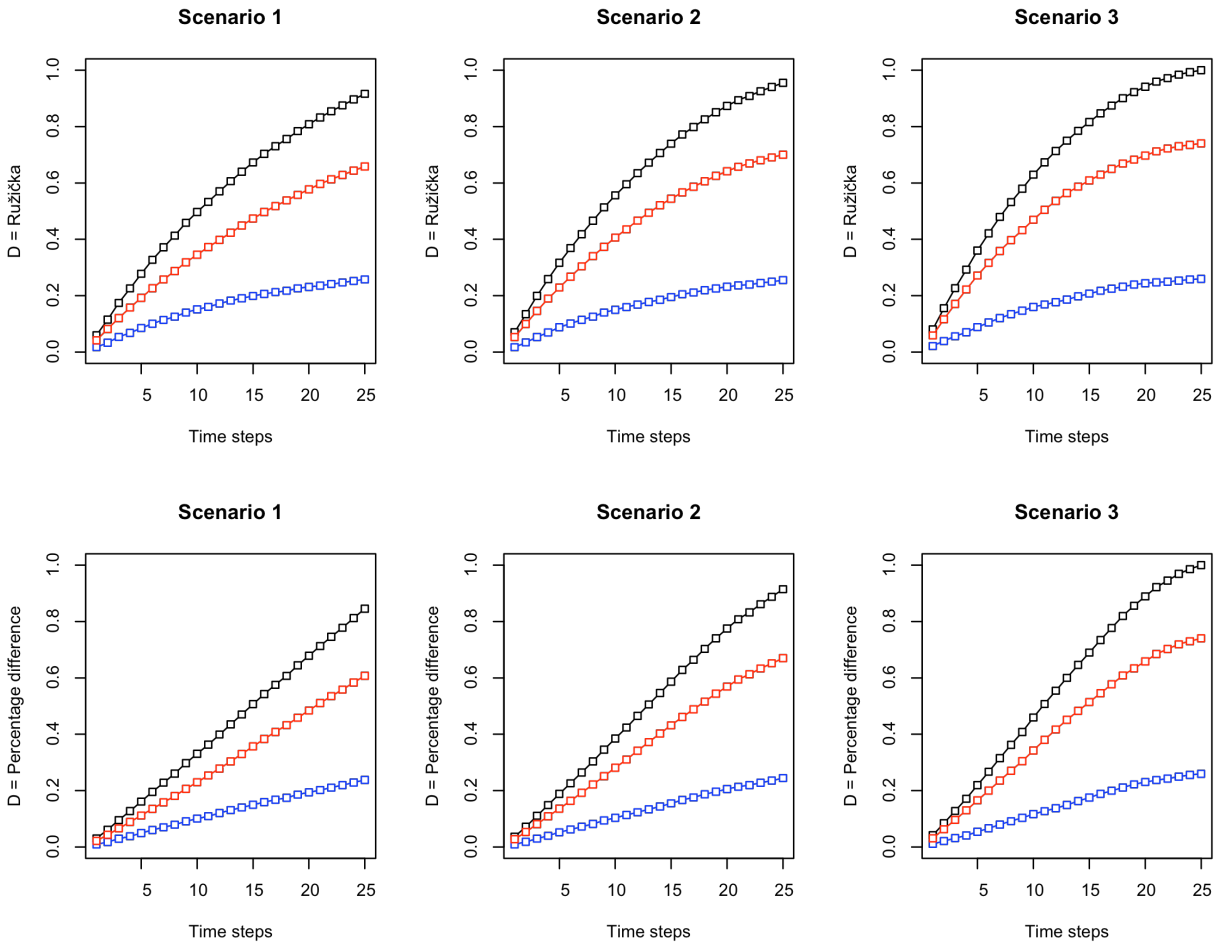


Figure S2.3 Simulation results for quantitative data. Podani-family indices decomposing the Ružička (D_R, upper row) and the percentage difference ($D_{\%diff}$, lower row) dissimilarities. In each graph, the initial species abundance data are compared to those at the given step along the simulated gradient (e.g. time). Black: dissimilarity values; blue: *Repl* index; red: *AbDiff* index.

R functions that generated the simulated data

```
carvalho.simul <- function(steps=25, p.init=100, scenario=2, a.min=5, val=1)
#
# Simulations using species presence-absence data to determine if the
# replacement and richness difference indices respond monotonically to
# gradients. See Carvalho et al. (2013, pp. 830:831) for details.
#
# steps : Number of (temporal) simulation steps after the initial vector.
# p.init : Number of species present at reference site 1.
# scenario={1,2,3} : how many species are lost at each step.
#         =1 : lose one species at each step.
#         =2 : lose two species at each step.
#         =3 : lose three species at each step.
# a.min : Minimum number of species in common, in scenario #3, between the
#         initial data (data row 1) and the last time step (data row 26).
# val : constant species abundance, e.g. 1 (binary) or 10 (quantitative).
#
# License: GPL-2
# Author:: Pierre Legendre, August 2013
{
# Modify p.init to insure there are species in common (a.min > 0) between the
# initial and last time steps in scenario #3.
p.init <- p.init+a.min
# Simulated data matrix
mat = matrix(0,(steps+1),(p.init+steps))
#
p = p.init    # Initial no. of species for computation of richness difference
loss = 0
gain = 0
mat[1,1:p] = val
for(i in 2:(steps+1)) {
      loss = loss+scenario
      gain = gain+1
      p = p-1
      # cat("[",i,",]","p",p," loss",loss," gain",gain,"\n")
      #
      if((p-loss)>0) mat[i,(loss+1):p] = val
      mat[i,(p.init+1):(p.init+gain)] = val
      }
mat
}
```

```
abundance.simul <- function(steps=25,p.init=100,scenario=2,val=10,cst.loss=1)
#
# Simulations using species abundance data to determine if the quantitative
# forms of the replacement and abundance difference indices respond
# monotonically to gradients. Method modified from that of Carvalho et al.
# (2013, pp. 830:831)
#
# Initial community: 'p.init' species, each with a runif() abundance.
# In the following (temporal) simulation steps:
#    Some species are lost (beginning of the vector).
#    Some species are replaced, with no loss of individuals.
#    The remaining initial species lose 1 individual each step (as in OCCAS).
#
# steps : Number of (temporal) simulation steps after the initial vector.
# p.init : Number of species present at reference site 1.
# scenario={1,2,3} : how many species are lost at each step.
#         =1 : lose one species at each step.
#         =2 : lose two species at each step.
#         =3 : lose three species at each step.
# val : Upper bound of values generated by runif().
# cst.loss : constant loss at each step for the initial species.
#
# License: GPL-2
# Author:: Pierre Legendre, August 2013
{
mat = matrix(0,(steps+1),(p.init+steps))
#
p = p.init     # number of non-zero data in initial species vector
loss = 0
gain = 0
mat[1,1:p] = floor(runif(p,1,(val+1)))+cst.loss*steps
#
for(i in 2:(steps+1)) {
# Beginning of vector 1:((i-1)*scenario): lost sp. (measured by RichDiff, red)
# Losses associated with replacements: (p.init-(i-2)): p.init)
# Gains  associated with replacements: (p.init+1): (p.init+(i-1)) (Repl, blue)
# => Maintain the number of individuals constant through all steps
      loss = loss+scenario
      gain = gain+1
      p = p-1
      # cat("[",i,",]","p",p," loss",loss," gain",gain,"\n")
      if((p-loss)>0) mat[i,(loss+1):p] = mat[(i-1),(loss+1):p]-cst.loss
      #
      lost.ind.repl <- sum(mat[(i-1),(p.init-(i-2)):p.init])
      n.per.new <- lost.ind.repl/gain
      # cat("lost.ind.repl =",lost.ind.repl," gain =",gain," n.per.new
=",n.per.new,"\n")
      mat[i,(p.init+gain)] = n.per.new   # Abundance of the (one) new species
      if(i>2) mat[i,(p.init+1):(p.init+(gain-1))]
            = mat[(i-1),(p.init+1):(p.init+(gain-1))]+n.per.new
      }
mat
}
```

```
# Appendix to:

# Legendre, P. (2014) Interpreting the replacement and richness difference
# components of beta diversity. Global Ecology and Biogeography, 23, xxx–xxx.

#                                   Appendix S3
#
# R function to compute the Podani- and Baselga-family decompositions of the
# Jaccard or Sørensen groups into replacement and richness/abundance difference
# (or nestedness) components, for species presence-absence or abundance data.


beta.div.comp <- function(mat, coef="J", quant=FALSE, save.abc=FALSE)
#
# Description --
#
# Podani-family and Baselga-family decompositions of the Jaccard and Sørensen
# dissimilarity coefficients into replacement and richness/abundance difference
# components, for species presence-absence or abundance data, as described
# in Legendre (2014).
#
# Usage --
#
# beta.div.comp(mat, coef="J", quant=FALSE, save.abc=FALSE)
#
# Arguments --
#
# mat : Data in matrix or data.frame form.
# coef : Family of coefficients to be computed --
#        "S" or "Sorensen": Podani family, Sørensen-based indices
#        "J" or "Jaccard" : Podani family, Jaccard-based indices
#        "BS" : Baselga family, Sørensen-based indices
#        "BJ" : Baselga family, Sørensen-based indices
#        "N" : Podani & Schmera (2011) relativized nestedness index.
#        The quantitative form in Sørensen family is the percentage difference.
#        The quantitative form in the Jaccard family is the Ruzicka index.
#
# quant=TRUE : Compute the quantitative form of the indices and D.
#      =FALSE: Compute the presence-absence form of the coefficients.
# save.abc=TRUE : Save the matrices of parameters a, b and c used in the
#      presence-absence calculations.
#
# Details --
#
#    For species presence-absence data, the distance coefficients are
# Jaccard=(b+c)/(a+b+c) and Sørensen=(b+c)/(2*a+b+c) with usual abc notation.
#
#    For species abundance data, the distance coefficients are
# the Ruzicka index = (B+C)/(A+B+C) and Odum's percentage difference
# (incorrectly called Bray-Curtis) = (B+C)/(2A+B+C), where
# A = sum of the intersections (or minima) of species abundances at two sites,
# B = sum at site 1 minus A, C = sum at site 2 minus A.
#
#    The binary (quant=FALSE) and quantitative (quant=TRUE) forms of the S and
# J indices return the same values when computed for presence-absence data.
#
```

```
# Value --
#
# repl : Replacement matrix, class = 'dist'.
# rich : Richness/abundance difference or nestedness matrix, class = 'dist'.
#        With options "BJ", "BS" and "N", 'rich' contains nestedness indices.
#        With option "N", the 'repl' and 'rich' values do not add up to 'D'.
# D    : Dissimilarity matrix, class = 'dist'.
# part : Beta diversity partitioning --
#        1. Total beta div. = sum(D.ij)/(n*(n-1)) (Legendre & De Cáceres 2013)
#        2. Total replacement diversity
#        3. Total richness/abundance difference diversity (or nestedness)
#        4. Total replacement div./Total beta div.
#        5. Total richness/abundance diff. div. (or nestedness)/Total beta div.
# Note : Name of the dissimilarity coefficient.
#
# References --
#
# Baselga, A. (2010) Partitioning the turnover and nestedness components of beta
# diversity. Global Ecology and Biogeography, 19, 134–143.
#
# Baselga, A. (2012) The relationship between species replacement, dissimilarity
# derived from nestedness, and nestedness. Global Ecology and Biogeography, 21,
# 1223–1232.
#
# Baselga, A. (2013) Separating the two components of abundance-based
# dissimilarity: balanced changes in abundance vs. abundance gradients. Methods
# in Ecology and Evolution, 4, 552–557.
#
# Carvalho, J.C., Cardoso, P., Borges, P.A.V., Schmera, D. & Podani, J. (2013)
# Measuring fractions of beta diversity and their relationships to nestedness:
# a theoretical and empirical comparison of novel approaches. Oikos, 122,
# 825–834.
#
# Legendre, P. (2014) Interpreting the replacement and richness difference
# components of beta diversity. Global Ecology and Biogeography, 23, xxx–xxx.
#
# Podani, J., Ricotta, C. & Schmera, D. (2013) A general framework for analyzing
# beta diversity, nestedness and related community-level phenomena based on
# abundance data. Ecological Complexity, 15, 52-61.
#
# Podani, J. & Schmera, D. (2011) A new conceptual and methodological framework
# for exploring and explaining pattern in presence-absence data. Oikos, 120,
# 1625–1638.
#
# License: GPL-2
# Author:: Pierre Legendre
{
coef <- pmatch(coef, c("S", "J", "BS", "BJ", "N"))
if(coef==5 & quant) stop("coef='N' and quant=TRUE: combination not programmed")
mat <- as.matrix(mat)
n <- nrow(mat)
if(is.null(rownames(mat))) noms <- paste("Site",1:n,sep="")
 else noms <- rownames(mat)
#
if(!quant) {      # Binary data provided, or make the data binary
 if(coef==   1) form="Podani family, Sorensen"
 if(coef==   2) form="Podani family, Jaccard"
```

```
if(coef==   3) form="Baselga family, Sorensen"
if(coef==   4) form="Baselga family, Jaccard"
if(coef==   5) form="Podani & Schmera (2011) relativized nestedness"
mat.b <- ifelse(mat>   0, 1, 0)
a <- mat.b %*% t(mat.b)
b <- mat.b %*% (   1 - t(mat.b))
c <- (   1 - mat.b) %*% t(mat.b)
min.bc <- pmin(b,c)
   #
if(coef==   1 || coef==2) {
    repl <-    2*min.bc   # replacement, turnover, beta-3
    rich <- abs(b-c)   # nestedness, richness diff., beta-rich
       #
       # Add the denominators
    if(coef==   1) {               # Sørensen-based components
        repl <- repl/(   2*a+b+c)
        rich <- rich/(   2*a+b+c)
        D <- (b+c)/(   2*a+b+c)
        } else if(coef==   2) {     # Jaccard-based components
        repl <- repl/(a+b+c)
        rich <- rich/(a+b+c)
        D <- (b+c)/(a+b+c)
        }
} else if(coef==   3) {     # Baselga 2010 components based on Sørensen
    D <- (b+c)/(   2*a+b+c)             # Sørensen dissimilarity
    repl <- min.bc/(a+min.bc)        # replacement, turnover
    rich <- D-repl                   # nestedness-resultant dissimilarity

} else if(coef==   4) {      # Baselga 2012 components based on Jaccard
    D <- (b+c)/(a+b+c)               # Jaccard dissimilarity
    repl <-    2*min.bc/(a+2*min.bc)    # replacement, turnover
    rich <- D-repl                   # nestedness-resultant dissimilarity
} else if(coef==   5) {      # rich = Podani N = nestdness based on Jaccard
    repl <-    2*min.bc/(a+b+c)
    D <- (b+c)/(a+b+c)
    rich <- matrix(   0,n,n)
    for(i in    2:n) {
        for(j in    1:(i-1)) {
        aa = a[i,j]; bb = b[i,j]; cc = c[i,j]
        if(a[i,j] ==    0)  rich[i,j] <- 0
            else  rich[i,j] <- (aa + abs(bb-cc))/(aa+bb+cc)
            }
        }
    }

rownames(repl) <- rownames(rich) <- rownames(D) <- noms
D <-    as.dist(D)
repl <-    as.dist(repl)
rich <-    as.dist(rich)
total.div <- sum(D)/(n*(n-   1))
repl.div <- sum(repl)/(n*(n-   1))
rich.div <- sum(rich)/(n*(n-   1))
part <- c(total.div,repl.div,rich.div,repl.div/total.div,rich.div/total.div)
   #
if(save.abc) {
res <- list(repl=repl, rich=rich, D=D, part=part, Note=form,
    a=   as.dist(a), b=as.dist(b), c=as.dist(c))
```

```
 } else {
 res <- list(repl=repl, rich=rich, D=D, part=part, Note=form)
 }
#
} else {      # Quantitative data
# Calculations based on individuals.within.species
 if(coef==   1) form<-"Podani family, percentage difference"
 if(coef==   2) form<-"Podani family, Ruzicka"
 if(coef==   3) form<-"Baselga family, percentage difference"
 if(coef==   4) form<-"Baselga family, Ruzicka"
    # Baselga (2013) notation:
    # A = W = sum of minima in among-site comparisons
    # B = site.1 sum - W = K.1 - W
    # C = site.2 sum - W = K.2 - W
 K <- vector(   "numeric", n)   # site (row) sums
 W <- matrix(   0,n,n)
 repl <- matrix(   0,n,n)
 rich <- matrix(   0,n,n)
 D <- matrix(   0,n,n)
 rownames(repl) <- rownames(rich) <- rownames(D) <- noms
 K <- apply(mat,   1,sum)         # Row sums
 for(i in    2:n) for(j in 1:(i-1)) W[i,j] <- sum(pmin(mat[i,], mat[j,]))
    #
    # Quantitative extensions of the S and J decompositions
 for(i in    2:n) {
     for(j in    1:(i-1)) {
         repl[i,j] <-    2*(min(K[i],K[j])-W[i,j]) # 2*min(B,C)
         rich[i,j] <- abs(K[i]-K[j])            # abs(B-C)
         }
     }
    #
    # Add the denominators
 if(coef==   1) {         # Sørensen-based (% difference) components
     for(i in    2:n) {
         for(j in    1:(i-1)) {                          # Baselga 2013 notation:
             repl[i,j] <- repl[i,j]/(K[i]+K[j])          #    2min(B,C)/(2A+B+C)
             rich[i,j] <- rich[i,j]/(K[i]+K[j])          # abs(B-C)/(   2A+B+C)
                # cat(K[i], K[j], W[i,j],"\n")
             D[i,j] <- (K[i]+K[j]-   2*W[i,j])/(K[i]+K[j])  # (B+C)/(2A+B+C)
             }
         }
     } else if(coef==   2) {    # Jaccard-based (Ruzicka) components
     for(i in    2:n) {
         for(j in    1:(i-1)) {                          # Baselga 2013 notation:
             repl[i,j] <- repl[i,j]/(K[i]+K[j]-W[i,j])   #    2min(B,C)/(A+B+C)
             rich[i,j] <- rich[i,j]/(K[i]+K[j]-W[i,j])   # abs(B-C)/(A+B+C)
                # cat(K[i], K[j], W[i,j],"\n")
         D[i,j]<-(K[i]+K[j]-   2*W[i,j])/(K[i]+K[j]-W[i,j]) # (B+C)/(A+B+C)
             }
         }
     }
    #
    # Baselga (2013): quantitative extensions of the Baselga (2010) indices
 if(coef==   3) {   # Baselga (2013) indices decomposing percentage difference
     for(i in    2:n) {
         for(j in    1:(i-1)) {
             repl[i,j] <- (min(K[i],K[j])-W[i,j])/min(K[i],K[j])
```

```
            rich[i,j] <- abs(K[i]-K[j])*W[i,j]/((K[i]+K[j])*min(K[i],K[j]))
               # cat(K[i], K[j], W[i,j],"\n")
            D[i,j] <- (K[i]+K[j]-   2*W[i,j])/(K[i]+K[j])
            }
        }
    }
 if(coef==   4) {   # Decomposing Ruzicka in the spirit of Baselga 2013
    for(i in    2:n) {
        for(j in    1:(i-1)) {
            repl[i,j] <-
                   2*(min(K[i],K[j])-W[i,j])/(2*min(K[i],K[j])-W[i,j])
            rich[i,j] <- abs(K[i]-K[j])*W[i,j]/
                ((K[i]+K[j]-W[i,j])*(   2*min(K[i],K[j])-W[i,j]))
               # cat(K[i], K[j], W[i,j],"\n")
            D[i,j] <- (K[i]+K[j]-   2*W[i,j])/(K[i]+K[j]-W[i,j])
            }
        }
    }
   #
 repl <-    as.dist(repl)
 rich <-    as.dist(rich)
 D <-    as.dist(D)
 repl.div <- sum(repl)/(n*(n-   1))
 rich.div <- sum(rich)/(n*(n-   1))
 total.div <- sum(D)/(n*(n-   1))
 part <- c(total.div,repl.div,rich.div,repl.div/total.div,rich.div/total.div)
   #
 res <- list(repl=repl, rich=rich, D=D, part=part, Note=form)
}
res
}
```

```
# Appendix to:

# Legendre, P. (2014) Interpreting the replacement and richness difference
# components of beta diversity. Global Ecology and Biogeography, 23, xxx–xxx.

#                                   Appendix S4
#
# R function to compute the dbRDA F-test of significance between response data
# represented by a Euclidean or non-Euclidean dissimilarity matrix and a matrix
# of explanatory variables, following McArdle and Anderson (2001).

dbRDA.D <- function(D, X, nperm=999, option=3, compute.eig=FALSE, coord=FALSE,
rda.coord=2, positive.RDA.values=FALSE)
#
# Description --
#
# Compute the dbRDA F-test of significance between response data represented by
# a Euclidean or non-Euclidean dissimilarity matrix and a matrix of explanatory
# variables, using the method of McArdle and Anderson (2001).
#
# Usage --
#
# dbRDA.D(D, X, nperm=999, option=3, compute.eig=FALSE, coord=FALSE,
#         rda.coord=2, positive.RDA.values=FALSE)
#
# Arguments --
#
# D : Distance matrix representing the response data. D may be non-Euclidean.
# X : Matrix of explanatory variables for the RDA, class 'data.frame' or
#     'matrix'. Factors must be recoded as dummy variables or Helmert contrasts.
# nperm : Number of permutations for the test of significance.
#
# option=1 : Original McArdle-Anderson (2001) equation 4. Slow, not recommended.
# option=2 : McArdle-Anderson equation, simplified.
# option=3 : Least-squares after orthogonalizing X.
# SSY = sum(diag(G)), where G is the Gower-centred distance matrix,
# SSYhat = sum(diag(H %*% G %*% H)), where H is the projector matrix.
#
# Option=1 -- The original F statistic of McArdle and Anderson (2001), eq. 4:
#     F = SSYhat / sum(diag(I.minus.H %*% G %*% I.minus.H))
#     Degrees of freedom are added to this equation in the output list.
# Option=2 -- Simplified equation:
#     F = SSYhat/(SSY-SSYhat)
# Option=3 -- Orthogonalize matrix X by PCA before computing H. No inversion.
#     Compute SSYhat as above, then F = SSYhat/(SSY-SSYhat)
# Opt. 2 and 3 are equivalent; they require half the computing time of option 1.
#
# compute.eig=TRUE : the eigenvalues and eigenvectors of D are computed.
#    => Do NOT use with very large matrices (slow).
# coord=TRUE : compute the principal coordinates corresponding to the
#    positive eigenvalues of D. Requires that compute.eig=TRUE.
# rda.coord : Number of RDA ordination coordinates to compute, for example 2.
# positive.RDA.values=TRUE : store only positive RDA eigenvalues in output list.
#                     =FALSE: store all RDA eigenvalues in output list.
#
#
```

```
# Details --
#
# Compute the dbRDA F-test of significance. The response is represented by a
# Euclidean or non-Euclidean dissimilarity matrix; X is a matrix of explanatory
# variables, as in regular RDA.
#
# The F-statistic is obtained without prior computation of the eigenvalues and
# eigenvectors of the dissimilarity matrix, hence no correction has to be made
# to eliminate the negative eigenvalues. Three computation methods are
# available, all derived from McArdle and Anderson (2001).
#
# The eigenvalues and eigenvectors of D are computed if compute.eig=TRUE.
# If coord=TRUE, the principal coordinates corresponding to the positive
# eigenvalues of D are computed.
#
# The function may fail to produce a meaningful RDA test of significance and
# ordination axes if D is extremely non-Euclidean. This is the case with some
# forms of genomic distances.
#
# Value --
#
# F : F-statistic.
# Rsquare : R-square and adjusted R-square statistics.
# P.perm : Permutational p-value of RDA R-square (test based on F).
# SS.total : Trace of matrix G, equal to the total sum of squares of Y and the
#            sum of the eigenvalues of D.
# PCoA.values : Eigenvalues (if they are computed, i.e. if compute.eig=TRUE).
# PCoA.vectors : Principal coordinates for the positive eigenvalues of D.
# RDA.values : RDA eigenvalues.
# RDA.rel.values : RDA relative eigenvalues.
# RDA.cum.values : RDA cumulative relative eigenvalues.
# RDA.coord : Ordination coordinates of objects on selected RDA axes.
#
# References --
#
# Legendre, P. (2014) Interpreting the replacement and richness difference
# components of beta diversity. Global Ecology and Biogeography, 23, xxx–xxx.
#
# Legendre, P. & Legendre, L. (2012) Numerical ecology, 3rd English edition.
# Elsevier Science BV, Amsterdam.
#
# McArdle, B.H. & Anderson, M.J. (2001) Fitting multivariate models to
# community data: a comment on distance-based redundancy analysis.
# Ecology, 82, 290–297.
#
# Example -- Six sites from the mite data available in the vegan package.
#
# library(vegan)
# Load function dbRDA.D()
# data(mite)
# data(mite.env)
# sel = c(14,24,31,41,49,64)
# mite.BC = vegdist(mite[sel,], "bray")   # Two negative eigenvalues
# res = dbRDA.D(mite.BC, mite.env[sel,1:2], nperm=999, compute.eig=TRUE)
# plot(res$RDA.coord)
# text(res$RDA.coord, labels=rownames(mite.env[sel6,]), pos=3)
#
```

```
# License: GPL-2
# Author:: Pierre Legendre, March 2013
{
 D <-    as.matrix(D)
 X <-    as.matrix(X)
 n <- nrow(D)
 epsilon <- .Machine$double.eps
#
# Gower centring, matrix formula. Legendre & Legendre (2012), equation 9.42
 One <- matrix(   1,n,n)
 mat <- diag(n) - One/n
 G <- -   0.5 * mat %*% (D^2) %*% mat
 SSY <- sum(diag(G))
    # LCBD <- diag(G)
#
# Principal coordinate analysis after eigenvalue decomposition of D
 if(compute.eig) {
     eig <- eigen(G, symmetric=TRUE)
     values <- eig$values     # All eigenvalues
     vectors <- eig$vectors   # All eigenvectors, scaled    to lengths 1
     if(coord) {
         select <- which(values > epsilon)
         princ.coord <- vectors[,select] %*% diag(sqrt(values[select]))
         } else { princ.coord <- NA }
     } else {
     values <- princ.coord <- NA
     }
#
# Compute projector matrix H ("hat" matrix in the statistical literature)
 X.c <- scale(X, center=TRUE, scale=FALSE)   # Centre matrix X
 m <- qr(X.c, tol=   1e-6)$rank                  # m = rank of X.c
 cat(   "Rank of X centred =",m,"\n")
 if(m==   1) {
     H <- (X.c[,   1] %*% t(X.c[,1]))/((t(X.c[,1]) %*% X.c[,1])[1,1])
     } else {
     if(option<   3) {
            # if(det(t(X.c)%*%X.c)<epsilon) stop ('Collinearity detected in X')
         if(m < ncol(X.c)) stop (   'Collinearity detected in X')
         H <- X.c %*% solve(t(X.c) %*% X.c) %*% t(X.c)
            #
            # option=3: compute projector H from orthogonalized X; no inversion
         } else {
         X.eig <- eigen(cov(X.c))
         k <- length(which(X.eig$values > epsilon))
         X.ortho <- X.c %*% X.eig$vectors[,   1:k]  # F matrix of PCA
         XprX <- t(X.ortho) %*% X.ortho
         H <- X.ortho %*% diag(diag(XprX)^(-   1)) %*% t(X.ortho)
         }
 }
#
# Compute the F statistic: McArdle & Anderson (2001), equation 4 modified
 HGH <- H %*% G %*% H
 SSYhat <- sum(diag(HGH))
    #
 if(option==   1) {
     I.minus.H <- diag(n) - H
     den1 <- sum(diag(I.minus.H %*% G %*% I.minus.H))
```

```
    F <- SSYhat/den1     # F statistic without the degrees of freedom
    Rsquare <- F/(F+   1)
 } else {
    F <- SSYhat/(SSY-SSYhat)  # F statistic without the degrees of freedom
    Rsquare <- SSYhat/SSY     # or equivalent: Rsquare <- F/(F+   1)
 }
 RsqAdj <-    1-((1-Rsquare)*(n-1)/(n-1-m))
#
# Permutation test of F
 if(nperm >    0) {
    nGE=   1
    for(i in    1:nperm) {
        order <- sample(n)
        Gperm <- G[order, order]
        H.Gperm.H <- H %*% Gperm %*% H
        SSYhat.perm <- sum(diag(H.Gperm.H))
           #
        if(option==   1) {
            den <- sum(diag(I.minus.H %*% Gperm %*% I.minus.H))
            F.perm <- SSYhat.perm/den
        } else {
            F.perm <- SSYhat.perm/(SSY-SSYhat.perm)
        }
        if(F.perm >= F) nGE=nGE+   1
        }
    P.perm <- nGE/(nperm+   1)
    } else { P.perm <- NA }
#
# Compute RDA ordination coordinates
 if(rda.coord >    0) {
    HGH.eig <- eigen(HGH, symmetric=TRUE)
       # kk <- length(which(HGH.eig$values > epsilon))
    RDA.values <- HGH.eig$values
    rel.eig <- RDA.values/SSY
    cum.eig <- cumsum(rel.eig)
    kk <- length(which(rel.eig > epsilon))
    if(positive.RDA.values) {
        RDA.values <- RDA.values[   1:kk]
        rel.eig <- rel.eig[   1:kk]
        cum.eig <- cum.eig[   1:kk]
        }
    k <- min(rda.coord, kk)
    if(k >=    2) {
    RDA.coord <-sweep(HGH.eig$vectors[,   1:k],2,sqrt(RDA.values[1:k]),FUN="*")
        } else {
        RDA.coord <- NA
        cat(   "k =",k," -- Fewer than two RDA eigenvalues > 0\n")
        }
    } else { RDA.values <- rel.eig <- cum.eig <- RDA.coord <- NA }
#
list(F=F*(n-m-1)/m, Rsquare=c(Rsquare,RsqAdj), P.perm=P.perm, SS.total=SSY,
PCoA.values=values, PCoA.vectors=princ.coord, RDA.values=RDA.values/(n-1),
RDA.rel.values=rel.eig, RDA.cum.values=cum.eig, RDA.coord=RDA.coord)
}
```

```
# Appendix to:
# Legendre, P. (2014) Interpreting the replacement and richness difference
# components of beta diversity. Global Ecology and Biogeography, 23, xxx–xxx.

#                                   Appendix S5
#
# R function to compute LCBD indices from a dissimilarity matrix (D) or from
# beta diversity component matrices (Repl, RichDiff/AbDiff or Nes).

LCBD.comp <- function(x, sqrt.x=TRUE)
#
# Description --
#
# Computes LCBD indices (Legendre and De Cáceres 2013) from a dissimilarity
# matrix (D) or beta div. component matrices (Repl, RichDiff/AbDiff or Nes).
#
# Arguments --
#
# x : D or beta diversity component matrix, class=dist.
# sqrt.x : Take sqrt() of components before computing LCBD.comp. Use
#     sqrt.x=TRUE for the replacement and richness/abundance difference indices
#     computed by beta.div.comp(), as well as for the corresponding D matrices.
#
# Reference --
#
# Legendre, P. & De Cáceres, M. (2013) Beta diversity as the variance of
# community data: dissimilarity coefficients and partitioning. Ecology
# Letters 16: 951–963.
#
# License: GPL-2
# Author:: Pierre Legendre, August 2013
{
### Internal function
centre <- function(D,n)
    # Centre a square matrix D by matrix algebra
    # mat.cen = (I - 11'/n) D (I - 11'/n)
 {    One <- matrix(  1,n,n)
     mat <- diag(n) - One/n
     mat.cen <- mat %*% D %*% mat
 }
###
n <- nrow(as.matrix(x))

if(sqrt.x) {
    # x = sqrt(x)
 SStotal <- sum(x)/n        # eq.    8
 BDtotal <- SStotal/(n-   1)   # eq. 3
 G <- centre(   as.matrix(-0.5*x), n)     # Gower-centred matrix
 } else {
 SStotal <- sum(x^   2)/n      # eq. 8
 BDtotal <- SStotal/(n-   1)   # eq. 3
 G <- centre(   as.matrix(-0.5*x^2), n)   # Gower-centred matrix
 }
LCBD <- diag(G)/SStotal   # Legendre & De Caceres (2013), eq. 10b
out <- list(SStotal_BDtotal=c(SStotal,BDtotal), LCBD=LCBD, D=x)
}
```

Appendix to:

Legendre, P. (2014) Interpreting the replacement and richness difference components of beta diversity. *Global Ecology and Biogeography*, **23**, xxx–xxx.

Appendix S6

Principal coordinate ordinations

Principal coordinate (PCoA) ordinations of the Podani-family replacement and richness difference indices based on the Jaccard (Fig. S6.1) and Sørensen (Fig. S6.2) dissimilarities for the case study data. Detailed properties of these ordinations will be investigated in another study.

The ordinations in Fig. S6.1 are based upon replacement (**Repl**$_J$) and richness difference (**RichDiff**$_J$) matrices that are not Euclidean. Ordinations along the first principal coordinates are Euclidean but, as in all multivariate ordinations, the distances among sampling units are approximate. The ordination of the *RichDiff*$_J$ indices in Fig. S6.1b shows a less precise ordering of the sampling units than Fig. S6.2b, which is based upon the *RichDiff*$_S$ indices.

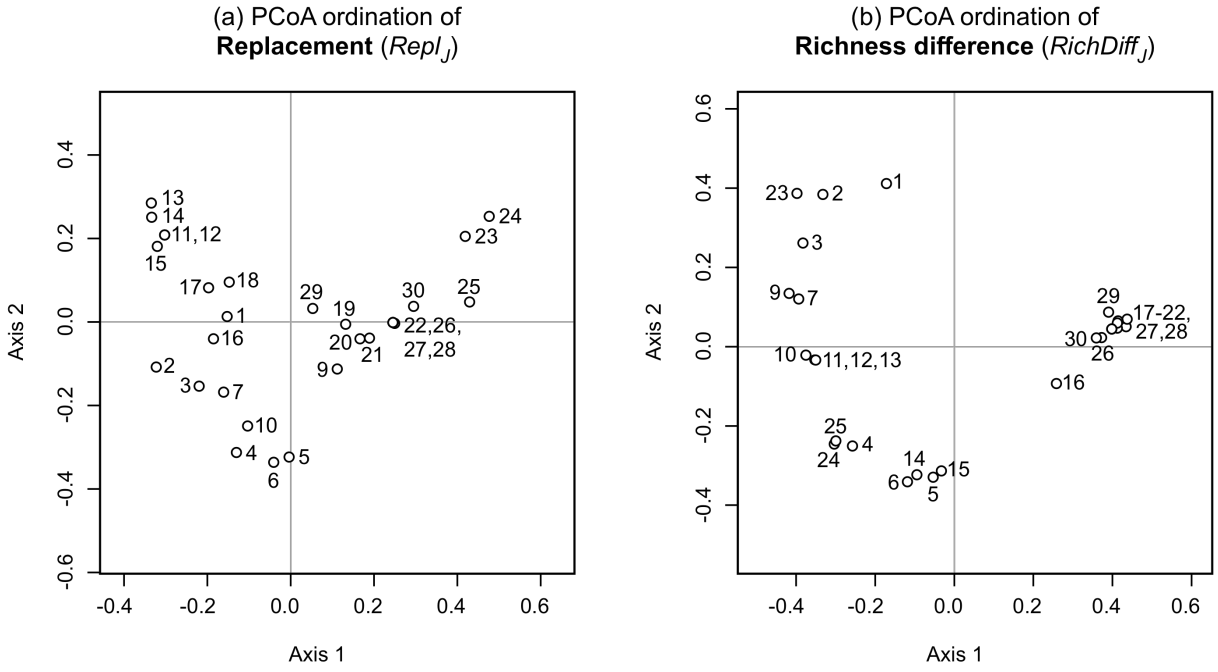


Figure S6.1 Principal coordinate ordinations of the Podani-family replacement and richness difference indices based on Jaccard dissimilarity for the fish case study data. PCoA of square-root transformed (a) replacement (*Repl*$_J$) and (b) richness difference (*RichDiff*$_J$) indices. These matrices are not Euclidean even after square-root transformation.

Of the four indices for species presence-absence data represented by ordinations in this appendix, the **RichDiff**$_S$ matrix is the only one that has the mathematical property of being Euclidean (Appendix S1, Table S1.4). The ordination in two dimensions (Fig. S6.2b) displays a precise ordering of the sampling units along the richness gradient, along which the sites form a curved one-dimensional ordination (i.e. a curved line). Site 1 (left of the graph) has the lowest richness with only 1 species whereas site 29 (right) is the richest with 26 species; 27 species were captured along the river. Sites 23-25 had reduced richness due to agricultural pollution; this caused reversal of the ordering of the sites by richness, compared to their geographical sequence.

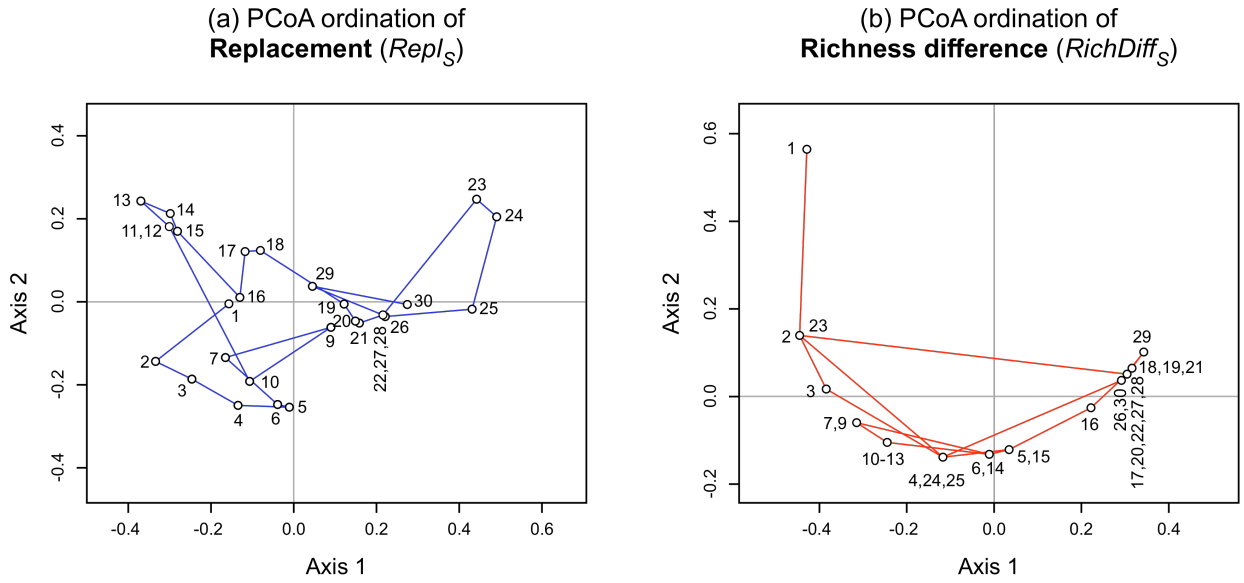


Figure S6.2 Principal coordinate ordinations of the Podani-family replacement and richness difference indices based on the Sørensen dissimilarity for the fish case study data. (a) PCoA of square-rooted replacement (*Repl*$_S$) and (b) of untransformed richness difference (*RichDiff*$_S$) indices. The **Repl**$_S$ matrix is not Euclidean even after square-root transformation whereas the **RichDiff**$_S$ matrix is Euclidean without transformation. Colour lines join the sites in their geographic sequence along the course of the river from site 1 (headwaters) to site 30 near the junction with the Saône River.