Journal of Vegetation Science **26** (2015) 1061–1067

Measuring similarity among plots including similarity among species: an extension of traditional approaches

Carlo Ricotta & Sandrine Pavoine

Keywords
Abundance-based coefficients; Contingency table; Interspecies resemblances; Presence/absence coefficients; Species ordinariness

Received 11 November 2014
Accepted 24 February 2015
Co-ordinating Editor: Otto Wildi

Ricotta, C. (corresponding author, carlo.ricotta@uniroma1.it)[1],
Pavoine, S. (pavoine@mnhn.fr)[2,3]

[1]Department of Environmental Biology, University of Rome 'La Sapienza', Piazzale Aldo Moro 5, 00185 Rome, Italy;
[2]Centre d'Ecologie et des Sciences de la Conservation (CESCO UMR7204), Sorbonne Universités, MNHN, CNRS, UPMC, CP51, 55-61 rue Buffon, Paris, France;
[3]Mathematical Ecology Research Group, Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK

Abstract

Aims: Presence/absence coefficients of (dis)similarity are generally based on various combinations of the matching/mismatching components of the classical 2 × 2 contingency table. While a few extensions of the contingency table to species abundance data were proposed, a generalization of such coefficients that includes species resemblances is still lacking.

Methods: In this paper we introduce a new family of plot-to-plot (dis)similarity coefficients, which include species abundances and interspecies resemblances in the calculation of the matching/mismatching components of the contingency table. Within this family, the behaviour of five similarity coefficients is examined by graphical comparisons based on simulated data.

Conclusions: This new family of similarity coefficients can equally apply to functional or phylogenetic resemblance among species, thus providing a unified framework for comparing species turnover among plots.

Introduction

Ecological studies often rely on similarity coefficients between pairs of species assemblages (or communities, plots, sites, quadrats, etc.) to explore the mechanisms driving community assembly. However, although similarity may appear as a simple and unambiguous notion, when we look for a suitable numerical definition, we find that it is independent of any single way of measurement. Ecologists have developed a rich arsenal of coefficients for expressing the similarity between two species plots and the choice among them is governed by several criteria, the foremost of which is whether composition is described in terms of species presence and absence scores or species abundances (Tamás et al. 2001).

Presence/absence coefficients of similarity are usually formulated based on the matching/mismatching components of the well-known 2 × 2 contingency table. These components express the number of species present (a) in both plots, (b) in the first plot but not in the second, (c) in the second plot but not in the first, and (d) the species absent from both plots but found in other plots, such that the sum $S = a + b + c + d$ is the total number of species in the entire collection of plots. Once the values of the matching/mismatching components are determined, any of a large number of similarity and dissimilarity measures for presence/absence data can be calculated, such as the simple matching coefficient (Sokal & Michener 1958) $(a + d)/(a + b + c + d)$ or the Jaccard (1900) index of similarity $(a)/(a + b + c)$. For an exhaustive list of presence/absence similarity measures, together with a thorough discussion of the pros and cons of including double absences (i.e. component d) in the calculation of plot-to-plot similarity coefficients, see Legendre & Legendre (2012) and Podani (2000).

Unlike presence/absence coefficients, abundance-based measures are usually expressed in terms of the original raw data scores, and this different formalism makes it difficult to generalize presence/absence coefficients to the quantitative case. Among the few attempts to redefine the 2 × 2 contingency table to account for species abundances (i.e. Tamás et al. 2001; Chao et al. 2006) we focus here on the approach of Tamás et al. (2001). Let $\mathbf{X} = \{x_{ik}\}$ denote a matrix of S species × N plots with the general element x_{ik} representing the abundance of species i in plot k. According to Tamás et al. (2001), for a pair of plots k and h, the

Doi: 10.1111/jvs.12329

matching/mismatching components of the contingency table can be defined as follows: using capital letters to represent the abundance-based components, A is the amount of species abundances in which the two plots agree, summed over all species

$$A = \sum_{i}^{S} \min\{x_{ik}, x_{ih}\} \quad (1)$$

B is the sum of species abundances by which plot k exceeds plot h

$$B = \sum_{i}^{S} (\max\{x_{ik}, x_{ih}\} - x_{ih}) \quad (2)$$

and C is the sum of species abundances by which plot h exceeds plot k

$$C = \sum_{i}^{S} (\max\{x_{ik}, x_{ih}\} - x_{ik}) \quad (3)$$

To clarify the meaning of A, B and C in a more intuitive manner, imagine species abundances are measured using the number of individuals in each plot. In this case, A is the number of individuals of each species shared by both plots k and h, while B and C summarize the number of individuals that are unique to plots k and h, respectively (Baselga 2013).

Finally, D is defined as the sum of differences between the highest abundance value of each species in the entire collection of plots and the highest abundance value in the pair of plots under scrutiny. First, the maximum abundance value of each species in the entire plot collection is determined; according to Tamás et al. (2001) this value represents the 'potential abundance' that the species could reach in the study area. The 'mutual absence' of a given species is then obtained as the potential species abundance minus the maximum abundance reached by the species in plots k and h. The summation of the mutual absences over all species in the $S \times N$ matrix of species' abundances in plots gives the component D:

$$D = \sum_{i}^{S} (\max_{k}\{x_{ik}\} - \max\{x_{ik}, x_{ih}\}) \quad (4)$$

Although mutual absences are generally not present in the formulation of abundance-based similarity measures, the component D is included in Tamás et al. (2001) to allow full correspondence between the matching/mismatching components of the 2 × 2 contingency table for presence/absence data and its abundance-based counterpart.

While the work of Tamás et al. (2001) allowed us to generalize the matching/mismatching components of the usual 2 × 2 contingency table to species abundances, it has been recently recognized that similarity measures should also incorporate information about the degree of ecological differences between species (i.e. Nipperess et al. 2010; Villéger et al. 2013; Chiu et al. 2014; Pavoine & Ricotta 2014; Ricotta et al. 2015). Obviously, the notion of interspecies differences is multifaceted, including functioning (Diáz & Cabido 2001), morphology (Roy & Foote 1997) or evolutionary history (Clarke et al. 2006). In a phylogenetic context, Ferrier et al. (2007) proposed a method for calculating the (dis)similarity between the species presences and absences in two plots based on the amount of total branch length unique to those plots. Accordingly, two well-known coefficients of phylogenetic dissimilarity, the UniFrac measure (Lozupone & Knight 2005) and the PhyloSor measure (Bryant et al. 2008), have since been special cases of the family of measures of Ferrier et al. (2007). By adopting the approach of Tamás et al. (2001), Nipperess et al. (2010) further extended this method to incorporate species abundance data into phylogenetic dissimilarity measures. Irrespective of the way interspecies resemblances are measured, the main objective of this study is to generalize the work of Tamás et al. (2001) to introduce a new family of plot-to-plot similarity coefficients, which include interspecies resemblances in the calculation of the matching/mismatching components of a 2 × 2 contingency table (hereafter ISR-based measures, where ISR stands for InterSpecies Resemblances).

Methods

Including interspecies resemblances in the components of the contingency table

Regardless of their nature, interspecies differences are usually represented by a square $S \times S$ symmetric matrix $\mathbf{D} = \{d_{ij}\}$ of (functional, morphological or phylogenetic) dissimilarities between a pair of species i and j with $d_{ij} = d_{ji}$ and $d_{ii} = 0$. If the dissimilarity coefficient d is in the range [0, 1], it is possible to define a corresponding similarity coefficient s as the complement of d: $s = 1 - d$. For any dissimilarity measure with an upper bound $d_{\max} > 1$, division by $d_{\max}$ produces a normalized dissimilarity measure in the range [0, 1]. For dissimilarity measures that do not possess an upper bound, such as phylogenetic distances, it is still possible to locally normalize the values d_{ij} in the range [0, 1] by dividing each term by the maximum value found in the data set.

As in Leinster & Cobbold (2012), based on species abundances and similarities, we can define the (absolute) abundance of species similar to i in plot k (Z_{ik}) as:

Doi: 10.1111/jvs.12329

$$Z_{ik} = \sum_{j}^{S(kh)} x_{jk} s_{ij} \quad (5)$$

where $S(kh)$ is the number of species in the pool of plots k and h (i.e. the species for which $\min\{x_{ik}, x_{ih}\} > 0$). Neglecting within-species differences (i.e. setting $s_{ii} = 1$ by definition) we always find that the relative abundance of all species similar to i (including i itself) cannot be lower than the abundance of species i ($x_{ik} \leq Z_{ik} \leq \sum_i x_{ik}$, i.e. the total species abundance in plot k). Z_{ik} is large if most species in the assemblage are similar to i. On the other hand, if species i is very dissimilar from the other species, $Z_{ik} \approx x_{ik}$. The quantity Z_{ik} thus measures the ordinariness of species i. Starting from the species ordinariness Z_{ik}, we can now expand the method of Tamás et al. (2001) to calculate a generalized version of the matching/mismatching components of the 2 × 2 contingency table including information on interspecies resemblances. This is obtained simply by substituting the quantities Z_{ik} for x_{ik} in Eqs 1–3:

$$A = \sum_{i}^{S(kh)} \min\{Z_{ik}, Z_{ih}\} \quad (6)$$

$$B = \sum_{i}^{S(kh)} (\max\{Z_{ik}, Z_{ih}\} - Z_{ih}) \quad (7)$$

$$C = \sum_{i}^{S(kh)} (\max\{Z_{ik}, Z_{ih}\} - Z_{ik}) \quad (8)$$

As for Eq. 5, the sum in Eqs 6–8 is over all species $S(kh)$. In other words, it is over all species that are actually present in at least one of the two plots h and k, without reference to the species absent from both plots but found in other plots of the $S \times N$ matrix of species' abundances in plots (see Leinster & Cobbold 2012; Ricotta et al. 2015).

According to this restrictive definition of the A, B and C components, a formal definition of the D component makes only little sense. This is because the calculation of the A, B and C components in Eqs 6–8 is based on a plot-level species pool, whereas the calculation of the term $\max_k\{Z_{ik}\}$ of the D component (see Eq. 4) usually refers to the matrix-level species pool. Nonetheless, for the sake of completeness and in analogy to Tamás et al. (2001), a less restrictive definition of the A, B and C components that takes into account all species in the $S \times N$ matrix, together with the corresponding D component, is shown in Appendix S1. In Appendix S1 an alternative definition of the new A, B, C and D components, which is based on species relative abundances instead of absolute abundances is also shown.

Starting from this generalized definition of the 2 × 2 contingency table, virtually all presence/absence similarity coefficients can be generalized to include ecological differences among species, thus providing a new family of measures. Note that for zero interspecies similarities (i.e. $s_{ij} = 0$ for $i \neq j$) Eqs 6–8 recover the A, B and C components of Tamás et al. (2001), while for zero interspecies similarities *and* unit abundances Eqs 6–8 recover the a, b and c matching/mismatching components of the traditional 2 × 2 contingency table.

Case study

The relationships between selected similarity coefficients calculated from the generalized contingency table and their traditional presence/absence counterparts were investigated by graphical evaluation of artificial data representing an ideal ecological gradient (Gower & Legendre 1986; Tamás et al. 2001; Podani & Miklós 2002). The basis for the comparative evaluation is the $S \times N$ matrix in Table 1, in which nine plots (P1–P9) are characterized in terms of 15 quantitative variables (the abundances of species S1–S15). The matrix was constructed such that all species have a unimodal response of varying length and intensity (abundance) to a hypothetical one-dimensional gradient, such that the species optima are located at different positions along the gradient. In some cases this unimodal response is truncated at one extreme of the gradient (see e.g. Species S1 or S14 in Table 1). The comparison of plot P1 with itself and with the remaining plots provides nine similarity values. We used profile diagrams to show the effect of species turnover on the similarities.

In this paper we compared the presence/absence version of five traditional similarity coefficients, the Jaccard, Sokal-Sneath, Sørensen, Ochiai and Kulczynski indices (formulas in Table 2) with their abundance-based analogues (calculated with Eqs 1–3) and with three different ISR-based analogues (all calculated with Eqs 6–8 but with different interspecies similarity matrices).

For calculation of the presence/absence coefficients, the abundance data of Table 1 were converted to binary presence and absence data. For calculation of the ISR-based coefficients, we used three different interspecies similarity matrices (**M1**–**M3**; see Appendix S2). In matrix **M1**, interspecies similarities reflect the species ecological differences along the artificial gradient of Table 1. To this end, interspecies similarities were set roughly proportional to the distance between the locations of the species optima and to the difference between their optimal abundances. For example, in Table 1 the optimum of species S9 is located in plot P4, while the optimum of species S12 is located in plot P8. Accordingly, the distance between their optima along

Doi: 10.1111/jvs.12329

Table 1. Artificial data matrix for the graphical comparison of the indices in Table 2. The data represent an ideal ecological gradient with simplified unimodal species responses of varying length and abundance. The matrix was generated randomly and then rearranged according to a sequential location of species optima from P1 to P9. Non-zero species abundances are shown in grey. The locations of all species optima are shown in bold.

Species	Plots								
	P1	P2	P3	P4	P5	P6	P7	P8	P9
S1	**2**	1	0	0	0	0	0	0	0
S2	**2**	1	0	0	0	0	0	0	0
S3	**3**	2	1	0	0	0	0	0	0
S4	**4**	3	2	1	0	0	0	0	0
S5	0	**1**	0	0	0	0	0	0	0
S6	0	0	**1**	0	0	0	0	0	0
S7	2	3	**4**	3	2	1	0	0	0
S8	0	1	2	**3**	2	1	0	0	0
S9	1	2	3	**4**	3	2	1	0	0
S10	0	0	0	1	**2**	1	0	0	0
S11	0	0	0	0	0	**1**	0	0	0
S12	0	0	0	0	0	0	1	**2**	1
S13	0	0	0	0	1	2	3	4	**5**
S14	0	0	0	0	0	0	1	2	**3**
S15	0	0	0	0	0	0	0	1	**2**

Table 2. Traditional similarity coefficients for presence/absence data used in this study. Only coefficients ignoring the value of the d component were considered.

Index Name	Formula	Source
Jaccard	$\frac{a}{a+b+c}$	Jaccard (1900)
Sørensen	$\frac{2a}{2a+b+c}$	Sørensen (1948)
Ochiai	$\frac{a}{\sqrt{(a+b)(a+c)}}$	Ochiai (1957)
Kulczynski	$\frac{1}{2}\left[\frac{a}{(a+b)}+\frac{a}{(a+c)}\right]$	Sokal & Sneath (1963)
Sokal-Sneath	$\frac{a}{a+2b+2c}$	Sokal & Sneath (1963)

the ideal gradient of Table 1 is Od = 4. Similarly, the difference in their optimal abundances is Oa = 4 − 2 = 2. Based on these quantities, the interspecies similarity s_{ij} between species S9 and S12 was calculated as: $s_{ij} = 1 - (\mathrm{Od} + \mathrm{Oa})/10$. Therefore, for the specific case of S9 and S12, $s_{ij} = 1 - (4 + 2)/10 = 0.4$. Note that if Od + Oa > 10, $s_{ij} = 0$ by definition, such that the species i and j are considered maximally dissimilar from each other. Note also that, according to this way of calculating interspecies distances, species S1 and S2 in Table 1 are maximally similar to each other, meaning that for this pair of species $s_{ij} = 1$. Finally, in matrix **M2** interspecies similarities were generated assigning to each pair of species a value $0 \leq s_{ij} \leq 1$ drawn at random from an even distribution, while in **M3** all interspecies similarities were uniformly set to $s_{ij} = 0.33$.

For each similarity index in Table 2 we calculated (1) its presence/absence version, (2) its abundance-based version according to Eqs 1–3 and (3) three ISR-based versions (using the interspecies similarity matrices **M1**–**M3** in Appendix S2). All ISR-based indices were calculated according to the distinct definitions of the matching/mismatching components proposed in this paper (i.e. Eqs 6–8 and Eqs A1–A3, A5–A7, and A10–A12 in Appendix S1). All calculations were performed with the new R script available in Appendix S3.

Results

For the artificial ecological gradient in Table 1, the analyses performed with the various definitions of the matching/mismatching components proposed in this paper yielded qualitatively similar results (see Appendix S4). Therefore, we report here only results from the ISR-based indices calculated with Eqs 6–8. Figure 1 illustrates the IRS-based version of the indices in Table 2 calculated with Matrix **M1**. In Fig. 1, the profiles of all similarity coefficients show essentially the same pattern, meaning that although all coefficients summarize plot-to-plot similarity from different viewpoints, they generally show a consistent behaviour in response to the artificial ecological gradient in Table 1. The profiles of the Sørensen, Ochiai and Kulczynski coefficients have very similar shapes. This is not surprising as for identical b and c components all three indices reduce to the same expression $a/(a + b)$.

Figure 2a–c illustrates the graphical comparison among the presence/absence versions of the Jaccard,

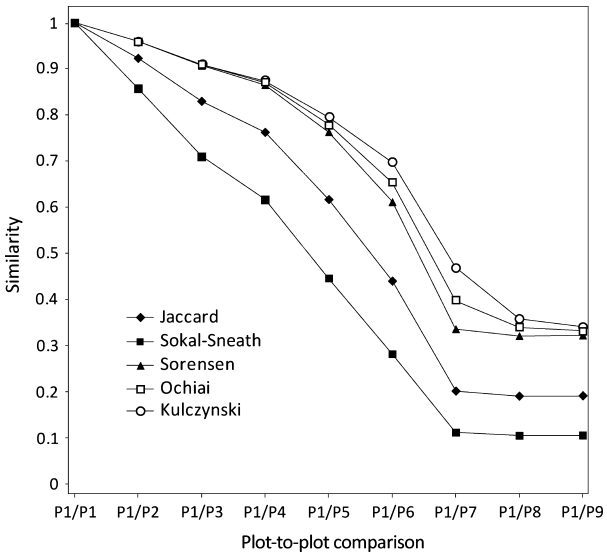


Fig. 1. Response of the ISR-based similarity functions (see Table 2) calculated with Matrix M1 to the artificial ecological gradient in Table 1. Plot P1 is compared to all plots P1–P9 in Table 1.

Doi: 10.1111/jvs.12329

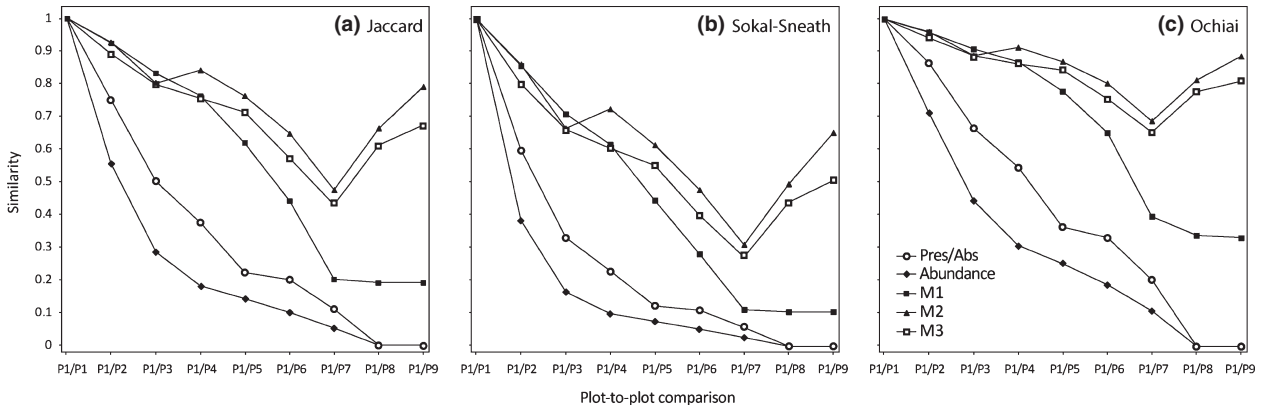


Fig. 2. Response of the presence/absence (Pres/Abs), abundance-based and ISR-based generalizations of the Jaccard, Sokal-Sneath and Ochiai similarity coefficients to the artificial ecological gradient in Table 1. Plot P1 is compared to all plots P1–P9 in Table 1. M1, M2 and M3 denote the ISR-based coefficients calculated with matrices M1, M2 and M3, respectively.

Sokal-Sneath and Ochiai coefficients with their abundance-based and ISR-based analogues. Due to their similarity with the Ochiai index, the Sørensen and Kulczynski coefficients are not shown here, but see Appendix S4. In all cases, the curves of the presence/absence indices and of the abundance-based indices show a non-linear decreasing pattern in response to the artificial gradient of Table 1. The same decreasing behaviour, although with a different curve shape, is shown by the ISR-based measures calculated with matrix **M1**. That is, if interspecies similarities s_{ij} reflect to some degree the species sorting along the gradient of Table 1, the ISR-based index is able to rank the plot-to-plot similarities correctly. However, due to the non-zero values of most interspecies similarities in matrix **M1**, the plot-to-plot similarity values of the ISR-based indices are systematically larger than the corresponding presence/absence and abundance-based indices.

In contrast, the ISR-based indices calculated with matrices **M2** and **M3** rank the plot-to-plot similarity values differently, with a number of distinct local maxima and minima that are absent from the other index profiles. Hence, unlike for presence/absence and abundance-based indices, the similarity values of the ISR-based coefficients depend on the complex interplay of species abundances *and* interspecies similarities. Note here that the similar shapes of the ISR-based index profiles calculated with matrix **M2** and **M3** mean that, on average, a random distribution of interspecies similarities (matrix **M2**) produces the same results as the uniform distribution of s_{ij} values of matrix **M3**.

Discussion

In this paper we developed a methodological framework for measuring the similarity among plots including the similarity among species within plots. The proposed framework is very flexible in the calculation of interspecies similarities, which can be obtained in a variety of ways from phylogenetic or functional data sets containing a combination of scale types (i.e. quantitative, ordinal, nominal and binary variables). This flexibility is particularly desirable, as it is now well established that a single measure is unable to summarize all aspects of plot-to-plot similarity. Juhász-Nagy (1993); see also Podani et al. 2005) first imagined the measurement of an ecological variable, such as a similarity coefficient, as an iterative process involving successive levels of increasing abstraction (i.e. from presence/absence-based measures to abundance-based measures, and to ISR-based measures). At each step of this 'complex, plural and dynamic approach' (*sensu* Juhász-Nagy), the view angle is restricted to zoom in on a specific aspect of the given problem, increasing at the same time the number of possible viewpoints.

In this paper, the extension of presence/absence indices to measures including interspecies resemblances, offers a number of different options that are shown in the main text (Eqs 6–8) and in Appendix S1, such that practitioners should determine which definition of the generalized contingency table is more adequate within the context of their analyses. Nonetheless, the suggested index generalizations were not proposed as a mere addition to the dozens of available similarity measures. Rather, they enable plot-to-plot comparisons at several levels of increasing biotic complexity while excluding confounding effects that may be due to differences in the way the indices are computed, and that may mask the ecological processes under scrutiny (Tamás et al. 2001).

Comparing the new indices, Fig. 1 shows that while the new ISR-based measures tend to rank plots in the same order, nonetheless the index values can be highly variable

Doi: 10.1111/jvs.12329

depending on how the three *A*, *B* and *C* matching/mismatching components are combined. Therefore, to compare two species assemblages, the same index needs to be used. This is a well-known property of the standard indices that use *a*, *b*, *c* matching/mismatching components. Indeed once an index has been chosen (e.g. the Jaccard index), care should be taken in interpreting and communicating the results of the index. While most similarity indices proposed for combining the *a*, *b*, *c* matching/mismatching components vary between the same limits (0 and 1), they usually do not provide comparable similarity values for the same pair of plots. This renders the automatic conversion of the index values to a qualitative scale, such as very low, low, moderate, high and very high (dis)similarity based on a 0.2 interval scale, impossible, thus making cross-study comparisons with different indices highly misleading (Bloom 1981). This well-known property of standard indices also applies to the new ISR-based measures. On the other hand, as the different indices have slightly different properties, they summarize plot-to-plot similarity from different viewpoints (Pavoine & Ricotta 2014).

Depending on interspecies resemblances and unlike conventional presence/absence or abundance-based indices, ISR-based indices usually provide non-zero similarity even for a pair of plots with no species in common, as different species may share some degree of similarity due to common phylogenetic ancestry and/or common functional characteristics. This is a very useful property if one wishes to calculate e.g. the functional similarity among plots under similar environmental conditions in different regions of the world. In addition, it has been recently shown that dissimilarities among communities result from two different processes: species turnover (or replacement) and species gain and loss, also called richness difference or nestedness (Lennon et al. 2001; Baselga 2010). While the debate concerning the best ways of additively partitioning dissimilarity coefficients is on-going (e.g. Carvalho et al. 2013; Legendre 2014), Podani & Schmera (2011) and Carvalho et al. (2012) independently proposed to decompose the Jaccard dissimilarity $1 - (a)/(a + b + c) = (b + c)/(a + b + c)$ into two additive portions: species replacement 2 min $\{b,c\}/(a + b + C)$ and richness difference $|b - c|/(a + b + c)$. Podani et al. (2013) next proposed an extension of the Jaccard dissimilarity decomposition to abundance data. Based on the generalization of the matching/mismatching components of the contingency table proposed in this paper, the additive dissimilarity decomposition can be further extended to ISR-based dissimilarities using a unified algebraic framework.

To conclude, we hope the proposed indices will help shed light on new aspects of ecological data that would remain undetected using standard similarity measures that do not include (phylogenetic and functional) information on the species.

References

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

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.

Bloom, S.A. 1981. Similarity indices in community studies: potential pitfalls. *Marine Ecology Progress Series* 5: 125–128.

Bryant, J.A., Lamanna, C., Morlon, H., Kerkhoff, A.J., Enquist, B.J. & Green, J.L. 2008. Microbes on mountainsides: contrasting elevational patterns of bacterial and plant diversity. *Proceedings of the National Academy of Sciences of the United States of America* 105: 11505–11511.

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.

Chao, A., Chazdon, R.L., Colwell, R.K. & Shen, T.J. 2006. Abundance-based similarity indices and their estimation when there are unseen species in samples. *Biometrics* 62: 361–371.

Chiu, C.-H., Jost, L. & Chao, A. 2014. Phylogenetic beta diversity, similarity, and differentiation measures based on Hill numbers. *Ecological Monographs* 84: 21–44.

Clarke, K.R., Somerfield, P.J. & Chapman, M. 2006. On resemblance measures for ecological studies, including taxonomic dissimilarities and a zero-adjusted Bray-Curtis coefficient for denuded assemblages. *Journal of Experimental Marine Biology and Ecology* 330: 55–80.

Diáz, S. & Cabido, M. 2001. Vive la différence: plant functional diversity matters to ecosystem processes. *Trends in Ecology & Evolution* 16: 646–655.

Ferrier, S., Manion, G., Elith, J. & Richardson, K. 2007. Using generalized dissimilarity modelling to analyse and predict patterns of beta diversity in regional biodiversity assessment. *Diversity and Distributions* 13: 252–264.

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

Jaccard, P. 1900. Contribution au problème de l'immigration post-glaciaire de la flore alpine. *Bulletin de la Société Vaudoise des Sciences Naturelles* 36: 87–130.

Juhász-Nagy, P. 1993. Notes on compositional diversity. *Hydrobiologia* 249: 173–182.

Doi: 10.1111/jvs.12329

Legendre, P. 2014. Interpreting the replacement and richness difference components of beta diversity. *Global Ecology and Biogeography* 23: 1324–1334.

Legendre, P. & Legendre, L. 2012. *Numerical ecology*. Elsevier Science, Amsterdam, NL.

Leinster, T. & Cobbold, C.A. 2012. Measuring diversity: the importance of species similarity. *Ecology* 93: 477–489.

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.

Lozupone, C.A. & Knight, R. 2005. UniFrac: a new phylogenetic method for comparing microbial communities. *Applied and Environmental Microbiology* 71: 8228–8235.

Nipperess, D.A., Faith, D.P. & Barton, K. 2010. Resemblance in phylogenetic diversity among ecological assemblages. *Journal of Vegetation Science* 21: 809–820.

Ochiai, A. 1957. Zoogeographic studies on the soleoid fishes found in Japan and its neighbouring regions. *Bulletin of the Japanese Society of Scientific Fisheries* 22: 526–530.

Pavoine, S. & Ricotta, C. 2014. Functional and phylogenetic similarity among communities. *Methods in Ecology and Evolution* 5: 666–675.

Podani, J. 2000. *Introduction to the exploration of multivariate biological data*. Backhuys, Leiden, NL.

Podani, J. & Miklós, I. 2002. Resemblance coefficients and the horseshoe effect in principal coordinates analysis. *Ecology* 83: 3331–3343.

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

Podani, J., Csontos, P., Tamás, J. & Miklós, I. 2005. A new multivariate approach to studying temporal changes of vegetation. *Plant Ecology* 181: 1–16.

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.

Ricotta, C., Bacaro, G. & Pavoine, S. 2015. A cautionary note on some phylogenetic dissimilarity measures. *Journal of Plant Ecology* 8: 12–16.

Roy, K. & Foote, M. 1997. Morphological approaches to measuring biodiversity. *Trends in Ecology & Evolution* 12: 277–281.

Sokal, R. & Michener, C. 1958. A statistical method for evaluating systematic relationships. *University of Kansas Scientific Bulletin* 38: 1409–1438.

Sokal, R.R. & Sneath, P.H.A. 1963. *Principles of numerical taxonomy*. Freeman, San Francisco, CA, US.

Sørensen, T. 1948. A method of establishing groups of equal amplitude in plant sociology based on similarity of species content. *Kongelige Danske Videnskabernes Selskabs Biologiske Skrifter* 5: 1–34.

Tamás, 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.

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

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Appendix S1. Definition of some alternative versions of the ISR-based matching/mismatching components of the generalized 2 × 2 contingency table.

Appendix S2. Interspecies similarity matrices **M1**–**M3** used for the calculation of the ISR-based versions of indices in Table 2 (their associated dissimilarity matrices, i.e. 1-M1, 1-M2 and 1-M3, respectively, are given as text files in Appendices S6 to S8).

Appendix S3. R script for the calculation of the similarity indices.

Appendix S4. Manual for using the R script and examples. The examples are based on Table 1 in the main text (see Appendix S5) and on the dissimilarity matrices 1-M1, 1-M2 and 1-M3 (see Appendices S6 to S8).

Appendix S5. Table 1 in text format.

Appendix S6. Dissimilarity matrix 1-M1 in text format.

Appendix S7. Dissimilarity matrix 1-M2 in text format.

Appendix S8. Dissimilarity matrix 1-M3 in text format.

Appendix S1

Definition of some alternative versions of the ISR-based matching/mismatching components of the generalized 2×2 contingency table.

Unlike in the main text, if the absolute abundances of species similar to i are calculated as $Z_{ik} = \sum_j^S x_{jk} s_{ij}$ taking the sum over all species in the matrix of S species × N plots, we can define a new contingency table with components

$$A_S = \sum_i^S \min\{Z_{ik}, Z_{ih}\} \tag{A1}$$

$$B_S = \sum_i^S \left(\max\{Z_{ik}, Z_{ih}\} - Z_{ih}\right) \tag{A2}$$

$$C_S = \sum_i^S \left(\max\{Z_{ik}, Z_{ih}\} - Z_{ik}\right) \tag{A3}$$

Here, the subscript S means that the sum in Eqs. (A1–A3) is taken over all S species of the S×N matrix. This enables us to define a D_S component of the generalized contingency table in analogy to Tamás et al. (2001; see Eq. 4 in the main text):

$$D_S = \sum_i^S \left(\max_k\{Z_{ik}\} - \max\{Z_{ik}, Z_{ih}\}\right) \tag{A4}$$

Besides dealing with absolute abundances, a definition of the matching/mismatching components of the generalized contingency table based on relative abundances is also possible. In this case, the A, B and C components are defined as:

$$A_S = \sum_i^S \min\{z_{ik}, z_{ih}\} \tag{A5}$$

$$B_S = \sum_i^S \left(\max\{z_{ik}, z_{ih}\} - z_{ih}\right) \tag{A6}$$

$$C_S = \sum_i^S \left(\max\{z_{ik}, z_{ih}\} - z_{ik}\right) \tag{A7}$$

where $z_{ik} = \sum_j^S p_{jk} s_{ij}$ is the ordinariness of species i in plot k, p_{ik} is the corresponding relative abundance, such that $0 \le p_{ik} \le 1$ and $\sum_i^{S(k)} p_{ik} = 1$, and S(k) are the species in plot k. The lower-case letter z_{ik} is used here for denoting the species ordinariness computed from relative abundances p_i.

In analogy to Eq. (A4), we can further define the corresponding D_S component as:

$$D_S = \sum_i^S \left(\max_k\{z_{ik}\} - \max\{z_{ik}, z_{ih}\}\right) \tag{A8}$$

As an alternative, since the quantity z_{ik} is comprised in the range [0, 1], we can also substitute the term $\max_k \{z_{ik}\}$ in Eq. (A8) with the theoretical upper bound for z_{ik}. Accordingly, the component D_S can also be defined as:

$$D_S = \sum_i^S \left(1 - \max\{z_{ik}, z_{ih}\}\right) \quad \text{(A9)}$$

such that, like for the presence/absence case, A+B+C+D = S.
Like for absolute abundances, if in Eqs. (A5–A7) the sum is taken over the species S(kh) that are actually present in at least one of the two plots k and h, we can define an additional contingency table with components:

$$A = \sum_i^{S(kh)} \min\{z_{ik}, z_{ih}\} \quad \text{(A10)}$$

$$B = \sum_i^{S(kh)} \left(\max\{z_{ik}, z_{ih}\} - z_{ih}\right) \quad \text{(A11)}$$

$$C = \sum_i^{S(kh)} \left(\max\{z_{ik}, z_{ih}\} - z_{ik}\right) \quad \text{(A12)}$$

where $z_{ik} = \sum_j^{S(kh)} p_{jk} s_{ij}$. Eqs. (A10–A12) are computed based on the plot-level species pool S(kh). However, due to the presence of an upper bound for the quantity z_{ik}, in analogy to Eq. (A9), we can define a D component as:

$$D_S = \sum_i^{S(kh)} \left(1 - \max\{z_{ik}, z_{ih}\}\right) \quad \text{(A13)}$$

Note that the different formulas for the D components are included here for completeness, and to allow a one-to-one correspondence between the presence/absence measures of similarity that comprise mutual absences and the newly proposed families of ISR-based similarities.
According to Eqs. (A1–A13) and (6–9) in the main text, virtually all presence/absence (dis)similarity coefficients can be generalized to include ecological resemblances among species. For example, the dissimilarity index recently proposed by Ricotta et al. (2014) is actually the classical Sørensen index of dissimilarity (b+c)/(2a+b+c) for presence and absence data calculated according to Eqs. (A10–A12).
In this view, although the various definitions of the generalized contingency table proposed in this paper start from different assumptions (i.e. absolute vs. relative abundances and plot-level vs. matrix-level species pools), nonetheless they use the same methodological approach, thus maintaining a consistent general structure. Practitioners should determine which definition of the generalized contingency table is more adequate within the context of their analyses.
Dealing with relative abundances the quantity $z_{ik} = \sum_j^{S(k)} p_{jk} s_{ij}$ calculated for one single species assemblage k is the expected ordinariness of species i in k (independently of any other plot). That is, the expected similarity between an individual of species i and an individual

chosen at random from plot k. Based on this quantity Leinster & Cobbold (2012) defined the average ordinariness of a given species assemblage as the expected similarity between two individuals chosen at random with replacement from the assemblage $z = \sum_i^{S(k)} p_{ik} \sum_j^{S(k)} p_{jk} s_{ij}$. Intuitively, average ordinariness z is inversely related to diversity. For instance, a standard measure of diversity is given by the complement $1 - z$, which is known as the Rao (1982) quadratic diversity $Q = \sum_i^{S(k)} p_{ik} \sum_j^{S(k)} p_{jk} d_{ij}$ (i.e. the expected *dis*similarity between two individuals chosen at random with replacement from a given community). Leinster & Cobbold (2012) also showed that the reciprocal of average ordinariness $1/z$ is the species equivalent of Q (i.e. the theoretical species richness of a perfectly even community with the same Rao diversity as the original community; see Ricotta & Szeidl 2009). Hence, in a sense, species ordinariness z_i is located at the crossroad of diversity theory and multivariate dissimilarity measures.

References

Leinster, T. & Cobbold, C.A. 2012. Measuring diversity: the importance of species similarity. *Ecology* 93: 477–89.

Rao, C.R. 1982. Diversity and dissimilarity coefficients: a unified approach. *Theoretical Population Biology* 21: 24–43.

Ricotta, C., Bacaro, G. & Pavoine, S. 2015. A cautionary note on some phylogenetic dissimilarity measures. *Journal of Plant Ecology* 8: 12–16.

Ricotta, C. & Szeidl, L. 2009. Diversity partitioning of Rao's quadratic entropy. *Theoretical Population Biology* 76: 299–302.

Tamás, 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.

Appendix S2

Interspecies similarity matrices **M1**–**M3** used for the calculation of the ISR-based versions of indices in Table 2.

In matrix **M1**, interspecies similarities s_{ij} reflect the species ecological differences along the artificial gradient of Table 1; for additional details, see the main text. In matrix **M2**, interspecies similarities were distributed randomly, and in matrix **M3,** all interspecies similarities were uniformly set to $s_{ij} = 0.33$ with $s_{ii} = 1$ by definition.

Matrix **M1**

	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15
S1	1	1	0.9	0.8	0.8	0.7	0.6	0.6	0.5	0.6	0.4	0.3	0	0.1	0.2
S2	1	1	0.9	0.8	0.8	0.7	0.6	0.6	0.5	0.6	0.4	0.3	0	0.1	0.2
S3	0.9	0.9	1	0.9	0.7	0.6	0.7	0.7	0.6	0.5	0.3	0.2	0	0.2	0.1
S4	0.8	0.8	0.9	1	0.6	0.5	0.8	0.6	0.7	0.4	0.2	0.1	0.1	0.1	0
S5	0.8	0.8	0.7	0.6	1	0.9	0.6	0.6	0.5	0.6	0.6	0.3	0	0.1	0.2
S6	0.7	0.7	0.6	0.5	0.9	1	0.7	0.7	0.6	0.7	0.7	0.4	0	0.2	0.3
S7	0.6	0.6	0.7	0.8	0.6	0.7	1	0.8	0.9	0.6	0.4	0.3	0.3	0.3	0.2
S8	0.6	0.6	0.7	0.6	0.6	0.7	0.8	1	0.9	0.8	0.6	0.5	0.3	0.5	0.4
S9	0.5	0.5	0.6	0.7	0.5	0.6	0.9	0.9	1	0.7	0.5	0.4	0.4	0.4	0.3
S10	0.6	0.6	0.5	0.4	0.6	0.7	0.6	0.8	0.7	1	0.8	0.7	0.3	0.5	0.6
S11	0.4	0.4	0.3	0.2	0.6	0.7	0.4	0.6	0.5	0.8	1	0.7	0.3	0.5	0.6
S12	0.3	0.3	0.2	0.1	0.3	0.4	0.3	0.5	0.4	0.7	0.7	1	0.6	0.8	0.9
S13	0	0	0	0.1	0	0	0.3	0.3	0.4	0.3	0.3	0.6	1	0.8	0.7
S14	0.1	0.1	0.2	0.1	0.1	0.2	0.3	0.5	0.4	0.5	0.5	0.8	0.8	1	0.9
S15	0.2	0.2	0.1	0	0.2	0.3	0.2	0.4	0.3	0.6	0.6	0.9	0.7	0.9	1

Matrix **M2**

	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15
S1	1	0.39	0.57	0.45	0.62	0.02	0.15	0.13	0.98	0.59	0.37	0.34	0.93	0.46	0.46
S2	0.39	1	0.09	0.30	0.11	0.11	0.14	0.65	0.93	0.54	0.16	0.31	0.20	0.69	0.11
S3	0.57	0.09	1	0.20	0.39	0.13	0.39	0.29	0.14	0.15	0.65	0.80	0.42	0.30	0.03
S4	0.45	0.30	0.20	1	0.57	0.01	1.00	0.72	0.71	0.08	0.43	0.70	0.10	0.88	1.00
S5	0.62	0.11	0.39	0.57	1	0.78	0.20	0.65	0.21	0.95	0.23	0.51	0.08	0.44	0.05
S6	0.02	0.11	0.13	0.01	0.78	1	0.77	0.56	0.18	0.70	0.15	0.14	0.69	0.86	0.54
S7	0.15	0.14	0.39	1.00	0.20	0.77	1	0.43	0.58	0.85	0.09	0.08	0.75	0.38	0.90
S8	0.13	0.65	0.29	0.72	0.65	0.56	0.43	1	0.43	0.54	0.81	0.92	0.26	0.19	0.20
S9	0.98	0.93	0.14	0.71	0.21	0.18	0.58	0.43	1	0.03	0.43	0.09	0.97	0.36	0.95
S10	0.59	0.54	0.15	0.08	0.95	0.70	0.85	0.54	0.03	1	0.02	0.01	0.46	0.73	0.39
S11	0.37	0.16	0.65	0.43	0.23	0.15	0.09	0.81	0.43	0.02	1	0.82	0.97	0.47	0.13
S12	0.34	0.31	0.80	0.70	0.51	0.14	0.08	0.92	0.09	0.01	0.82	1	0.55	0.69	0.20
S13	0.93	0.20	0.42	0.10	0.08	0.69	0.75	0.26	0.97	0.46	0.97	0.55	1	0.43	0.59
S14	0.46	0.69	0.30	0.88	0.44	0.86	0.38	0.19	0.36	0.73	0.47	0.69	0.43	1	0.25
S15	0.46	0.11	0.03	1.00	0.05	0.54	0.90	0.20	0.95	0.39	0.13	0.20	0.59	0.25	1

Matrix **M3**

	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15
S1	1	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33
S2	0.33	1	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33
S3	0.33	0.33	1	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33
S4	0.33	0.33	0.33	1	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33
S5	0.33	0.33	0.33	0.33	1	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33
S6	0.33	0.33	0.33	0.33	0.33	1	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33
S7	0.33	0.33	0.33	0.33	0.33	0.33	1	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33
S8	0.33	0.33	0.33	0.33	0.33	0.33	0.33	1	0.33	0.33	0.33	0.33	0.33	0.33	0.33
S9	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	1	0.33	0.33	0.33	0.33	0.33	0.33
S10	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	1	0.33	0.33	0.33	0.33	0.33
S11	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	1	0.33	0.33	0.33	0.33
S12	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	1	0.33	0.33	0.33
S13	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	1	0.33	0.33
S14	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	1	0.33
S15	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	1

```
dissABC<-function(abundances, dissimilarities, option = 1:4, method =
c("J", "S", "O", "K", "SS"))
{

     if(inherits(dissimilarities,     "dist"))     dissimilarities     <-
as.matrix(dissimilarities)
     if(any(dissimilarities>1)){
        warning("Phylogenetic dissimilarities are not in the range 0-1.
They have been normalized by the maximum")
        dissimilarities <- dissimilarities/max(dissimilarities)
    }
    if(any(!colnames(abundances) %in%rownames(dissimilarities))) stop("At
least one species in the matrix of abundances is missing in the matrix of
dissimilarities")
    if(any(!colnames(abundances) %in%colnames(dissimilarities))) stop("At
least one species in the matrix of abundances is missing in the matrix of
dissimilarities")
    dissimilarities       <-       dissimilarities[colnames(abundances),
colnames(abundances)]
    dataset <- t(abundances)
    similarities <- 1-as.matrix(dissimilarities)
    if(option[1]%in%c(1, 2)) rel_abu <- dataset
    else{
        total <- colSums(dataset)
        rel_abu <- sweep(dataset, 2, total, "/")
    }
    num.plot<-dim(dataset)[2]
    num.sp <- dim(dataset)[1]
    names<-list(colnames(dataset), colnames(dataset))
    sim.matrix<-matrix(0, nrow=num.plot, ncol=num.plot, dimnames=names)
    for (i in 2:num.plot) {
    for (j in 1:(i-1)) {
         if(option[1]%in%c(1,3)) {
            garde <- (1:num.sp)[(rel_abu[, j]+rel_abu[, i])>0]
            sim <- similarities[garde, garde]
            x <- rel_abu[, j]
            x <- x[garde]
            y <- rel_abu[, i]
            y <- y[garde]
            mat_folk <- sim*x
            mat_folk2 <- sim*y
        }
         else{
            mat_folk <- similarities*rel_abu[, j]
            mat_folk2 <- similarities*rel_abu[, i]
         }
        Zik <- colSums(mat_folk)
        Zih <- colSums(mat_folk2)
        tabZ <- rbind.data.frame(Zik, Zih)
        A <- sum(sapply(tabZ, min))
        B <- sum(sapply(tabZ, max)-Zih)
        C <- sum(sapply(tabZ, max)-Zik)
        if(method == "J") index <- A/(A+B+C)
        else if(method == "S") index <- 2*A/(2*A+B+C)
        else if(method == "O") index <- A/(sqrt(A+B)*sqrt(A+C))
        else if(method == "K") index <- 0.5*(A/(A+B)+A/(A+C))
        else if(method == "SS") index <- A/(A+2*B+2*C)
        sim.matrix[i, j] <- index
    }
    }
```

```
    sim.matrix <- sim.matrix + t(sim.matrix)
    sim.matrix <- sim.matrix + diag(rep(1, num.plot))
    return(sim.matrix)
}
```

Appendix S4. Manual for using the R script and examples.

Manual for using the R function "dissABC" for calculating plot-to-plot ISR-based similarities (the R script of the function is available in Appendix S3). This program is free software: you can redistribute it and/or modify it under the terms of the GNU General Public License http://www.gnu.org/licenses/.

Disclaimer: users of this code are cautioned that, while due care has been taken and it is believed accurate, it has not been rigorously tested and its use and results are solely the responsibilities of the user.

Description: given a matrix of *S* species' relative or absolute abundance values × *N* plots, together with an *S* × *S* (phylogenetic or functional) dissimilarity matrix, this function calculates a matrix with the values of the ISR-based similarities for each pair of plots, as proposed in the main text and Appendix S1.

Dependencies: none.

Usage: dissABC(abundances, dissimilarities, option = 1:4, method = c("J", "S", "O", "K", "SS"))

Arguments
abundances: a matrix (or data frame) of *N* plots × *S* species containing the absolute abundances of all species in the plots. Columns are species and plots are rows. Column labels (species names) should be assigned as in the object '*dissimilarities'*.

dissimilarities: an *S* × *S* matrix (or data frame) of (phylogenetic or functional) dissimilarities among species rescaled in the range [0, 1] or an object of class 'dist' [obtained by functions like vegdist in package vegan (Oksanen et al. 2013), gowdis in package FD (Laliberté & Shipley 2011), or dist.ktab in package ade4 for functional dissimilarities (Dray et al. 2007), or functions like cophenetic.phylo in package ape (Paradis et al. 2004) or distTips in package adephylo (Jombart & Dray 2010) for phylogenetic dissimilarities]. Note that dissimilarities among species need first to be rescaled in the range [0, 1]. If the dissimilarities are outside the range 0-1, a warning message is displayed.

option: a numeric, either 1, 2, 3, or 4 (if several are given only the first one is considered). If 1, equations 6-8 of the main text are used for calculating components A, B, C. If 2, equations A1-A3 from Appendix S1 are used. If 3, equations A5-A7 from Appendix S1 are used. If 4, equations A10-A12 are used.

method: a character, either "J", "S", "O", "K", or "SS" (if several are given only the first one is considered). "J"=the Jaccard index is used; "S"=the Sørensen index; "O"=the Ochiai index; "K"=the Kulczynski index; "SS"=the Sokal-Sneath index.

Output: the function returns a matrix with the values of the proposed ISR-based similarities among plots.

Example

Load Appendix S3 in the R console. The appendix contains the R function dissABC.

```
source("PATH THROUGH APPENDIX S3.txt")
```

In the instruction above, replace PATH THROUGH APPENDIX S3 by the actual path through the Appendix file in your computer. Alternatively use:

```
source(file.choose())
```

Load Appendix S5 in the R console. The appendix contains Table 1 from the main text:

```
ab <- read.table("NAME OF APPENDIX S5.txt", h=T, row.names=1)
```

In the instruction above, replace ‘NAME OF APPENDIX S5’ by the actual name of this Appendix file.

```
ab
   S1 S2 S3 S4 S5 S6 S7 S8 S9 S10 S11 S12 S13 S14 S15
P1  2  2  3  4  0  0  2  0  1   0   0   0   0   0   0
P2  1  1  2  3  1  0  3  1  2   0   0   0   0   0   0
P3  0  0  1  2  0  1  4  2  3   0   0   0   0   0   0
P4  0  0  0  1  0  0  3  3  4   1   0   0   0   0   0
P5  0  0  0  0  0  0  2  2  3   2   0   0   1   0   0
P6  0  0  0  0  0  0  1  1  2   1   1   0   2   0   0
P7  0  0  0  0  0  0  0  0  1   0   0   1   3   1   0
P8  0  0  0  0  0  0  0  0  0   0   0   2   4   2   1
P9  0  0  0  0  0  0  0  0  0   0   0   1   5   3   2
```

ab contains species as columns and plots as rows.

Now load the matrices of dissimilarity among species. These matrices are contained in appendices S6 to S8. Use the same process as above to load the files:

```
D1 <- read.table("NAME OF APPENDIX S6.txt", h=T, row.names=1)
D2 <- read.table("NAME OF APPENDIX S7.txt", h=T, row.names=1)
D3 <- read.table("NAME OF APPENDIX S8.txt", h=T, row.names=1)
```

D1 contains similarities among species as defined by 1-M1 were M1 is the matrix of similarities given in Appendix S2. Similarly, D2 contains 1-M2; and D3 contains 1-M3, see for example:

```
1-D1
     S1  S2  S3  S4  S5  S6  S7  S8  S9 S10 S11 S12 S13 S14 S15
S1  1.0 1.0 0.9 0.8 0.8 0.7 0.6 0.6 0.5 0.6 0.4 0.3 0.0 0.1 0.2
S2  1.0 1.0 0.9 0.8 0.8 0.7 0.6 0.6 0.5 0.6 0.4 0.3 0.0 0.1 0.2
S3  0.9 0.9 1.0 0.9 0.7 0.6 0.7 0.7 0.6 0.5 0.3 0.2 0.0 0.2 0.1
S4  0.8 0.8 0.9 1.0 0.6 0.5 0.8 0.6 0.7 0.4 0.2 0.1 0.1 0.1 0.0
S5  0.8 0.8 0.7 0.6 1.0 0.9 0.6 0.6 0.5 0.6 0.6 0.3 0.0 0.1 0.2
S6  0.7 0.7 0.6 0.5 0.9 1.0 0.7 0.7 0.6 0.7 0.7 0.4 0.0 0.2 0.3
S7  0.6 0.6 0.7 0.8 0.6 0.7 1.0 0.8 0.9 0.6 0.4 0.3 0.3 0.3 0.2
S8  0.6 0.6 0.7 0.6 0.6 0.7 0.8 1.0 0.9 0.8 0.6 0.5 0.3 0.5 0.4
S9  0.5 0.5 0.6 0.7 0.5 0.6 0.9 0.9 1.0 0.7 0.5 0.4 0.4 0.4 0.3
S10 0.6 0.6 0.5 0.4 0.6 0.7 0.6 0.8 0.7 1.0 0.8 0.7 0.3 0.5 0.6
S11 0.4 0.4 0.3 0.2 0.6 0.7 0.4 0.6 0.5 0.8 1.0 0.7 0.3 0.5 0.6
S12 0.3 0.3 0.2 0.1 0.3 0.4 0.3 0.5 0.4 0.7 0.7 1.0 0.6 0.8 0.9
S13 0.0 0.0 0.0 0.1 0.0 0.0 0.3 0.3 0.4 0.3 0.3 0.6 1.0 0.8 0.7
S14 0.1 0.1 0.2 0.1 0.1 0.2 0.3 0.5 0.4 0.5 0.5 0.8 0.8 1.0 0.9
S15 0.2 0.2 0.1 0.0 0.2 0.3 0.2 0.4 0.3 0.6 0.6 0.9 0.7 0.9 1.0
```

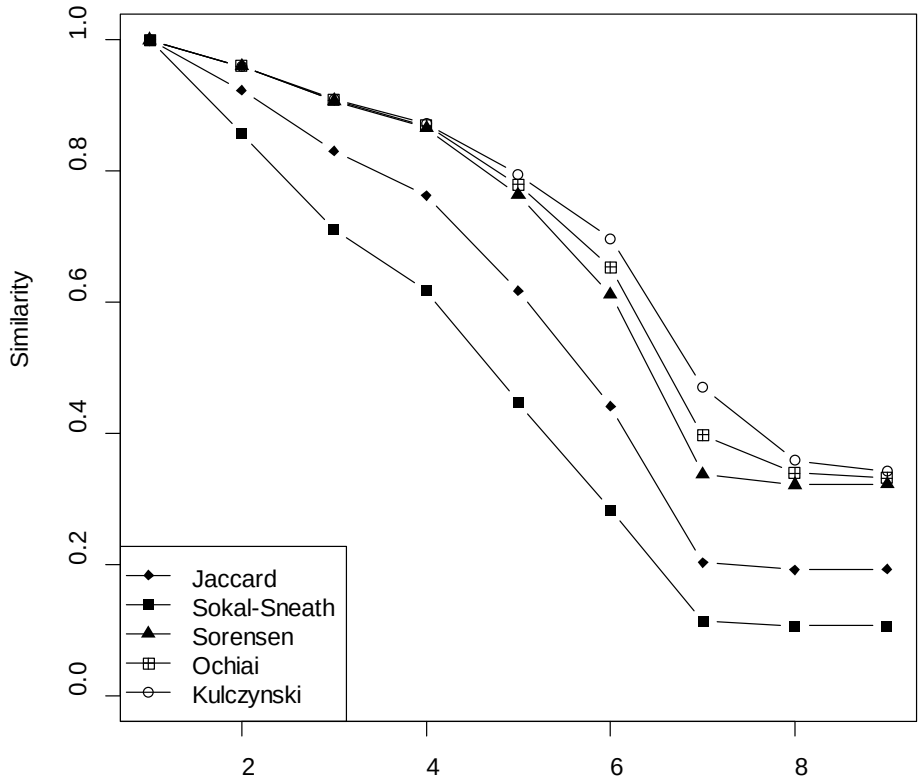
To obtain the dissimilarities among plots, one needs to choose the equations to be used for the components A, B, and C thanks to parameter "option" and the way the parameters will be combined as explained in Table 2 of the main text, thanks to parameter “method” (see above for details). For example, to use equations (6-8) from the main text, one needs to specify option=1, and to use the Jaccard index, one needs to specify method="J", leading to:

```
dissABC(ab, D1, method="J", option=1)
      P1    P2    P3    P4    P5    P6    P7    P8    P9
P1 1.000 0.923 0.830 0.763 0.618 0.441 0.202 0.192 0.192
P2 0.923 1.000 0.892 0.816 0.629 0.445 0.229 0.229 0.234
```

```
P3 0.830 0.892 1.000 0.925 0.717 0.515 0.307 0.302 0.303
P4 0.763 0.816 0.925 1.000 0.786 0.575 0.393 0.404 0.404
P5 0.618 0.629 0.717 0.786 1.000 0.737 0.515 0.512 0.499
P6 0.441 0.445 0.515 0.575 0.737 1.000 0.648 0.616 0.589
P7 0.202 0.229 0.307 0.393 0.515 0.648 1.000 0.624 0.509
P8 0.192 0.229 0.302 0.404 0.512 0.616 0.624 1.000 0.813
P9 0.192 0.234 0.303 0.404 0.499 0.589 0.509 0.813 1.000
```

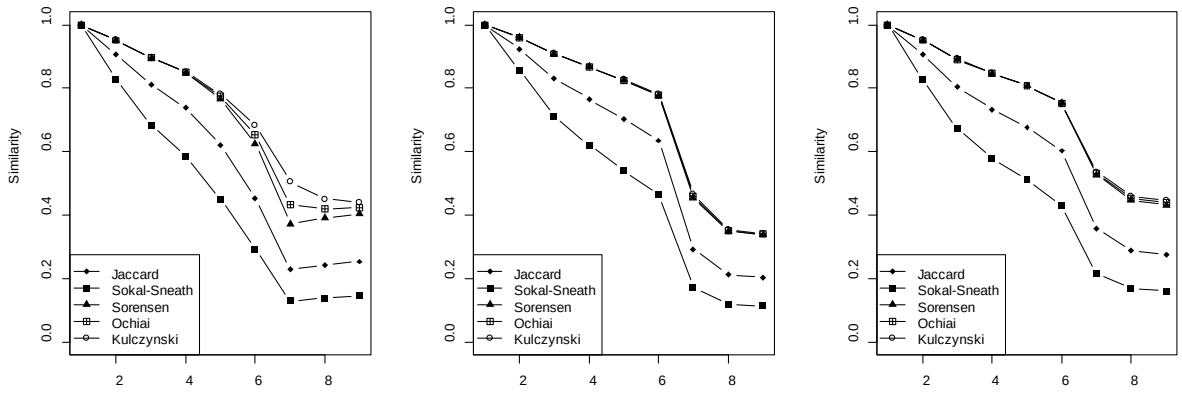
Figure 1 from the main text can be obtained as follows:

```
J <- as.matrix(dissABC(ab, D1, method="J", option=1))[, 1]
SS <- as.matrix(dissABC(ab, D1, method="SS", option=1))[, 1]
S <- as.matrix(dissABC(ab, D1, method="S", option=1))[, 1]
O <- as.matrix(dissABC(ab, D1, method="O", option=1))[, 1]
K <- as.matrix(dissABC(ab, D1, method="K", option=1))[, 1]

plot(1:9, J, xlab="number of the plots which plot 1 is compared to",
ylab="Similarity", type="b", ylim=c(0,1), pch=18)
lines(1:9, SS, type="b", pch=15)
lines(1:9, S, type="b", pch=17)
lines(1:9, O, type="b", pch=12)
lines(1:9, K, type="b", pch=1)
legend("bottomleft", c("Jaccard", "Sokal-Sneath", "Sorensen", "Ochiai",
"Kulczynski"), pch=c(18,15,17,12,1), lty=1)
```



With the other three options, the response of the ISI-based similarity functions (calculated with matrix M1) to the artificial gradient would have led to:

```
par(mfrow=c(1,3))

J <- as.matrix(dissABC(ab, D1, method="J", option=2))[, 1]
SS <- as.matrix(dissABC(ab, D1, method="SS", option=2))[, 1]
S <- as.matrix(dissABC(ab, D1, method="S", option=2))[, 1]
O <- as.matrix(dissABC(ab, D1, method="O", option=2))[, 1]
K <- as.matrix(dissABC(ab, D1, method="K", option=2))[, 1]

plot(1:9, J, xlab="number of the plots which plot 1 is compared to",
ylab="Similarity", type="b", ylim=c(0,1), pch=18)
lines(1:9, SS, type="b", pch=15)
lines(1:9, S, type="b", pch=17)
lines(1:9, O, type="b", pch=12)
lines(1:9, K, type="b", pch=1)
legend("bottomleft", c("Jaccard", "Sokal-Sneath", "Sorensen", "Ochiai",
"Kulczynski"), pch=c(18,15,17,12,1), lty=1)


J <- as.matrix(dissABC(ab, D1, method="J", option=3))[, 1]
SS <- as.matrix(dissABC(ab, D1, method="SS", option=3))[, 1]
S <- as.matrix(dissABC(ab, D1, method="S", option=3))[, 1]
O <- as.matrix(dissABC(ab, D1, method="O", option=3))[, 1]
K <- as.matrix(dissABC(ab, D1, method="K", option=3))[, 1]

plot(1:9, J, xlab="number of the plots which plot 1 is compared to",
ylab="Similarity", type="b", ylim=c(0,1), pch=18)
lines(1:9, SS, type="b", pch=15)
lines(1:9, S, type="b", pch=17)
lines(1:9, O, type="b", pch=12)
lines(1:9, K, type="b", pch=1)
legend("bottomleft", c("Jaccard", "Sokal-Sneath", "Sorensen", "Ochiai",
"Kulczynski"), pch=c(18,15,17,12,1), lty=1)


J <- as.matrix(dissABC(ab, D1, method="J", option=4))[, 1]
SS <- as.matrix(dissABC(ab, D1, method="SS", option=4))[, 1]
S <- as.matrix(dissABC(ab, D1, method="S", option=4))[, 1]
O <- as.matrix(dissABC(ab, D1, method="O", option=4))[, 1]
K <- as.matrix(dissABC(ab, D1, method="K", option=4))[, 1]

plot(1:9, J, xlab="number of the plots which plot 1 is compared to",
ylab="Similarity", type="b", ylim=c(0,1), pch=18)
lines(1:9, SS, type="b", pch=15)
lines(1:9, S, type="b", pch=17)
lines(1:9, O, type="b", pch=12)
lines(1:9, K, type="b", pch=1)
legend("bottomleft", c("Jaccard", "Sokal-Sneath", "Sorensen", "Ochiai",
"Kulczynski"), pch=c(18,15,17,12,1), lty=1)
```

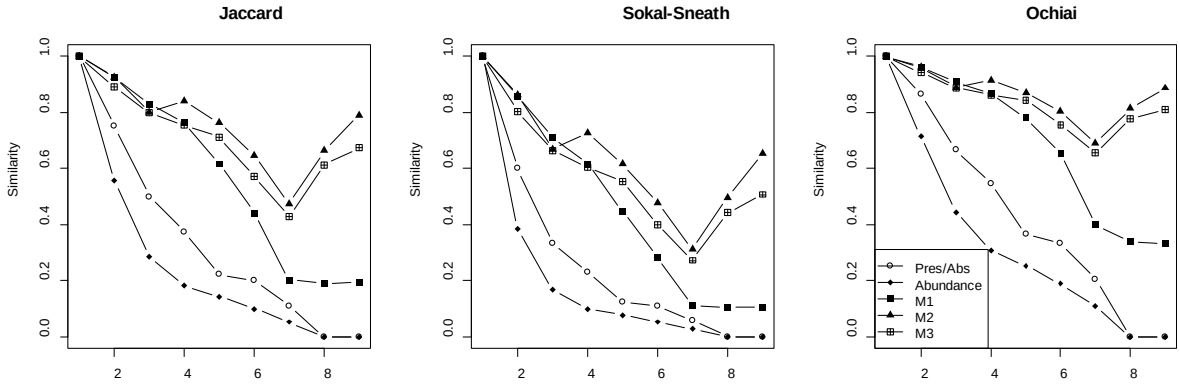


Equations used for the ISR-based components: Left= equations A1, A2, A3; Middle = A5, A6, A7, Right = A10, A11, A12.

Figure 2 from the main text can be obtained as follows:

```
par(mfrow=c(1,3))
J1 <- as.matrix(dissABC(ab, D1, method="J", option=1))[, 1]
J2 <- as.matrix(dissABC(ab, D2, method="J", option=1))[, 1]
J3 <- as.matrix(dissABC(ab, D3, method="J", option=1))[, 1]
U <- as.dist(matrix(1, 15, 15))
attributes(U)$Labels <- paste("S", 1:15, sep="")
Jab <- as.matrix(dissABC(ab, U, method="J", option=1))[, 1]
pres <- ab
pres[pres>0] <- 1
Jpa <- as.matrix(dissABC(pres, U, method="J", option=1))[, 1]

plot(1:9, Jpa, xlab="number of the plots which plot 1 is compared to",
ylab="Similarity", type="b", ylim=c(0,1), pch=1, main="Jaccard")
lines(1:9, Jab, type="b", pch=18)
lines(1:9, J1, type="b", pch=15)
lines(1:9, J2, type="b", pch=17)
lines(1:9, J3, type="b", pch=12)

SS1 <- as.matrix(dissABC(ab, D1, method="SS", option=1))[, 1]
SS2 <- as.matrix(dissABC(ab, D2, method="SS", option=1))[, 1]
SS3 <- as.matrix(dissABC(ab, D3, method="SS", option=1))[, 1]
SSab <- as.matrix(dissABC(ab, U, method="SS", option=1))[, 1]
SSpa <- as.matrix(dissABC(pres, U, method="SS", option=1))[, 1]

plot(1:9, SSpa, xlab="number of the plots which plot 1 is compared to",
ylab="Similarity", type="b", ylim=c(0,1), pch=1, main="Sokal-Sneath")
lines(1:9, SSab, type="b", pch=18)
lines(1:9, SS1, type="b", pch=15)
lines(1:9, SS2, type="b", pch=17)
lines(1:9, SS3, type="b", pch=12)

O1 <- as.matrix(dissABC(ab, D1, method="O", option=1))[, 1]
O2 <- as.matrix(dissABC(ab, D2, method="O", option=1))[, 1]
O3 <- as.matrix(dissABC(ab, D3, method="O", option=1))[, 1]
Oab <- as.matrix(dissABC(ab, U, method="O", option=1))[, 1]
Opa <- as.matrix(dissABC(pres, U, method="O", option=1))[, 1]
```

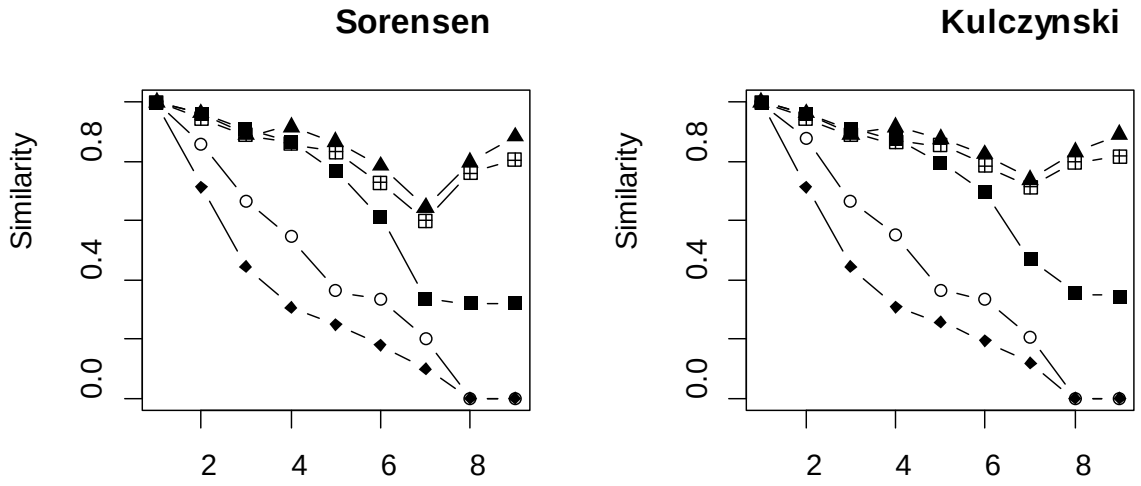
```
plot(1:9, Opa, xlab="number of the plots which plot 1 is compared to",
ylab="Similarity", type="b", ylim=c(0,1), pch=1, main="Ochiai")
lines(1:9, Oab, type="b", pch=18)
lines(1:9, O1, type="b", pch=15)
lines(1:9, O2, type="b", pch=17)
lines(1:9, O3, type="b", pch=12)


legend("bottomleft", c("Pres/Abs", "Abundance", "M1", "M2", "M3"),
pch=c(1,18,15,17,12), lty=1)
```



The same analysis with Sørensen and Kulczynski coefficients leads to:

```
par(mfrow=c(1,2))
S1 <- as.matrix(dissABC(ab, D1, method="S", option=1))[, 1]
S2 <- as.matrix(dissABC(ab, D2, method="S", option=1))[, 1]
S3 <- as.matrix(dissABC(ab, D3, method="S", option=1))[, 1]
Sab <- as.matrix(dissABC(ab, U, method="S", option=1))[, 1]
Spa <- as.matrix(dissABC(pres, U, method="S", option=1))[, 1]

plot(1:9, Spa, xlab="number of the plots which plot 1 is compared to",
ylab="Similarity", type="b", ylim=c(0,1), pch=1, main="Sorensen")
lines(1:9, Sab, type="b", pch=18)
lines(1:9, S1, type="b", pch=15)
lines(1:9, S2, type="b", pch=17)
lines(1:9, S3, type="b", pch=12)

K1 <- as.matrix(dissABC(ab, D1, method="K", option=1))[, 1]
K2 <- as.matrix(dissABC(ab, D2, method="K", option=1))[, 1]
K3 <- as.matrix(dissABC(ab, D3, method="K", option=1))[, 1]
Kab <- as.matrix(dissABC(ab, U, method="K", option=1))[, 1]
Kpa <- as.matrix(dissABC(pres, U, method="K", option=1))[, 1]

plot(1:9, Kpa, xlab="number of the plots which plot 1 is compared to",
ylab="Similarity", type="b", ylim=c(0,1), pch=1, main="Kulczynski")
lines(1:9, Kab, type="b", pch=18)
lines(1:9, K1, type="b", pch=15)
```

```
lines(1:9, K2, type="b", pch=17)
lines(1:9, K3, type="b", pch=12)
```



Same legend as above. Left Sørensen coefficient; Right: Kulczynski coefficient

References

Dray, S., Dufour, A.B. & Chessel, D. 2007. The ade4 package-II: Two-table and K-table methods. R News 7: 47-52.

Jombart, T. & Dray, S. 2010. Adephylo: exploratory analyses for the phylogenetic comparative method. Bioinformatics 26: 1907-1909.

Kembel, S.W., Cowan, P.D., Helmus, M.R., Cornwell, W.K., Morlon, H., Ackerly, D.D., Blomberg, S.P. & Webb, C.O. 2010. Picante: R tools for integrating phylogenies and ecology. Bioinformatics 26: 1463-1464.

Laliberté, E. & Shipley, B. 2011. FD: measuring functional diversity from multiple traits, and other tools for functional ecology. R package version 1.0-11. [http://cran.r-project.org/web/packages/FD/].

Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H. & Wagner, H. 2013. Vegan: Community Ecology Package. R package version 2.0-10. [http://cran.r-project.org/package/vegan].

Paradis, E., Claude, J. & Strimmer, K. 2004. APE: analyses of phylogenetics and evolution in R language. Bioinformatics 20: 289-290.

	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15
P1	2	2	3	4	0	0	2	0	1	0	0	0	0	0	0
P2	1	1	2	3	1	0	3	1	2	0	0	0	0	0	0
P3	0	0	1	2	0	1	4	2	3	0	0	0	0	0	0
P4	0	0	0	1	0	0	3	3	4	1	0	0	0	0	0
P5	0	0	0	0	0	0	2	2	3	2	0	0	1	0	0
P6	0	0	0	0	0	0	1	1	2	1	1	0	2	0	0
P7	0	0	0	0	0	0	0	0	1	0	0	1	3	1	0
P8	0	0	0	0	0	0	0	0	0	0	0	2	4	2	1
P9	0	0	0	0	0	0	0	0	0	0	0	1	5	3	2

	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15
S1	0	0	0.1	0.2	0.2	0.3	0.4	0.4	0.5	0.4	0.6	0.7	1	0.9	0.8
S2	0	0	0.1	0.2	0.2	0.3	0.4	0.4	0.5	0.4	0.6	0.7	1	0.9	0.8
S3	0.1	0.1	0	0.1	0.3	0.4	0.3	0.3	0.4	0.5	0.7	0.8	1	0.8	0.9
S4	0.2	0.2	0.1	0	0.4	0.5	0.2	0.4	0.3	0.6	0.8	0.9	0.9	0.9	1
S5	0.2	0.2	0.3	0.4	0	0.1	0.4	0.4	0.5	0.4	0.4	0.7	1	0.9	0.8
S6	0.3	0.3	0.4	0.5	0.1	0	0.3	0.3	0.4	0.3	0.3	0.6	1	0.8	0.7
S7	0.4	0.4	0.3	0.2	0.4	0.3	0	0.2	0.1	0.4	0.6	0.7	0.7	0.7	0.8
S8	0.4	0.4	0.3	0.4	0.4	0.3	0.2	0	0.1	0.2	0.4	0.5	0.7	0.5	0.6
S9	0.5	0.5	0.4	0.3	0.5	0.4	0.1	0.1	0	0.3	0.5	0.6	0.6	0.6	0.7
S10	0.4	0.4	0.5	0.6	0.4	0.3	0.4	0.2	0.3	0	0.2	0.3	0.7	0.5	0.4
S11	0.6	0.6	0.7	0.8	0.4	0.3	0.6	0.4	0.5	0.2	0	0.3	0.7	0.5	0.4
S12	0.7	0.7	0.8	0.9	0.7	0.6	0.7	0.5	0.6	0.3	0.3	0	0.4	0.2	0.1
S13	1	1	1	0.9	1	1	0.7	0.7	0.6	0.7	0.7	0.4	0	0.2	0.3
S14	0.9	0.9	0.8	0.9	0.9	0.8	0.7	0.5	0.6	0.5	0.5	0.2	0.2	0	0.1
S15	0.8	0.8	0.9	1	0.8	0.7	0.8	0.6	0.7	0.4	0.4	0.1	0.3	0.1	0

	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15
S1	0	0.61	0.43	0.55	0.38	0.98	0.85	0.87	0.02	0.41	0.63	0.66	0.07	0.54	0.54
S2	0.61	0	0.91	0.7	0.89	0.89	0.86	0.35	0.07	0.46	0.84	0.69	0.8	0.31	0.89
S3	0.43	0.91	0	0.8	0.61	0.87	0.61	0.71	0.86	0.85	0.35	0.2	0.58	0.7	0.97
S4	0.55	0.7	0.8	0	0.43	0.99	0	0.28	0.29	0.92	0.57	0.3	0.9	0.12	0
S5	0.38	0.89	0.61	0.43	0	0.22	0.8	0.35	0.79	0.05	0.77	0.49	0.92	0.56	0.95
S6	0.98	0.89	0.87	0.99	0.22	0	0.23	0.44	0.82	0.3	0.85	0.86	0.31	0.14	0.46
S7	0.85	0.86	0.61	0	0.8	0.23	0	0.57	0.42	0.15	0.91	0.92	0.25	0.62	0.1
S8	0.87	0.35	0.71	0.28	0.35	0.44	0.57	0	0.57	0.46	0.19	0.08	0.74	0.81	0.8
S9	0.02	0.07	0.86	0.29	0.79	0.82	0.42	0.57	0	0.97	0.57	0.91	0.03	0.64	0.05
S10	0.41	0.46	0.85	0.92	0.05	0.3	0.15	0.46	0.97	0	0.98	0.99	0.54	0.27	0.61
S11	0.63	0.84	0.35	0.57	0.77	0.85	0.91	0.19	0.57	0.98	0	0.18	0.03	0.53	0.87
S12	0.66	0.69	0.2	0.3	0.49	0.86	0.92	0.08	0.91	0.99	0.18	0	0.45	0.31	0.8
S13	0.07	0.8	0.58	0.9	0.92	0.31	0.25	0.74	0.03	0.54	0.03	0.45	0	0.57	0.41
S14	0.54	0.31	0.7	0.12	0.56	0.14	0.62	0.81	0.64	0.27	0.53	0.31	0.57	0	0.75
S15	0.54	0.89	0.97	0	0.95	0.46	0.1	0.8	0.05	0.61	0.87	0.8	0.41	0.75	0

	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15
S1	0	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66
S2	0.66	0	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66
S3	0.66	0.66	0	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66
S4	0.66	0.66	0.66	0	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66
S5	0.66	0.66	0.66	0.66	0	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66
S6	0.66	0.66	0.66	0.66	0.66	0	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66
S7	0.66	0.66	0.66	0.66	0.66	0.66	0	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66
S8	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0	0.66	0.66	0.66	0.66	0.66	0.66	0.66
S9	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0	0.66	0.66	0.66	0.66	0.66	0.66
S10	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0	0.66	0.66	0.66	0.66	0.66
S11	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0	0.66	0.66	0.66	0.66
S12	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0	0.66	0.66	0.66
S13	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0	0.66	0.66
S14	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0	0.66
S15	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0