Ecological Modelling 383 (2018) 41-51.

Exploring multiple presence-absence data structures in ecology

János Podani[1,2], Péter Ódor[3,8], Simone Fattorini[4,5], Giovanni Strona[6], Jani Heino[7] and Dénes Schmera[8,9]

1 Corresponding author. Department of Plant Systematics, Ecology and Theoretical Biology, Institute of Biology, Eötvös University, Pázmány P. s. 1.c, H-1117 Budapest, Hungary. Email: podani@ludens.elte.hu. ORCID: 0000-0002-1452-1486

2 MTA-ELTE-MTM Ecology Research Group, Eötvös University, Pázmány P. s. 1/C, H-1117, Budapest, Hungary

3 MTA Centre for Ecological Research, Institute of Ecology and Botany, Alkotmány u. 2-4, H-8237, Vácrátót, Hungary

4 Department of Life, Health and Environmental Sciences, University of L'Aquila, L'Aquila, Italy

5 CE3C – Centre for Ecology, Evolution and Environmental Changes/Azorean Biodiversity Group and University of Azores, Angra do Heroísmo, Portugal

6 European Commission, Joint Research Centre, Directorate D - Sustainable Resources – Bio-Economy Unit, Via Enrico Fermi 274 9, 21027 Ispra (VA), Italy

7 Finnish Environment Institute, Natural Environment Centre, Biodiversity, Paavo Havaksen Tie 3, FIN-90570 Oulu, Finland

8 MTA Centre for Ecological Research, GINOP Sustainable Ecosystem Group, Klebelsberg K. u. 3, H-8237 Tihany, Hungary

9 MTA Centre for Ecological Research, Balaton Limnological Institute, Klebelsberg K. u. 3, H-3237 Tihany, Hungary

Abstract

Ecological studies may produce presence-absence data sets for different taxonomic groups, with varying spatial resolution and temporal coverage. Comparison of these data is needed to extract meaningful information on the background ecological factors explaining community patterns, to improve our understanding of how beta diversity and its components vary among communities and biogeographical regions, and to reveal their possible implications for biodiversity conservation. A methodological difficulty is that the number of sampling units may be unequal: no method has been designed as yet to compare data matrices in such cases.

The problem is solved by converting presence-absence data matrices to simplex plots based on the decomposition of Jaccard dissimilarity into species replacement and richness difference fractions used together with the complementary similarity function. Pairs of simplex plots representing different data matrices are then compared by quantifying, for each of them, the relative frequency of points in small, pre-defined subregions of the simplex, and then calculating a divergence function between the two frequency distributions. Given more than two data matrices, classification and ordination techniques may be used to obtain a synthetic and informative picture of metacommunity structure.

We demonstrate the potential of our data analytical model by applying it to different case studies spanning different spatial scales and taxonomic levels (Mediterranean Island faunas; Finnish stream macroinvertebrate assemblages; Hungarian forest assemblages), and to a study of temporal changes in small islands (insect fauna in Florida). We conclude that, by accounting for various structural aspects simultaneously, the method permits a thorough ecological interpretation of presence-absence data. Furthermore, the examples illustrate succinctly how similarity, beta diversity and two of its additive components, species replacement and richness difference influence presence-absence patterns under different conditions.

Keywords: Beta diversity; Classification; Comparison; Ordination; Similarity; Simplex diagram

1. **Introduction**

Community data derived from field surveys have been routinely summarized in form of presence-absence matrices, with species (or other taxa) as rows and study objects (e.g., sites, plots, localities, etc.) as columns. A given study may produce two or more data matrices from the same region which differ from one another in taxonomic coverage, spatial resolution, the time of sampling, or any other ecologically meaningful factor. By definition, a meta-analysis attempting to summarize community level information from various and independent sources is also concerned with several data matrices. In all these cases, one is faced with the fundamental problem of comparing the *inherent structure of data matrices* with heterogeneous origin and properties. By structured presence-absence data we refer here to a matrix containing 0-s and 1-s which deviates from a random arrangement of scores by having, for example, any tendency of grouping, trends, or nestedness. These features are inherent, thanks to their independence from the actual ordering of rows and columns (see Podani & Schmera 2011). Such comparisons are essential to understand variations in beta diversity among communities and biogeographical regions, the ecological factors explaining these patterns, and their possible implications for biodiversity conservation. One possibility to tackle these issues is to perform classification or ordination on each data matrix and then to compare the resulting scatter plots, dendrograms or partitions. However, standard procedures available for this purpose can only be applied to cases in which the number of study objects is the same in all the data sets under evaluation (Podani 2000). The other way to proceed is to compare the data matrices directly, without multivariate analysis, but this methodology – in addition to equality in the number of objects – requires identical number of species as well. (Hubert & Golledge 1982; Zani 1986). That is, no universally applicable method has been developed as yet to compare the structure of data matrices that are unequal in size.

As a possible solution to this problem, we suggest a new analytical model that makes it possible to investigate multiple, heterogeneous datasets in a single framework. Essentially, the approach is based on the decomposition of Jaccard dissimilarities between pairs of objects into two additive components, namely species replacement (R) and richness difference (D), which, together with the complementary Jaccard similarity (S), are used to represent data structure as a point cloud in a ternary plot called SDR-simplex (Podani & Schmera 2011; Carvalho *et al*. 2012). Point clouds representing different data structures can then be compared on the basis of the relative frequencies of points (object pairs) in pre-defined subsections of the ternary plot. Since calculation of a frequency distribution is involved, we shall refer to this strategy as the *indirect* comparison of simplexes.

We suggest that the approach is equally useful to situations where the matrices to be compared represent the same set of objects (for example, when a given set of objects is surveyed for different organism groups, or when a taxonomic group is examined in the same sites several times to monitor temporal changes of community composition), even though in those cases comparisons of ordinations and classifications could also work. This is because decomposition of dissimilarity into additive terms allows separating the effect of major ecological driving forces – a possibility not available otherwise. Now, the simplex plots need not be partitioned; the shapes of point clouds can be *directly* compared by measuring the shift of the corresponding points in the two configurations.

Both indirect and direct comparisons may be performed on all possible pairs of matrices in a multiple dataset, yielding a dissimilarity matrix of SDR simplexes that can be then used in further analyses, such as classifications and ordinations. We emphasize here that this meta-analysis approach is more suited to exploratory analysis rather than to hypothesis-testing. In this paper, we describe in detail the technical aspects of our method, and illustrate its potential in ecological research, by reporting results for both artificial examples and empirical case studies.

2. Computational steps

2.1 The SDR-simplex

Jaccard's (1901) similarity coefficient is one of the oldest and most commonly used resemblance functions, computed for any two objects *j* and *k* as:

$$s_{jk} = a / (a+b+c), \qquad (1)$$

where a is the number of species occurring in both *j* and *k*, while b and c correspond to the number of species exclusive, respectively, to *j* and *k*. Its complement, Jaccard dissimilarity, is computed as:

$$\delta_{jk} = 1 - s_{jk} = (b+c) / (a+b+c). \qquad (2)$$

Dissimilarity can be partitioned into two additive fractions (Podani & Schmera 2011; Carvalho *et al*. 2012):

$$\delta_{jk} = d_{jk} + r_{jk} = | b–c | / (a+b+c) + 2\min\{b,c\} / (a+b+c), \qquad (3)$$

where $d_{jk} = | b–c | / (a+b+c)$ is the relative richness difference, while $r_{jk} = 2\min\{b,c\} / (a+b+c)$ is the relative species replacement with respect to objects *j* and *k*. In the latter, the numerator is the maximum fraction of the so-called species turnover which is equally shared by *j* and *k*. Since $s_{jk} + d_{jk} + r_{jk} = 1$, these three quantities may be used to define the relative position of the point representing object pair *jk* with respect to the three vertices (S–Similarity, D–richness Difference and R–species Replacement) of an equilateral triangle, the so-called SDR-simplex diagram (Podani & Schmera 2011). In the SDR-simplex, the distance of each point from a given vertex is inversely proportional to the corresponding fraction, that is, S, D or R. Similar ternary plots have been used in ecology as illustrations of C-S-R strategies of plants (Grime 1977), of feeding habits of fish (Fig. 6.9 in Stoffels 2013), and are even more widely used in population genetics (commonly referred to as "de Finetti diagram") to represent the genotype frequencies of diploid populations for a biallelic locus (Edwards 2000), and in geology to classify rocks and minerals on the basis of their fractional composition (Streckeisen 1976).

Let us first demonstrate the procedure for a pair of hypothetical objects *j* and *k* containing a total of 12 species with different nonzero values of a, b and c. If the objects have many species in common (a = 8), and species replacement and richness difference are equal ($2\min\{b,c\} =| b–c | = 2$), then the point representing this pair of objects in the ternary plot is positioned close to the S vertex, and with equal distance from D and R (Fig. 1a). If richness difference is high ($| b–c | = 9$) and similarity and replacement are the same ($a = 2\min\{b,c\} = 2$), then the point moves close to the D vertex (Fig. 1b). Analogously, if species replacement is the dominating phenomenon, with 4 species being replaced by other 4 ($2\min\{b,c\} = 8$), and the two objects sharing only 2 species, with a richness difference of 2 ($a = | b–c | = 2$), the point representing j,k in the ternary plot is positioned close to the R vertex, and with equal distance from D and S (Fig. 1c). When the three components are equal ($a = 2\min\{b,c\} = | b–c| = 4$), the corresponding point will fall onto the center of the triangle (Fig. 1d).

In a data matrix **X** containing m objects, the possible number of pairwise comparisons would be $w = (m^2 – m) / 2$, each corresponding to a point in the simplex. Notably, the shape of the point cloud in a simplex is unaffected by the actual arrangement of rows and columns in the matrix. "Extreme" structural patterns produce clear distributions of points in the triangle. If compositional similarity is high for all pairs, the point cloud will be near the S corner. When the objects have extreme richness difference, with low replacement and similarity, the points will be close to the D vertex. In cases when richness is similar but similarity is low, the points will be in the upper third of the diagram. These cases are extensions of the two-object situations explained above, and are not illustrated. However, there are further noted examples in which two of the three components contribute approximately equally to data structure, whereas the third is zero. Maximum beta diversity (anti-nestedness) in the data (with $s_{jk} = 0$ for all $j \neq k$), makes all points fall onto the left (D-R) side of the triangle (Fig. 1e), while maximum nestedness of objects (with $r_{jk} = 0$ for all pairs) forces all points to the bottom (D-S) side (Fig. 1f). In case of a perfect gradient (when the species richness is constant, the same number of species are lost and gained at each sampled step along that gradient, and $d_{jk} = 0$ for all pairs) all points are distributed on the S-R side (Fig. 1g). See Podani & Schmera (2011), for further examples of structural patterns and their simplex representations. The position of the centroid of the point cloud (calculated as the means of the s_{jk}, d_{jk} and r_{jk} values) will be used in a synthetic measure to compare the structure of comparable plots. Furthermore, these means multiplied by 100 quantify the percentage contributions of the three fractions to community pattern. In addition to these contributions, it is also useful to consider the percentage of presence scores in the data matrix, i.e., matrix fill, denoted here by q.

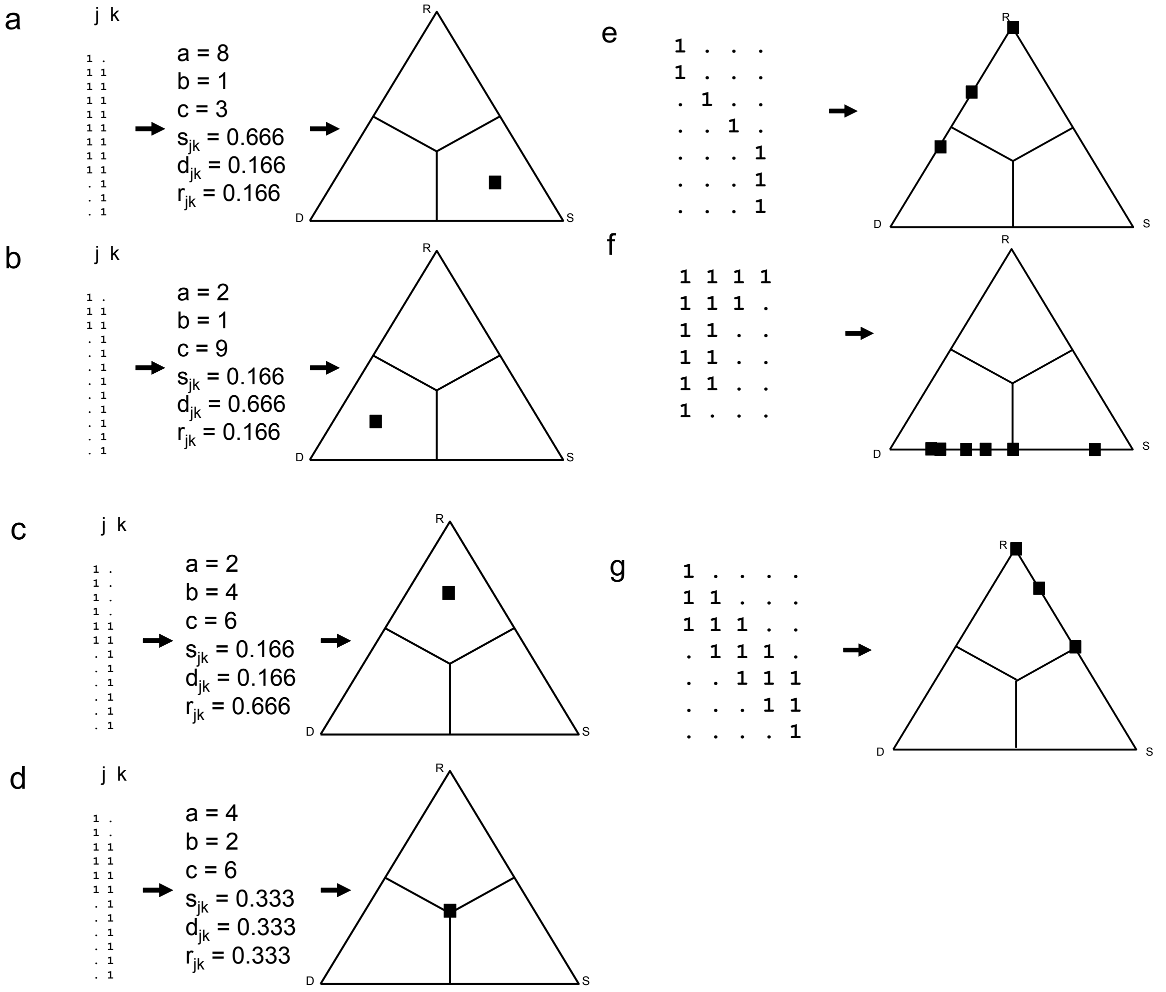

Fig. 1. (**a-d**) The three fractions and the position of the corresponding point in the ternary plot for a pair of artificial objects for various values of a, b and c. (**e**) A simple case of maximum beta diversity (anti-nestedness) for 4 objects (columns) and its representation in a ternary plot. (**f**) A case of maximum nestedness for 4 objects and the corresponding SDR-simplex. (**g**) Perfect gradient for 5 objects as is depicted by the simplex plot. Note that the number of object pairs, therefore the number of symbols is six in (**e**)-(**f**) and 10 in (**g**), but many of them overlap in the plot.

2.2 Indirect comparison of two SDR-simplexes for different sets of objects

Assume that we have another data matrix, **Y**, with p objects. Although p can differ considerably from m, the two corresponding point clouds in the ternary plots may have similar appearance, notwithstanding that the number of constituting points (w versus $z = [p^2 - p]/2$) will be clearly different. Similarities in the shape of point clouds should reflect similar structure in the two data matrices, as demonstrated by the artificial examples in Figure 2. In these sample matrices, the species and the objects are co-distributed to form an almost perfect modular pattern (with, in both cases, two large and clearly identifiable modules manifested as blocks of contiguous “1” values in the matrix). Converting the two data matrices into SDR-simplex triangles makes their structural similarity obvious. For both matrices, one set of points corresponds to within-block (near the S vertex) comparisons, while the other set of points represents between-block (near the R vertex) comparisons.

Although the visual examination of SDR simplexes is appealing and immediate, it is clear how simplexes’ geometrical properties offer a possible solution to the problem of comparing heterogeneous datasets, overcoming the difficulties posed by matrix-level comparisons. The spatial position of point sets A_X and B_Y within the corresponding triangle may be simplified into the relative frequency distributions F_X and G_Y, respectively, following the dissection of each triangle into small equilateral triangles identical in size, as shown in Fig. 3 (where 100 small triangles are used). Each value f_i in F_X is obtained as the number of points falling into small triangle *i*, divided by w, in the simplex diagram corresponding to **X**. The problem with points falling right onto the sides of small triangles is resolved in a systematic way (see Appendix S1 in Supplementary Information, for algorithmic details). Values in G_Y are derived in the same manner, dividing all g_i-s by z. Then, the divergence between the two distributions will provide the desired quantity:

$$E_{XY} = \sum_{i=1}^{100} \left| \frac{f_i}{w} - \frac{g_i}{z} \right|, \qquad (4)$$

that is the sum of the absolute differences between the corresponding relative frequencies over the 100 small triangles. Its maximum is 2, obtained when the two point clouds do not overlap, therefore E_{XY} may be divided by 2 to obtain a dissimilarity with a range of (0,1)

$$E'_{XY} = \frac{1}{2} \sum_{i=1}^{100} \left| \frac{f_i}{w} - \frac{g_i}{z} \right| . \qquad (5)$$

Although the above calculation may seem to provide a straightforward solution, the same E'_{XY} value of 1 may result for radically different situations: in fact, the maximum possible difference between two sets is recorded both when the point sets A_X and B_Y fall into close, but not overlapping positions, or when they fall right onto two different vertices of the triangle.

However, it should be considered that the D vertex is usually not occupied by actual data (because it corresponds to localities with no species, which, apart from few exceptions, are normally excluded from meta-community matrices). Thus, the maximum E'_{XY} value should more logically correspond to a situation where A_X corresponds to perfect species replacement (all points fall on the R vertex), and B_Y depicts complete similarity (all points coincide with the S vertex, see Fig. 3b). In the triangular representation, the distance between these vertices, that is, the maximum distance within the plot, is scaled appropriately to $\sqrt{2} = \sqrt{[(1-0)^2 + (0-1)^2 + (0-0)^2]}$.

To tackle this issue, we suggest including the distance between the centroids of point sets A_X and B_Y, abbreviated as c_{XY}, as a weighting factor in Eq. 5. The weighting factor is defined to be $t_{XY} = 1$ when the centroids coincide, and $t_{XY} = 1 + c_{XY}$ otherwise. Since weighting influences the maximum, it is advisable to rescale the quantity into the unit range using the maximum centroid distance, i.e., $\sqrt{2}$ to obtain a measure of dissimilarity between two simplex configurations:

$$\Delta_{XY} = \frac{(1+c_{XY})\sum_{i=1}^{100}\left|\frac{f_i}{w}-\frac{g_i}{z}\right|}{2(1+\sqrt{2})} \qquad (6)$$

In this, the maximum is 1, achieved only if the centroids fall onto the vertices, that is, in the unique situation described above. The minimum value, 0, is obtained when the centroids of the two point clouds coincide, *and* the relative frequency distributions F_X and G_Y are in perfect agreement. This does not imply, however, that zero Δ reflects complete identity of the two data structures being compared. Note that the two matrices need not be equal in size to yield $\Delta_{XY} = 0$, and that, additionally, the same result can be obtained when the differences between the two simplex configurations are small enough to produce identical relative frequencies. For these reasons, our measure is close to what the mathematicians call *pseudometric*, or *semi-distance* (Vialar 2016, p. 312). For the two matrices in Fig. 2, we obtain $\Delta_{XY} = 0.19$.

$$\mathbf{X}_{14,12} = \begin{array}{cccccccccccc}
1 & 1 & 1 & 1 & 1 & 1 & . & . & . & . & . & . \\
1 & 1 & 1 & . & 1 & 1 & . & 1 & . & . & . & . \\
1 & 1 & . & 1 & 1 & 1 & . & . & . & . & . & . \\
1 & 1 & 1 & 1 & 1 & 1 & . & . & . & . & . & . \\
1 & 1 & 1 & 1 & 1 & 1 & . & . & . & . & 1 & . \\
1 & 1 & 1 & 1 & . & . & . & . & . & . & . & . \\
. & 1 & 1 & 1 & 1 & 1 & . & . & . & . & . & . \\
1 & 1 & 1 & 1 & 1 & 1 & . & . & . & . & . & 1 \\
. & . & . & . & . & . & 1 & 1 & 1 & 1 & 1 & . \\
. & . & . & . & 1 & . & 1 & 1 & 1 & 1 & 1 & 1 \\
. & . & . & . & . & . & 1 & 1 & 1 & 1 & 1 & 1 \\
. & 1 & . & . & . & . & 1 & 1 & 1 & 1 & 1 & 1 \\
. & . & . & . & . & . & 1 & . & 1 & 1 & 1 & 1 \\
. & . & 1 & . & . & . & 1 & 1 & 1 & 1 & 1 & 1
\end{array} \rightarrow$$

$$\mathbf{Y}_{16,10} = \begin{array}{cccccccccc}
1 & . & 1 & 1 & 1 & 1 & . & . & . & . \\
1 & 1 & 1 & 1 & 1 & 1 & . & 1 & . & . \\
1 & 1 & . & 1 & 1 & 1 & . & . & . & . \\
. & 1 & 1 & 1 & 1 & 1 & . & . & . & . \\
1 & 1 & 1 & 1 & 1 & 1 & . & . & . & . \\
1 & 1 & 1 & 1 & 1 & . & . & . & . & . \\
1 & 1 & 1 & 1 & 1 & 1 & . & . & . & . \\
1 & 1 & . & 1 & 1 & 1 & . & . & . & . \\
. & . & . & 1 & . & . & 1 & 1 & 1 & 1 \\
. & . & . & . & . & . & 1 & 1 & 1 & 1 \\
. & . & . & . & . & . & 1 & 1 & 1 & 1 \\
. & . & . & . & . & . & 1 & 1 & 1 & 1 \\
1 & . & . & . & . & . & 1 & . & 1 & 1 \\
. & . & . & . & . & . & 1 & 1 & 1 & 1 \\
. & . & . & . & . & 1 & 1 & 1 & 1 & . \\
. & 1 & . & . & . & 1 & 1 & 1 & 1 & 1
\end{array} \rightarrow$$

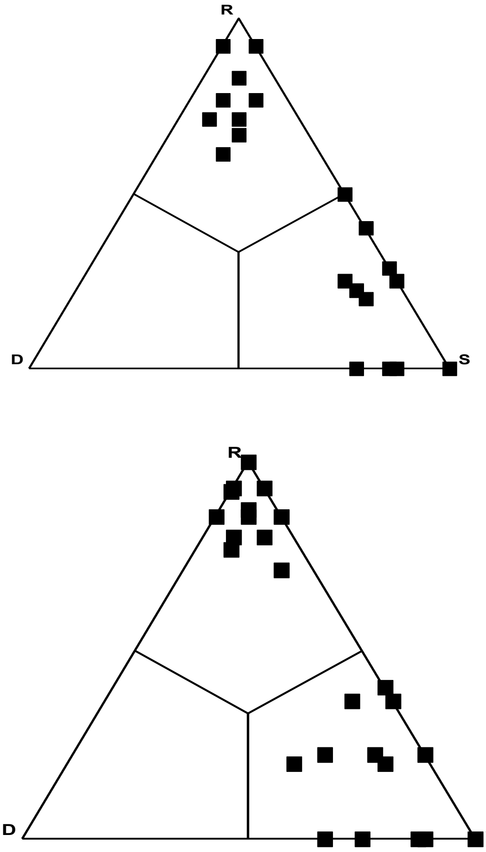

Fig. 2. Two matrices of different size with similar data structure, and the corresponding SDR-simplex diagrams.

2.3 Comparison of two data matrices for the same set of objects

A carefully designed study may provide several matrices representing, for example, the same meta-community sampled at different times (i.e., sequential snapshots of the same matrix) or the same set of objects sampled for different sets of organisms. In such cases, the above method may be considerably simplified into direct comparison, without dissecting the ternary plots into small triangles. Since every member of the set A_X has its counterpart in G_Y, we can calculate the relative shift in position between each pair of points in the ternary plot. A straightforward solution to measure the shift is offered by the additive fractions of the Jaccard coefficient. The desired measure takes the following form, where upper indices refer to the two matrices being compared:

$$M_{XY} = \frac{\sum_{j<k} \left|s_{jk}^{X} - s_{jk}^{Y}\right| + \left|d_{jk}^{X} - d_{jk}^{Y}\right| + \left|r_{jk}^{X} - r_{jk}^{Y}\right|}{2w}, \qquad (7)$$

which is a pairwise Manhattan distance standardized by the number of points (w) in the simplex to yield the range of (0,1). Eq. 7 quantifies the mean shift of the points in **X** with respect to **Y**. The minimum M_{XY} value, 0, indicates that the two sets of points coincide exactly. The maximum value, 1, which corresponds to a situation where the centroids of the two corresponding simplexes fall onto two different vertices of the triangle, represents, as discussed above, a rather theoretical situation (Fig. 3b).

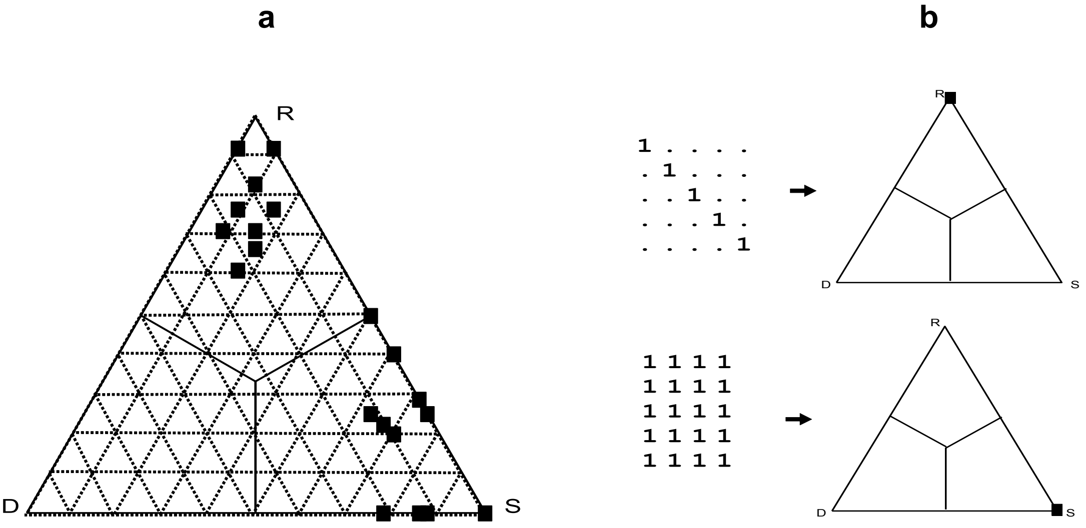

Fig. 3. (a) Dissecting the ternary plot into 100 small equilateral triangles, as superimposed onto the SDR-simplex on top in Fig. 2. (b) The rather theoretical situation when the two simplexes maximally differ: the upper one representing pure species replacement while the other corresponding to complete overall similarity.

2.4 Meta analysis and computer programs

If the study involves the comparison of *k* data matrices in every possible pair, a natural question is how to extract new information from the dissimilarity matrix of simplexes in order to explain underlying factors affecting ecological or biogeographical patterns in the starting datasets. That is, the next step is a sort of meta-analysis. The most straightforward approach to the issue is multivariate exploration, i.e., the application of some methods of numerical classification and ordination (Podani 2000). In the present case studies, we performed group average clustering (UPGMA, Sneath & Sokal 1973) and metric multidimensional scaling (or principal coordinates analysis, PCoA, Gower 1966), two procedures routinely used in biological data analysis. Interpretation of PCoA plots is enhanced by *a posteriori* superposition of arrows representing the three simplex fractions and matrix fill. The coordinates of these arrows are the correlations with the axes themselves, rescaled arbitrarily to fit the plotted area. Note that despite their similarity in graphical appearance, these ordination diagrams cannot be interpreted as conventional biplots.

The between-simplex dissimilarities were calculated by the SDR-DIST stand-alone WIN application written by the first author. Numerical results for the SDR simplex diagrams were prepared by program SDRSimplex (Podani & Schmera 2011). Cluster analysis and multidimensional scaling were performed and the simplex plots were drawn by using the SYN-TAX 2000 package (Podani 2001). All of these programs and their documentation can be downloaded free of charge from http://ramet.elte.hu/~podani. In addition, a thoroughly commented R script including all the functions needed to replicate in full the analysis performed in the present paper is provided in Appendix S3.

3. Case studies

We present here the main results for four different case studies, reporting a full description of datasets and a more thorough interpretation of results in Supplementary Information.

3.1 Biogeography of the central Aegean Islands.

We compiled 12 presence-absence matrices using published distributional data (see Appendix S2) for land snails, isopods, chilopods (centipedes), tenebrionid beetles, butterflies and reptiles from two island groups (Anatolian and Cycladic) in the Aegean Sea (east Mediterranean). Exploratory analysis of the 12 simplex diagrams via hierarchical classification identifies two large clusters (Fig. 4a). The group in the left includes matrices with the highest overall similarity (and the lowest beta diversity), namely butterflies in both island groups, and isopods and reptiles in the Anatolian islands. In the right group, data matrices have higher beta diversity (both data matrices for centipedes; snails and tenebrionids, and the isopods and reptiles in the Cyclades).

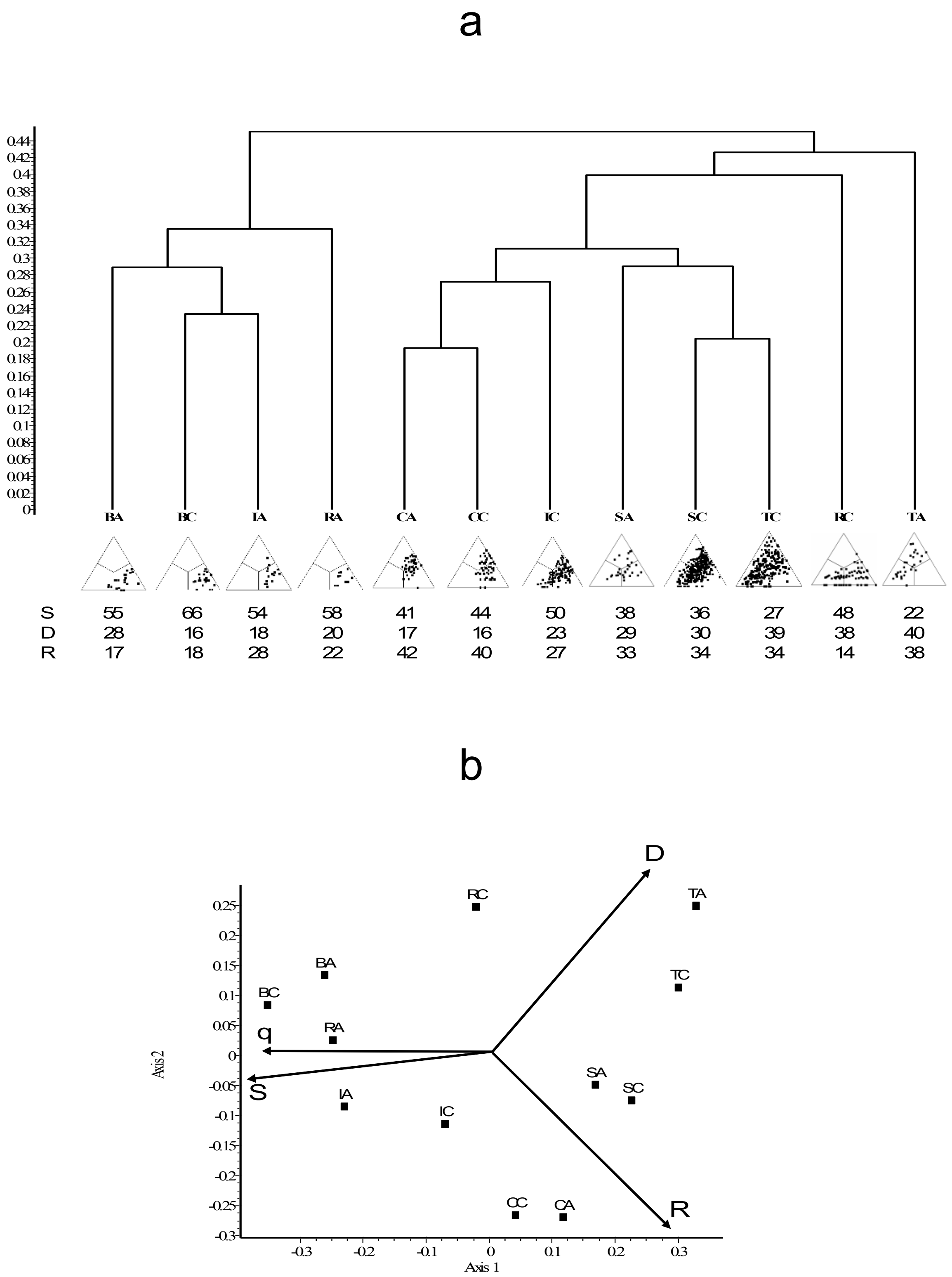

Fig. 4. (a) UPGMA dendrogram for the Mediterranean islands example. The SDR-simplex diagrams are shown in miniature under each label together with the percentage contributions from the S (right corner), D (left corner) and R (top corner) fractions. (b) PCoA ordination of faunas; scaling of arrows: 0.4 = unit correlation. Abbreviations as in Table 1.

Both the Anatolian and the Cycladic butterfly faunas are characterized by high species distributional overlaps, possibly related to high species dispersal ability, which may promote butterfly persistence in most islands of both archipelagos. A similar pattern and interpretation applies to reptiles in the Anatolian islands. Conversely, reptiles' tendency for nestedness (S+D=86%, the highest in this case study) in the Cyclades could possibly be a result of local extinctions, as reconstructed by Foufopoulos & Ives (1999) and Foufopoulos *et al.* (2011).

Isopods exhibit high species similarity in the Anatolian islands as well, but not in the Cyclades. The relative placement of Anatolian and Cycladic isopods in the dendrogram and in the ordination analysis reflects the similar positions of their centroids in the SDR plots (for IA – Anatolian Isopods: 54%, 18%, 28%, for IC – Cycladic Isopods: 50%, 23%, 27%). However, the larger variance of points in the Cyclades data prevents IA and IC from clustering together. Many species are common to most islands in both archipelagos, as reflected by many points falling into the S section of the simplex, possibly corroborating the idea that isopods are not as poor dispersers as generally thought (Tajoský *et al.* 2012). The higher level of nestedness in the Cyclades could be due to isopod distribution in a larger number of islands (16 islands, vs. 6 in the Anatolian group) which, in turn, may reflect potential interspecific differences in dispersal ability.

The distributional pattern of centipedes is highly consistent in the two island groups, with most communities tending towards richness agreement and comparable levels of species replacement and similarity, possibly due to centipedes' high dispersal ability (Simaiakis & Martínez-Morales 2010). Ternary plots reveal similar patterns for land snails in the two archipelagos, with most points scattered around the center, indicating randomness. This may reflect a combination of high population abundances (which may compensate for low dispersal ability), similarities in the interspecific dispersal abilities, and a certain degree of randomness in dispersal events.

Both ternary plots indicate a situation with large overlap between tenebrionid species composition within certain island groups, and large differences between the two island groups. These differences cannot be due to bias in sampling records, because occurrences of tenebrionids on the Aegean islands are well known (Pitta *et al.* 2017). Rather, this pattern may reflect a combination of current rare, overseas, long-distance dispersal and past colonization via land-bridges.

Table 1. Main features and abbreviations for data sets used in case studies.

Mediterranean islands			Finnish streams			Hungarian forests		
Fauna	Matrix size Species by islands	Matrix fill %	Fauna	Matrix size Species by sites	Matrix fill %	Group	Species richness	Matrix fill %
Butterflies - Anatolian Isl. BA	74 x 7	55	Muonionjoki (MU)	70 x 15	30	Birds	35	28
Butterflies – Cyclades BC	34 x 7	63	Norhern Ostrobothnia (NO)	64 x 15	22	Bryophytes – epiphytic (B-EP)	60	24
Chilopods - Anatolian Isl. CA	45 x 10	37	Käsivarsi (KÄ)	55 x 10	29	Bryophytes – understorey (B-S)	74	26
Chilopods - Cyclades CC	36 x 14	28	Tenojoki (TE)	66 x 15	31	Carabids (Carab)	34	22
Isopods – Anatolian Isl. IA	51 x 6	55	Koutajoki (KO)	96 x 14	24	Fungi – ectomycorrhizal (F_MYC)	293	14
Isopods – Cyclades IC	58 x 16	40	Southern coast (SC)	78 x 12	25	Fungi – terricolous saprotrophic (F_TS)	129	14
Reptiles - Anatolian Isl. RA	25 x 5	67	Upper Kymijoki (UK)	94 x 15	32	Fungi – wood inhabiting (F_W)	245	16
Reptiles - Cyclades RC	19 x 15	46	Southern Ostrobothnia (SO)	47 x 15	36	Herbs	132	16
Snails - Anatolian Isl. SA	96 x 8	41				Epiphytic lichens (Lich)	44	22
Snails - Cyclades SC	99 x 25	27				Saplings (Sapl)	40	24
Tenebrionids - Anatolian Isl. TA	79 x 8	27				Spiders	91	18
Tenebrionids - Cyclades TC	70 x 22	23						

3.2 The distribution of stream macroinvertebrates in eight regions in Finland.

Previous studies on northern streams' macroinvertebrate communities revealed significant regional differences in species composition and richness (Heino *et al.* 2002, 2003), evidence of community assembly processes (Heino *et al.* 2015a), and a strong latitudinal variation in community composition (Sandin 2003), possibly driven by climatic and environmental factors. Here we applied our innovative analytical framework to show how such macroecological patterns can be easily identified (and synthesized) by the SDR approach.

We compiled 8 data matrices summarizing the macroinvertebrate species composition in all the streams in eight northern Finnish regions (for details, see Table 1 and Appendix S2). Like in the previous case study, we compared indirectly every possible pair of simplexes, and then we used the resulting dissimilarity matrix to perform an UPGMA cluster analysis and a PCoA.

The dendrogram (Fig. 5a) splits the regions into two major clusters of equal size: the left cluster is characterized by relatively high similarity (35-44%) and by species replacement (36-48%) much higher than richness difference (15-20%). This pattern corresponds in the simplexes as points scattered around the centroid, and close to the right edge (richness agreement, S+R edge). This indicates that high richness differences between streams are rare within each region of this cluster. Although the position of the simplex centroid of Southern Ostrobothnia (SO) is close to that of Tenojoki (TE), the point cloud of the first one appears much more elongated, with some pairs exhibiting complete nestedness (on the bottom edge). Data matrices in the other (right) cluster have a much lower tendency for similarity (23-27%), with the R fraction dominating again beta diversity – with the exception of Northern Ostrobothnia (NO), where D and R have fairly equal contribution. In these cases, the point cloud moves closer to the beta diversity (left) edge of the ternary plot. The point cloud is most scattered for Northern Ostrobothnia (NO), reflecting the very high variability in species richness among streams in the region.

The ordination along axes 1-2 (accounting for 34% and 18% of variance, respectively) confirms the classification results, with a clear separation between the two clusters on the first axis. This is essentially an S versus beta diversity axis ($r_{1S} = -0.99$), with high correlations with matrix fill ($r_{qS} = 0.89$, $r_{1q} = -0.87$). The vertical axis corresponds to an increasing contrast between the two fractions of beta diversity. On the bottom (Northern Ostrobothnia, NO) contrast is low, i.e., D and R are similar, whereas on the top (Käsivarsi, KÄ) contrast is high, with species replacement (R) being three times higher than richness difference. This is also shown by the relatively high product moment correlation of 2nd axis scores with D (r_{2D} = –0.88) and R (r_{2R} = 0.72).

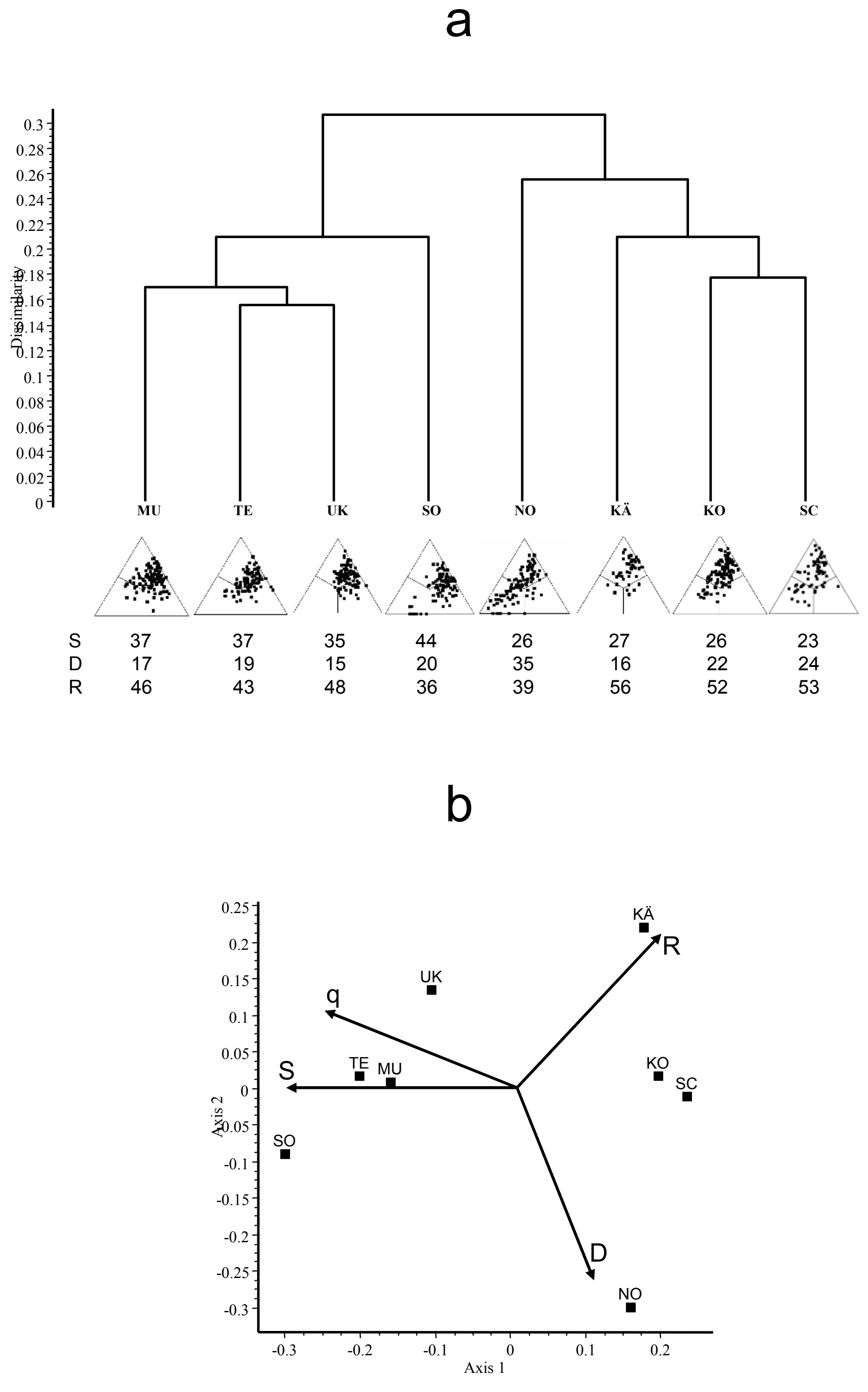

Fig. 5. (a) UPGMA dendrogram for the Finnish stream macroinvertebrates example. The SDR-simplex diagrams are shown in miniature under each label together with the percentage contributions from the S (right corner), D (left corner) and R (top corner) fractions, (b) PCoA ordination; scaling of arrows: 0.3 = unit correlation. Abbreviations as in Table 1.

Overall, the eight regions exhibited a fair similarity in metacommunity structure, mostly driven by between-stream similarity. Contrary to our expectation, however, we did not find strong evidence for latitudinal gradients in the S, D or R components. This finding, however, is in line with studies focusing on regional and local richness patterns in Finnish streams, showing clear among-region differences but no clear latitudinal gradients (Heino *et al.* 2003). Among-region differences emerged also from our analysis. The two main clusters (Fig. 5a) comprised both northern and southern regions, suggesting that latitude does not drive patterns in beta diversity. This may be a consequence of variation in environmental heterogeneity among the regions. In fact, heterogeneity in environmental conditions due to natural or anthropogenic influences may indeed affect differences in beta diversity among regions (Bini *et al.* 2014; Heino *et al.* 2015b). It is also possible that region-specific context dependency plays an important role in driving these patterns, since each region has different underlying environmental characteristics (e.g., water chemistry and physical habitat conditions), stream network configurations and, consequently, may show different metacommunity dynamics (Grönroos *et al.* 2013; Tonkin *et al.* 2016).

3.3 Community structure in Central European mixed forests.

The species composition of eleven organism groups (herbs, saplings, understorey bryophytes, epiphytic bryophytes, epiphytic lichens, terricolous saprotrophic fungi, wood-inhabiting fungi, ectomycorrhizal fungi, spiders, carabid beetles, and birds[1]) was surveyed in mixed temperate forests in West-Hungary (North 46°51’-55’, East 16°07’-23’, for map see Fig. S2.4).

In this case study, sampling was designed in a way that all organism groups were surveyed in all the 35 sample sites (see Appendix S2, for details). As a consequence, the comparison of point scatters within the ternary plots simplifies to the direct measurement of shift (Eq. 7). The matrix of resulting Manhattan distances was then subjected to UPGMA clustering and PCoA, as in the above two case studies.

---

[1] The latter three are monophyletic taxa, the first two are largely independent of taxonomy, the others represent combinations of ecological categories and paraphyletic taxa.

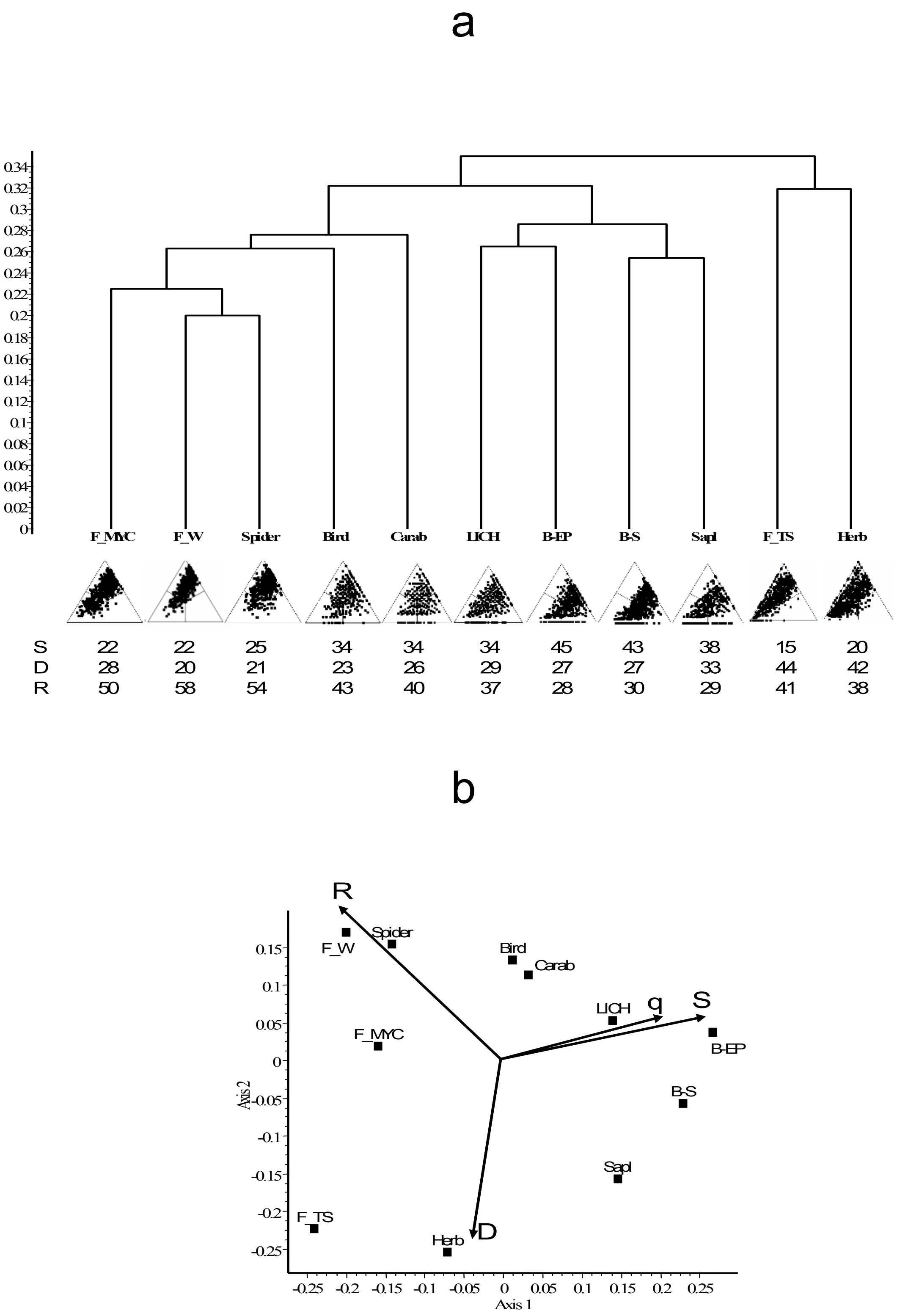

Fig. 6. (a) UPGMA dendrogram for the forest community example. The SDR-simplex diagrams are shown in miniature under each label together with the percentage contributions from the S (right corner), D (left corner) and R (top corner) fractions, (b) PCoA ordination for axes 1-2; scaling of arrows: 0.25 = unit correlation. Abbreviations as in Table 1.

The dendrogram (Fig. 6a) separates organism groups in which species replacement (R) is the dominant component (on the left: ectomycorrhizal fungi, wood inhabiting fungi, spiders, birds, carabids), from similarity (S) dominated groups (in the middle: epiphytic bryophytes, understory bryophytes and saplings), and beta diversity dominated groups for which the D and R components had similarly high importance (on the right: terricolous saprothrophic fungi, herbs). Epiphytic lichens had an intermediate position between the R and S groups. The ordination (Fig. 6b) agrees well with the agglomerative classification in revealing meta-structure of the organism group data. PCoA axis 1 (20%) represents a gradient from high levels of species replacement (R component, negative side) to relatively high similarity (S, positive side) as expressed by its high correlation with R ($r_{1R} = -0.83$) and S ($r_{1S} = 0.96$). Matrix fill has high correlations with similarity ($r_{qS} = 0.89$) and axis 1 ($r_{1q} = 0.82$). Ordination axis 2 (15%) is the most correlated with richness difference ($r_{2D} = -0.94$).

The overall picture on point clouds within the SDR triangles is that different organism groups develop greatly different community pattern in the same sites. The points in many cases form a narrow triangular shape with the tip near the D vertex and the base on the right, richness agreement (R-S) edge. Most of the organism groups with high beta diversity have a relatively high species richness (fungi groups, herbs, birds) and low level of matrix fill (fungi groups, herbs, spiders, Table 1). It means that in these organism groups (especially for fungi) the proportion of rare species (occurring only in one or two plots) is very high.

This study indicates that in some organism groups high beta diversity, while for others higher similarity is the dominant structural component. The high beta diversity and low nestedness of fungus assemblages compared to other sessile organism groups have been proved for saproxylic communities (Halme *et al.* 2013; Heilmann-Clausen *et al.* 2014). This high beta diversity of the fungal groups could be related to the methodology of sampling which focused on sporocarp inventory. Similarly to this study, the species replacement component of beta diversity for breeding bird assemblages had higher importance than richness difference (low level of nestedness effect, Si *et al.* 2015). In vascular plant and beetle assemblages of European beech forests, species replacement had much stronger effect on beta diversity than nestedness (Gossner *et al.* 2013), while for epixylic and epiphytic bryophytes in this community nestedness had higher importance (Táborska *et al.* 2017). For forest bryophytes and lichens, especially for epiphytic and epixylic assemblages, a high degree of nestedness is detected in many forest types (Hylander & Dynesius 2006; Nascimbene *et al.* 2010). Spider and carabid beetle communities of this study had similar metacommunity structure with relatively even S-D-R distribution. For spiders, the dominance of species replacement and low level nestedness has been shown (Carvalho & Cardoso 2014).

The similar environmental drivers only partly support groups based on the simplex structure. Within the beta diversity dominated clusters, the species composition of the three fungus groups and spiders is determined mainly by tree species composition, and forest microclimate (Samu et al. 2014, Kutszegi et al. 2015). However, herbs also exhibiting high beta diversity are influenced mainly by other variables (e.g., light conditions, tree species diversity and shrub layer density, Márialigeti *et al.* 2009, 2016; Tinya *et al.* 2009). Epiphytic bryophytes and lichens also share some common environmental drivers such as shrub density and humid microclimate (Király *et al.* 2013; Ódor *et al.* 2013). Birds are related to different drivers than the other organism groups like tree size, dead wood amount and understory cover (Mag & Ódor 2015).

3.4 Community recovery of arthropods in Florida: Rey's defaunation experiment.

In order to test the equilibrium theory of insular biogeography, Rey (1981) investigated the arthropod fauna of six islets in northern Florida. These islets are from 56 m^2 to 1023 m^2 in area, and are located at a distance of 29 m to 1752 m from the mainland. The vegetation of the study area was a monodominant stand of smooth cordgrass (*Spartina alternifolia* Loisel.). At the beginning of the experiment, the terrestrial arthropods of the islets were killed by insecticides. The arthropod fauna was then recorded in weekly intervals for a year to monitor the recolonization process. Details are given in the original publication by Rey (1981). We took the species by islands data from 20 points of time (weeks 9-26, 31 and, finally, 53), made available as a supplement to the nestedness calculator program package developed by Atmar & Patterson (1995).

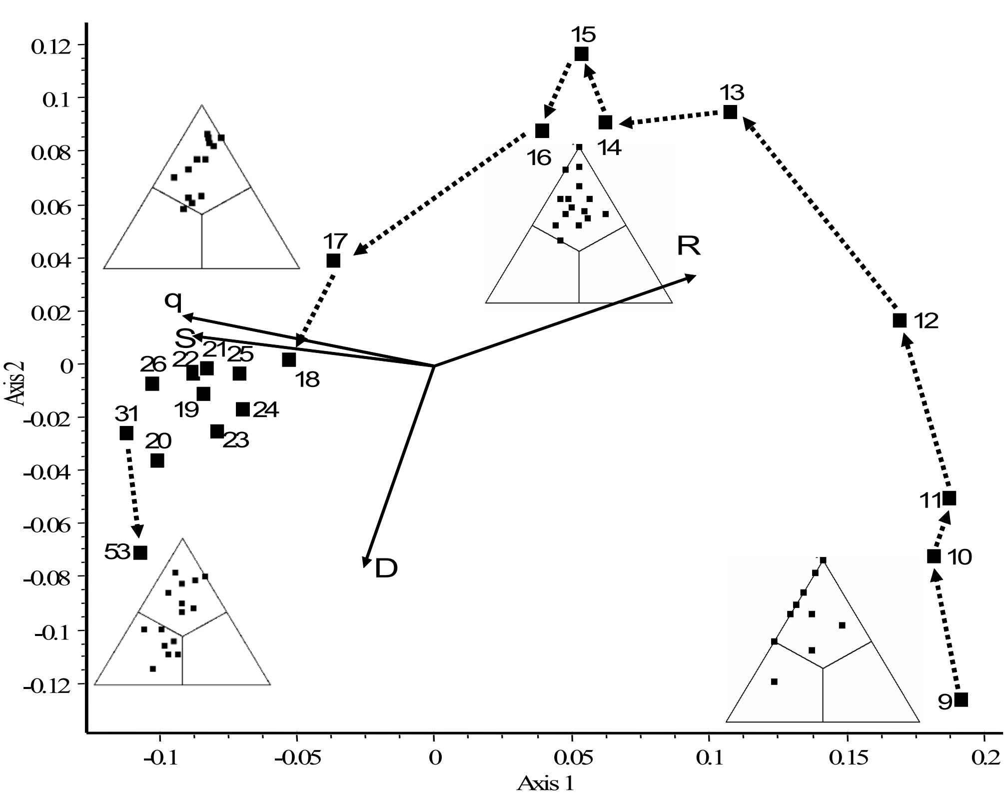

Fig. 7. Principal coordinates ordination of 20 data matrices from the island recolonization experiment in northern Florida. Simplex plots are shown for weeks 9, 14, 25 and 53. Note the S (right corner), D (left corner) and R (top corner) fractions in each plot. Dotted arrows connect subsequent weeks, but those are omitted for weeks 18 to 31 for clarity. Scaling of solid arrows: 0.1 = unit correlation. See Table 2, for centroids and matrix fill percentages.

This survey provided data suitable for demonstrating the performance of our procedure in the analysis of pattern development for the same set of localities (i.e., islets) over time. We therefore used Eq. 7, the direct method to express pattern dissimilarity between points of time. We feel that hierarchical classification is less relevant to this situation, and suggest that the problem of monitoring temporal changes is sufficiently approached by PCoA. The results are shown in Fig. 7 for the first 2 ordination axes (accounting for 54% and 16% of variance, respectively). Four simplex plots, from weeks 9, 14, 25 and 53, are superimposed to the ordination near the points they represent.

Table 2. Main data properties and SDR centroids for data from Rey's island recolonization experiment.

Week	Species richness	S	D	R	Matrix fill % q
9	23	5	26	69	21
10	25	6	23	71	23
11	26	7	25	68	24
12	26	10	23	67	27
13	27	15	17	68	28
14	28	16	19	65	30
15	29	15	15	70	30
16	30	13	19	68	28
17	27	16	23	61	33
18	30	15	24	61	32
19	31	15	27	58	33
20	33	17	30	53	33
21	34	19	26	55	34
22	35	19	26	55	33
23	37	19	24	57	32
24	39	12	22	59	32
25	41	12	22	59	31
26	41	21	24	55	31
31	35	26	21	53	36
52	33	23	31	46	35

The arrangement of simplex diagrams in the ordination follows an obvious horseshoe pattern, otherwise very typical for community data with a single dominant environmental gradient to which all species respond in a unimodal fashion (Podani & Miklós 2002). Thus, the presence of the horseshoe in the present case appears to indicate a one-directional temporal trend regarding changes of community pattern on the islands. In this, with the exception of week 15, we cannot see seasonal variations and strong fluctuations that were otherwise observed for various diversity statistics by Rey. As also shown by Podani & Schmera (2011), there is a fairly monotonous increase of matrix fill, S and nestedness over time; therefore, replacement and beta diversity exhibit the opposite behaviour while changes of richness difference appear less consistent (Table 2). This is expressed quite well with the product moment correlation coefficients calculated between axis 1 and S, R and q ($r_{1S} = -.81$, $r_{1R} = .88$, $r_{1q} = -.92$) of which the correlation with matrix fill is the highest. However, the second axis may also be interpreted in terms of the SDR values, namely the correlation with richness difference is $r_{2D} = -.76$, showing some less obvious trend that richness difference first decreases and then increases over time. This suggests that in this case study the horseshoe pattern is not a mathematical consequence of a long, unidimensional background gradient, but the manifestation of two, largely independent changes of community pattern.

As Rey (1981) reported, the first species appeared c. 4-5 weeks after treatment and then the total number of species started to increase rapidly. Re-occurrence of species on particular islets was accidental, however. Therefore, this pioneer stage is characterized by very high beta diversity as is indicated by many points lying on the left side of the SDR plot: the corresponding pairs of islands had no single species in common (week 9). Then, extinctions and immigrations dominated until week 18 with further increases in species number. This second period is shown by the increased concentration of points inside the upper third of the triangle (the species replacement sector). As Rey observed, species richness reached the original levels in approximately 20 weeks and oscillated around these values until the end of the experiment. Our analysis reveals that not only species richness but community pattern was also oscillating in this period, as shown by little changes in the ordination. The simplex plot for the last week (53) shows recovery to the original state: richness difference has increased as expected, and half of the points moved into the richness difference sector of the triangle. It is the manifestation of the classical species/area relationship. That is, by the end of the experiment the normal conditions are observed again because the islets differ considerably in size so that they maintain different levels of alpha diversity.

4. Discussion

Methods of multivariate analysis have been commonly used in ecology and biogeography to reveal non-random structure implied in presence-absence data matrices. Comparison of their results is a common practice whenever interest lies in evaluating the relative importance of user decisions made during sampling and data processing, in quantifying the effect of choosing between descriptor variables, and in the discovery of underlying factors affecting these results (Podani 1989, Lengyel & Podani 2014 and references therein). Generally, such studies are performed on (dis)similarity matrices (e.g., matrix correlation, Mantel 1967, Sneath & Sokal 1973, Hubert 1983), dendrograms (Rohlf 1974), partitions (Arabie & Boorman 1973), and ordinations (Gower 1971, 1975); but see Podani (2000, Chapter 9) for a more complete account of the issue. Application of these types of results in comparisons always implies some loss of information. Calculating dissimilarities from raw data usually ignores the relationships between variables by reducing the comparison to a single number.

Obtaining dendrograms, partitions and ordinations from such square matrices involves further information loss or even distortion. A possible solution would be the direct comparison of data matrices, for which very few methods have been suggested (Hubert & Golledge 1982, Zani 1986). However, all approaches to the problem are constrained by that the number of objects (plots, sites, localities, etc.), and sometimes the number of variables (e.g., species) are fixed, and that every result (all dendrograms, for example) relate to the same set of study objects.

Large-scale studies are often concerned with several data matrices derived from different studies, which rarely agree in the number of objects and variables used. For example, we may want to evaluate how species from different taxonomic groups are distributed in space in the same geographic region, such that the data are obtained by experts of the given taxa in different surveys. Such studies are of central importance to deepen our understanding of beta diversity in communities and biogeographical regions, the ecological factors explaining these relationships and their consequences in environmental conservation. No method is available yet to perform comparisons in such studies.

The novel data analysis model introduced here has been designed to satisfy this requirement. The method involves three major computational steps. First, matrices are decomposed into three fractions of Jaccard dis(similarity) for every pair of objects. Then, a 2D simplex plot is determined for each matrix in which a given point represents a pair of objects. Finally, a pairwise divergence measure is calculated for two matrices based on the relative frequency distribution of points in pre-defined subsections of the simplex. This algorithmic summary reveals the fundamental assumption of our approach: the shape of the point cloud in the two-dimensional simplex space is informative on data structure, because the component terms, namely similarity, richness divergence and replacement reflect different aspects of community organization.

The third step simplifies to the calculation of Manhattan distance based on the three fractions when the two matrices under comparison relate to the same set of objects. In this case, direct comparison of dissimilarity matrices derived from the data could appear as a simpler solution. Yet, the simplex approach maintains its advantage in that data patterns are evaluated through the use of fractions of dissimilarity having different ecological interpretations, offering insights not available otherwise. Our results clearly show that similar levels of dissimilarity can reveal different underlying patterns when the different components are considered.

Our approach is particularly useful for the simultaneous evaluation of more than two data sets. In this case, as a fourth step of the analysis, a dissimilarity matrix is computed by calculating the divergence between each pair of simplexes, and then subjected to clustering and/or ordination. This approach is useful for categorization of data sets and for evaluating the effect of underlying factors mentioned above.

We illustrated the potential of our method, and particularly the usefulness of both classifications and PCoA ordination techniques to investigate general biogeographical and ecological patterns in multiple datasets, using four case studies representing different ecological and biogeographical situations: faunas from eastern Mediterranean islands, stream macroinvertebrate assemblages in Finland, different organism groups in a mixed deciduous forest in West-Hungary, and a temporal sequence corresponding to a recolonization process in defaunated small islets in Florida, USA. The clusters identified by classification were, in general, consistent with the results of PCoA ordinations. Minor differences between the dendrograms and the corresponding ordination plots are likely due to the fact that clustering attempts to concentrate all information into a single tree, whereas ordination separates total variance into orthogonal components (so that the first two dimensions cover only a fraction of total variation).

Superimposing correlations of the three SDR components and matrix fill as arrows in a biplot-like ordination diagram demonstrated the success of ordinations in decomposing total variation into interpretable orthogonal fractions. In all cases, the strongest component determining data pattern was similarity. It had high correlations with the most important ordination axis, which, in turn, was always highly correlated with matrix fill. The first axis corresponded always to the contrast between S and beta diversity, but with different contributions to the latter by its two additive components, richness difference and species replacement. D was dominating in case of the Mediterranean island faunas, whereas in the Finnish streams example R was slightly more influential than D. In the forest and recolonization studies, the effect of species replacement was even higher, so that the first axis could be interpreted purely as a contrast between S and R. In all the four cases, the second axis was always associated with the two beta diversity fractions. This was perhaps the weakest in the recolonization study in which the second axis purely attributed to richness difference. Since this fraction first decreased and then increased over time, the temporal trajectory took an arched shape. This provided a more straightforward interpretation of the horseshoe effect than usual in ecological ordinations.

We emphasize that in the comparison of data sets the entire point cloud, and therefore the distribution of points in the partition set of each simplex matters. Thus, this should be in general preferred to evaluating the results in terms of the three centroids, since the two approaches could lead to slightly different interpretations. This was obvious in several examples, when the clustering tendency of certain study objects was not in agreement with similarities in SDR centroids. Furthermore, it should be highlighted that the approach we propose is mainly exploratory. On the one hand, our protocol provides a straightforward method to improve our understanding of the ecological determinants of species distribution patterns. On the other hand, however, the main purpose of SDR simplexes, divergence measures, and the meta-analysis of resulting dissimilarity matrices is that of revealing underlying structural properties and external influential factors. We expect with good reason that the approach has potential in any field of science which produces several data matrices whose comparison is expected to give additional insight into the general problem investigated.

Acknowledgements

In the case of the forest dataset, the field inventory and identification of organism groups is appreciated for Botos Istvánné, Francesco Bortignon, Bálint Dima, Gábor Lengyel, Marilena Dalle Vedove, Ildikó Király, Gergely Kutszegi, Zsuzsa Mag, Sára Márialigeti, István Mazál, Zsolt Merényi, Juri Nascimbene, Balázs Németh, Ferenc Samu, Irén Siller, Győző Szél, Katalin Takács, Flóra Tinya, Gábor Turcsányi, Torda Varga. The forest study was supported by the Hungarian Science Foundation (OTKA 79158), National Research, Development and Innovation Office (GINOP-2.3.2-15-2016-00019) and Őrség National Park Directorate. The authors confirm that there is no conflict of interest among them.

Supplementary information

Additional supporting information may be found in the online version of this article at the publisher’s web-site:

Appendix S1 Conversion of simplex data into a frequency distribution.

Appendix S2 Data sources.

Appendix S3 R script for the comparison of multiple data matrices.

References

Arabie, P. & Boorman S. A. (1973) Multidimensional scaling of measures of distance between partitions. *J. Mathematical Psychology,* **10**, 148–203.

Atmar, W. & Patterson, B.D. (1995) *The Nestedness Temperature Calculator: A Visual BASIC Program, Including 294 Presence-absence Matrices.* Chicago Field Museum, Chicago.

Bini, L.M., Landeiro, V.L., Padial, A.A., Siqueira, T. & Heino, J. (2014) Nutrient enrichment is related to two facets of beta diversity for stream invertebrates across the United States. *Ecology*, **95**, 1569–1578.

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 & Biogeography* **21**, 760–771.

Carvalho, J.C. & Cardoso, P. (2014) Drivers of beta diversity in Macaronesian spiders in relation to dispersal ability. *Journal of Biogeography*, **41,** 1859–1870.

Edwards, A.W.F. (2000) *Foundations of Mathematical Genetics.* 2nd ed. Cambridge University Press, Cambridge.

Foufopoulos, J. & Ives, A.R. (1999) Reptile extinctions on land-bridge islands: life-history attributes and vulnerability to extinction. *American Naturalist*, **153,** 1–25.

Foufopoulos, J., Kilpatrick, A.M. & Ives, A.R. (2011) Climate change and elevated extinction rates of reptiles from Mediterranean islands. *American Naturalist,* **177,** 119–129.

Gossner, M.M., Getzin, S., Lange, M., Pasalic, E., Turke, M., Wiegand, K. & Weisser, W.W. (2013) The importance of heterogeneity revisited from a multiscale and multitaxa approach. *Biological Conservation,* **166,** 212–220.

Gower, J.C. (1966) Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika,* **53,** 325–338.

Gower, J. C. (1971) Statistical methods of comparing different multivariate analyses of the same data. In: F.R. Hodson, D.G. Kendall & P. Tautu (eds), *Mathematics in the*

Archaeological and Historical Sciences. Edinburgh University Press, Edinburgh. pp. 138-149.

Grönroos, M., Heino, J., Siqueira, T., Landeiro, V.L., Kotanen, J. & Bini, L.M. (2013) Metacommunity structuring in stream networks: roles of dispersal mode, distance type and regional environmental context. *Ecology & Evolution,* **3,** 4473–4487.

Grime, J.P. (1977) Evidence for the existence of three primary strategies in plants and its relevance to ecological and evolutionary theory. *American Naturalist,* **111,** 1169–1194.

Halme, P., Ódor, P., Christensen, M., Piltaver, A., Veerkamp, M., Walleyn, R., Siller, I. & Heilmann-Clausen, J. (2013) The effects of habitat degradation on metacommunity structure of wood-inhabiting fungi in European beech forests. *Biological Conservation,* **168,** 24–30.

Heilmann-Clausen, J., Aude, E., van Dort, K.W., Christensen, M., Piltaver, A., Veerkamp, M. T., Walleyn, R., Siller, I., Standovár, T. and Ódor, P. (2014) Communities of wood-inhabiting bryophytes and fungi on dead beech logs in Europe - reflecting substrate quality or shaped by climate and forest conditions? *Journal of Biogeography*, **41,** 2269–2282.

Heino, J., Melo, A.S. & Bini, L. M. (2015b) Reconceptualising the beta diversity-environmental heterogeneity relationship in running water systems. *Freshwater Biology,* **60,** 223–235.

Heino, J., Muotka, T. & Paavola, R. (2003) Determinants of macroinvertebrate diversity in headwater streams: regional and local influences. *Journal of Animal Ecology,* **72,** 425–434.

Heino, J., Muotka, T., Paavola, R., Hämäläinen, H. & Koskenniemi, E. (2002) Correspondence between regional delineations and spatial patterns in macroinvertebrate assemblages of boreal headwater streams. *Journal of the North American Benthological Society,* **21,** 397–413.

Heino, J., Nokela, T., Soininen, J., Tolkkinen, M., Virtanen, L. & Virtanen, R. (2015a) Elements of metacommunity structure and community-environment relationships in stream organisms. *Freshwater Biology,* **60,** 973–988.

Hubert, L.J. & Golledge, R.G. (1982) Comparing rectangular data matrices. *Environmental Planning A,* **14,** 1087–1095.

Hubert, L. J. (1983) Inference procedures for the evaluation and comparison of proximity matrices. In: J. Felsenstein (ed.), *Numerical Taxonomy*. Springer, Berlin. pp. 209-228.

Hylander, K. & Dynesius, M. (2006) Causes of the large variation in bryophyte species richness and composition among boreal streamside forests. *Journal of Vegetation Science,* **17,** 333–346.

Jaccard, P. (1901. Étude comparative de la distribution florale dans une portion des Alpes et des Jura. *Bulletin de la Société Vaudoise des Sciences Naturelles*, **37,** 547–579.

Király, I., Nascimbene, J., Tinya, F. & Ódor, P. (2013) Factors influencing epiphytic bryophyte and lichen species richness at different spatial scales in managed temperate forests. *Biodiversity & Conservation,* **22,** 209–223.

Kutszegi, G., Siller, I., Dima, B., Takács, K., Merényi, Z., Varga, T., Turcsányi, G., Bidlo, A. & Ódor, P. (2015) Drivers of macrofungal species composition in temperate forests, West Hungary: functional groups compared. *Fungal Ecology,* **17,** 69–83.

Lengyel, A. & Podani, J. (2015) Assessing the relative importance of methodological decisions in classifications of vegetation data. *Journal of Vegetation Science,* **26**, 804-815.

Mag, Z. & Ódor, P. (2015) The effect of stand-level habitat characteristics on breeding bird assemblages in Hungarian temperate mixed forests. *Community Ecology,* **16,** 156–166.

Mantel, N. (1967) The detection of disease clustering and a generalized regression approach. *Cancer Research, 2*7, 209-220.

Márialigeti, S., Németh, B., Tinya, F. & Ódor, P. (2009) The effects of stand structure on ground-floor bryophyte assemblages in temperate mixed forests. *Biodiversity & Conservation,* **18,** 2223–2241.

Márialigeti, S., Tinya, F., Bidlo, A. & Ódor, P. (2016) Environmental drivers of the composition and diversity of the herb layer in mixed temperate forests in Hungary. *Plant Ecology,* **217,** 549–563.

Nascimbene, J., Marini, L. & Nimis, P. L. (2010) Epiphytic lichen diversity in old-growth and managed Picea abies stands in Alpine spruce forests. *Forest Ecology & Management,* **260,** 603–609.

Ódor, P., Király, I., Tinya, F., Bortignon, F. & Nascimbene, J. (2013) Patterns and drivers of species composition of epiphytic bryophytes and lichens in managed temperate forests. *Forest Ecology & Management,* **306,** 256–265.

Pitta, E., Kassara, C., Trichas, A., Sfenthourakis, S. & Chatzaki, M. (2017) Community variation of spiders, beetles and isopods in three small island groups of the Aegean Sea: the interplay between history and ecology. *Journal of Biogeography*, **44,** 1077–1087.

Podani, J. (1989) Comparison of classifications and ordinations of vegetation data. *Vegetatio,* **83,** 111-128.

Podani, J. (2000) *Introduction to the Exploration of Multivariate Biological Data.* Backhuys, Leiden.

Podani, J. (2001) *SYN-TAX 2000. Computer Programs for Data Analysis in Ecology and Systematics. User's Manual.* Scientia, Budapest.

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.

Rey, J. R. (1981) Ecological biogeography of arthropods on Spartina islands in Northwest Florida. *Ecological Monographs,* **51,** 237–265.

Rohlf, F. J. (1974) Methods of comparing classifications. *Annual Reviews in Ecology and Systematics,* **5,** 101-113.

Samu, F., Lengyel, G., Szita, E., Bidlo, A. & Ódor, P. (2014) The effect of forest stand characteristics on spider diversity and species composition in deciduous-coniferous mixed forests. *Journal of Arachnology,* **42,** 135–141.

Sandin, L. (2003) Benthic macroinvertebrates in Swedish streams: community structure, taxon richness, and environmental relations. *Ecography,* 26, 269–282.

Si, X.F., Baselga, A. & Ding, P. (2015) Revealing beta-diversity patterns of breeding bird and lizard communities on inundated land-bridge islands by separating the turnover and nestedness components. *Plos One*, **10,** e0127692.

Simaiakis, S.M. & Martínez-Morales, M.A. (2010) Nestedness in centipede (Chilopoda) assemblages on continental islands (Aegean, Greece). *Acta Oecologia,* **36,** 282–290.

Sneath, P.H.A. & Sokal, R.R. (1973) *Numerical Taxonomy.* 2^{nd} ed. Freeman, San Francisco.

Stoffels, R. J. (2013) *Trophic Ecology. Ecology of Australian Freshwater Fishes.* (ed. by P. Humphries and K. Walker), Chapter 6. CSIRO Publishing, Canberra.

Streckeisen, A. (1976) To each plutonic rock its proper name. *Earth Science Reviews,* **12,** 1–33.

Táborská, M., Procházková, J., Lengyel, A., Vrška, T., Hort, L. & Ódor, P. (2017) Wood-inhabiting bryophyte communities are influenced by different management intensities in the past. *Biodiversity & Conservation,* **26,** 2893-2909.

Tajovský, K., Hošek, J., Hofmeister, J. & Wytwer, J. (2012) Assemblages of terrestrial isopods (Isopoda, Oniscidea) in a fragmented forest landscape in Central Europe. *ZooKeys,* **176,** 189–198.

Tinya, F., Márialigeti, S., Király, I., Németh, B. & Ódor, P. (2009) The effect of light conditions on herbs, bryophytes and seedlings of temperate mixed forests in Őrség, Western Hungary. *Plant Ecology,* **204,** 69–81.

Tonkin, J.D., Heino, J., Sundermann, A., Haase, P. & Jähnig, S. (2016) Context dependency in biodiversity patterns of central German stream metacommunities. *Freshwater Biology,* **61,** 607–620.

Vialar, T. (2016) *Handbook of Mathematics.* Paris West University, Nanterre.

Zani, S. (1986) Some measures for the comparison of data matrices. Proceedings of the XXXIII Meeting of the Italian Statistical Society, pp 157–169. Bari, Italy.