

A multiple-site similarity measure independent of richness

Andrés Baselga^{1,*}, Alberto Jiménez-Valverde¹ and Gilles Niccolini²

¹Departamento de Biodiversidad y Biología Evolutiva, Museo Nacional de Ciencias Naturales—CSIC, c/José Gutiérrez Abascal 2, 28006 Madrid, Spain

²Laboratoire Universitaire d'Astrophysique de Nice, UMR 6525 Université de Nice-Sophia Antipolis, Bâtiment Hippolyte Fizeau, Parc Valrose 06108, Nice cedex 02, France

*Author for correspondence (baselga@mncn.csic.es).

The Diserud–Ødegaard multiple-site similarity index makes use of data on species shared by two or more sites, but produces equal similarity values in two different circumstances: species loss and true species turnover. We developed a new multiple-site similarity measure, which is independent of richness and performs better than the Diserud–Ødegaard index under conditions of equal richness between sites, because it discriminates between situations in which shared species are distributed evenly among sites or concentrated in few pairs of sites. We conducted several simulations to assess the relative performance of both the indices. The use of the new measure is recommended, enabling the simultaneous analysis of turnover and richness gradients based on two independent measures.

Keywords: beta diversity; incidence; multiple-site similarity; species richness

1. INTRODUCTION

Beta diversity, a measure of variability in site-assemblage composition (Whittaker 1960), may be caused by two different phenomena: nestedness and species turnover (Harrison *et al.* 1992). Nestedness of species assemblages occurs when the biotas of sites with lower numbers of species are subsets of the biotas at richer sites (Wright & Reeves 1992), reflecting a process of species loss. It is the consequence of any factor that promotes the orderly disaggregation of assemblages (Gaston & Blackburn 2000). Contrary to species loss, species turnover implies the replacement of some species by others and is a consequence of environmental as well as spatial and historical differences among sites (Qian *et al.* 2005). Patterns of biodiversity that may be revealing different processes (Williams *et al.* 1999) must be discerned and, regarding beta diversity, this implies that nestedness and species turnover must be disentangled (Baselga in press).

Diserud & Ødegaard (2007) proposed a multiple-site similarity measure, with conceptual and methodological advantages over traditional approaches based

on average pairwise similarities (e.g. Koleff & Gaston 2002; Gaston *et al.* 2007). A multiple-site index avoids (i) the loss of information concerning the number of species shared among three or more sites and (ii) the lack of independence between pairwise similarities due to the repetition of each site in several pairs (Diserud & Ødegaard 2007). The index is an extension of the widely used Sørensen similarity index, thereby inheriting its sensitivity to variation in richness, which in many cases is an undesirable property (Wilson & Shmida 1984; Koleff *et al.* 2003).

If an index that is not independent of richness is used, when the selected sites span a richness gradient, then the differences in composition due to differences in richness (nestedness) cannot be distinguished from differences in composition that are independent of richness. In such a scenario (figure 1), species loss cannot be distinguished from true species turnover (Harrison *et al.* 1992). The aim of this paper is to provide a new multiple-site similarity measure that is independent of patterns of richness but retains the advantages of the Diserud–Ødegaard index.

2. MATERIAL AND METHODS

We derive the new multiple-site similarity measure from the Simpson index in a fashion to ensure its independence from patterns of richness. The Simpson index for similarity between two sites (Sim) is given by

$$\text{Sim} = \frac{a_{ij}}{\min(b_{ij}, b_{ji}) + a_{ij}}, \quad (2.1)$$

where a_{ij} is the number of species common to both sites; b_{ij} is the number of species that occur in the site i but not in the site j ; and b_{ji} is the number of species that occur in the site j but not in the site i .

To extend this index to situations involving multiple sites, we start with three sites. Following the same notation of equation (2.1), in this case the term a_{ij} can be substituted by the number of shared species counts exceeding the number of species shared by the first pair of sites

$$\left[\sum_{i < j} a_{ij} - a_{ijk} \right],$$

a_{ijk} being the number of species common to the three sites, and the term $\min(b_{ij}, b_{ji})$ can be substituted by the sum of minimum values of species not shared between each pair of sites

$$\left[\sum_{i < j} \min(b_{ij}, b_{ji}) \right].$$

Thus, a three-site similarity measure can be

$$M_{\text{Sim}} = \frac{\sum_{i < j} a_{ij} - a_{ijk}}{\left[\sum_{i < j} \min(b_{ij}, b_{ji}) \right] + \left[\sum_{i < j} a_{ij} - a_{ijk} \right]}. \quad (2.2)$$

Considering a general case with n sites, the new multiple-site similarity measure can be formulated as

$$M_{\text{Sim}} = \frac{\sum_{i < j} a_{ij} - \sum_{i < j < k} a_{ijk} + \sum_{i < j < k < l} a_{ijkl} - \dots}{\left[\sum_{i < j} \min(b_{ij}, b_{ji}) \right] + \left[\sum_{i < j} a_{ij} - \sum_{i < j < k} a_{ijk} + \sum_{i < j < k < l} a_{ijkl} - \dots \right]}, \quad (2.3)$$

where a_{ij} is the number of species shared by sites i and j ; a_{ijk} is the number of species shared by sites i , j and k , etc.; and b_{ij} and b_{ji} are the number of species exclusive from sites i and j , respectively, when compared by pairs. Note that for two sites this reduces to the Simpson index (equation (2.1)). Finally, the notation of the new measure can be simplified using the inclusion–exclusion principle (Erickson 1996) to substitute the term $\left[\sum_{i < j} a_{ij} - \sum_{i < j < k} a_{ijk} + \sum_{i < j < k < l} a_{ijkl} - \dots \right]$ by $\left[\sum_i S_i - S_T \right]$, where S_i is the total number of species in site i and S_T is the total

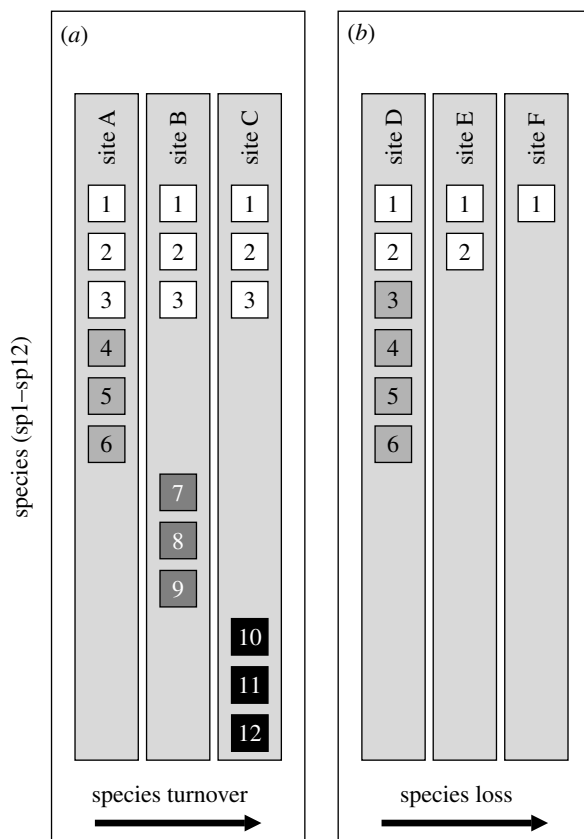


Figure 1. Hypothetical example involving two islands ((a) island 1 and (b) island 2) and three sampling sites in each. Sites A–C have the same richness (six species each) with three species common to all three sites, each site also harbouring three taxa exclusive to it. Sites D–F form a richness gradient in which species are lost from site D (six species) to E (two) and F (one). These biotas are completely nested. In both islands, the Diserud–Ødegaard index yields exactly the same similarity of 0.5, although it is clear that species turnover produces much more dissimilar assemblages than species loss.

number of species in all sites considered together,

$$M_{\text{sim}} = \frac{\sum_i S_i - S_T}{\left[\sum_{i < j} \min(b_{ij}, b_{ji}) \right] + \left[\sum_i S_i - S_T \right]} \quad (2.4)$$

An R function (R Development Core Team 2006) to compute M_{sim} from presence–absence tables is provided in the electronic supplementary material. The function is called `Simpson.multi(x)`, where x is a data frame in which sites are rows and species are columns.

The new measure was compared with the Diserud–Ødegaard index to assess their respective performances under three different conditions defined by richness or degree of species overlap among sites:

Test 1. Equal and constant richness, increasing species overlap: three sites have the same richness (S_i) and the number of species shared by pairs of sites and by the three sites increases from 0 to S_i in intervals of 10, considering all possible combinations of a_{ij} and a_{ijk} . This scenario is replicated for five different values of S_i (60, 70, 80, 90 and 100 species).

Test 2. Equal and increasing richness, increasing species overlap: in the initial case, richness of the three sites is equal to 1, and increases 1 species in each site in subsequent cases (to $S_i=100$); thus in all 100 cases richness remains equal among sites. Species are added randomly from a fixed pool of 100 species; thus the number of shared species increases with richness, with some random fluctuations. This scenario is replicated five times to increase the combinations due to the random inclusion of species.

Test 3. Increasing richness differences, equal species overlap: in the initial case, richness of the three sites is equal ($S_i=15$), five species are shared among three sites ($a_{123}=5$) and five species are shared

between each pair of sites ($a_{12}=a_{13}=a_{23}=5$). New species exclusive of site 1 are added sequentially (to $S_1=50$, $S_2=S_3=15$), increasing richness differences between this site and the others but maintaining the number of species shared between the three sites as identical. This scenario is replicated for four other combinations of initial richness and overlap: initial S_i increasing from 16 to 19; a_{123} decreasing from 4 to 1; and a_{12} ($=a_{13}=a_{12}$) increasing from 6 to 9, respectively.

Our expectations are a high correlation between both the measures under conditions simulated by tests 1 and 2, but a complete independence between the new measure and the Diserud–Ødegaard index under conditions simulated by test 3, in which the Diserud–Ødegaard measure should correlate with the standard deviation of richness, contrary to the situation for the new index.

3. RESULTS

The performance of the new measure compared with the Diserud–Ødegaard index yielded clear results. In conditions of equal richness among all sites, both the indices highly correlated (figure 2a,c). Moreover, test 1 revealed another unexpected advantage of the new measure, which discriminated between situations that were considered to be equivalent by the Diserud–Ødegaard index: columns of points in the scatterplot (figure 2a) indicate that the new measure yielded different similarity values for different combinations of shared species, computing a higher similarity for balanced situations in which the shared species are more evenly distributed among sites, and not concentrated in only two very similar sites (figure 2b), which is in accord with the common concept of similarity.

Test 2 yielded similar statistical results (figure 2c), confirming equivalent performance of both the indices when the increment of diversity is equivalent in all sites. This equivalence between measures is lost when the increment in richness is not equal in all sites, as simulated in test 3. Under these conditions, the two indices are independent (figure 2d) because the Diserud–Ødegaard measure reflects the magnitude of richness differences (Pearson's correlation between Diserud–Ødegaard similarity and s.d. of S_i , $-0.988 \geq r \geq -0.992$, $p \leq 0.001$ in all replicates), whereas the new index is completely independent of richness gradients (figure 2e).

4. DISCUSSION

The disparate performance of different similarity measures is well known for the traditional pairwise similarity indices (Wilson & Shmida 1984; Harrison *et al.* 1992; Williams *et al.* 1999; Koleff *et al.* 2003). The selection of a particular index should not be based on subjective preferences or on previous widespread usage. Rather, the appropriateness of the measure to test the addressed hypothesis (Willig *et al.* 2003) should guide the choice of measure (Baselga *in press*). The same applies for the multiple-site similarity measures. The Diserud–Ødegaard index performs appropriately in samples of similar richness, but incorporates richness differences as if they were compositional differences. Depending on the hypothesis tested, this characteristic may be assumed and even desired, but it should be taken into account that under richness gradient conditions the Diserud–Ødegaard measure may consider two different circumstances as equivalent. This could lead to confusing results if we try to compare multiple-site

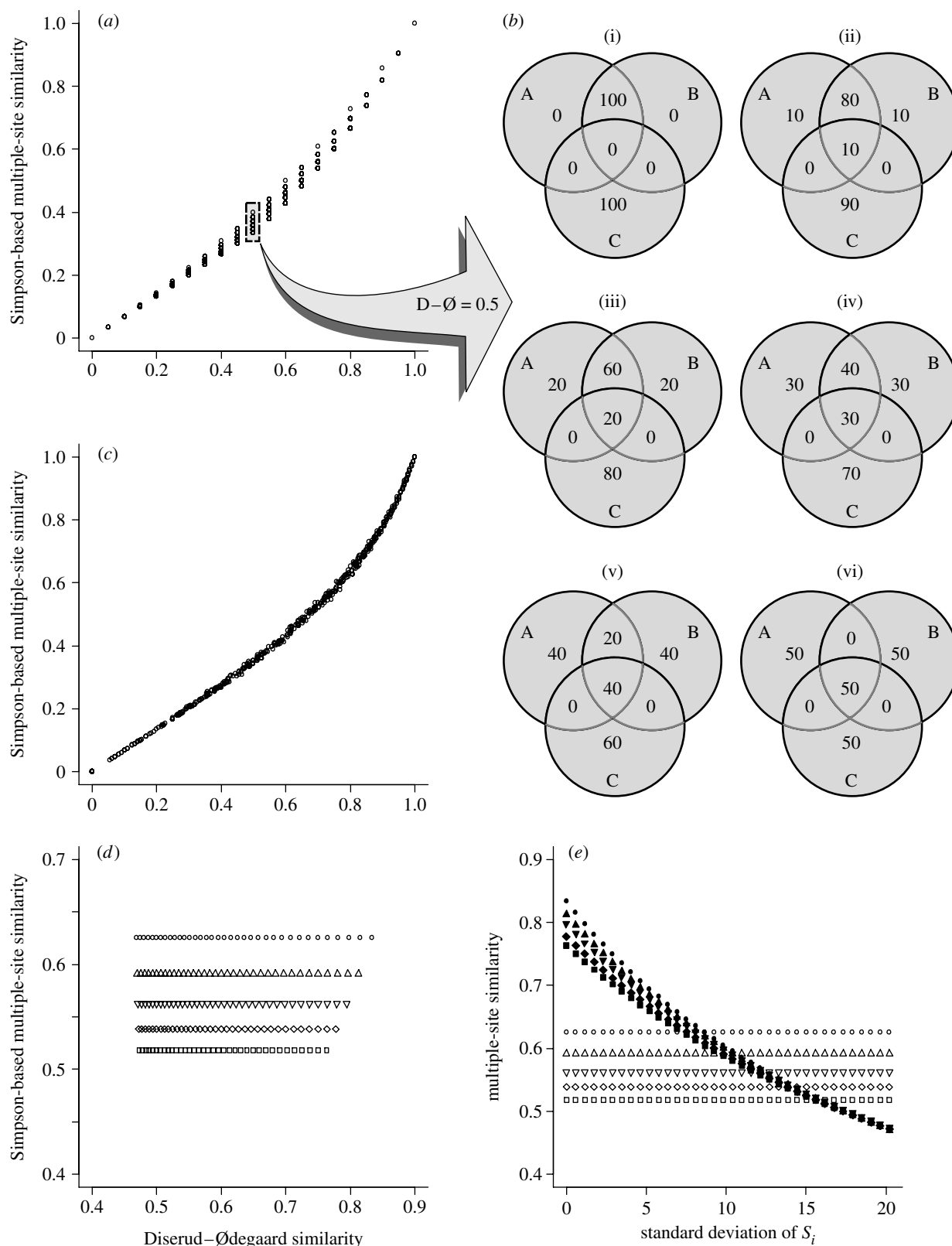


Figure 2. Relative performance of the Diserud-Ødegaard (D-Ø) measure and the Simpson-based multiple-site index under three different scenarios. (a) Test 1: equal and constant richness, increasing species overlap. Only one of the five replicates is represented provided that all of them yielded exactly the same pattern ($r=0.984$, $p \leq 0.001$). (b) Venn diagrams showing six different situations (M_{sim} : (i) 0.333, (ii) 0.349, (iii) 0.357, (iv) 0.370, (v) 0.385 and (vi) 0.400) in which the Diserud-Ødegaard index yields the same value of similarity (0.5), whereas the Simpson-based multiple-site index discriminates balanced cases (in which shared species are evenly distributed among sites A, B and C) from unbalanced cases (in which shared species are concentrated in any of the shared fractions). (c) Test 2: richness increasing simultaneously in three sites (thus remaining equal among them), increasing species overlap ($r=0.967$, $p \leq 0.001$). The scatterplot combines the five replicates. (d) Test 3: increasing richness differences among sites, equal species overlap. Each symbol identifies each of the five replicates. (e) Relationship of each similarity measure with the dispersion of richness among sites (s.d.) as simulated in test 3. Symbols identify each of the five replicates for the Diserud-Ødegaard index (filled symbols) and the Simpson-based multiple-site index (open symbols).

similarities of datasets with different richness gradients among the samples.

The discrimination of species turnover from species loss is a critical characteristic since richness gradients are pervasive (i.e. latitudinal gradients; see Willig *et al.* (2003) for a review). In the cases illustrated by figure 1, the new index yields two different results ($M_{\text{sim}} = 0.4$ versus $M_{\text{sim}} = 1.0$), identifying completely nested biotas as entirely similar. For this reason, in the analysis of beta diversity among biotas, richness gradients caused by nestedness and pure species turnover should be simultaneously analysed, disentangling their patterns by means of independent but complementary measures. This partition is crucial for the complete understanding of central biogeographic, ecological and conservation issues. In biogeography, areas where biotas are replaced are defined as borders between biogeographic regions and are identified as peaks of biotic dissimilarity, and thus must be distinguished from impoverished zones, because each case is generated by different historical or environmental factors (Williams *et al.* 1999). In ecology, assigning the different biodiversity patterns to their respective biological phenomena is essential to analyse the causality of the processes underlying biodiversity. Regarding conservation purposes, similarity indices can be used to identify areas of maximum species turnover and thus maximize biodiversity of protected areas (Wiersma & Urban 2005). If richness gradients are present and the chosen similarity index is not independent of richness, maximizing dissimilarity (complementarity) between reserves could support selection of rich and poor areas with exactly the same species (nested biotas) as dissimilar. This is exactly the opposite of the conservation objective.

- Baselga, A. In press. Disentangling distance decay of similarity from richness gradients: response to Soininen *et al.* 2007. *Ecography*. (doi:10.1111/j.2007.0906-7590.05191.x)
- Diserud, O. H. & Ødegaard, F. 2007 A multiple-site similarity measure. *Biol. Lett.* **3**, 20–22. (doi:10.1098/rsbl.2006.0553)

- Erickson, M. J. 1996 *Introduction to combinatorics*. New York, NY: Wiley.
- Gaston, K. J. & Blackburn, T. M. 2000 *Pattern and process in macroecology*. Oxford, UK: Blackwell Science Ltd.
- Gaston, K. J. *et al.* 2007 Spatial turnover in the global avifauna. *Proc. R. Soc. B* **274**, 1567–1574. (doi:10.1098/rspb.2007.0236)
- Harrison, S., Ross, S. J. & Lawton, J. H. 1992 Beta-diversity on geographic gradients in Britain. *J. Anim. Ecol.* **61**, 151–158. (doi:10.2307/5518)
- Koleff, P. & Gaston, K. J. 2002 The relationship between local and regional species richness and spatial turnover. *Global Ecol. Biogeogr.* **11**, 363–375. (doi:10.1046/j.1466-822x.2002.00302.x)
- Koleff, P., Gaston, K. J. & Lennon, J. K. 2003 Measuring beta diversity for presence-absence data. *J. Anim. Ecol.* **72**, 367–382. (doi:10.1046/j.1365-2656.2003.00710.x)
- Qian, H., Ricklefs, R. E. & White, P. S. 2005 Beta diversity of angiosperms in temperate floras of eastern Asia and eastern North America. *Ecol. Lett.* **8**, 15–22. (doi:10.1111/j.1461-0248.2004.00682.x)
- R Development Core Team 2006 *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. See <http://www.r-project.org>.
- Whittaker, R. H. 1960 Vegetation of the Siskiyou Mountains, Oregon and California. *Ecol. Monogr.* **30**, 280–338.
- Wiersma, Y. F. & Urban, D. L. 2005 Beta diversity and nature reserve system design in the Yukon, Canada. *Conserv. Biol.* **19**, 1262–1272. (doi:10.1111/j.1523-1739.2005.00099.x)
- Williams, P. H., Klerk, H. M. & Crowe, T. M. 1999 Interpreting biogeographical boundaries among Afro-tropical birds: spatial patterns in richness gradients and species replacement. *J. Biogeogr.* **26**, 459–474. (doi:10.1046/j.1365-2699.1999.00294.x)
- Willig, M. R., Kaufman, D. M. & Stevens, R. D. 2003 Latitudinal gradients of biodiversity: pattern, process, scale, and synthesis. *Annu. Rev. Ecol. Evol. Syst.* **34**, 273–309. (doi:10.1146/annurev.ecolsys.34.012103.144032)
- Wilson, M. V. & Shmida, A. 1984 Measuring beta diversity with presence absence data. *J. Ecol.* **72**, 1055–1064. (doi:10.2307/2259551)
- Wright, D. H. & Reeves, J. H. 1992 On the meaning and measurement of nestedness of species assemblages. *Oecologia* **92**, 416–428. (doi:10.1007/BF00317469)