



Interpreting the replacement and richness difference components of beta diversity

Pierre Legendre*

Département de Sciences Biologiques,
Université de Montréal, Montréal, QC,
Canada

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 formulae for decomposing dissimilarities into replacement, richness difference and nestedness indices. These formulae are presented for species presence–absence and abundance using a unified algebraic framework. The indices decomposing beta play different roles in ecological analysis than do 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.

Keywords

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

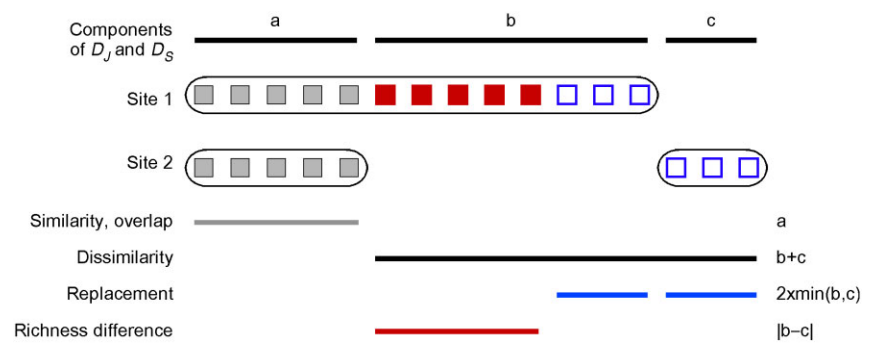
*Correspondence: Pierre Legendre,
Département de Sciences Biologiques,
Université de Montréal, CP 6128, Succursale
Centre-ville, Montréal, QC, Canada H3C 3J7.
E-mail: pierre.legendre@umontreal.ca

INTRODUCTION

Since Whittaker (1972), ecological dissimilarities have been used to measure beta diversity between sampling units. Koleff *et al.* (2003) reviewed 24 beta-diversity indices proposed in the literature while 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.

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



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 (e.g. Williams, 1996; Baselga, 2010, 2012; Podani & Schmera, 2011; 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 analysed along spatial or environmental gradients. It implies the simultaneous gain and loss of species due to environmental filtering, competition and historical events (Leprieur *et al.*, 2011). The replacement component of dissimilarity measures (Appendix S1 in Supporting Information) may thus reflect the influence on community structure of the variables controlling ecological gradients.

Richness difference refers to the fact that one 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 (e.g. physical barriers).

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). However, how the concept of nestedness can be translated into an index is a subject open to discussion as there is no unequivocal way to do it.

In a 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) con-

tribute 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).

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

The main contributions of this paper are as follows.

1. It demonstrates that the sums of the replacement and richness difference indices corresponding to all pairwise comparisons over a study area form a proper decomposition of the total beta diversity over that area.
2. It describes how the differences among sites measured by these indices can be analysed and interpreted using descriptive or experimental factors, or explanatory environmental variables. They can also be used to produce ordinations.
3. It shows how to work out the local contributions of replacement and richness difference to total beta diversity. These contributions can be mapped to facilitate interpretation. The calculations are illustrated using an ecological case study. (R software is presented in the Supporting Information.)

METHODS

Many papers have appeared, mostly during the past 4 years, describing or discussing replacement, richness (or abundance) difference and nestedness indices (Williams, 1996; Lennon *et al.*, 2001; Cardoso *et al.*, 2009; Baselga, 2010, 2012, 2013a; Podani & Schmera, 2011; Schmera & Podani, 2011; Carvalho *et al.*, 2012, 2013; Podani *et al.*, 2013); index notations differed among papers. Because the present paper is limited in page space, a brief history of the development of these indices is presented in Appendix S1 together with a summary of the criticisms formulated against the suggested indices.

That discussion shows that all authors agree on the dissimilarity coefficients that seem to be the most useful bases for decomposition. First, 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 $2\min(B,C)$ used as the replacement numerator for quantitative data (see Appendix S1 for the meanings of B and C). 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 (formulae 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 Appendix S1, 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 ordina-

tion, compared with using the numerator values only. None of the denominators proposed up to now has all the optimal qualities. The discussions about over- or under-estimation of species replacement by indices from 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 with Podani and Schmera's relativized nestedness index, not with 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 $RichDiff_s$ and $AbDiff_{\%diff}$ matrices are Euclidean, meaning that the data points can be fully represented in Euclidean space by principal coordinates 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, none of the indices described in Appendix S1 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 in Appendix S1 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 in Appendix S1 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 $\mathbf{D}^{(0.5)} = [D_{hi}^{0.5}]$ Euclidean (with the meaning described in ‘Replacement, richness difference and nestedness indices’; Legendre & Legendre, 2012, Tables 7.2 and 7.3). For that reason, in the variance approach in the previous paragraph, Legendre & De Cáceres (2013) recommended taking the square root of the distances before computing total beta diversity, BD_{Total} with these four coefficients:

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

$$= \sum_{h=1}^{n-1} \sum_{i=h+1}^n D_{hi} / [n(n-1)]$$

as shown by Whittaker (1972) (Whittaker’s formula gives twice the value of BD_{Total} produced by equation 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

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

Because these equations are similar to equation 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

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

$$RichDiff_{Prop} = RichDiff_{Total} / BD_{Total}.$$

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

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

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

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 in a 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 in 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 variation in *Repl* and *RichDiff*

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 species data by canonical analysis. That method consists of the following steps: first, a 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, the \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 (with the meaning described above). Using the complete set of principal coordinates computed from $\mathbf{D}^{(0.5)}$ in RDA 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 (**RichDiff/AbDiff**) and nestedness (**Nes**) 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 **RichDiff**_s and **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

$$\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \quad (5)$$

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 PCoA (Gower, 1966) before eigenvalue decomposition:

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

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

$$F = \frac{\text{tr}(\mathbf{HGH})/m}{\text{tr}[(\mathbf{I} - \mathbf{H})\mathbf{G}(\mathbf{I} - \mathbf{H})]/(n - m - 1)} \quad (7)$$

where $\text{tr}()$ 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-Euclidean nature.

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 PCoA, the sites with high LCBD values are those found far from the multivariate centroid of the graphs; LCBD is actually the squared distance of a site to the data centroid, which is the point of origin of the multivariate ordination graph.

Large LCBD values indicate sites that have strongly different species compositions compared with a mean site. For conservation biology, large LCBD values may indicate sites that have unusual species combinations and high conservation value, or degraded and species-poor sites that are good candidates for ecological restoration. They may also correspond to special ecological conditions or result from the effect of invasive species on

communities. LCBD values can be mapped to facilitate interpretation, as shown in Legendre & De Cáceres (2013).

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

Here the calculation of LCBD indices is extended to replacement and richness difference. These new LCBD indices, $Repl_{LCBD}$ and $RichDiff_{LCBD}$, measure how exceptional each site is, when compared with the other sites, in terms of replacement or richness (or abundance) difference. In an ordination of the replacement or richness difference indices, the sites with high values of $Repl_{LCBD}$ and $RichDiff_{LCBD}$ are those that are far from the multivariate centroid of the graph. Like total beta LCBD, $Repl_{LCBD}$ and $RichDiff_{LCBD}$ are the diagonal values of the Gower-centred dissimilarity matrices computed during PCoA 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 indices 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 & Schmera (2011) provide a series of examples that can be used to familiarize oneself with the interpretation of SDR simplices.

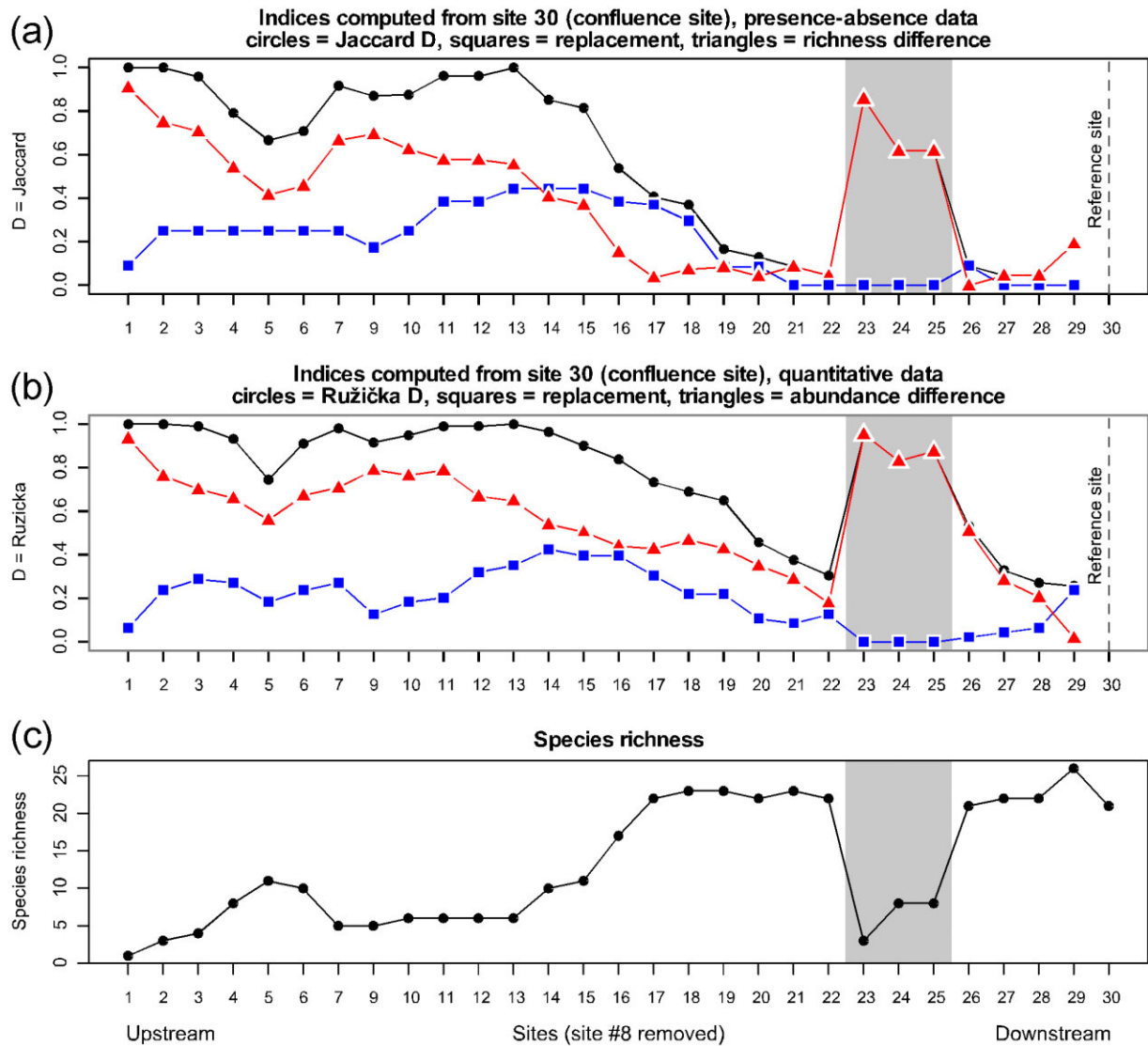


Figure 2 (a) Jaccard dissimilarity (D , circles), Replacement (squares) and Richness difference (triangles) indices for presence–absence data, comparing sites 1–29 in the Case study with site 30. (b) The same but using species abundance data and the Ružička dissimilarity index. (c) Species richness at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

CASE STUDY

Freshwater fish were collected by Verneaux (1973) in the Doubs River, a tributary of the Saône that runs near the French–Swiss border in the Jura Mountains in eastern France. In his paper, Verneaux proposed the use of 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 geographic coordinates of the site, and environmental data (source: <http://adn.biol.umontreal.ca/~numeralecology/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, respectively (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 five 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 with 28% for species replacement ($Repl_{prop}$). In the **Repl** and **RichDiff** matrices, 27 pairs of sites had richness difference values of 0 compared with 150 pairs that had replacement values of 0. With presence–absence data, a 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

P. Legendre

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

Ward hierarchical clustering of the sites based upon the matrix of environmental variables showed a first division of the sites between the upper (sites 1 to 22) and lower courses of the river (sites 23 to 30). The lower course is characterized by lower altitude and higher water discharge, higher hardness and biological oxygen demand, and lower oxygen concentrations. A test of significance of the variation of the D_j , $Repl_j$ and $RichDiff_j$ matrices against a factor representing the upper and lower groups of sites was carried out to illustrate the test of significance (equation 7) of the variation in these matrices by a factor, using the function in Appendix S4. Note that the indices derived from the community composition data subjected to the test are independent of the environmental data used in the cluster analysis that produced the factor. The binary factor significantly explained the variation between the two groups for D_j and $Repl_j$ (P -values of 0.005 and 0.001, respectively, after 999 random permutations) but not for $RichDiff_j$ (P -value of 0.943). This result was expected from the ordinations shown in Appendix S6, where sites 1–22 are clearly separated from sites 23–30 in the replacement ordinations but not in the richness difference ordinations (Figs S6.1 & S6.2). Tests based upon the quantitative forms of the indices produced similar results.

Following that, the variation among sites in the upper course of the river only (sites 1 to 22) was studied in more detail using the environmental variables. In order to identify the environmental variables that best explained each index, a PCoA of each of the Jaccard-based matrices (with indices square-rooted) was carried out, with Lingoes correction for the negative eigenvalues in the case of the $Repl_j$ and $RichDiff_j$ matrices; matrix $D_j^{(0.5)}$ was Euclidean and did not produce negative eigenvalues. The principal coordinates were used as input response variables into a forward selection procedure in RDA (function `forward.sel()` of the R package `packfor`; Dray *et al.*, 2012). The selected explanatory variables were {slope, hardness, nitrate and O_2 } for D_j , { O_2 } for $Repl_j$ and {slope, hardness and nitrate} for $RichDiff_j$. The tests of significance using principal coordinates corrected for

negative eigenvalues were only approximate (McArdle & Anderson, 2001). An exact test of significance was then computed for each matrix and its set of explanatory variables, using the function in Appendix S4, which implements the F -test described in equation 7 in a permutation testing procedure. The variation in the D_j , $Repl_j$ and $RichDiff_j$ matrices explained by the selected environmental variables (those listed above) was highly significant in all cases (P -values = 0.001) with high values of R_{adj}^2 of 0.36, 0.44 and 0.47, respectively. Thus the $Repl$ and $RichDiff$ components of the species variation were significantly related to different environmental variables.

LCBD indices were computed for the replacement and richness difference indices 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 $Repl_{LCBD}$ and $RichDiff_{LCBD}$ 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 quieter 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 (one to four 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 (Fig. 4a, b) and for the Ružička and percentage difference indices on the other hand (Fig. 4c, d), are very similar. The graphs show that the among-site variation is dominated by richness difference [mean points (large dots) along the $RichDiff$ or $AbDiff$ axis near 0.5 in Fig. 4(a) and (c), and near 0.4 in 4(b) and 4(d), compared with lower mean values along the $Repl$ axis], confirming the values found above for $RichDiff_{prop} = 72\%$ of total beta diversity versus 28% for $Repl_{prop}$. Remember that these

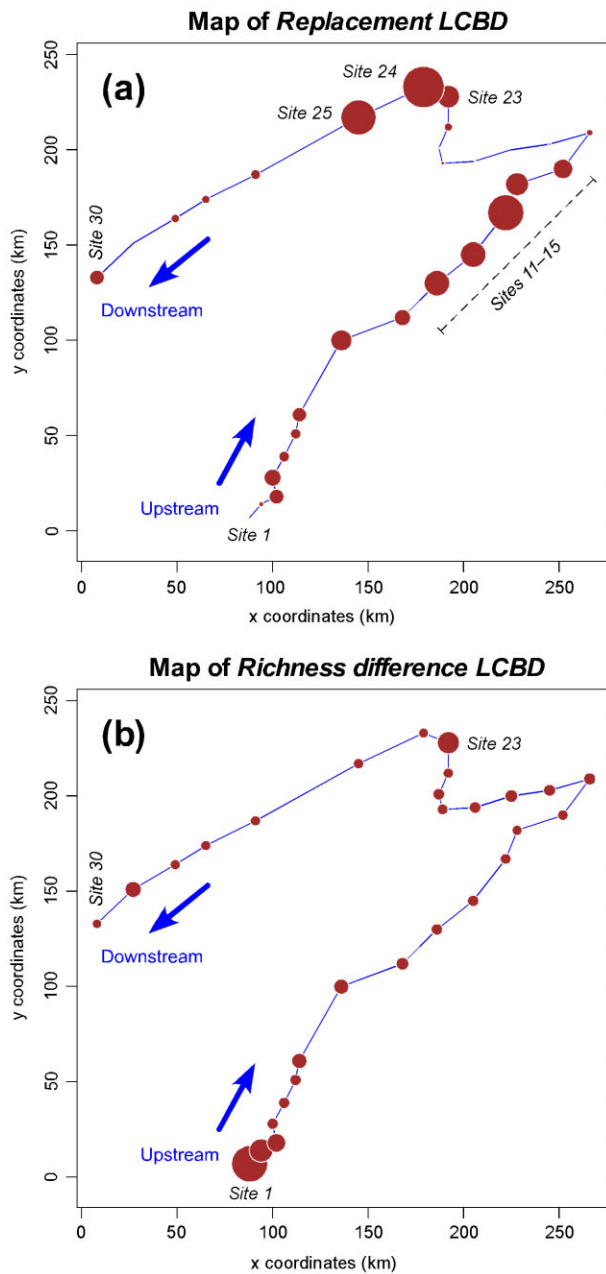


Figure 3 Schematic maps of the Doubs River (line) showing the Podani family (a) Replacement LCBD ($Repl_{LCBD}$) and (b) Richness difference LCBD ($RichDiff_{LCBD}$) of the binary fish assemblage data at the 29 study sites decomposing the Jaccard dissimilarity. Circle sizes are proportional to LCBD values.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

two statistics use the sum of the D values as their denominator (equation 4a,b), and that sum is $n(n-1)/2$ minus the sum of the S values used in the triangular plots. That is why the $RichDiff_{Prop}$ and $Repl_{Prop}$ statistics are larger than the means of $RichDiff$ and $Repl$ shown by large dots along the graph margins. Note also the many triplets (points in the triangular graphs) along the Similarity axis that have $Repl$ values of 0.

DISCUSSION

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

The present paper focussed on two families of indices and used a unified algebraic framework for their computation. These indices have interesting properties: the numerators of the indices are chosen to estimate the ecological phenomena of interest (replacement and richness/abundance difference) and the denominators are added afterwards to normalize the indices. The sum of the replacement and richness difference components is always equal to the corresponding dissimilarity, for binary (Jaccard or Sørensen) or quantitative (Ružička or percentage difference) indices. The indices react monotonically to species composition (presence–absence) or abundance gradients, as shown by numerical simulations. An R function provided with this paper (Appendix S3) allows readers to check empirically that when the quantitative indices are computed on presence–absence data, they produce the same results as the binary versions of the indices; that fact can also be checked algebraically.

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

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

Computing indices is a first step; interpreting their variation is the next necessary step. This paper described statistical tools and methods for the interpretation of these new indices. That is its main contribution. We showed that the replacement and richness difference indices can be summed over the study sites in an area, and the sums can be divided by the sum of the corresponding distances. These ratios estimate the relative contributions of the two ecological processes to beta diversity over the study area. This is because for the Jaccard, Sørensen, Ružička and percentage difference dissimilarities, which are non-Euclidean, the sum of all pairwise dissimilarities over a study

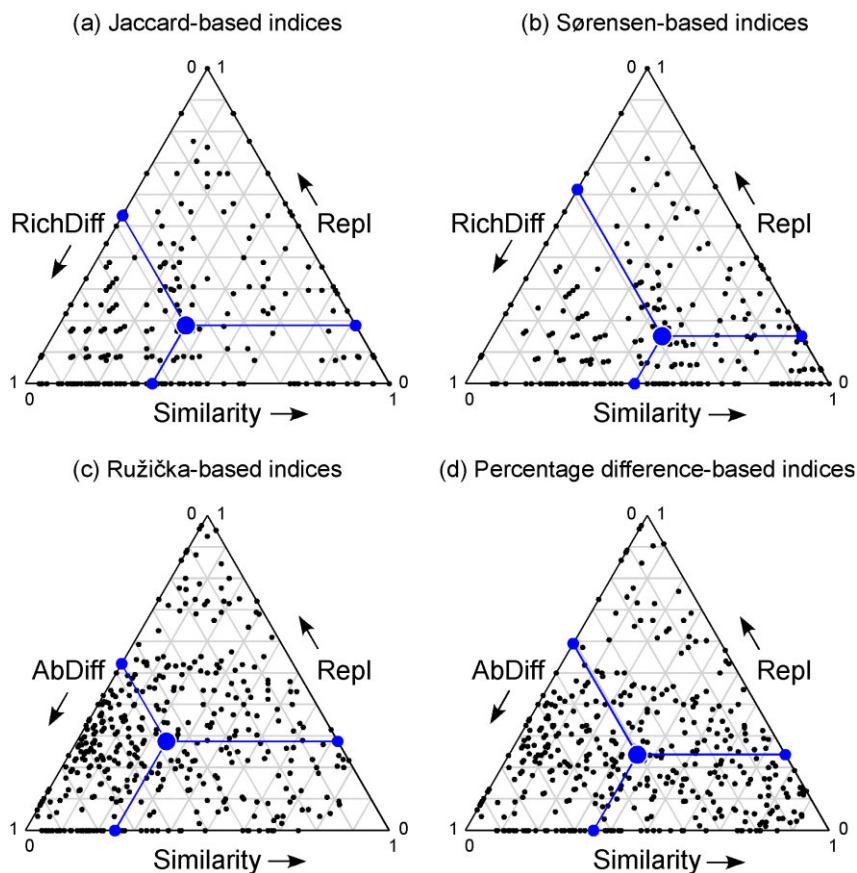


Figure 4 Triangular plots (simplices) of the relationships among the 406 pairs of sites for the Doubs River fish data. Each point (black dot) represents a pair of sites. Its position is determined by a triplet of values from the $S = (1 - D)$ (similarity), **Repl** (replacement), **RichDiff** and **AbDiff**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

area, divided by $n(n - 1)$, estimates the beta diversity of the area. Different communities in the same area, as well as a community of interest in two or more study areas, may be compared using these ratios. Some communities may be dominated by replacement processes, others by richness/abundance difference processes, and this gives ecologists insights into the influence of environmental variation on communities and the importance of species interactions.

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

The Case study illustrates how the calculations are done. The last methodological section showed that local contributions can be computed for the replacement and richness difference components of beta diversity, and that these contributions can be plotted on a map of the study 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 comprising 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 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, 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 data 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 grateful 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 referee, who sent useful comments that greatly helped in improving the manuscript. This research was funded by a Natural Sciences and Engineering Research Council of Canada (NSERC) research grant to P. Legendre.

REFERENCES

- Anderson, M.J., Ellingsen, K.E. & McArdle, B.H. (2006) Multivariate dispersion as a measure of beta diversity. *Ecology Letters*, **9**, 683–693.
- Atmar, W. & Patterson, B.D. (1993) The measure of order and disorder in the distribution of species in fragmented habitat. *Oecologia*, **96**, 373–382.
- 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. (2013a) Separating the two components of abundance-based dissimilarity: balanced changes in abundance vs. abundance gradients. *Methods in Ecology and Evolution*, **4**, 552–557.
- Baselga, A. (2013b) Multiple site dissimilarity quantifies compositional heterogeneity among several sites, while average pairwise dissimilarity may be misleading. *Ecography*, **36**, 124–128.
- Borcard, D., Gillet, F. & Legendre, P. (2011) *Numerical ecology with R*. Use R! series. Springer Science, New York.
- 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.
- Cardoso, P., Rigal, F., Carvalho, J.C., Fortelius, M., Borges, P.A.V., Podani, J. & Schmera, D. (2014) Partitioning taxon, phylogenetic and functional beta diversity into replacement and richness difference components. *Journal of Biogeography*, **41**, 749–761.
- 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.
- Dobrovolski, R., Melo, A.S., Cassemiro, F.A.S. & Diniz-Filho, J.A.F. (2012) Climatic history and dispersal ability explain the relative importance of turnover and nestedness components of beta diversity. *Global Ecology and Biogeography*, **21**, 191–197.
- Dray, S. [with contributions from Legendre, P. & Blanchet, G.] (2012) *Packfor: forward selection with permutation (Canoco p. 46)*. R package version 0.0-8. Available at: https://r-forge.r-project.org/R/?group_id=195.
- Gower, J.C. (1966) Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika*, **53**, 325–338.
- Harrison, S., Ross, S.J. & Lawton, J.H. (1992) Beta-diversity on geographic gradients in Britain. *Journal of Animal Ecology*, **61**, 151–158.
- 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. & Anderson, M.J. (1999) Distance-based redundancy analysis: testing multispecies responses in multifactorial ecological experiments. *Ecological Monographs*, **69**, 1–24.
- 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 edn. Elsevier Science BV, Amsterdam.

- Legendre, P., Borcard, D. & Peres-Neto, P.R. (2005) Analyzing beta diversity: partitioning the spatial variation of community composition data. *Ecological Monographs*, **75**, 435–450.
- 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., Tedesco, P.A., Hugué, B., Beauchard, O., Dürr, H.H., Brosse, S. & Oberdorff, T. (2011) Partitioning global patterns of freshwater fish beta diversity reveals contrasting signatures of past climate changes. *Ecology Letters*, **14**, 325–334.
- Leprieur, F., Albouy, C., De Bortoli, J., Cowman, P.F., Bellwood, D.R. & Mouillot, D. (2012) Quantifying phylogenetic beta diversity: distinguishing between 'true' turnover of lineages and phylogenetic diversity gradients. *PLoS ONE*, **7**, e42760, doi:10.1371/journal.pone.0042760
- McArdle, B.H. & Anderson, M.J. (2001) Fitting multivariate models to community data: a comment on distance-based redundancy analysis. *Ecology*, **82**, 290–297.
- Pelissier, R., Couteron, P., Dray, S. & Sabatier, D. (2003) Consistency between ordination techniques and diversity measurements: two strategies for species occurrence data. *Ecology*, **84**, 242–251.
- Podani, J. & Schmera, D. (2011) A new conceptual and methodological framework for exploring and explaining pattern in presence–absence data. *Oikos*, **120**, 1625–1638.
- 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.
- Schmera, D. & Podani, J. (2011) Comments on separating components of beta diversity. *Community Ecology*, **12**, 153–160.
- Verneaux, J. (1973) Cours d'eau de Franche-Comté (Massif du Jura). Recherches écologiques sur le réseau hydrographique du Doubs – Essai de biotypologie. *Annales Scientifiques de l'Université de Franche-Comté, Biologie Animale*, **3**, 1–260.
- Villéger, S., Grenouillet, G. & Brosse, S. (2013) Decomposing functional β -diversity reveals that low functional β -diversity is driven by low functional turnover in European fish assemblages. *Global Ecology and Biogeography*, **22**, 671–681.
- Whittaker, R.H. (1952) A study of summer foliage insect communities in the Great Smoky Mountains. *Ecological Monographs*, **22**, 1–44.
- 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 B: Biological Sciences*, **263**, 579–588.

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 & Anderson (2001).

Appendix S5 R function to compute LCB_D 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.

BIOSKETCH

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

Editor: José Alexandre Diniz-Filho