

Partitioning abundance-based multiple-site dissimilarity into components: balanced variation in abundance and abundance gradients

Andrés Baselga*

Departamento de Zoología, Genética y Antropología Física, Facultad de Biología, Universidad de Santiago de Compostela, c/Lope Gómez de Marzoa s/n, 15782 Santiago de Compostela, Spain

Summary

1. Abundance-based assemblage dissimilarity can be partitioned into components accounting for (i) balanced variation in abundance, whereby the individuals of some species in one site are substituted by the same number of individuals of different species in another site, and (ii) abundance gradients, whereby some individuals are lost from one site to the other.
2. Such a method was available for pairwise dissimilarity, but not for multiple-site dissimilarity.
3. New multiple-site extensions of Bray–Curtis and Ruzicka indices based on the formulation of multiple-site analogues of matching components (i.e. the intersection and the relative complements among multiple sites in terms of species abundances) are introduced.
4. These new indices can be partitioned into balanced-variation and abundance-gradients components.
5. An example assessing multiple-site dissimilarity of birds and butterflies in a mosaic of habitats is shown to illustrate the usefulness of comparing incidence- and abundance-based multiple-site dissimilarity and its components to characterize biotic heterogeneity across multiple sites.
6. The method may be generally useful when the question of interest is the overall abundance-based dissimilarity across multiple units (in space, time or other), as separating the balance-variation (i.e. substitution) and abundance-gradients (i.e. subsets) components of dissimilarity can shed light on the processes behind variation in species abundances.

Key-words: abundance, beta diversity, Bray–Curtis, multiple-site dissimilarity, replacement, subset, turnover

Introduction

The number of co-occurrences in three, four, and five (and so on) sites has an obvious impact on the total amount of dissimilarity in a set of multiple sites (or times, or, more generally, units). Therefore, averaging pairwise dissimilarity measures across multiple pairs of sites is a suboptimal approach for estimating the overall dissimilarity among more than three sites, that is, multiple-site dissimilarity (Diserud & Ødegaard 2007; Baselga 2013a).

Classical measures of strict sense beta diversity (i.e. gamma divided by alpha, Whittaker 1960; Jost 2007) do account for patterns of co-occurrence in more than two sites and are, thus, effective measures of multiple-site dissimilarity. These classical beta diversity measures can be standardized to be independent of the number of sites involved in the calculation, becoming multiple-site extensions of well-known dissimilarity indices (Chao, Chiu & Hsieh 2012). Moreover, because multiplicative beta diversity can be computed using Hill numbers (Jost 2007) when diversity is computed with Hill numbers of order $q > 0$,

beta diversity accounts for variation in species abundance in addition to variation in species occurrence. Likewise, the multiple-site dissimilarity indices derived from beta diversity of $q > 0$ (i.e. the multiple-site extension of Horn and Morisita–Horn indices, see Chao, Chiu & Hsieh 2012) do account for variation in species composition in terms of species abundances and occurrences. However, all these beta diversity measures and derived dissimilarity indices operate with relative abundances instead of absolute abundances. In consequence, these measures are unable to quantify whether some assemblages are subsets of others in terms of species abundances (Baselga 2013b). Some dissimilarity indices have been extended to account for absolute abundances (Chiu, Jost & Chao 2014) and, more recently, a multiple-site extension of Bray–Curtis dissimilarity has been introduced (Chao & Chiu 2016). However, no solution is available to separate the components of abundance-based multiple-site dissimilarity caused by (i) balanced variation in abundance and (ii) abundance gradients in which one assemblage is a subset of another (Fig. 1).

These components are analogous to the replacement and nestedness-resultant components of incidence-based dissimilarity (Baselga 2010, 2012; Baselga & Leprieur 2015). The properties of the indices accounting for these components have

*Correspondence author. E-mail: andres.baselga@usc.es

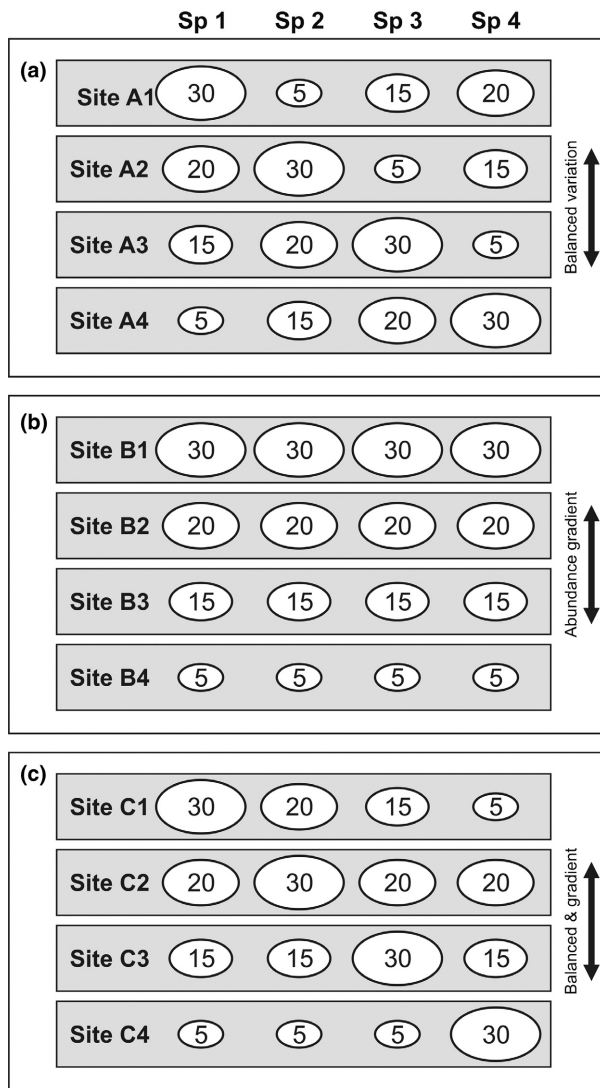


Fig. 1. Idealized examples involving 4 sites and 4 species in each situation (a–c), showing that abundance-based multiple-site dissimilarity may be linked to two antithetic phenomena. In all situations the total number of individuals (280), the mean abundance per site or per species (70) are exactly the same and, in consequence, multiple-site dissimilarity is identical in all situations. As a result, a previous multiple-site extension of Bray–Curtis dissimilarity (i.e. $\Delta_{B\text{Cnorm}}$, Chao & Chiu 2016) and the unpartitioned Bray–Curtis measures developed in latter text yield exactly the same value in the three situations (a–c). However, in situation A, the total species abundance is the same in the four sites (70) and variation in species abundances is perfectly balanced, that is, abundance increases in some species are exactly the same as abundance decreases in other species, whereas in situation b all species decrease its abundance following a gradient from B1 to B4. In situation c, both patterns are mixed, as the variation in species abundances is only partially balanced and there are still differences in total abundance among sites. These contrasting patterns can be distinguished by partitioning total abundance-based multiple-site dissimilarity using the indices developed in this paper, which yield different results in the three situations (a–c).

been discussed in the aforementioned papers, but the basic requirements could be summarized as follows. The replacement component (i) must be independent or richness difference

(i.e. not mathematically constrained by it), and (ii) must yield the maximum value (1) when there are no species present in more than one site (property P5 in Legendre & De Cáceres 2013). In turn, the nestedness-resultant component (i) must yield the minimum value (0) when there are no species in common, and (ii) must account for richness differences conditioned to the degree of nestedness (i.e. the extent to which the poorest assemblage is a subset of the richest assemblage). A method to separate these antithetic components of abundance-based dissimilarity has been developed for pairwise-dissimilarity based on the Bray–Curtis index (Baselga 2013b) and then extended to the Ruzicka index (Legendre 2014). In this case, the aforementioned basic requirements can be expressed as follows. The balanced variation component (i) must be independent or total abundance difference (i.e. not mathematically constrained by it), and (ii) must yield the maximum value (1) when no species is present in more than one site (property P5 in Legendre & De Cáceres 2013). In turn, the abundance-gradient component (i) must yield the minimum value (0) when there are no species in common, and (ii) must account for abundance differences conditioned to the degree to which the assemblage with lower abundance is a subset of the assemblage with higher abundance. The purpose of this paper is to implement a method allowing the separation of the balanced-variation and abundance-gradient components of abundance-based multiple-site dissimilarity.

Multiple-site matching components and novel indices

The partition of pairwise abundance-based dissimilarity has been introduced in a previous paper (Baselga 2013b). The partition relies in the fact that the intersection (*a*) and complement (*b* and *c*) matching components used to build incidence-based dissimilarity (or similarity) indices, can be generalized to account for abundance (Tamas, Podani & Csontos 2001). The intersection (*A* component) and the relative complements (*B* and *C* components) in terms of species abundances can be formulated as follows:

$$A = \sum_i^S \min(x_{ij}, x_{ik}) \quad \text{eqn 1a}$$

$$B = \sum_i^S x_{ij} - \min(x_{ij}, x_{ik}) \quad \text{eqn 1b}$$

$$C = \sum_i^S x_{ik} - \min(x_{ij}, x_{ik}) \quad \text{eqn 1c}$$

where x_{ij} is the abundance of species *i* on site *j*, and x_{ik} is the abundance of species *i* on site *k*, and *S* is the total number of species. Therefore, *A* is the number of individuals of each species that exists in both sites *j* and *k*, whereas *B* and *C* are the number of individuals that are unique to sites *j* and *k* respectively. This formulation is useful because it makes obvious the relationship between the Sørensen (β_{Sor}) and Bray–Curtis (β_{bc}) indices because

Table 1. Formulation of all the novel abundance-based multiple-site indices

Index	Formula
β_{BC}	$\frac{2[T_{\text{AB}} - \sum_i^S \max(x_{ij}, x_{ik}, x_{it}, x_{im}, x_{in}, \dots)] + \sum_{j < k} \min(x_{ij} - \min(x_{ij}, x_{ik})), \sum_i^S [x_{ik} - \min(x_{ij}, x_{ik})]] + \sum_{j < k} \max(\sum_i^S [x_{ij} - \min(x_{ij}, x_{ik})], \sum_i^S [x_{ik} - \min(x_{ij}, x_{ik})])}{[T_{\text{AB}} - \sum_i^S \max(x_{ij}, x_{ik}, x_{it}, x_{im}, x_{in}, \dots)] + \sum_{j < k} \min(x_{ij} - \min(x_{ij}, x_{ik})), \sum_i^S [x_{ik} - \min(x_{ij}, x_{ik})]] + \sum_{j < k} \max(\sum_i^S [x_{ij} - \min(x_{ij}, x_{ik})], \sum_i^S [x_{ik} - \min(x_{ij}, x_{ik})])}$
β_{BCBAL}	$\frac{\sum_{j < k} \min(\sum_i^S [x_{ij} - \min(x_{ij}, x_{ik})], \sum_i^S [x_{ik} - \min(x_{ij}, x_{ik})])}{[T_{\text{AB}} - \sum_i^S \max(x_{ij}, x_{ik}, x_{it}, x_{im}, x_{in}, \dots)] + \sum_{j < k} \min(x_{ij} - \min(x_{ij}, x_{ik})), \sum_i^S [x_{ik} - \min(x_{ij}, x_{ik})]}$
β_{BCGRA}	$\frac{2[T_{\text{AB}} - \sum_i^S \max(x_{ij}, x_{ik}, x_{it}, x_{im}, x_{in}, \dots)] + \sum_{j < k} \min(x_{ij} - \min(x_{ij}, x_{ik})), \sum_i^S [x_{ik} - \min(x_{ij}, x_{ik})]] - \sum_{j < k} \min(x_{ij} - \min(x_{ij}, x_{ik})), \sum_i^S [x_{ik} - \min(x_{ij}, x_{ik})]}{[T_{\text{AB}} - \sum_i^S \max(x_{ij}, x_{ik}, x_{it}, x_{im}, x_{in}, \dots)] + \sum_{j < k} \min(x_{ij} - \min(x_{ij}, x_{ik})), \sum_i^S [x_{ik} - \min(x_{ij}, x_{ik})]] + \sum_{j < k} \max(\sum_i^S [x_{ij} - \min(x_{ij}, x_{ik})], \sum_i^S [x_{ik} - \min(x_{ij}, x_{ik})])}$ $\times \frac{[T_{\text{AB}} - \sum_i^S \max(x_{ij}, x_{ik}, x_{it}, x_{im}, x_{in}, \dots)]}{[T_{\text{AB}} - \sum_i^S \max(x_{ij}, x_{ik}, x_{it}, x_{im}, x_{in}, \dots)] + \sum_{j < k} \min(x_{ij} - \min(x_{ij}, x_{ik})), \sum_i^S [x_{ik} - \min(x_{ij}, x_{ik})]}$
β_{RUZ}	$\frac{\sum_{j < k} \min(\sum_i^S [x_{ij} - \min(x_{ij}, x_{ik})], \sum_i^S [x_{ik} - \min(x_{ij}, x_{ik})]) + \sum_{j < k} \max(\sum_i^S [x_{ij} - \min(x_{ij}, x_{ik})], \sum_i^S [x_{ik} - \min(x_{ij}, x_{ik})])}{[T_{\text{AB}} - \sum_i^S \max(x_{ij}, x_{ik}, x_{it}, x_{im}, x_{in}, \dots)] + \sum_{j < k} \min(x_{ij} - \min(x_{ij}, x_{ik})), \sum_i^S [x_{ik} - \min(x_{ij}, x_{ik})]] + \sum_{j < k} \max(\sum_i^S [x_{ij} - \min(x_{ij}, x_{ik})], \sum_i^S [x_{ik} - \min(x_{ij}, x_{ik})])}$
β_{RUZBAL}	$\frac{2 \sum_{j < k} \min(\sum_i^S [x_{ij} - \min(x_{ij}, x_{ik})], \sum_i^S [x_{ik} - \min(x_{ij}, x_{ik})])}{[T_{\text{AB}} - \sum_i^S \max(x_{ij}, x_{ik}, x_{it}, x_{im}, x_{in}, \dots)] + 2 \sum_{j < k} \min(x_{ij} - \min(x_{ij}, x_{ik})), \sum_i^S [x_{ik} - \min(x_{ij}, x_{ik})]}$
β_{RUZGRA}	$\frac{\sum_{j < k} \max(\sum_i^S [x_{ij} - \min(x_{ij}, x_{ik})], \sum_i^S [x_{ik} - \min(x_{ij}, x_{ik})]) - \sum_{j < k} \min(x_{ij} - \min(x_{ij}, x_{ik})), \sum_i^S [x_{ik} - \min(x_{ij}, x_{ik})]}{[T_{\text{AB}} - \sum_i^S \max(x_{ij}, x_{ik}, x_{it}, x_{im}, x_{in}, \dots)] + \sum_{j < k} \min(x_{ij} - \min(x_{ij}, x_{ik})), \sum_i^S [x_{ik} - \min(x_{ij}, x_{ik})]] + \sum_{j < k} \max(\sum_i^S [x_{ij} - \min(x_{ij}, x_{ik})], \sum_i^S [x_{ik} - \min(x_{ij}, x_{ik})])}$ $\times \frac{[T_{\text{AB}} - \sum_i^S \max(x_{ij}, x_{ik}, x_{it}, x_{im}, x_{in}, \dots)]}{[T_{\text{AB}} - \sum_i^S \max(x_{ij}, x_{ik}, x_{it}, x_{im}, x_{in}, \dots)] + 2 \sum_{j < k} \min(x_{ij} - \min(x_{ij}, x_{ik})), \sum_i^S [x_{ik} - \min(x_{ij}, x_{ik})]}$

$$\beta_{\text{sor}} = \frac{b + c}{2a + b + c} \quad \text{eqn 2}$$

and,

$$\beta_{\text{bc}} = \frac{\sum_i^S |x_{ij} - x_{ik}|}{\sum_i^S [x_{ij} - x_{ik}]} = \frac{\sum_i^S [x_{ij} + x_{ik} - 2 \times \min(x_{ij}, x_{ik})]}{\sum_i^S [x_{ij} + x_{ik}]} = \frac{B + C}{2A + B + C} \quad \text{eqn 3}$$

Likewise, the Ruzicka index of dissimilarity (β_{ruz}) can be formulated by using the same abundance-based matching components (see Legendre 2014), showing its relationship with the Jaccard index (β_{jac}), because

$$\beta_{\text{jac}} = \frac{b + c}{a + b + c} \quad \text{eqn 4}$$

and,

$$\beta_{\text{ruz}} = \frac{B + C}{A + B + C} \quad \text{eqn 5}$$

These indices can be partitioned into components of dissimilarity just by differentiating the maximum and minimum of B and C matching components as already shown elsewhere (Baselga 2013b; Legendre 2014).

It should be noted that the intersection component A of the pairwise Bray–Curtis dissimilarity index can be described as the amount of abundance that is shared by the two assemblages. So, even if individuals themselves are not common to both sites, if species i has 10 individuals in locality 1 and 5 individuals in locality 2, the amount of abundance that is common to both sites is 5 individuals, while the complements are 5 and 0 respectively. This can also be seen as a way to weight the shared species presences by their respective abundances: in the example above species i contributes with 1 to the incidence-based intersection component (a) and with 0 to the relative complements (b, c), but contributes with 5 to the abundance-based intersection component (A , that is, the agreement in abundance between both sites) and with 5 and 0 to the relative components (B and C respectively, that is, the disagreement in abundance between both sites). In turn, if species j has 100 individuals in locality 1 and 50 individuals in locality 2, it would contribute the same to the incidence-based components, but would weight more in the abundance-based indices given that the difference in abundance is 10 times larger.

To build multiple-site analogues of these abundance–abundance based matching components, we need to follow the same strategy as for incidence-based indices (Diserud & Ødegaard 2007). For a pair of sites, the intersection matching component A can be formulated in the following way:

$$A = \sum_i^S \min(x_{ij}, x_{ik}) = \sum_i^S x_{ij} + \sum_i^S x_{ik} - \sum_i^S \max(x_{ij}, x_{ik}) \quad \text{eqn 6}$$

The multiple-site intersection component accounting for abundances can be formulated because we can generalize the

intersection, using the inclusion-exclusion principle (Erickson 1996):

$$\sum_i^S \sum_j^N x_{ij} = \sum_i^S \max(x_{ij}, x_{ik}, x_{il}, x_{im}, x_{in} \dots) + \sum_i^S \left[\sum_{j < k} \min(x_{ij}, x_{ik}) - \sum_{j < k < l} \min(x_{ij}, x_{ik}, x_{il}) + \sum_{j < k < l < m} \min(x_{ij}, x_{ik}, x_{il}, x_{im}) - \sum_{j < k < l < m < n} \min(x_{ij}, x_{ik}, x_{il}, x_{im}, x_{in}) \dots \right] \quad \text{eqn 7}$$

The expression before the equality is the sum of abundances of all species in all sites, that is, the total abundance in the data set [T_{AB}]. The first term after the equality [$\sum_i^S \max(x_{ij}, x_{ik}, x_{il}, x_{im}, x_{in} \dots)$] is the sum across species of all the respective species maximum abundances across sites. The remaining terms account for the multiple-site intersection in abundance. Therefore, the multiple-site intersection (A_M) is the sum across species of the abundance of each species that is common to pairs of sites, minus the abundance of each species that is common to sets of three sites, plus the abundance of each species that is common to sets of four sites, and so on, and can be expressed as:

$$A_M = \sum_i^S \left[\sum_{j < k} \min(x_{ij}, x_{ik}) - \sum_{j < k < l} \min(x_{ij}, x_{ik}, x_{il}) + \sum_{j < k < l < m} \min(x_{ij}, x_{ik}, x_{il}, x_{im}) - \sum_{j < k < l < m < n} \min(x_{ij}, x_{ik}, x_{il}, x_{im}, x_{in}) \dots \right] \quad \text{eqn 8a}$$

And using equation 7, the A_M component can be simplified as follows:

$$A_M = T_{\text{AB}} - \sum_i^S \max(x_{ij}, x_{ik}, x_{il}, x_{im}, x_{in} \dots) \quad \text{eqn 8b}$$

In other words, A_M is the total abundance in the data set minus the sum across species of the maximum abundance of each species across sites. This quantity equals zero when the intersection in abundances equals total abundance (i.e. no abundance is common to two or more sites, implying that no species is present in more than one site) and reaches its maximum when all the species have exactly the same abundance in all the sites. Importantly, it accounts for patterns of intersection among more than two sites, so it is a proper multiple-site extension of the A matching component.

Having the multiple-site extension of the A matching component, it is easy to build a multiple-site Bray–Curtis similarity measure, using the same approach as Diserud & Ødegaard (2007), that is, dividing the multiple-site intersection in

abundances by the total sum of abundances. This measure ranges between 0 (when all the species have exactly the same abundance in all the sites) and $(N-1)/N$ (when no species is present in more than one site), so it has to be normalized by multiplying by $N/(N-1)$ to make it range between 0 and 1:

$$S_{BC-1} = \frac{N}{N-1} \times \frac{T_{AB} - \sum_i^S \max(x_{ij}, x_{ik}, x_{il}, x_{im}, x_{in}, \dots)}{T_{AB}} = \frac{N \times A_M}{(N-1) \times T_{AB}} \quad \text{eqn 9a}$$

Expressed as a dissimilarity index, the new multiple-site extension of the Bray–Curtis dissimilarity is thus:

$$\beta_{BC-1} = \frac{[N \times \sum_i^S \max(x_{ij}, x_{ik}, x_{il}, x_{im}, x_{in}, \dots)] - T_{AB}}{(N-1) \times T_{AB}} \quad \text{eqn 9b}$$

I use the notation β_{BC-1} for this index to stress that this measure is the abundance-based extension of β_{-1} (Harrison, Ross & Lawton 1992). The problem of this approach is that it is not possible to separate the components of abundance-based multiple-site dissimilarity associated to (i) balanced variation in abundance and, (ii) abundance gradients.

To do this, the same approach as for the incidence-based matching components (Baselga 2010) must be taken, using the *B* and *C* multiple-site analogue components. Given the need for separating the maximum and minimum complements in order to account for subset patterns (i.e. unidirectional abundance gradients, in this case), it is not possible to generalize the formulation, so the sum of pairwise *B* and *C* components must be used. Therefore,

$$B_M = \sum_{j < k} \min \left(\sum_i^S [x_{ij} - \min(x_{ij}, x_{ik})], \sum_i^S [x_{ik} - \min(x_{ij}, x_{ik})] \right) \quad \text{eqn 10a}$$

$$C_M = \sum_{j < k} \max \left(\sum_i^S [x_{ij} - \min(x_{ij}, x_{ik})], \sum_i^S [x_{ik} - \min(x_{ij}, x_{ik})] \right) \quad \text{eqn 10b}$$

Note that B_M is the sum of all the minimum complements across all pairs of sites, and C_M is the sum of all the maximum complements across all pairs of sites. Using these formulations as multiple-site analogues of matching components it is possible to build multiple-site extensions of Bray–Curtis and Ruzicka indices, as well as of their components of dissimilarity caused by (i) balanced variation in abundance and, (ii) abundance gradients in which one assemblage is a subset of another (Baselga 2013b), because

$$\beta_{BC} = \beta_{BC.BAL} + \beta_{BC.GRA} \equiv \frac{B_M + C_M}{2A_M + B_M + C_M} = \frac{B_M}{A_M + B_M} + \frac{C_M - B_M}{2A_M + B_M + C_M} \times \frac{A_M}{A_M + B_M} \quad \text{eqn 11}$$

and

$$\beta_{RUZ} = \beta_{RUZ.BAL} + \beta_{RUZ.GRA} \equiv \frac{B_M + C_M}{A_M + B_M + C_M} = \frac{2B_M}{A_M + 2B_M} + \frac{C_M - B_M}{A_M + B_M + C_M} \times \frac{A_M}{A_M + 2B_M} \quad \text{eqn 12}$$

The new indices (Table 1) thus allow discriminating the different patterns exemplified in Fig. 1. For the Bray–Curtis family, for example, the partitioned indices properly show that in the situation depicted in Fig. 1a the variation in abundance is perfectly balanced ($\beta_{BC} = 0.50$, $\beta_{BC.BAL} = 0.50$, $\beta_{BC.GRA} = 0.00$); in Fig. 1b the variation in abundance follows a unidirectional gradient ($\beta_{BC} = 0.50$, $\beta_{BC.BAL} = 0.00$, $\beta_{BC.GRA} = 0.50$); and in Fig. 1c the variation in abundance is a mixture of both previous patterns ($\beta_{BC} = 0.50$, $\beta_{BC.BAL} = 0.36$, $\beta_{BC.GRA} = 0.14$). Besides this key feature, the new indices β_{BC} and β_{RUZ} (and their components) are truly multiple-site indices accounting for information involving comparisons among more than two sites (Fig. 2). This is attained because the intersection component A_M is a true multiple-site matching component accounting for the agreement in abundances among multiple sites.

In principle, a similar approach could be used to decompose the multiple-site Horn and Morisita–Horn measures (i.e. dissimilarity indices related to beta diversity of order $q = 1$ and $q = 2$ in Hill numbers) into balance-variation and abundance-gradients components. However, the problem is that these indices operate with relative abundances, making it impossible to define subset patterns. A promising way forward is to expand the conventional concept of alpha so that beta diversity can quantify the differentiation among *N* sets of vectors for any measure of species importance, including absolute abundances (Chiu, Jost & Chao 2014). However, we still need to find the way to define the minimum and maximum relative complements in terms of abundances, so no satisfactory solution is still available and should be aimed for in the future.

New functions in R package betapart

The new indices can be computed, using the R package *betapart* (Baselga & Orme 2012). The following new functions have been implemented in version 1.4 of the package (Baselga *et al.* 2017) and are also available in Appendix S1 (Supporting Information):

1 `betapart.core.abund(x)` computes the basic quantities needed for computing abundance-based dissimilarity, both the pairwise and multiple-site indices. The input (*x*) is the table of abundances of species (columns) in sites (rows). The result is a list including pairwise matrices of matching components between sites. As these matrices are used in several functions and can be time consuming to calculate for large matrices, precalculating them using the `betapart.core.abund` function can markedly improve the speed of subsequent analyses. The function returns a new object of class ‘*betapart.abund*’ containing these quantities, which can be used as the input (*x*) to the following functions.

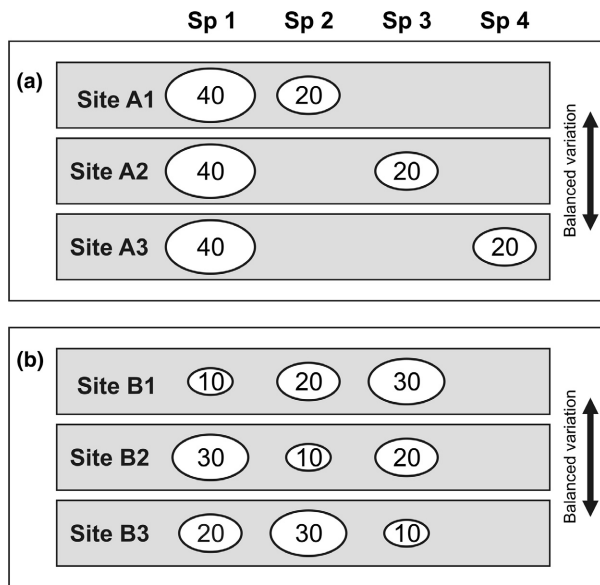


Fig. 2. Idealized examples involving 3 sites and 3 or 4 species in each situation (a and b), showing that the new indices are truly multiple-site measures of dissimilarity. Using the mean of pairwise dissimilarities would yield exactly the same result in both situations: $\text{mean}(\beta_{bc}) = 0.33$ in both situations a and b, because all pairs are equally dissimilar ($\beta_{bc} = 0.33$). In contrast, the multiple-site indices identify situation a as more dissimilar than situation b ($\Delta_{BCnorm} = 0.33$ and $\Delta_{BCnorm} = 0.25$; $\beta_{BC-1} = 0.33$ and $\beta_{BC-1} = 0.25$; $\beta_{BC} = 0.43$ and $\beta_{BC} = 0.40$ in a and b respectively).

2 `beta.multi.abund(x, index.family)`. This function computes the total dissimilarity across all sites in the species per sites abundance table, along with its components of (i) balanced variation in abundance and, (ii) abundance gradients. The input `x` may be an abundance table or a `beta.part.abund` object. The argument `index.family` selects whether the Bray–Curtis or Ruzicka index is used as a measure of total dissimilarity (β_{BC} or β_{RUZ}) and the respective components of balanced variation ($\beta_{BC.BAL}$ or $\beta_{RUZ.BAL}$) and abundance gradients ($\beta_{BC.GRA}$ and $\beta_{RUZ.GRA}$). The function returns three values, which are (i) the balanced variation in abundance and (ii) abundance gradients components, and (iii) the sum of both, that is, the total abundance-based multiple-site dissimilarity across the sites.

3 `beta.pair.abund(x, index.family)`. This function computes the same three dissimilarity metrics as for the previous function and can again be set using the argument `index.family` to use the Bray–Curtis or Ruzicka index of total dissimilarity. Rather than returning three single values as in the previous function, `beta.pair.abund` returns three matrices containing the pairwise between-site values of each component of dissimilarity. The dissimilarity matrices yielded by `beta.pair` are objects of class `dist` and can be submitted to further analyses as, for example, Mantel tests, non-metric multidimensional scaling, and cluster analysis, using other R packages as `vegan` (Oksanen *et al.* 2015) or `cluster` (Maechler *et al.* 2015).

4 `beta.sample.abund(x, index.family, sites, samples)` will resample the three abundance-based multiple-

site dissimilarities for a subset of sites of the original data frame. The arguments of this function are (`x`), the table of abundances of species (columns) in sites (rows); `sites`, the number of sites for which dissimilarity must be computed; and `samples`, the number of random samples used to calculate the distribution of dissimilarity measures. The function returns a data frame containing the individual sampled measures along with vectors of the means and standard deviations across samples of each measure.

5 `multi.bray.minus1(x)` computes the unpartitioned multiple-site extension of Bray–Curtis dissimilarity [β_{BC-1} , see Eq. 9(b)] that I derived following the approach taken by Diserrud & Odegaard (Diserrud & Ødegaard 2007) for the presence-absence equivalent, that is, multiple-site, the extension of Sørensen dissimilarity (β_{-1}).

6 `multi.bray.chao(x)` computes the unpartitioned multiple-site extension of Bray–Curtis dissimilarity [Δ_{BCnorm}] introduced by Chao & Chiu (Chao & Chiu 2016).

Comparisons to alternative approaches

I have compared the behaviour of the new indices [β_{BC-1} , β_{BC} , $\beta_{BC.BAL}$, and $\beta_{BC.GRA}$] against a previous multiple-site extension of the Bray–Curtis index [Δ_{BCnorm} , Chao & Chiu 2016] and the mean of pairwise Bray–Curtis dissimilarities [$\text{mean}(\beta_{bc})$]. To do this, I created an artificial series of presence-absence matrices (10 sites and 10 species) following two gradients of replacement and nestedness respectively (36 matrices in total). These were converted into quantitative matrices by multiplying each matrix by 4 different abundance matrices: (i) decreasing abundance from site to site (100 to 10); (ii) decreasing occupancy from species to species (100 to 10); (iii) equal abundance (100); and (iv) equal abundance (10). This yielded a set of quantitative 144 matrices spanning all degrees of abundance-based dissimilarity in terms of balanced variation and abundance gradients. To increase the variability of abundance variation, I created up to 10 000 matrices by randomly sampling these 144 matrices and multiplying each of their abundance values by a random value (drawn from sets of three): (i) 1.0, 1.1 and 1.2 (samples from 145 to 3000), (ii) 1, 2 and 3 (samples from 3001 to 6000), and (iii) 1, 50 and 100 (samples from 6000 to 10 000). This created a long series with contrasting patterns in the variation of abundances. All indices were computed for each of the 10 000 quantitative matrices of 10 species and 10 sites ($N = 10$), as well as for as for subsets of these matrices including only 5 species and 5 sites ($N = 5$) and 3 species and 3 sites ($N = 3$). The respective correlations between indices were measured with the Spearman's rank coefficient (ρ).

In the first place, as shown in Fig. 3, all unpartitioned indices were correlated and the correlation decreased with the number of localities (i.e. from $N = 3$ to 10). This was expected as all the alternative multiple-site extensions of Bray–Curtis dissimilarity converge at $N = 2$ (i.e. pairwise Bray–Curtis dissimilarity), and differences in behaviour start to be relevant to increasing N . I thus focus on results for $N = 10$. The three alternative multiple-site extensions of Bray–Curtis dissimilarity

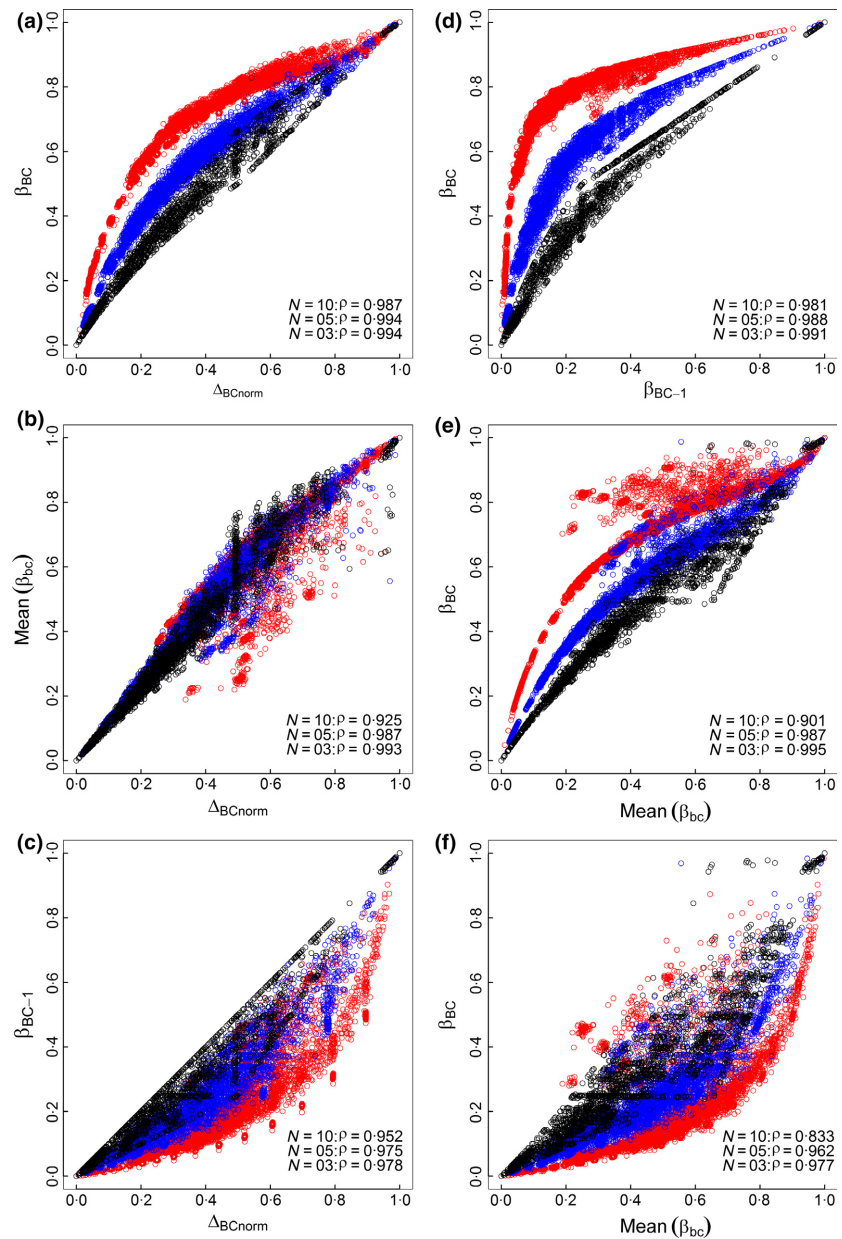


Fig. 3. Scatter plots showing the relationship between alternative indices for simulated data sets of 10 sites and 10 species ($N = 10$, red), 5 sites and 5 species ($N = 5$, blue), and 3 sites and 3 species ($N = 3$, black). The Spearman rank correlation values are shown. (a) Δ_{BCnorm} vs. β_{BC} , (b) Δ_{BCnorm} vs. $\text{mean}(\beta_{bc})$, (c) Δ_{BCnorm} vs. β_{BC-1} , (d) β_{BC-1} vs. β_{BC} , (e) $\text{mean}(\beta_{bc})$ vs. β_{BC} , and (f) $\text{mean}(\beta_{bc})$ vs. β_{BC-1} .

(i.e. Δ_{BCnorm} , β_{BC-1} , and β_{BC}) were very highly correlated ($0.99 > \rho > 0.95$), and their correlation with the average of pairwise Bray–Curtis dissimilarities [$\text{mean}(\beta_{bc})$] was slightly lower in all cases ($0.93 > \rho > 0.83$). These correlation values must be interpreted in relative terms, as the simulated data sets do not necessarily cover uniformly all the possible configurations of abundance tables, so the correlation values could be driven by particular regions of the range of variation in abundance-based dissimilarity that could be more frequent in the simulated data sets just because of the particular way they were generated. However, they point to a general congruence among different approaches to extend Bray–Curtis dissimilarity to multiple-site data sets, and less so between multiple-site indices and the average of pairwise Bray–Curtis dissimilarities. The reasons for differences in behaviour among different multiple-site extensions of Bray–Curtis dissimilarity should be analysed in detail in

future studies, but the task is not simple because abundance-based indices are affected by both the (i) configuration of presence and absences (i.e. replacement and nestedness) and (ii) the pattern in the variation in abundances (balanced and gradients), as shown elsewhere for abundance-based pairwise dissimilarity (Baselga 2013b; Fig. 1). However, a preliminary assessment suggests that different indices account differently for patterns of substitution (replacement and balanced variation in abundance) and subsets (nestedness and abundance gradients, see Fig. S1).

In the second place, as shown in Fig. 4, the correlation between unpartitioned indices and the (i) balanced variation ($\beta_{BC,BAL}$) and (ii) abundance gradient ($\beta_{BC,GRA}$) components of β_{BC} was much lower (i.e. moderate in the case of $\beta_{BC,BAL}$ and very low in the case of $\beta_{BC,GRA}$). This was not unexpected and reveals that the unpartitioned indices Δ_{BCnorm} and

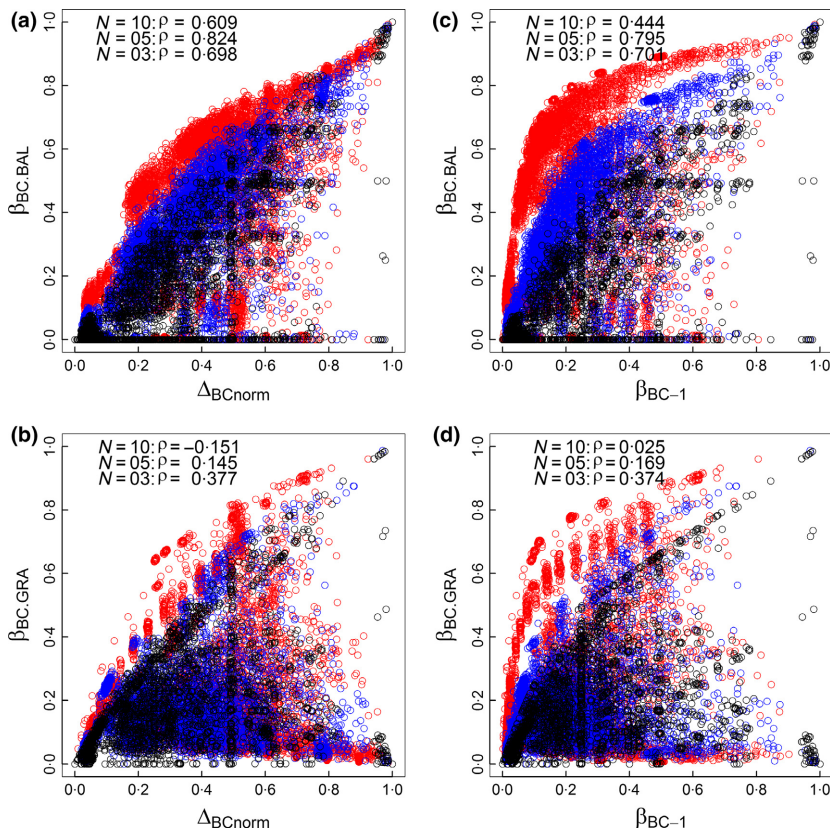


Fig. 4. Scatter plots showing the relationship between unpartitioned multiple-site extensions of Bray–Curtis dissimilarity and the indices proposed here to separate the balanced variation and abundance gradient components of multiple-site Bray–Curtis dissimilarity. Indices were computed for simulated data sets of 10 sites and 10 species ($N = 10$, red), 5 sites and 5 species ($N = 5$, blue), and 3 sites and 3 species ($N = 3$, black). The Spearman rank correlation values are shown. (a) Δ_{BCnorm} vs. $\beta_{BC,BAL}$. (b) Δ_{BCnorm} vs. $\beta_{BC,GRA}$. (c) β_{BC-1} vs. $\beta_{BC,BAL}$. (d) β_{BC-1} vs. $\beta_{BC,GRA}$.

β_{BC-1} were unable to discriminate whether dissimilarity was associated to balanced variation in abundance or to abundance gradients.

A simple case study

To exemplify the use of the new partitioned abundance-based multiple-site indices, I analysed the data in Myers *et al.* (2015a). In short, the data set includes the abundances of birds and butterflies in multiple plots with different habitat characteristics induced by soil variation in Iowa, USA. These plots ($n = 48$) were the result of an experimental design that included four vegetation treatments of varying tallgrass prairie species richness, replicated on three soil types (see Myers *et al.* 2015a for details). The bird and butterfly data sets were downloaded from Dryad (Myers *et al.* 2015b). My analyses are by no means a re-analysis of the original study, but just a simplified example aiming to show (i) the usefulness of separating the two components of abundance-based multiple-site dissimilarity, and (ii) the usefulness of comparing incidence- and abundance-based multiple-site dissimilarity patterns, which could potentially provide different information. With this aim, using the function `beta.multi.abund`, I first computed the overall abundance-based multiple-site Bray–Curtis dissimilarity and its components across the 48 plots for both birds and butterflies. Second, to assess the significance of differences between birds and butterflies, I estimated the distributions of abundance-based multiple-site dissimilarity indices computed in samples of 10 plots

using the function `beta.sample.abund`. To assess whether a dissimilarity index was larger in one group than in the other, P -values were computed by assessing the proportion of samples for which the opposite result (compared to the observed one) was obtained by chance. Finally, the same calculations were done for the corresponding incidence-based indices (i.e. Sørensen dissimilarity and its components), by transforming the abundance tables into presence-absence tables, and using the function `beta.sample`. This comparison allowed assessing whether differences in dissimilarity patterns between birds and butterflies were more related to (i) differences in species replacement and nestedness (incidence-based patterns) or (ii) differences in balanced abundance variation and abundance gradients (abundance-based patterns).

The multiple-site Bray–Curtis dissimilarity (β_{BC}) across the total 48 plots was 0.95 for birds and 0.93 for butterflies. The abundance balanced variation ($\beta_{BC,BAL}$) and abundance gradients ($\beta_{BC,GRA}$) components were 0.92 and 0.03 for birds, and 0.78 and 0.15 for butterflies respectively. When the indices were computed for 1000 samples of 10 plots, the mean value of total dissimilarity for birds was $\beta_{BC} = 0.84$, with components $\beta_{BC,BAL} = 0.75$ and $\beta_{BC,GRA} = 0.09$. In turn, for butterflies total dissimilarity was $\beta_{BC} = 0.76$, with components $\beta_{BC,BAL} = 0.46$ and $\beta_{BC,GRA} = 0.30$. The empirical comparison of the distributions of the dissimilarity indices across samples (Fig. 5a) revealed that total dissimilarity (β_{BC}) and its component of balanced variation in abundance ($\beta_{BC,BAL}$) were significantly higher in birds than in butterflies ($P = 0.006$ and $P < 0.001$ respectively), whereas the abundance gradients

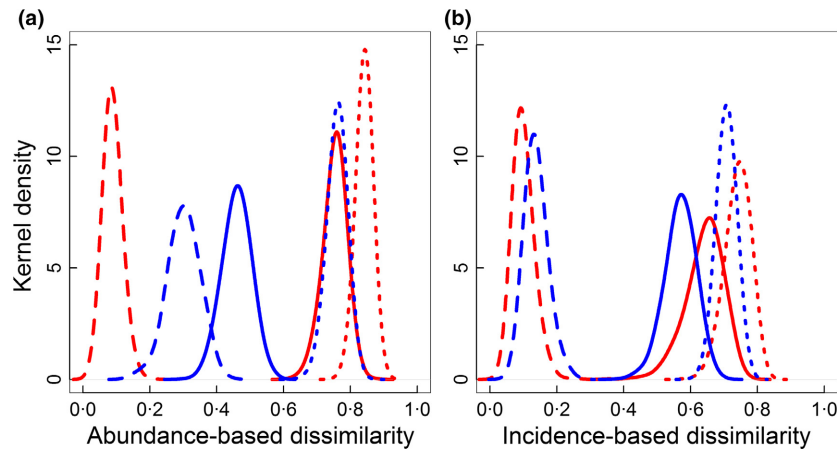


Fig. 5. Density plots representing the distribution dissimilarity indices across 1000 samples of 10 plots with different habitat characteristics (data from Myers *et al.* 2015a). Components of multiple-site dissimilarity were computed for birds (red) and butterflies (blue) and for (a) abundance based dissimilarity, where dotted line is total dissimilarity, solid line is the component of dissimilarity due to balanced variation in abundance and dashed line is the component of due to abundance gradients, and (b) incidence-based dissimilarity, where dotted line is total dissimilarity, solid line is the component of dissimilarity due to species replacement and dashed line is the component of due to nestedness.

component was significantly higher in butterflies than in birds ($P = 0.001$). In the case of incidence-based indices computed for 1000 samples of 10 plots, the mean value of total dissimilarity (β_{SOR}) was 0.74 for birds and 0.71 for butterflies. When partitioned into components, the replacement component (β_{SIM}) was 0.64 in birds and 0.57 in butterflies, and the nestedness-resultant component (β_{SNE}) was 0.09 in birds and 0.13 in butterflies. None of these differences was significant (Fig. 5b, $P > 0.147$ in all cases).

Thus, using presence–absence data, birds and butterflies presented very similar levels of faunal dissimilarity across multiple plots, which in both groups was mostly caused by species replacement instead of nestedness. These two components of dissimilarity were also very similar for both groups. In turn, when the variation in abundances was assessed, it turned out that faunal dissimilarity was significantly higher in birds than in butterflies and, more strikingly, dissimilarity was mostly related to balanced variation in abundances (i.e. substitution) in the case of birds, while the contribution of abundance gradients (i.e. subsets) was much more relevant to butterflies (i.e. mean ratio $\beta_{\text{BC,BAL}}/\beta_{\text{BC}}$ for samples of 10 plots was 0.90 in birds and 0.60 in butterflies). In other words, while differences in habitat characteristics induced by soil variation affected birds and butterflies similarly in terms of species replacement and richness gradients, the variation in abundances was markedly different in both groups: in birds variation in species abundances was higher and balanced, while in butterflies variation in abundances was lower and with more marked abundance gradients.

General applications

Incidence- and abundance-based dissimilarity indices provide markedly different information (Anderson *et al.* 2011; Barwell, Isaac & Kunin 2015). The former account for patterns of species replacement (i.e. substitution) and nestedness (i.e. subsets), while the latter account for patterns of variation in

abundances, either balanced (i.e. substitution) or in gradients (i.e. subsets). In both cases, the substitution and subset components of total dissimilarity can be separated (Baselga 2010, 2012, 2013b; Legendre 2014; Baselga & Leprieur 2015). In the case of abundance-based indices, the partitioning method was only available for pairwise dissimilarity measures but, as argued elsewhere (Baselga, Jiménez-Valverde & Niccolini 2007; Diserud & Ødegaard 2007; Baselga 2013a) when the question of interest is the quantification of overall dissimilarity among more than two sites, multiple-site dissimilarity measures are a better methodological approach. Therefore, the method provided here may be useful for research questions aiming to quantify abundance-based dissimilarity (and its components) in sets of multiple units (spatial, temporal or other). For example, this approach could be implemented when assessing the variation in dissimilarity patterns among different taxonomic groups, as in the simplified example above (or as in Gómez-Rodríguez, Freijeiro & Baselga 2015; for a more complex analysis, using incidence-based dissimilarity). Likewise, analyses assessing the variation in dissimilarity (and its components) among regions (e.g. Gaston *et al.* 2007; Kraft *et al.* 2011; Leprieur *et al.* 2011; Baselga, Gómez-Rodríguez & Lobo 2012) could benefit from the method proposed here, as it simultaneously allows quantifying multiple-site dissimilarity and its components while accounting for abundance data.

Acknowledgements

I acknowledge the insightful comments and suggestions done by the Associate Editor and three anonymous reviewers, which helped substantially, improve previous versions of this work. This research was funded by the Spanish Ministry of Economy and Competitiveness and the European Regional Development Fund (ERDF) through grants CGL2013-43350-P and CGL2016-76637-P.

Data accessibility

All data used in the case study are openly available from Myers *et al.* (2015b): Data from: Birds and butterflies respond to soil-induced habitat heterogeneity in experimental plantings of tallgrass prairie species managed as agro energy crops

in Iowa, USA. Dryad Digital Repository. <http://dx.doi.org/10.5061/dryad.4ft7n>.

References

- Anderson, M.J., Crist, T.O., Chase, J.M. *et al.* (2011) Navigating the multiple meanings of beta diversity: a roadmap for the practicing ecologist. *Ecology Letters*, **14**, 19–28.
- Barwell, L.J., Isaac, N.J.B. & Kunin, W.E. (2015) Measuring β -diversity with species abundance data. *Journal of Animal Ecology*, **84**, 1112–1122.
- 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) Multiple site dissimilarity quantifies compositional heterogeneity among several sites, while average pairwise dissimilarity may be misleading. *Ecography*, **36**, 124–128.
- Baselga, A. (2013b) 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., Gómez-Rodríguez, C. & Lobo, J.M. (2012) Historical legacies in world amphibian diversity revealed by the turnover and nestedness components of beta diversity. *PLoS ONE*, **7**, e32341.
- Baselga, A., Jiménez-Valverde, A. & Niccolini, G. (2007) A multiple-site similarity measure independent of richness. *Biology Letters*, **3**, 642–645.
- Baselga, A. & Leprieux, F. (2015) Comparing methods to separate components of beta diversity. *Methods in Ecology and Evolution*, **6**, 1069–1079.
- Baselga, A. & Orme, C.D.L. (2012) betapart: an R package for the study of beta diversity. *Methods in Ecology and Evolution*, **3**, 808–812.
- Baselga, A., Orme, C.D.L., Villéger, S., De Bortoli, J. & Leprieux, F. (2017) betapart: Partitioning beta diversity into turnover and nestedness components. R package version 1.4. <http://CRAN.R-project.org/package=betapart>
- Chao, A. & Chiu, C.H. (2016) Bridging two major approaches (the variance framework and diversity decomposition) to beta diversity and related similarity and differentiation measures. *Methods in Ecology and Evolution*, **8**, 919–928.
- Chao, A., Chiu, C.-H. & Hsieh, T.C. (2012) Proposing a resolution to debates on diversity partitioning. *Ecology*, **39**, 2037–2051.
- Chiu, C.H., Jost, L. & Chao, A. (2014) Phylogenetic beta diversity, similarity, and differentiation measures based on Hill numbers. *Ecological Monographs*, **84**, 21–44.
- Diserud, O.H. & Ødegaard, F. (2007) A multiple-site similarity measure. *Biology Letters*, **3**, 20–22.
- Erickson, M.J. (1996) *Introduction to Combinatorics*. Wiley, New York, NY, USA.
- Gaston, K.J., Davies, R.G., Orme, C.D.L. *et al.* (2007) Spatial turnover in the global avifauna. *Proceedings of the Royal Society B-Biological Sciences*, **274**, 1567–1574.
- Gómez-Rodríguez, C., Freijeiro, A. & Baselga, A. (2015) Dispersal and ecological traits explain differences in beta diversity patterns of European beetles. *Journal of Biogeography*, **42**, 1526–1537.
- Harrison, S., Ross, S.J. & Lawton, J.H. (1992) Beta-diversity on geographic gradients in Britain. *Journal of Animal Ecology*, **61**, 151–158.
- Jost, L. (2007) Partitioning diversity into independent alpha and beta components. *Ecology*, **88**, 2427–2439.
- Kraft, N.J.B., Comita, L.S., Chase, J.M. *et al.* (2011) Disentangling the Drivers of beta Diversity Along Latitudinal and Elevational Gradients. *Science*, **333**, 1755–1758.
- Legendre, P. (2014) Interpreting the replacement and richness difference components of beta diversity. *Global Ecology and Biogeography*, **23**, 1324–1334.
- Legendre, P. & De Cáceres, M. (2013) Beta diversity as the variance of community data: dissimilarity coefficients and partitioning. *Ecology Letters*, **16**, 951–963.
- Leprieux, 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.
- Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M. & Hornik, K. (2015) Cluster Analysis Basics and Extensions. R package version 2.0.1. <http://cran.r-project.org/> (accessed 24 May 2015).
- Myers, M.C., Mason, J.T., Hoks, B.J., Cambardella, C.A. & Pfrimmer, J.D. (2015a) Birds and butterflies respond to soil-induced habitat heterogeneity in experimental plantings of tallgrass prairie species managed as agroenergy crops in Iowa, USA. *Journal of Applied Ecology*, **52**, 1176–1187.
- Myers, M.C., Mason, J.T., Hoks, B.J., Cambardella, C.A. & Pfrimmer, J.D. (2015b) Data from: Birds and butterflies respond to soil-induced habitat heterogeneity in experimental plantings of tallgrass prairie species managed as agroenergy crops in Iowa, USA. *Dryad Digital Repository*, <http://dx.doi.org/10.5061/dryad.4ft7n>
- Oksanen, J., Blanchet, G., Kindt, R. *et al.* (2015) vegan: Community Ecology Package. R package version 2.3-0, <http://cran.r-project.org/> (accessed 23 July 2015).
- 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**, 280–338.

Received 1 July 2016; accepted 19 October 2016

Handling Editor: Anne Chao

Supporting Information

Additional Supporting Information may be found online in the supporting information tab for this article:

Fig. S1. Scatter plots showing the relationship between alternative indices for simulated data sets.

Appendix S1. R functions implementing the methods introduced in this paper (to be included in R package betapart 1.4).