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 $r_1p_1 = r_2p_2$,

$$\beta_{-3} = \frac{r_1 + r_2 - r_1 + r_2 - 2r_2p_2}{r_1 + r_2 - r_2p_2} = \frac{2r_2(1-p_2)}{r_1 + r_2 - r_2p_2}$$

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

$$\beta_{-3} = \frac{2r_2(1-p_2)}{\frac{r_2p_2}{p_1} + r_2 - r_2p_2} = \frac{2r_2(1-p_2)}{\frac{r_2p_2}{p_1} + \frac{r_2p_1}{p_1} - \frac{r_2p_1p_2}{p_1}} = \frac{2p_1(1-p_2)}{p_1 + p_2 - p_1p_2}$$

If $r_1 < r_2$,

$$\beta_{-3} = \frac{r_1 + r_2 + r_1 - r_2 - 2r_1p_1}{r_1 + r_2 - r_2p_2} = \frac{2r_1(1-p_1)}{r_1 + r_2 - r_2p_2}$$

Since $r_2 = r_1p_1/p_2$,

$$\beta_{-3} = \frac{2r_1(1-p_1)}{r_1 + \frac{r_1p_1}{p_2} - \frac{r_1p_1p_2}{p_2}} = \frac{2r_1(1-p_1)}{\frac{r_1p_2}{p_2} + \frac{r_1p_1}{p_2} - \frac{r_2p_1p_2}{p_2}} = \frac{2p_2(1-p_1)}{p_1 + p_2 - p_1p_2}$$

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

$\beta_{-3} = \dfrac{2\min(p_1,p_2)(1-\max(p_1,p_2)}{p_1 + p_2 - p_1p_2}$, which is independent of r_1 and r_2.

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

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

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

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

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) (*$Repl_{BJ}$* 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.

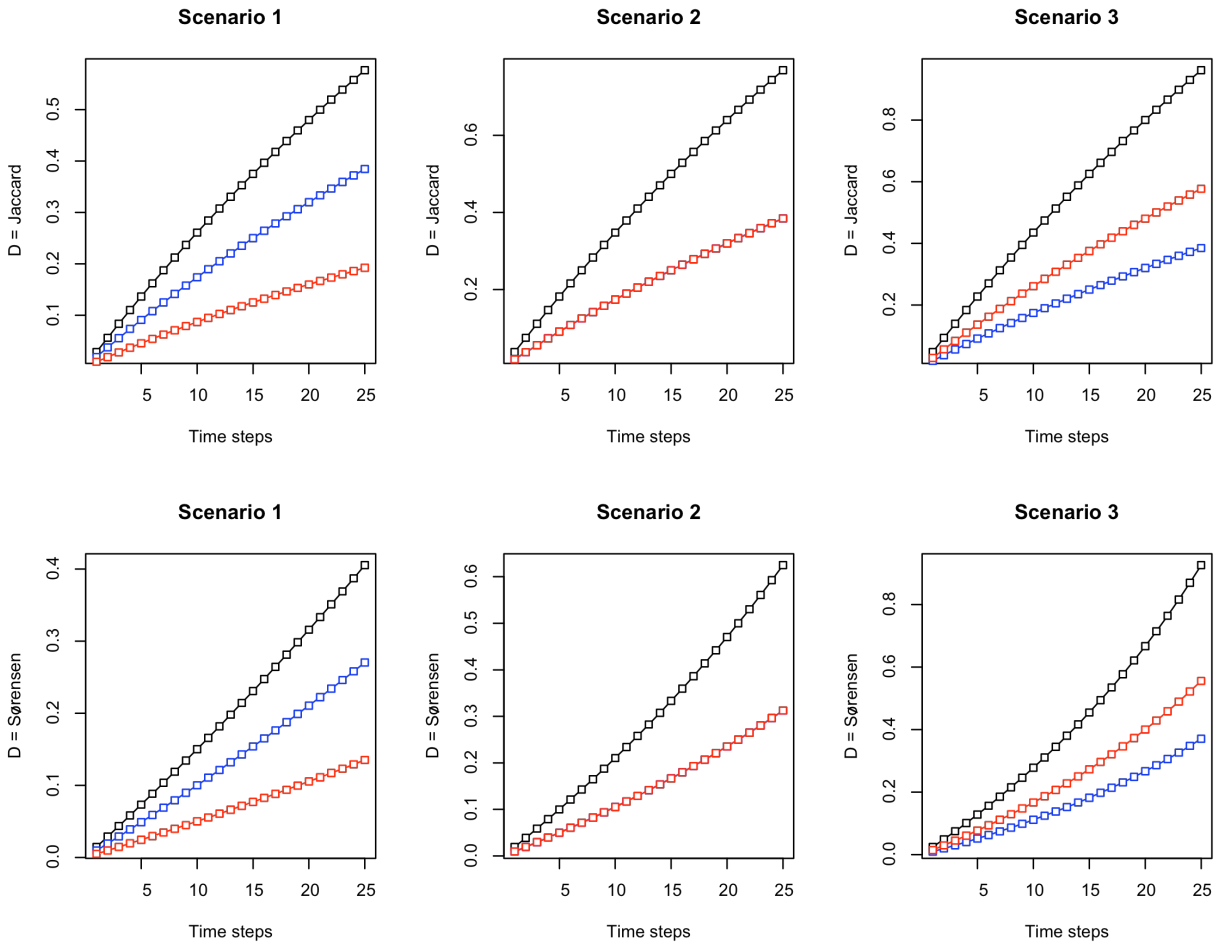

Figure S2.1 Simulation results for binary data. Podani & Schmera (2011) indices decomposing the Jaccard (D_J, upper row) and the Sørensen (D_S, lower row) dissimilarities. In each graph, the *initial* species data are compared to those at the given step along the simulated gradient (e.g. time). Black: 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.

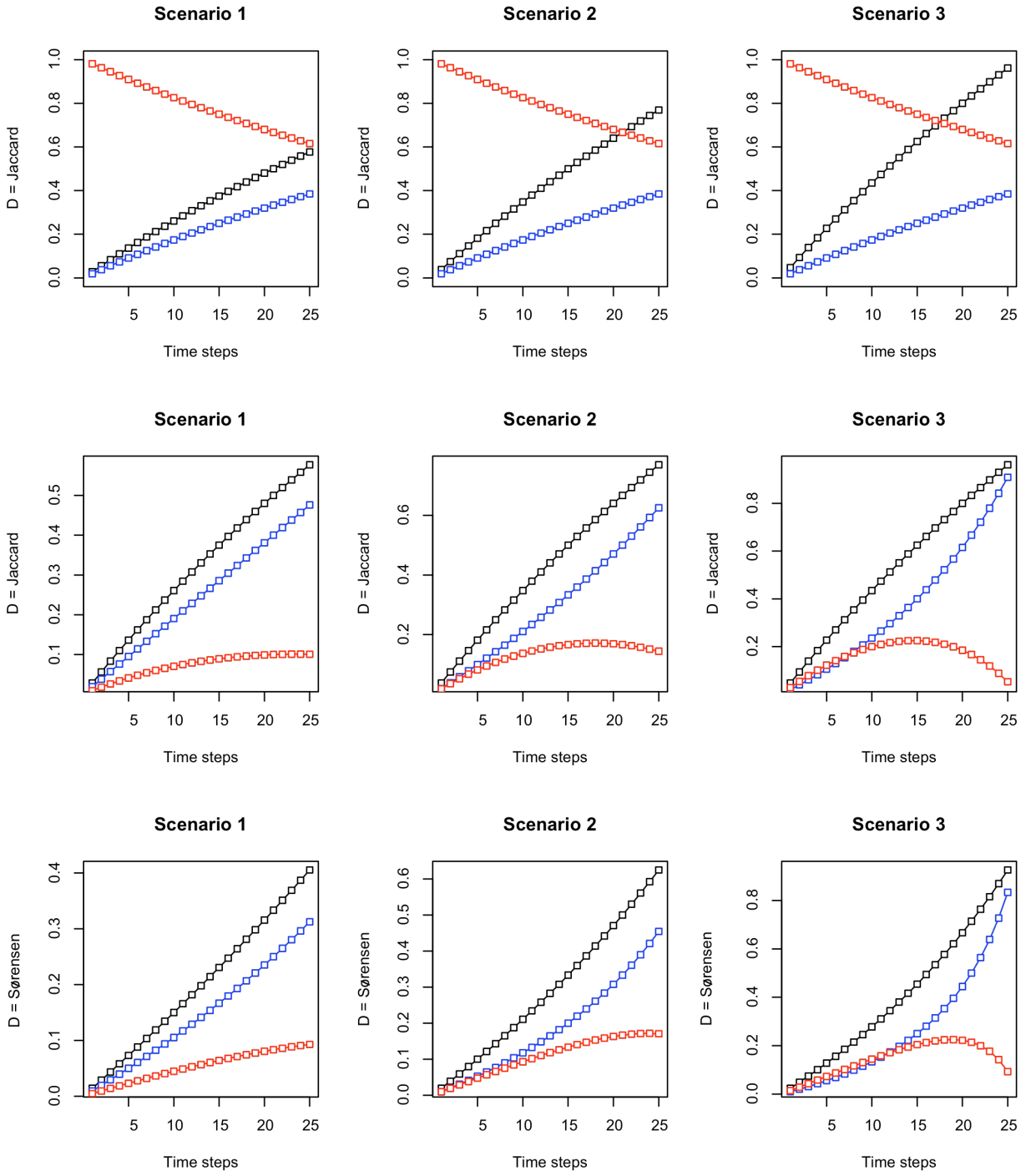

Figure S2.2 Simulation results for binary data. Upper row: Podani & Schmera (2011) decomposition of D_J into *Repl* and nestedness (N_{rel}); middle row: Baselga-family decomposing D_J; lower row: Baselga-family decomposing D_S. Black: dissimilarity values; blue: *Repl* indices; red: *nestedness* indices.

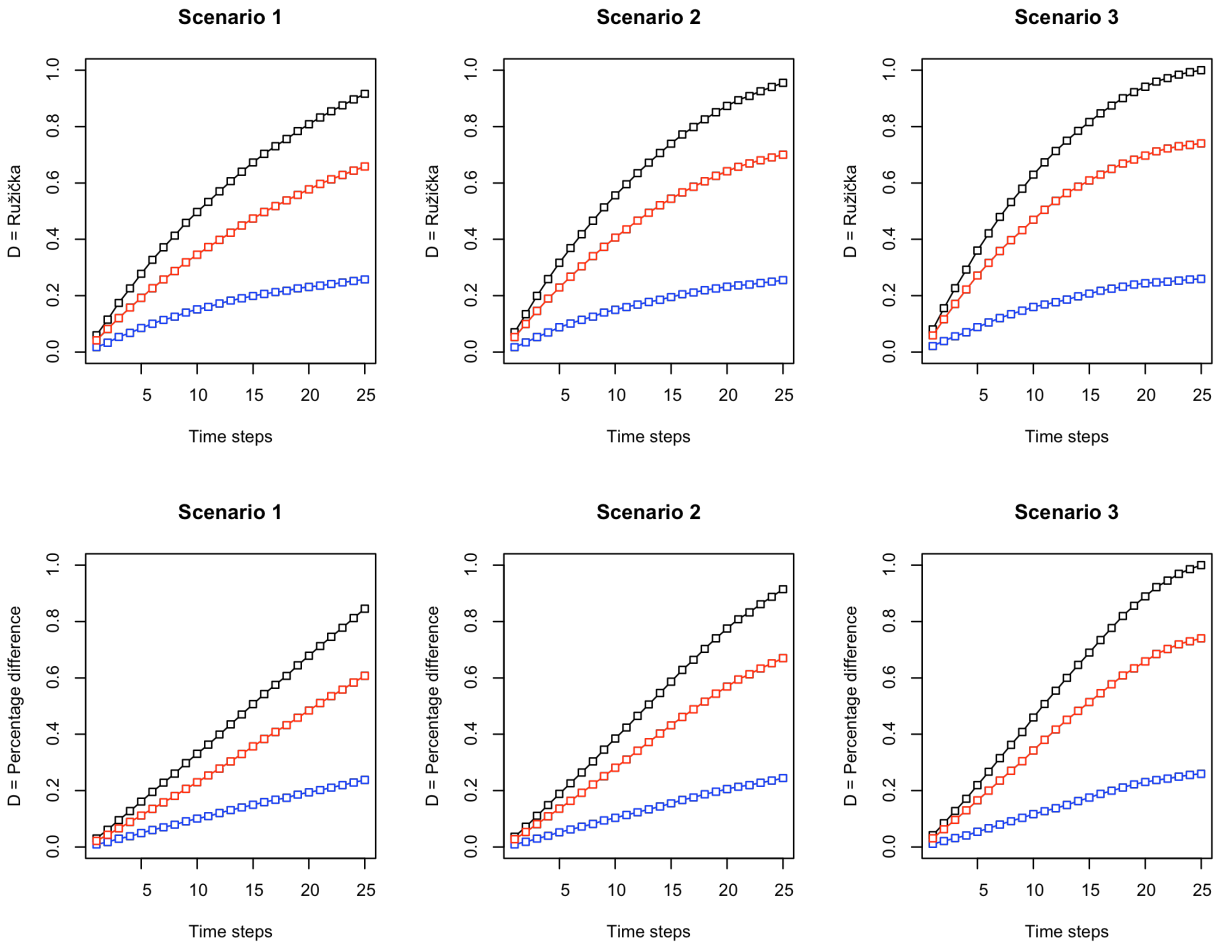

Figure S2.3 Simulation results for quantitative data. Podani-family indices decomposing the Ružička (D_R, upper row) and the percentage difference ($D_{\%diff}$, lower row) dissimilarities. In each graph, the initial species abundance data are compared to those at the given step along the simulated gradient (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.

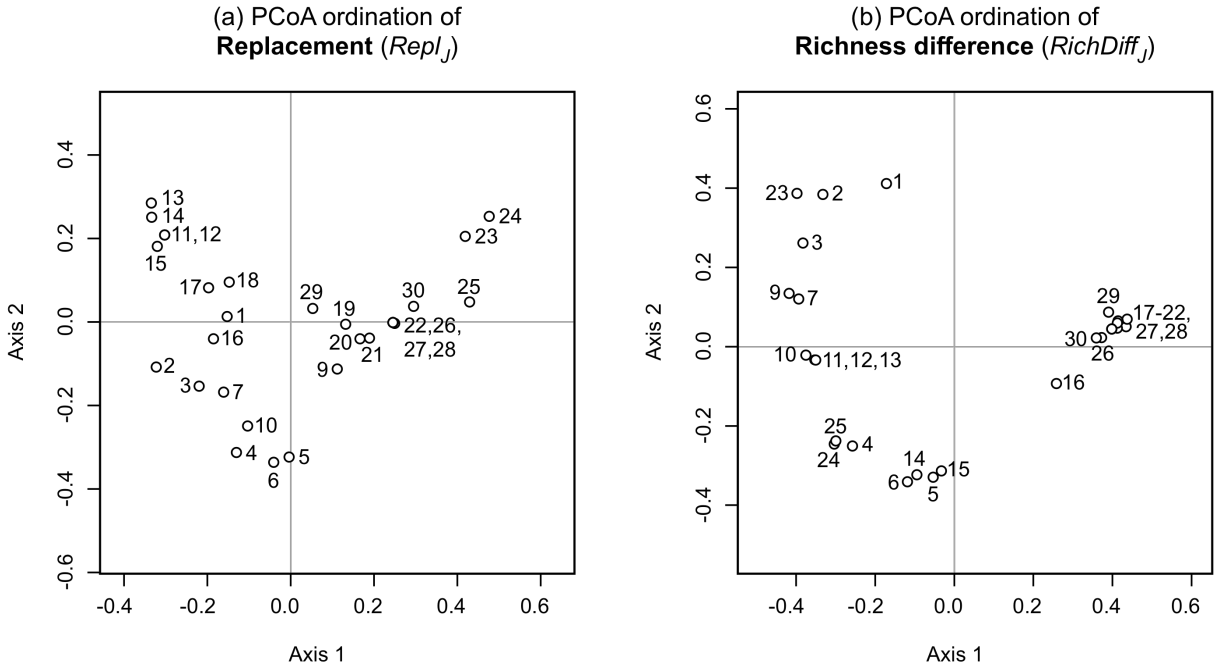

Figure S6.1 Principal coordinate ordinations of the Podani-family replacement and richness difference indices based on Jaccard dissimilarity for the fish case study data. PCoA of square-root transformed (a) replacement (*Repl*$_J$) and (b) richness difference (*RichDiff*$_J$) indices. These matrices 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.

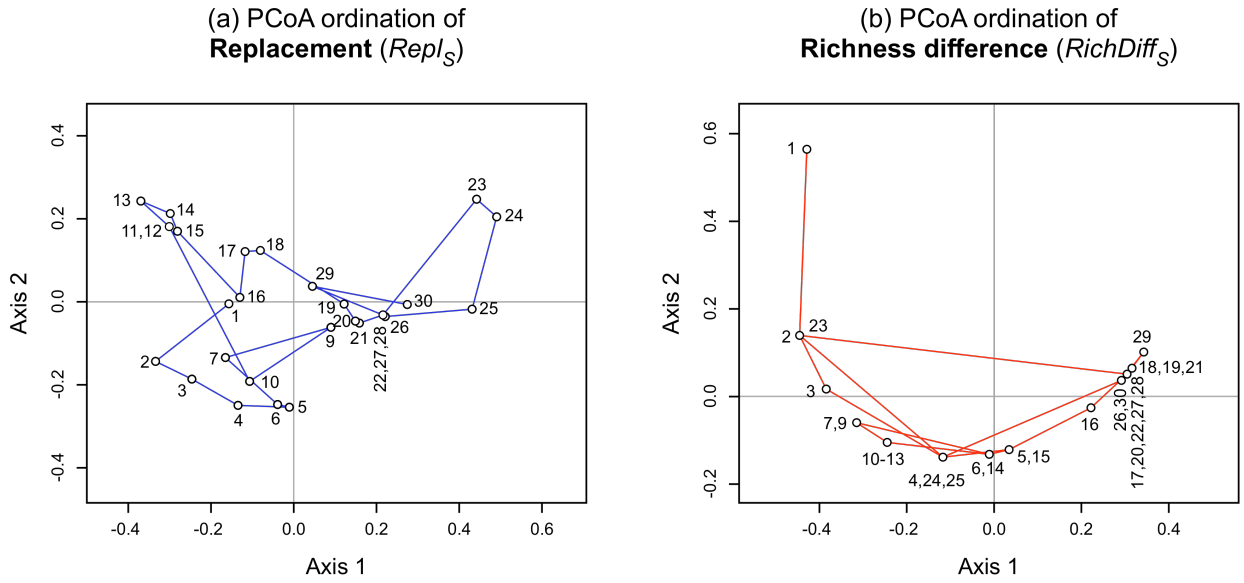

Figure S6.2 Principal coordinate ordinations of the Podani-family replacement and richness difference indices based on the Sørensen dissimilarity for the fish case study data. (a) PCoA of square-rooted replacement (*Repl*$_S$) and (b) of untransformed richness difference (*RichDiff*$_S$) indices. The **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.