

A family of functional dissimilarity measures for presence and absence data

Carlo Ricotta¹, János Podani² & Sandrine Pavoine³

¹Department of Environmental Biology, University of Rome "La Sapienza", Piazzale Aldo Moro 5, 00185, Rome, Italy

²Department of Plant Systematics, Ecology and Theoretical Biology, MTA-ELTE-MTM Ecology Research Group, Institute of Biology, L. Eötvös University, Pázmány P. s. 1/C, H-1117, Budapest, Hungary

³Centre d'Ecologie et des Sciences de la Conservation (CESCO UMR7204), Sorbonne Universités, MNHN, CNRS, UPMC, CP51, 55-61 rue Buffon, Paris, France

Keywords

Between-species dissimilarities, contingency table, functional excess, functional turnover, species richness.

Correspondence

Carlo Ricotta, Department of Environmental Biology, University of Rome "La Sapienza", Piazzale Aldo Moro 5, 00185, Rome, Italy.
Tel: +390649912504;
E-mail: carlo.ricotta@uniroma1.it

Funding Information

Funding was provided by the University of Rome 'La Sapienza' (C26A15CCWF).

Received: 19 February 2016; Revised: 6 May 2016; Accepted: 12 May 2016

Ecology and Evolution 2016; **6(15)**: 5383–5389

doi: 10.1002/ece3.2214

Introduction

Dissimilarity coefficients between pairs of species assemblages (communities, plots, relevés, sites, quadrats, etc.) provide a helpful tool for exploring the complex ecological mechanisms that drive community assembly. Most of these measures summarize different facets of plot-to-plot dissimilarity based either on species presence/absence scores or on species abundances, thus implicitly assuming that all species are equally and maximally distinct from one another (Ricotta et al. 2015).

In particular, given two plots X and Y , dissimilarity coefficients for presence and absence data are usually formulated using the matching/mismatching components a , b , and c of a 2×2 contingency table: a is the number of species present in both plots, b is the number of species present only in plot X , and c is the number of species present only in plot Y , such that $a + b + c$ is the total

Abstract

Plot-to-plot dissimilarity measures are considered a valuable tool for understanding the complex ecological mechanisms that drive community composition. Traditional presence/absence coefficients are usually based on different combinations of the matching/mismatching components of the 2×2 contingency table. However, more recently, dissimilarity measures that incorporate information about the degree of functional differences between the species in both plots have received increasing attention. This is because such "functional dissimilarity measures" capture information on the species' functional traits, which is ignored by traditional coefficients. Therefore, functional dissimilarity measures tend to correlate more strongly with ecosystem-level processes, as species influence these processes via their traits. In this study, we introduce a new family of dissimilarity measures for presence and absence data, which consider functional dissimilarities among species in the calculation of the matching/mismatching components of the 2×2 contingency table. Within this family, the behavior of the Jaccard coefficient, together with its additive components, species replacement, and richness difference, is examined by graphical comparisons and ordinations based on simulated data.

number of species in the two plots. Using the values of a , b , and c , a large number of coefficients can be calculated, such as the Jaccard (1900) index of dissimilarity $J = (b + c)/(a + b + c)$, or the Sørensen (1948) index $S = (b + c)/(2a + b + c)$. For a comprehensive inventory of presence/absence dissimilarity coefficients, see Podani (2000) and Legendre and Legendre (2012). A common feature of most presence/absence indices is that they are usually expressed in terms of a ratio where the numerator, that is the "operational part" of the index *sensu* Legendre (2014), estimates the amount of plot-to-plot dissimilarity depending on the purpose of the study, while the denominator scales the index to values between 0 and 1. Note that the fourth component of the contingency table d , which represents joint absences (i.e., the species absent from both plots being compared but found in other plots), is only rarely used in community ecology for the calculation of dissimilarity measures. Therefore, in

this study we will not discuss dissimilarity coefficients, which include mutual absences in their formulation.

More recently, dissimilarity measures that incorporate information about the degree of functional differences between the species in both assemblages have received considerable attention (Izsák and Price 2001; Champely and Chessel 2002; Pavoine et al. 2004; Ricotta and Bacaro 2010; Chiu and Chao 2014; Chiu et al. 2014; Pavoine and Ricotta 2014; Ricotta and Pavoine 2015). This is because such “functional dissimilarity measures” capture information on the species’ functional traits, which is ignored by traditional dissimilarity measures. Functional traits are morphological, physiological, and phenological attributes, which impact individual fitness via their effects on growth, reproduction, and survival (Violle et al. 2007). Therefore, measures of functional dissimilarity tend to correlate more strongly with ecosystem-level processes, such as productivity, regulation of biogeochemical fluxes, or resilience to perturbations, as species influence these processes via their traits (Mason and de Bello 2013).

The observed relationship between functional dissimilarity and ecosystem functioning raises the question of how to measure functional dissimilarity in meaningful ways. As ecological data are often multivariate with high dimensionality, no single measure summarizes adequately all aspects of functional dissimilarity. Thus, in order to assess relevant differences in the functional organization of species assemblages, a multifaceted approach is needed. While the notion of functional dissimilarity is independent of any particular method of measurement, a number of basic criteria have been proposed that functional dissimilarity measures should meet to behave reasonably in ecological research (e.g., Pavoine and Ricotta 2014). In this study, we first briefly summarize such basic requirements. Then, we introduce a new family of dissimilarity measures for presence and absence data, which consider functional dissimilarities among species in the calculation of the matching/mismatching components of the 2×2 contingency table. These presence/absence measures may be useful whenever data on species abundances are either unknown or irrelevant, for example, for large-scale environmental protection purposes. In these circumstances, presence/absence measures represent the adequate choice for quantifying the functional dissimilarity among plots.

Methods

Regardless of any specific measure of dissimilarity, in most cases the information available for summarizing the functional organization of a given set of plots is an $N \times K$ matrix of species presences and absences (i.e., usually 0/1 scores) for N species and K plots, together with an $N \times \tau$ matrix with values for τ selected functional traits for each species. As most plot-to-plot functional dissimilarity

indices are built on pairwise functional dissimilarities between species, this latter matrix is first transformed to an $N \times N$ matrix of pairwise functional dissimilarities d_{ij} between species i and j in the interval 0–1 with $d_{ii} = 0$ and $d_{ij} = d_{ji}$. Note that for any dissimilarity index d_{ij} falling into the unit range, its similarity counterpart s_{ij} can be simply calculated as $s_{ij} = 1 - d_{ij}$, so that $s_{ij} + d_{ij} = 1$.

In this context, to coherently frame the notion of plot-to-plot dissimilarity, many authors have proposed a number of basic requirements that a good index should meet to reasonably behave in ecological research (see Anderson et al. 2006; Clarke et al. 2006; Legendre and De Cáceres 2013; Pavoine and Ricotta 2014). Among them, the primary requirements that are generally accepted as necessary for a meaningful (functional) plot-to-plot dissimilarity index D in the range 0–1 are related to its extreme values: (1) For two identical assemblages, D takes the value zero, denoting maximum similarity, and (2) D takes the value one, denoting maximum dissimilarity, only for two completely distinct assemblages. This latter criterion is satisfied for assemblages with no species in common and with zero functional similarities between the species in the first assemblage and those in the second assemblage (see, e.g., Pavoine and Ricotta 2014).

While both requirements are straightforward, they have a rich corollary of implications that may help in generalizing traditional presence/absence dissimilarity measures among plots to include functional dissimilarities among species. For instance, requirement (2) implies that the dissimilarity component of the generalized index depends on how the dissimilarities between the species in the first assemblage and those in the second assemblage are distributed. Accordingly, in the calculation of a generalized index D of functional dissimilarity between two plots X and Y , the traditional mismatching components b and c of the 2×2 contingency table can be expressed as follows to include functional differences among species:

$$B = \sum_{i \in X, i \notin Y} \min_{j \in Y} \{d_{ij}\} \quad (1)$$

and

$$C = \sum_{j \in Y, j \notin X} \min_{i \in X} \{d_{ij}\} \quad (2)$$

where B and C are the generalized counterparts of the traditional mismatching components b and c and the summations in equations (1) and (2) are taken over all species that are present only in plot X or Y , respectively (i.e., the species contributing to b and c of the contingency table). The components B and C thus represent the functional uniqueness of plot X compared with plot Y and vice versa.

Among the possible options for calculating how different a species of one assemblage is from the species of the other assemblage, the minimum functional dissimilarity

(see, e.g., Izsák and Price 2001) has been chosen. Accordingly, for the calculation of the component B we used the minimum functional dissimilarity $\min_{j \in Y} \{d_{ij}\}$ between species i found only in plot X and all species in plot Y . Likewise, for the calculation of the component C we used the minimum functional dissimilarity $\min_{i \in X} \{d_{ij}\}$ between species j found only in plot Y and all species in plot X . This choice is in agreement with the core of requirement (1) that a species in plot X should be compared to its functional nearest neighbor in plot Y such that for two identical assemblages the index takes the value zero. Taking the minimum functional dissimilarity also ensures that, given two plots with no species in common but for which each species in plot X has a functionally similar species in plot Y , the dissimilarity between X and Y is close to zero. Also, if a species that appears only in plot X is functionally identical to a species shared by X and Y , then its contribution to B is zero.

The quantity $\min_{j \in Y} \{d_{ij}\}$ can be interpreted as the (functional) fraction of species i that is not shared by the species in plot Y , such that $0 \leq B \leq b$ and $0 \leq C \leq c$. According to equation 1, $B = 0$ if all species that are present only in plot X are functionally identical to at least one species in Y . At the other extreme, $B = b$ if all species that are present only in plot X are maximally dissimilar from every species in Y .

Note that, due to the relationship $s_{ij} = 1 - d_{ij}$, the generalized components B and C can also be expressed as: $B = \sum_{i \in X, i \notin Y} \min_{j \in Y} \{1 - s_{ij}\} = b - \sum_{i \in X, i \notin Y} \max_{j \in Y} \{s_{ij}\}$ and $C = c - \sum_{j \in Y, j \notin X} \max_{i \in X} \{s_{ij}\}$, where $\max_{j \in Y} \{s_{ij}\}$ is the maximum functional similarity between species i found only in plot X and all species in plot Y .

Based on the above definitions, virtually all presence/absence dissimilarity coefficients expressed as a normalized ratio $D = \text{Operational Part}/\text{Scaling Factor}$ can be generalized to include functional differences among species by substituting the traditional mismatching components b and c with their generalized counterparts B and C in the operational part of the index; for example, the generalized expression of the Jaccard dissimilarity takes the following form:

$$J = \frac{B + C}{a + b + c} \quad (3)$$

while the generalized expression of the Sørensen index becomes:

$$S = \frac{B + C}{2a + b + c} \quad (4)$$

The generalized Jaccard dissimilarity can be interpreted as the functional fraction of species in X and Y that is not shared by the two plots being compared. Note that in

equations (3) and (4) the scaling factors (denominators) are the same as in the traditional presence/absence measures. In this way, the generalized dissimilarities conform to requirement (2) that the indices take their maximum value only for two completely distinct assemblages for which $B = b$ and $C = c$. In this case, equations (3) and (4) recover the traditional expressions of the Jaccard and the Sørensen dissimilarity, respectively.

The differences $b - B = \sum_{i \in X, i \notin Y} \max_{j \in Y} \{s_{ij}\}$ and $c - C = \sum_{j \in Y, j \notin X} \max_{i \in X} \{s_{ij}\}$ can be interpreted as the functional fraction of species present only in plot X that is shared by the species in plot Y , and vice versa. Therefore, both terms increase the similarity between X and Y , such that the generalized counterpart of the traditional matching component a of the 2×2 contingency becomes

$$A = a + (b - B) + (c - C) \quad (5)$$

which ensures that $A \geq a$ and $A + B + C = a + b + c$, meaning that the generalized matching/mismatching components A , B , and C can be expressed in terms of functional species, or functional richness sensu Villéger et al. (2008). According to equation (5), the similarity complement of the Jaccard index can be thus expressed as $1 - J = A/(a + b + c)$.

Worked Example

The potential of the proposed approach for highlighting the relationships between community composition and ecosystem functioning was examined by comparing the generalized Jaccard dissimilarity coefficient (Eq. 3) with its traditional presence/absence counterpart in virtual plant communities along a simulated ecological gradient. The artificial data of Ricotta and Pavoine (2015) for 15 species (S1–S15) and 9 plots (P1–P9) were converted to presence/absence scores (Table 1; see also Appendix S1). The corresponding matrix of pairwise functional dissimilarities d_{ij} between all species pairs is given in Appendix S2. The original species \times plots matrix was generated with unimodal response of all species to a one-dimensional gradient with varying amplitude (length) and intensity (abundance), while the dissimilarity matrix was constructed such that the interspecies dissimilarities reflect the species ecological differences along the simulated gradient in Table 1. For details, see Ricotta and Pavoine (2015).

To highlight the behavior of both indices (traditional vs. generalized Jaccard index), we used profile diagrams: For each index, we calculated the dissimilarity of plot $P1$ with itself and with the remaining plots. This operation provides nine dissimilarity values whose graphical illustration shows the effect of changes in community composition along the simulated gradient.

Table 1. Artificial data matrix composed of 15 species (S1–S15) and 9 plots (P1–P9) for the graphical comparison of the indices used in the worked example. The data are the same as in Ricotta and Pavoine (2015) converted to presence/absence scores. Species presences are highlighted in gray.

Species	Plots								
	P1	P2	P3	P4	P5	P6	P7	P8	P9
S1	1	1	0	0	0	0	0	0	0
S2	1	1	0	0	0	0	0	0	0
S3	1	1	1	0	0	0	0	0	0
S4	1	1	1	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	1	1	1	1	1	1	0	0	0
S8	0	1	1	1	1	1	0	0	0
S9	1	1	1	1	1	1	1	0	0
S10	0	0	0	1	1	1	0	0	0
S11	0	0	0	0	0	1	0	0	0
S12	0	0	0	0	0	0	1	1	1
S13	0	0	0	0	1	1	1	1	1
S14	0	0	0	0	0	0	1	1	1
S15	0	0	0	0	0	0	0	1	1

In addition, it has been recently argued that the overall dissimilarity between two species assemblages is actually driven by two different processes, as species assemblages can differ in richness (i.e., one assemblage has more species than the other) and composition (i.e., some species are replaced by others; Baselga 2010). Therefore, we used the additive decomposition proposed by Podani and Schmera (2011) for partitioning the Jaccard dissimilarity coefficient into species replacement (or turnover J_{Repl}) and richness difference (J_{Rich}) such that $J = J_{\text{Repl}} + J_{\text{Rich}}$. The turnover component J_{Repl} summarizes how many species in one plot are replaced by a different species in the other plot, normalized by the total species richness of both plots ($a + b + c$). As one replacement always involves two species (Carvalho et al. 2013), the total number of replaced species is $2 \min\{b, c\}$. Hence, $J_{\text{Repl}} = 2 \min\{b, c\} / (a + b + c)$. By contrast, J_{Rich} summarizes the difference in species richness between both plots, normalized in the same way, such that $J_{\text{Rich}} = |b - c| / (a + b + c)$. Therefore, in this study we also compared the additive components of the classical Jaccard dissimilarity for presence and absence data with their functional analogues $J_{\text{Repl}} = 2 \min\{B, C\} / (a + b + c)$ and $J_{\text{Rich}} = |B - C| / (a + b + c)$. In this case, J_{Repl} corresponds to the amount of functional richness unique to site X that is replaced by the functional richness at site Y , or functional turnover. Likewise, the generalized version of J_{Rich} represents the difference in functional richness between plots X and Y , or, in other words, functional excess.

All calculations were performed with the new R scripts available in Appendix S3 (see Appendix S4 for a guide through the R scripts).

The entire dissimilarity matrices among plots were also subjected to principal coordinates analysis (PCoA) to obtain a compositional ordination with the traditional Jaccard index and a functional ordination with the generalized Jaccard index. Comparison of the two results may give insight into the mechanisms by which a traditional ordination based on compositional dissimilarities only is changed if mismatches are represented by functional dissimilarities between species. The ordinations were calculated by the SYN-TAX 2000 package (Podani 2001).

Results

Figure 1 shows the graphical comparison between the compositional (presence/absence) and the functional versions of the Jaccard dissimilarity, together with their additive components, species replacement/functional turnover, and richness difference/functional excess.

In agreement with requirement (1), for all indices, the comparison of plot $P1$ with itself results in zero dissimilarity. The graphical profiles of the compositional and functional versions of J and J_{Repl} both show an increasing pattern, which reflects the taxonomic and functional turnover along the simulated gradient of Table 1. By contrast, the graphical profiles of both versions of J_{Rich} show a more irregular pattern reflecting differences in species richness and functional excess between plot $P1$ and all other plots.

The two ordinations of the nine sites are shown in Figure 2. As expected, the ordination based on presence/absence scores exhibits a conspicuous horseshoe or arch effect, reflecting the simulated unimodal response of the species to the underlying gradient. The eigenvalues of the first three ordination axes were 44.5, 24.2, and 9.9, respectively. In the functional ordination, the arch is less pronounced, meaning that the replacement of species presences and absences by the functional species dissimilarities had a detrending effect on the gradient. This is shown by the greatly increased first eigenvalue (72.5) and the lower subsequent eigenvalues (14.8 and 4.4 for the second and the third axes, respectively).

Discussion

In this study, we proposed a new family of functional dissimilarity measures based on the generalization of the matching/mismatching components a , b , and c of the traditional contingency table to include functional dissimilarities among species. Unlike most indices used to date (e.g., Rao 1982; Izsák and Price 2001; Chiu and Chao

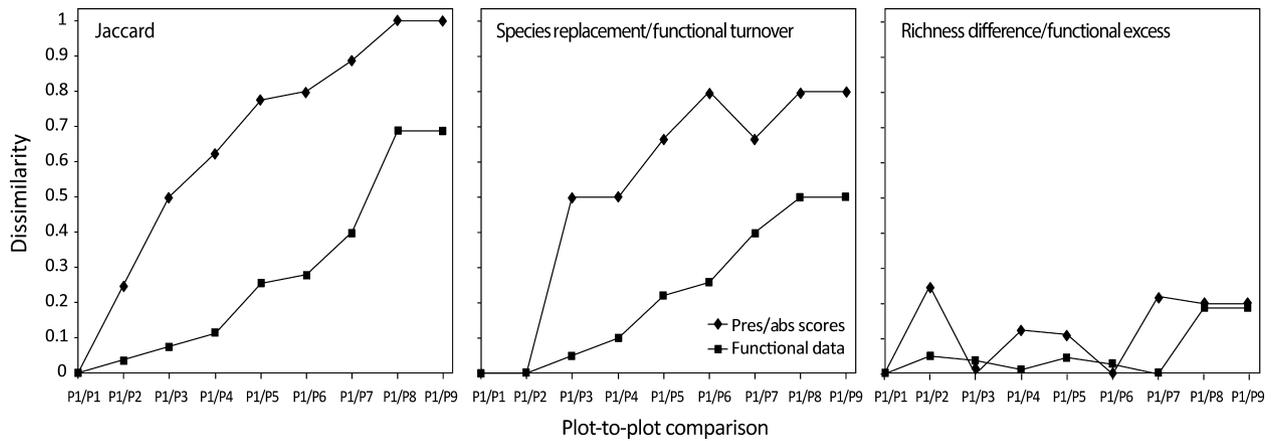


Figure 1. Graphical profiles showing the response to the simulated ecological gradient (Table 1) of the compositional (presence/absence scores only) Jaccard dissimilarity index, its functional generalization and their additive components, species replacement/functional turnover, and richness difference/functional excess. Plot *P1* is compared to itself and to all other plots in Table 1.

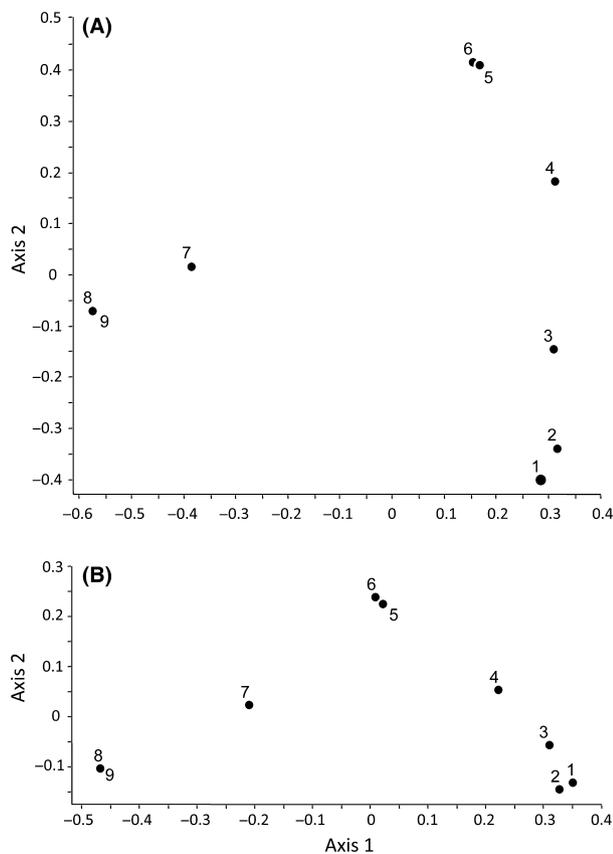


Figure 2. Principal coordinates ordinations of the artificial presence/absence scores in Table 1 based on (A) the traditional Jaccard index and (B) the generalized Jaccard index. Note the obvious detrending effect of incorporating functional information in the analysis.

2014; Pavoine and Ricotta 2014), which are expressed in terms of average or minimum functional distances between pairs of plots (or as a derived function of these

distances), the operational part and the scaling factor of the proposed family of measures are both expressed in terms of species numbers.

This approach is flexible enough to enable practitioners to generalize a large number of existing presence/absence dissimilarity measures or to “construct” new measures depending on the context of their analyses. In addition, relaxing requirement (2) that the index should take the value one *only* for maximally distinct assemblages, it will be possible to build an even larger family of indices. Here, it is worth noting that while the values of the functional mismatching components B and C are by definition not higher than the values of their traditional presence/absence analogues b and c , this does not necessarily hold for their “combinations”; for example, the absolute difference $|B - C|$ can be larger than $|b - c|$. Accordingly, as shown in Figure 1, the value of the functional version of J_{Rich} can be larger than its presence/absence counterpart, meaning that in this case the differences in functional richness between both plots are larger than the corresponding differences in taxonomic richness. This complementarity of viewpoints brought by the compositional and the functional approaches is also important for ordination studies, because, as shown by the worked example, extreme horseshoe effects may be reduced when the analysis shifts from pure compositional dissimilarities to functional dissimilarities.

Apart from a few exceptions (e.g., the indices proposed by Izsák and Price 2001), most of the functional dissimilarity indices published so far were developed for abundance data (e.g., Champely and Chessel 2002; Schmera et al. 2009; Ricotta and Bacaro 2010; Baiser and Lockwood 2011; Chiu and Chao 2014; Pavoine and Ricotta 2014; Sonnier et al. 2014). Among them, the indices dealing solely with relative abundances can only be adapted

to cope with presence and absence data by assuming that all N species present in a given plot have an abundance equal to $1/N$. However, in some cases, the incorporation of these abundances in the dissimilarity functions contrasts with the idea that the presence/absence scores (usually 0 and 1) represent a binary variable for which 1 stands for presence and 0 stands for absence without any reference to species abundances.

On the contrary, the indices that can deal with absolute abundances (e.g., Chiu and Chao 2014; Pavoine and Ricotta 2014) can be directly used with presence/absence scores by assuming that all species have unit abundance when present in a plot. But, as far as we are aware, none of them were expressed in terms of species richness while respecting requirements (1) and (2); for example, the indices proposed by Chiu and Chao (2014) can be expressed in terms of species richness (using functional Hill numbers) but have a different definition of two maximally distinct plots, whereas the indices of Pavoine and Ricotta (2014) satisfy requirements (1) and (2) but are expressed in terms of average functional dissimilarity among species. All these measures thus complete each other into a multifaceted approach to the calculation of plot-to-plot dissimilarity.

Finally, although we proposed this new family of dissimilarity measures in a functional context, the proposed approach is aimed at summarizing the dissimilarity between pairs of plots based on any between-species dissimilarity measure of choice. Therefore, the same approach can be extended to any other ecologically meaningful measure of dissimilarity among species, such as phylogenetic dissimilarities rescaled to the range [0–1]. This is a very desirable property of our approach as it allows to summarize relevant aspects of plot-to-plot dissimilarity from different, equally relevant, standpoints.

Conflict of Interest

None declared.

References

- Anderson, M. J., K. E. Ellingsen, and B. H. McArdle. 2006. Multivariate dispersion as a measure of beta diversity. *Ecol. Lett.* 9:683–693.
- Baiser, B., and J. L. Lockwood. 2011. The relationship between functional and taxonomic homogenization. *Glob. Ecol. Biogeogr.* 20:134–144.
- Baselga, A. 2010. Partitioning the turnover and nestedness components of beta diversity. *Glob. Ecol. Biogeogr.* 19:134–143.
- Carvalho, J. C., P. Cardoso, P. A. V. Borges, D. Schmera, and J. Podani. 2013. Measuring fractions of beta diversity and their relationships to nestedness: a theoretical and empirical comparison of novel approaches. *Oikos* 122:825–834.
- Champely, S., and D. Chessel. 2002. Measuring biological diversity using Euclidean metrics. *Environ. Ecol. Stat.* 9:167–177.
- Chiu, C.-H., and A. Chao. 2014. Distance-based functional diversity measures and their decomposition: a framework based on Hill numbers. *PLoS One* 9:e100014.
- Chiu, C.-H., L. Jost, and A. Chao. 2014. Phylogenetic beta diversity, similarity, and differentiation measures based on Hill numbers. *Ecol. Monogr.* 84:21–44.
- Clarke, K. R., P. J. Somerfield, and M. Chapman. 2006. On resemblance measures for ecological studies, including taxonomic dissimilarities and a zero-adjusted Bray-Curtis coefficient for denuded assemblages. *J. Exp. Mar. Biol. Ecol.* 330:55–80.
- Izsák, C., and R. G. Price. 2001. Measuring β -diversity using a taxonomic similarity index, and its relation to spatial scale. *Mar. Ecol. Prog. Ser.* 215:69–77.
- 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.
- Legendre, P. 2014. Interpreting the replacement and richness difference components of beta diversity. *Glob. Ecol. Biogeogr.* 11:1324–1334.
- Legendre, P., and M. De Cáceres. 2013. Beta diversity as the variance of community data: dissimilarity coefficients and partitioning. *Ecol. Lett.* 16:951–963.
- Legendre, P., and L. Legendre. 2012. *Numerical ecology*. Elsevier, Amsterdam.
- Mason, N. W. H., and F. de Bello. 2013. Functional diversity: a tool for answering challenging ecological questions. *J. Veg. Sci.* 24:777–780.
- Pavoine, S., and C. Ricotta. 2014. Functional and phylogenetic similarity among communities. *Methods Ecol. Evol.* 5:666–675.
- Pavoine, S., A.-B. Dufour, and D. Chessel. 2004. From dissimilarities among species to dissimilarities among communities: a double principal coordinate analysis. *J. Theor. Biol.* 228:523–537.
- Podani, J. 2000. Introduction to the exploration of multivariate biological data. Backhuys, Leiden. Freely available at: <http://ramet.elte.hu/~podani/books.html>.
- Podani, J. 2001. SYN-TAX 2000. Computer programs for data analysis in ecology and systematics. Scientia, Budapest. Freely available at: <http://ramet.elte.hu/~podani/>.
- Podani, J., and D. Schmera. 2011. A new conceptual and methodological framework for exploring and explaining pattern in presence-absence data. *Oikos* 120:1625–1638.
- Rao, C. R. 1982. Diversity and dissimilarity coefficients: a unified approach. *Theor. Popul. Biol.* 21:24–43.
- Ricotta, C., and G. Bacaro. 2010. On plot-to-plot dissimilarity measures based on species functional traits. *Community Ecol.* 11:113–119.

- Ricotta, C., and S. Pavoine. 2015. Measuring similarity among plots including similarity among species: an extension of traditional approaches. *J. Veg. Sci.* 26:1061–1067.
- Ricotta, C., G. Bacaro, and S. Pavoine. 2015. A cautionary note on some phylogenetic dissimilarity measures. *J. Plant Ecol.* 8:12–16.
- Schmera, D., T. Erős, and J. Podani. 2009. A measure for assessing functional diversity in ecological communities. *Aquat. Ecol.* 43:157–167.
- Sonnier, G., S. E. Johnson, K. L. Amatangelo, D. A. Rogers, and D. M. Waller. 2014. Is taxonomic homogenization linked to functional homogenization in temperate forests? *Glob. Ecol. Biogeogr.* 23:894–902.
- Sørensen, T. 1948. A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application in analysis of the vegetation on Danish commons. *Biologiske Skrifter* 5:1–34.
- Villéger, S., N. W. H. Mason, and D. Mouillot. 2008. New multidimensional functional diversity indices for a multifaceted framework in functional ecology. *Ecology* 89:2290–2301.
- Violle, C., M.-L. Navas, D. Vile, E. Kazakou, C. Fortunel, I. Hummel, et al. 2007. Let the concept of trait be functional! *Oikos* 116:882–892.

Supporting Information

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

Appendix S1. Input data used for the artificial example.

Appendix S2. Interspecies dissimilarity matrix used for the calculation of the generalized functional dissimilarity indices of the artificial example.

Appendix S3. R scripts for the calculation of the dissimilarity indices.

Appendix S4. Manual associated with the R scripts.

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

	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

```

PADDis <- function (comm, dis, method = NULL, diag = FALSE, upper =
FALSE)
{
  METHODS <- c("components", "Jaccard", "Sorensen", "SokalSneath",
"Ochiai", "Simpson")
  if (!(inherits(comm, "data.frame") | inherits(comm, "matrix")))
    stop("comm is not a data.frame or a matrix")
  if (!inherits(dis, "dist"))
    stop("dis is not an object of class 'dist'")
  df <- as.matrix(comm)
  dis <- as.matrix(dis)
  if(nrow(dis)!=ncol(df))
    stop("the size of dis is not equal to the number of columns in
comm")
  if(is.null(colnames(df)) | is.null(rownames(dis)))
    warning("Species (attributes) in dis was been assumed to be in
the same order as species (columns) in comm")
  else
    {if (!all(c(colnames(df)%in%rownames(dis),
rownames(dis)%in%colnames(df))))
      warning("Species (attributes) in dis was been assumed to be in
the same order as species (columns) in comm")
    else
      df <- df[, rownames(dis)]}
  if (!is.numeric(df))
    stop("df must contain numeric values")
  if (any(df < 0))
    stop("non negative value expected in df")
  nlig <- nrow(df)
  nsp <- ncol(df)
  d.names <- row.names(df)
  if (is.null(d.names))
    d.names <- 1:nlig
  df <- as.matrix(1 * (df > 0))
  if (is.null(method)) {
    cat("0 = list of all components (a, b, c, A, B, C)\n")
    cat("1 = Jaccard index (1901)\n")
    cat("d1 = (B+C)/(a+b+c)\n")
    cat("2 = Czekanowski (1913) or Sorensen (1948)\n")
    cat("d2 = (B+C)/(2*a+b+c)\n")
    cat("3 = Sokal & Sneath (1963)\n")
    cat("d3 = 2(B+C)/(a+2(b+c))\n")
    cat("4 = Ochiai (1957)\n")
    cat("d4 = [sqrt((A+B)(A+C))-A]/sqrt((a+b)(a+c))\n")
    cat("5 = Simpson (1943) \n")
    cat("d5 = min(B,C)/(a+min(b,c))\n")
    cat("Select an integer (0-5): ")
    method <- as.integer(readLines(n = 1))
  }
  a <- df %*% t(df)
  b <- df %*% (1 - t(df))
  c <- (1 - df) %*% t(df)
  B <- C <- matrix(0, nlig, nlig)
  for (i in 2:nlig){
    for(j in 1:(i-1)){
      compb <- (1:nsp)[df[i,]>1e-8 & df[j, ]<1e-8]
      compY <- (1:nsp)[df[j, ]>1e-8]
      bY <- dis[compb, compY]
      if(is.null(compb)) B[i, j] <- 0
      else {if(length(compb)==1) B[i, j] <- min(bY)
      else B[i, j] <- sum(apply(bY, 1, min)) }
    }
  }
}

```

```

        compc <- (1:nsp)[df[j,]>1e-8 & df[i, ]<1e-8]
        compX <- (1:nsp)[df[i, ]>1e-8]
        cX <- dis[compc, compX]
        if(is.null(compc)) C[i, j] <- 0
        else {if(length(compc)==1) C[i, j] <- min(cX)
              else C[i, j] <- sum(apply(cX, 1, min)) }
    }
  }
  Bf <- B + t(C)
  Cf <- C + t(B)
  B <- Bf
  C <- Cf
  colnames(B) <- colnames(C) <- rownames(B) <- rownames(C) <-
rownames(df)
  A <- a + (b - B) + (c - C)
  if (method == 0) {
    return(list(a = a, b = b, c = c, A = A, B = B, C = C))
  }
  if (method == 1) {
    d <- (B+C)/(a + b + c)
  }
  else if (method == 2) {
    d <- (B+C)/(2 * a + b + c)
  }
  else if (method == 3) {
    d <- (2*(B+C))/(a + 2 * (b + c))
  }
  else if (method == 4) {
    d <- (sqrt(A+B)*sqrt(A+C)-A)/sqrt((a + b) * (a + c))
  }
  else if (method == 5) {
    minBC <- (B<=C)*B+(B>C)*C
    minbc <- (b<=c)*b+(b>c)*c
    d <- minBC / (a + minbc)
  }
  else stop("Non convenient method")
  d <- as.dist(d)
  attr(d, "Size") <- nlig
  attr(d, "Labels") <- d.names
  attr(d, "Diag") <- diag
  attr(d, "Upper") <- upper
  attr(d, "method") <- METHODS[method+1]
  attr(d, "call") <- match.call()
  class(d) <- "dist"
  return(d)
}

DJac <- function (comm, dis, diag = FALSE, upper = FALSE)
{
  if (!(inherits(comm, "data.frame") | inherits(comm, "matrix")))
    stop("comm is not a data.frame or a matrix")
  if (!inherits(dis, "dist"))
    stop("dis is not an object of class 'dist'")
  df <- as.matrix(comm)
  dis <- as.matrix(dis)
  if(nrow(dis)!=ncol(df))
    stop("the size of dis is not equal to the number of columns in
comm")
  if(is.null(colnames(df)) | is.null(rownames(dis)))

```

```

warning("Species (attributes) in dis was been assumed to be in
the same order as species (columns) in comm")
else { if (!all(c(colnames(df)%in%rownames(dis),
rownames(dis)%in%colnames(df))))
warning("Species (attributes) in dis was been assumed to be in
the same order as species (columns) in comm")
else
df <- df[, rownames(dis)]}
if (!is.numeric(df))
stop("df must contain numeric values")
if (any(df < 0))
stop("non negative value expected in df")
nlig <- nrow(df)
nsp <- ncol(df)
d.names <- row.names(df)
if (is.null(d.names))
d.names <- 1:nlig
df <- as.matrix(1 * (df > 0))
a <- df %*% t(df)
b <- df %*% (1 - t(df))
c <- (1 - df) %*% t(df)
B <- C <- matrix(0, nlig, nlig)
for(i in 2:nlig){
for(j in 1:(i-1)){
compb <- (1:nsp)[df[i,]>1e-8 & df[j, ]<1e-8]
compY <- (1:nsp)[df[j, ]>1e-8]
bY <- dis[compb, compY]
if(is.null(compb)) B[i, j] <- 0
else {if(length(compb)==1) B[i, j] <- min(bY)
else B[i, j] <- sum(apply(bY, 1, min)) }
compc <- (1:nsp)[df[j,]>1e-8 & df[i, ]<1e-8]
compX <- (1:nsp)[df[i, ]>1e-8]
cX <- dis[compc, compX]
if(is.null(compc)) C[i, j] <- 0
else {if(length(compc)==1) C[i, j] <- min(cX)
else C[i, j] <- sum(apply(cX, 1, min)) }
}
}
Bf <- B + t(C)
Cf <- C + t(B)
B <- Bf
C <- Cf
A <- a + (b - B) + (c - C)
j <- (B+C)/(a + b + c)
jrich <- abs(B-C)/(a+b+c)
jrepl <- j-jrich
J <- as.dist(j)
JRich <- as.dist(jrich)
JRepl <- as.dist(jrepl)
attr(J, "Size") <- attr(JRich, "Size") <- attr(JRepl, "Size") <- nlig
attr(J, "Labels") <- attr(JRich, "Labels") <- attr(JRepl, "Labels")
<- d.names
attr(J, "Diag") <- attr(JRich, "Diag") <- attr(JRepl, "Diag") <- diag
attr(J, "Upper") <- attr(JRich, "Upper") <- attr(JRepl, "Upper") <-
upper
attr(J, "method") <- "Jaccard"
attr(JRich, "method") <- "Component of richness difference"
attr(JRepl, "method") <- "Component of replacement or turnover"
class(J) <- "dist"
class(JRepl) <- "dist"

```

```

class(JRich) <- "dist"
return(list(J=J, JRepl = JRepl, JRich = JRich))
}

Jac <- function (comm, diag = FALSE, upper = FALSE)
{
  if (!(inherits(comm, "data.frame") | inherits(comm, "matrix")))
    stop("comm is not a data.frame or a matrix")
  df <- as.matrix(comm)
  if (!is.numeric(df))
    stop("df must contain numeric values")
  if (any(df < 0))
    stop("non negative value expected in df")
  nlig <- nrow(df)
  d.names <- row.names(df)
  if (is.null(d.names))
    d.names <- 1:nlig
  df <- as.matrix(1 * (df > 0))
  a <- df %*% t(df)
  b <- df %*% (1 - t(df))
  c <- (1 - df) %*% t(df)
  j <- (b+c)/(a + b + c)
  jrich <- abs(b-c)/(a+b+c)
  jrepl <- j-jrich
  J <- as.dist(j)
  JRich <- as.dist(jrich)
  JRepl <- as.dist(jrepl)
  attr(J, "Size") <- attr(JRich, "Size") <- attr(JRepl, "Size") <- nlig
  attr(J, "Labels") <- attr(JRich, "Labels") <- attr(JRepl, "Labels")
  <- d.names
  attr(J, "Diag") <- attr(JRich, "Diag") <- attr(JRepl, "Diag") <- diag
  attr(J, "Upper") <- attr(JRich, "Upper") <- attr(JRepl, "Upper") <-
upper
  attr(J, "method") <- "Jaccard"
  attr(JRich, "method") <- "Component of richness difference"
  attr(JRepl, "method") <- "Component of replacement or turnover"
  class(J) <- "dist"
  class(JRepl) <- "dist"
  class(JRich) <- "dist"
  return(list(J=J, JRepl = JRepl, JRich = JRich))
}

```

Appendix 4. Manual associated with the R scripts

R functions "PADDis", "DJac" and "Jac" for calculating the dissimilarity coefficients introduced in the main text.

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 N plots \times S species' incidence (0/1) values, and an object of class 'dist' containing the (functional) dissimilarities among species, the function PADDis calculates the matching/mismatching components a , b , c , A , B , C , for each pair of plots. It can also calculate the coefficients J and S developed in the main text + other coefficients from the same family, see below.

Dependencies: none. However, the scripts were developed by modifying the function 'dist.binary' of the package ade4 (Dray & Dufour 2007).

Usage: PADDis (comm, dis, method = NULL, diag = FALSE, upper = FALSE)

DJac (comm, dis, diag = FALSE, upper = FALSE)

Jac (comm, diag = FALSE, upper = FALSE)

Arguments

<i>comm</i>	A matrix of N plots \times S species containing the incidence (0/1) of all species in the N plots. Columns are species and plots are rows
<i>dis</i>	An object of class 'dist' containing the (functional) dissimilarities among species
<i>method</i>	An integer between 0 and 5. If NULL the choice is made with a console message. See details.
<i>diag</i>	a logical value indicating whether the diagonal of the distance matrix should be printed by 'print.dist'
<i>upper</i>	a logical value indicating whether the upper triangle of the distance matrix should be printed by 'print.dist'

Details: If method=0, then the function PADDis returns 6 matrices corresponding to the a, b, c, A, B, and C values per pair of plots. Otherwise, it returns an object of class 'dist' corresponding to the dissimilarities among plots calculated with the following formulas:

$$\frac{B+C}{a+b+c} \# \text{generalized Jaccard dissimilarity, with method} = 1$$

$$\frac{B+C}{2a+b+c} \# \text{generalized Sørensen dissimilarity, with method} = 2$$

$$\frac{2(B+C)}{a+2(b+c)} \# \text{generalized Sokal \& Sneath dissimilarity, with method} = 3$$

$$\frac{\sqrt{A+B}\sqrt{A+C} - A}{\sqrt{a+b}\sqrt{a+c}} \# \text{generalized Ochiai dissimilarity, with method} = 4$$

$$\frac{\min\{B, C\}}{a + \min\{b, c\}} \# \text{generalized Simpson dissimilarity, with method} = 5$$

DJac and Jac use the additive decomposition of the Jaccard index. DJac takes into account the (functional) dissimilarities among species while Jac doesn't. Formulas used by DJac thus are:

$$J = \frac{B+C}{a+b+c}$$

$$J_{Repl} = \frac{2 \min\{B, C\}}{a+b+c}$$

$$J_{Rich} = \frac{|B-C|}{a+b+c}$$

and those used by Jac are:

$$J = \frac{b+c}{a+b+c}$$

$$J_{Repl} = \frac{2 \min\{b, c\}}{a+b+c}$$

$$J_{Rich} = \frac{|b-c|}{a+b+c}$$

Example

Load the data sets contained in Appendices 1 and 2 and name them Com and Dis, respectively. For that one can use the following instructions:

```
Com <- read.table(file.choose(), sep="\t", row.names=1, h=TRUE)
# here select Appendix 1.
```

```
Dis <- read.table(file.choose(), sep="\t", row.names=1, h=TRUE)
# here select Appendix 2.
```

```
Dis<- as.dist(Dis)
```

Load the functions contained in Appendix 3. For that, the following instructions can be used:

```
source(file.choose())
# here select Appendix 3
```

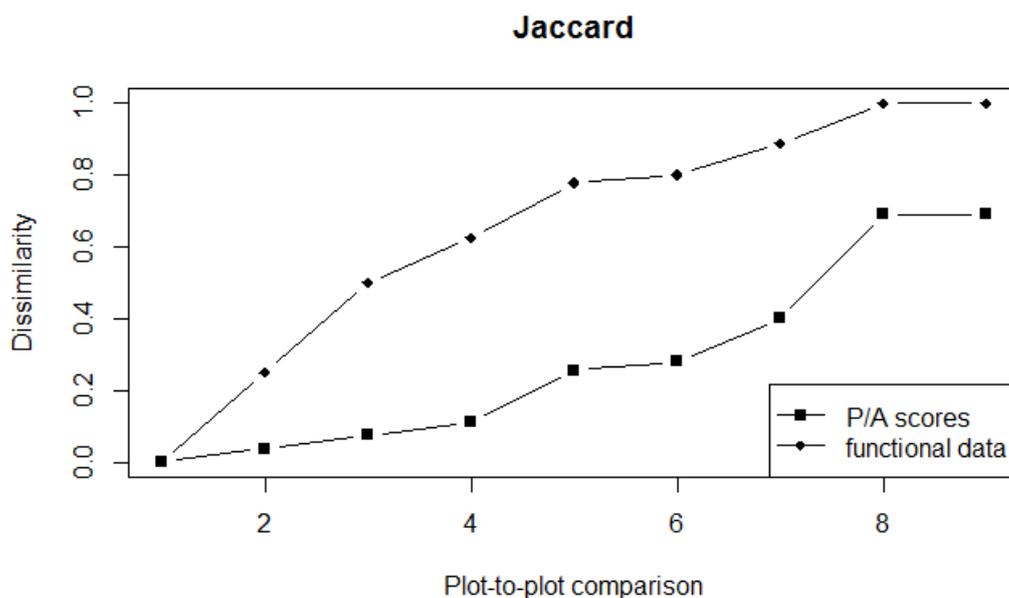
The following instructions were used to obtain Figure 1 in the main text:

```
J <- Jac(Com)
DJ <- DJac(Com, Dis)

plot(c(as.matrix(DJ$J)[1,]), ylab="Dissimilarity", xlab="Plot-to-plot
comparison", pch=15, type="b", ylim=c(0,1), main="Jaccard")

lines(c(as.matrix(J$J)[1,]), type="b", pch=18)

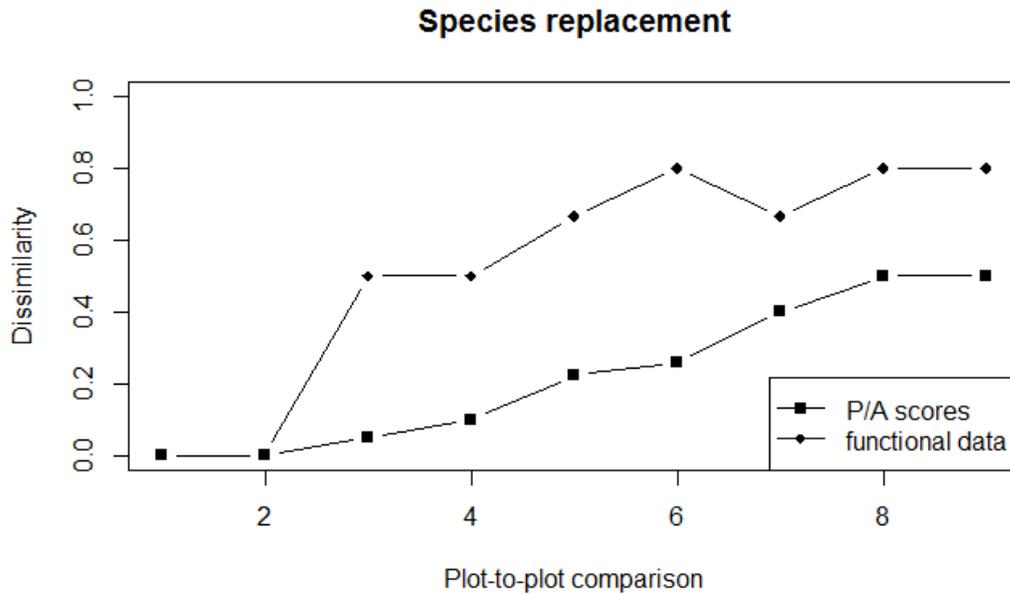
legend("bottomright", legend=c("P/A scores", "functional data"),
pch=c(15,18), lty=1)
```



```
plot(c(as.matrix(DJ$JRepl)[1,]), ylab="Dissimilarity", xlab="Plot-to-plot
comparison", pch=15, type="b", ylim=c(0,1), main="Species replacement")
```

```
lines(c(as.matrix(J$JRepl)[1,]), type="b", pch=18)
```

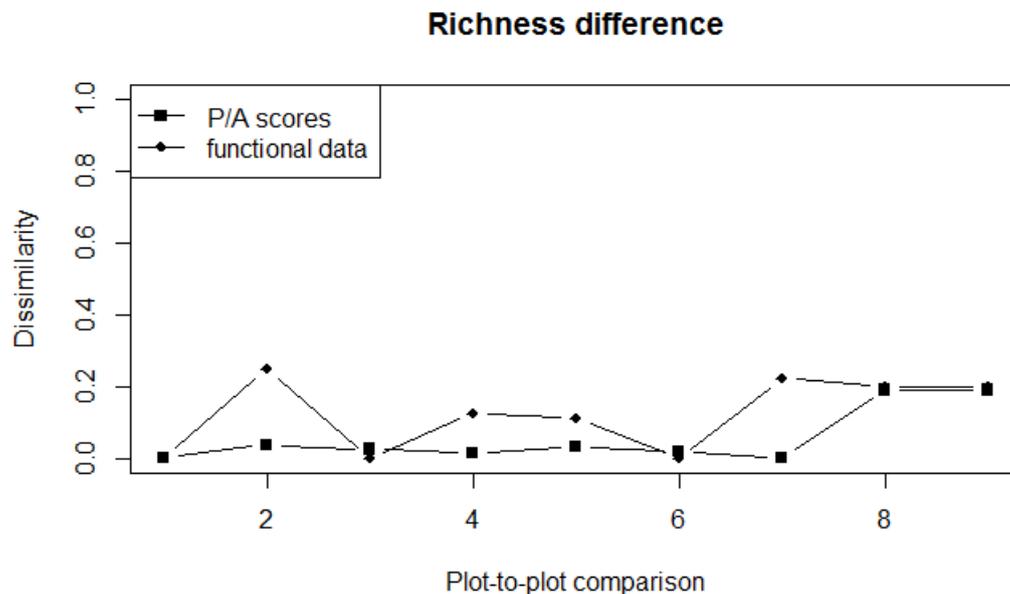
```
legend("bottomright", legend=c("P/A scores", "functional data"),
pch=c(15,18), lty=1)
```



```
plot(c(as.matrix(DJ$JRich)[1,]), ylab="Dissimilarity", xlab="Plot-to-plot
comparison", pch=15, type="b", ylim=c(0,1), main="Richness difference")
```

```
lines(c(as.matrix(J$JRich)[1,]), type="b", pch=18)
```

```
legend("topleft", legend=c("P/A scores", "functional data"), pch=c(15,18),
lty=1)
```



Use the following instruction to obtain all components:

PADDis(Com, Dis)

0 = list of all components (a, b, c, A, B, C)

1 = Jaccard index (1901)

d1 = $(B+C)/(a+b+c)$

2 = Czekanowski (1913) or Sorensen (1948)

d2 = $(B+C)/(2*a+b+c)$

3 = Sockal & Sneath (1963)

d3 = $2(B+C)/(a+2(b+c))$

4 = Ochiai (1957)

d4 = $[\sqrt{(A+B)(A+C)}-A]/\sqrt{(a+b)(a+c)}$

5 = Simpson (1943)

d5 = $\min(B,C)/(a+\min(b,c))$

Select an integer (0-5): 0

\$a

	P1	P2	P3	P4	P5	P6	P7	P8	P9
P1	6	6	4	3	2	2	1	0	0
P2	6	8	5	4	3	3	1	0	0
P3	4	5	6	4	3	3	1	0	0
P4	3	4	4	5	4	4	1	0	0
P5	2	3	3	4	5	5	2	1	1
P6	2	3	3	4	5	6	2	1	1
P7	1	1	1	1	2	2	4	3	3
P8	0	0	0	0	1	1	3	4	4
P9	0	0	0	0	1	1	3	4	4

\$b

	P1	P2	P3	P4	P5	P6	P7	P8	P9
P1	0	0	2	3	4	4	5	6	6
P2	2	0	3	4	5	5	7	8	8
P3	2	1	0	2	3	3	5	6	6
P4	2	1	1	0	1	1	4	5	5
P5	3	2	2	1	0	0	3	4	4
P6	4	3	3	2	1	0	4	5	5
P7	3	3	3	3	2	2	0	1	1
P8	4	4	4	4	3	3	1	0	0
P9	4	4	4	4	3	3	1	0	0

\$c

	P1	P2	P3	P4	P5	P6	P7	P8	P9
P1	0	2	2	2	3	4	3	4	4
P2	0	0	1	1	2	3	3	4	4
P3	2	3	0	1	2	3	3	4	4
P4	3	4	2	0	1	2	3	4	4
P5	4	5	3	1	0	1	2	3	3
P6	4	5	3	1	0	0	2	3	3
P7	5	7	5	4	3	4	0	1	1
P8	6	8	6	5	4	5	1	0	0
P9	6	8	6	5	4	5	1	0	0

\$A

	P1	P2	P3	P4	P5	P6	P7	P8	P9
P1	6.0	7.7	7.4	7.1	6.7	7.2	5.4	3.1	3.1
P2	7.7	8.0	8.6	7.9	7.5	8.1	7.0	4.2	4.2
P3	7.4	8.6	6.0	6.4	6.4	7.1	6.1	3.7	3.7
P4	7.1	7.9	6.4	5.0	5.2	6.0	5.8	4.2	4.2
P5	6.7	7.5	6.4	5.2	5.0	5.8	6.0	5.1	5.1
P6	7.2	8.1	7.1	6.0	5.8	6.0	6.7	5.8	5.8
P7	5.4	7.0	6.1	5.8	6.0	6.7	4.0	4.3	4.3

P8 3.1 4.2 3.7 4.2 5.1 5.8 4.3 4.0 4.0
P9 3.1 4.2 3.7 4.2 5.1 5.8 4.3 4.0 4.0

§B

	P1	P2	P3	P4	P5	P6	P7	P8	P9
P1	0.0	0.0	0.2	0.5	1.3	1.3	1.8	4.4	4.4
P2	0.3	0.0	0.3	0.9	1.7	1.7	2.4	5.6	5.6
P3	0.4	0.1	0.0	0.4	0.8	0.8	1.3	4.1	4.1
P4	0.4	0.2	0.2	0.0	0.2	0.2	0.8	3.0	3.0
P5	1.0	0.8	0.8	0.6	0.0	0.0	0.5	2.1	2.1
P6	1.5	1.2	1.1	0.8	0.2	0.0	0.8	2.4	2.4
P7	1.8	1.6	1.6	1.4	0.5	0.5	0.0	0.6	0.6
P8	2.5	2.2	2.2	1.8	0.8	0.8	0.1	0.0	0.0
P9	2.5	2.2	2.2	1.8	0.8	0.8	0.1	0.0	0.0

§C

	P1	P2	P3	P4	P5	P6	P7	P8	P9
P1	0.0	0.3	0.4	0.4	1.0	1.5	1.8	2.5	2.5
P2	0.0	0.0	0.1	0.2	0.8	1.2	1.6	2.2	2.2
P3	0.2	0.3	0.0	0.2	0.8	1.1	1.6	2.2	2.2
P4	0.5	0.9	0.4	0.0	0.6	0.8	1.4	1.8	1.8
P5	1.3	1.7	0.8	0.2	0.0	0.2	0.5	0.8	0.8
P6	1.3	1.7	0.8	0.2	0.0	0.0	0.5	0.8	0.8
P7	1.8	2.4	1.3	0.8	0.5	0.8	0.0	0.1	0.1
P8	4.4	5.6	4.1	3.0	2.1	2.4	0.6	0.0	0.0
P9	4.4	5.6	4.1	3.0	2.1	2.4	0.6	0.0	0.0

References

- Dray, S. & Dufour, A.B. (2007) The ade4 package: implementing the duality diagram for ecologists. *Journal of Statistical Software* 22: 1-20.
- 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.
- 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.
- Simpson, G.G. (1943) Mammals and the nature of continents. *American Journal of Science* 241: 1–31.
- Sokal, R.R. & Sneath, P.H.A. (1963) *Principles of Numerical Taxonomy*. W.H. Freeman, San Francisco.
- Sørensen, T. (1948) A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application in analysis of the vegetation on Danish commons. *Biologiske Skrifter* 5: 1–34.