Combinatorial functional diversity: an information theoretical approach

J. Podani[1,5], C. Ricotta[2], J.G. Pausas[3] and D. Schmera[4]

[1] Department of Plant Systematics, Ecology and Theoretical Biology, and Ecology Research Group of HAS,
Institute of Biology, L. Eötvös University, Pázmány P. s. 1/C, H-1117 Budapest, Hungary
[2] Department of Environmental Biology, University of Rome "La Sapienza", Piazzale Aldo Moro 5,
I-00185 Rome, Italy
[3] Centro de Investigaciones sobre Desertificación, Consejo Superior de Investigaciones Cientificas (CIDE-CSIC),
Ctra. Nàquera Km. 4.5 (IVIA), 46113 Montcada, Valencia, Spain
[4] Section of Conservation Biology, University of Basel, St. Johanns-Vorstadt 10, CH-4056 Basel, Switzerland, and
Balaton Limnological Institute, Centre for Ecological Research, Hungarian Academy of Sciences,
Klebelsberg K. u. 3, H-8237 Tihany, Hungary
[5] Corresponding author. Email: podani@ludens.elte.hu

Abstract. A new approach to the measurement of functional diversity based on two-state nominal traits is developed from the florula diversity concept of P. Juhász-Nagy. For evaluating functional diversity of an assemblage, first a traits by species matrix is compiled. Various information theory functions are used to examine structural properties in this matrix, including the frequency distribution of trait combinations. The method is illustrated by actual examples, the first from plant communities prone to fire in Spain, and the second from running water invertebrate assemblages in Hungary. The results suggest that of the various functions used the standardized joint entropy, termed combinatorial functional evenness supplies most meaningful results. In plant communities, high fire recurrence decreased combinatorial functional evenness, while this measure for freshwater assemblages was uncorrelated with stream width and negatively correlated with the degree of human impact. Stream width is negatively correlated with the number of manifested functional combinations. In both case studies, combinatorial functional evenness has an inverse relationship to species richness – i.e., fewer species have a larger chance to produce equiprobable functional combinations.

Abbreviations: CFD–Combinatorial functional diversity, CFE–Combinatorial functional evenness, CFR–Combinatorial functional richness, FA–Functional associatum, FD–Functional diversity, FH–Functional heterogeneity.

Introduction

Functional traits are the main ecological attributes by which different species and their communities influence ecosystem processes (de Bello et al. 2010). Therefore, measures of functional diversity tend to correlate more strongly than those of traditional species-diversity with ecosystem functions (Petchey and Gaston 2006). The observed trait-environment relationships raise the question of how to measure functional diversity in biologically and statistically meaningful ways (Lavorel et al. 2008, Villéger et al. 2008, Ricotta and Moretti 2011).

According to classical diversity theory, different diversity indices measure different aspects of the partition of abundance between species (Magurran 1988). For this reason, Whittaker (1965) considers that the partition of abundance cannot be adequately summarized by one statistic only, but should be characterized by both species richness and the community 'dominance concentration'. Peet (1974) distinguished two groups of diversity indices: Type I measures, which are most affected by species richness, and Type II measures, which are most sensitive to changes in the abundance of the dominant species. Likewise, a number of authors have introduced several measures that summarize the degree to which abundances are evenly divided among the species of a given community, and called them 'evenness' (see Taillie 1979, Smith and Wilson 1996, Ricotta 2003, Tuomisto 2012, for review).

Although the concept of functional evenness appeared much earlier (e.g., Troussellier and Legendre 1981), Mason et al. (2005) were the first who popularized the distinction of the primary components of functional diversity: functional richness, functional evenness and functional divergence. Functional richness was defined as the amount of niche space filled by the species in the community, functional evenness as the equitability of the distribution of abundance in trait space, whereas functional divergence as deviation in the abundance of the species from the center of gravity in trait space (Mason et al. 2005, Villéger et al. 2008). According to Cornwell et al. (2006) and Villéger et al. (2008), functional richness is estimated as the multidimensional convex hull volume of trait space occupied by the community (but see

Podani 2009, for criticism), functional evenness index is based on the minimum spanning tree, which links all the species in the multidimensional functional space, while functional divergence is defined as the 'species deviance from the mean distance to the center of gravity weighted by relative abundance' (see Mouchet et al. 2010). Ricotta and Moretti (2011) proposed to adopt the 'community-weighted mean trait value' and the Rao coefficient to describe two complementary aspects of community structure, like the mean and the dispersion of functional traits within a given species assemblage.

When dealing with the measurement of functional diversity, the issue of data type becomes a major factor. There are two choices to be made: 1) The data type by which species performance within the assemblage is expressed (presence-absence, abundance-dominance on the ordinal scale, and abundance on the ratio scale) and 2) the data type by which functional traits are measured (nominal with two states, multistate nominal, ordinal, ratio-scale, and mixed, see Podani and Schmera 2006). Most studies derive the traits of species from external databases (e.g. Kattge et al. 2011, Pavoine et al. 2013), whereas others quantify traits using the collected individuals (e.g., Mouillot et al. 2008).

It is often assumed that the combination of functional traits matters in most cases (Ackerly 2004, McGill et al. 2006, Gross et al. 2007, Kühner and Kleyer 2008, Roscher et al. 2012). These statements are made *without* incorporating any direct study of trait combinations, however. The only exception according to our knowledge is the field of aquatic ecology. Heino (2005) calculated the functional richness (number of the combinations of multistate nominal variables), functional diversity (Shannon diversity of the combinations of multistate nominal variables) and functional evenness (Shannon evenness of the combinations of multistate nominal variables), whereas others only counted the combination of unique traits (Poff et al. 2006, Erős et al. 2009, Schmera et al. 2009) for quantifying functional diversity. In addition, evaluating the multiple, rather than pairwise association among the traits is another possibility not yet exhausted in ecology.

In the present paper, indices of functional diversity which rely upon the frequency distribution of trait combinations and measure association of traits and their heterogeneity are introduced. The ideas behind the presented methodology date back to the early 1980's when P. Juhász-Nagy developed his information theoretical models to evaluate scale dependence of plant pattern in a multivariate context. First, the fundamental concepts are summarized, and equations are presented. We shall focus on the simplest situation when each trait can have two states, noting that the essence of the approach can be extended to multistate nominal functional characters as well. The methodology is illustrated by two actual examples, the first demonstrating how the frequency of fires affects trait combinations in Mediterranean woody plant communities. The other study reveals relationships between the functional traits of invertebrate assemblages and two major habitat features in rivers and streams.

Definitions

Let s be the number of species in the assemblage, and let t denote the number of traits. Every trait is a nominal variable with two states, coded by 0 and 1. For example, 0 may refer to a C3 plant while 1 to a C4 plant. Thus, the assemblage is described by the trait data matrix **X** containing t rows and s columns such that x_{ij} is the value of trait i for species j. It is important that none of the traits can be a direct mathematical derivative of other traits presented in the same matrix. That is, we cannot use a trait expressing whether a species is of the C3 type or not (1 versus 0), and another trait to refer to the C4 type in a similar way, because these two states are complements of each other. Of course, we cannot exclude correlated traits because, in general, correlations are natural phenomena of biological collections. The marginal total for row i is denoted by $x_{i.}$ (= the number of 1s), so that the number of zeros in the same row is $s - x_{i.}$. Constant traits, i.e., those with the same value for all species are ignored, because these do not influence diversity measurement anyway.

Between-trait relationships in **X** can be summarized using a $2 \times 2 \times \ldots \times 2$ contingency table of dimensionality t. The number of non-empty cells in this table gives the number of manifested combinations of t traits, to which we shall refer as *combinatorial functional richness* (CFR), in accordance with Mason et al.'s (2005) definitions. We should note that the same concept is also termed as the *number of unique trait combinations* (Erős et al. 2009, Schmera et al. 2009). However, we can extract much more than this simple richness statistic from the contingency table. First of all, we can measure how heterogeneous is the assemblage for each trait and for all traits taken as a whole. The corresponding statistic will be called the *functional heterogeneity* (FH). Furthermore, CFR obviously neglects the frequency distribution of combinations, i.e., how often each trait combination occurs in the community. *Combinatorial functional diversity* (CFD), as defined here, will be based on this distribution. Last but not least, the contingency table is a good starting point to evaluate the multiple association among traits as well (*functional associatum, FA*). We shall use the information theory methods of Juhász-Nagy (1984,1993, see also Juhász-Nagy and Podani 1983) to derive combinatorial functional diversity, which results from a simple mathematical relationship between the other two quantities according to:

$$CFD = FH - FA \tag{1}$$

Functional heterogeneity – FH

The overall heterogeneity of a community in functional traits is illustrated by the following data on a hypothetical species assemblage with 8 species as described by three traits

Trait 1: 1 0 0 0 0 0 0 0
Trait 2: 1 0 1 0 0 1 1 0
Trait 3: 1 1 1 1 1 0 1 1

Trait 1 is fairly homogeneous because it has score 1 only for one species. The situation is similar for trait 3 whose 0 state appears only in one species. These two traits must contribute to overall trait heterogeneity with the same value, because coding is immaterial, i.e., there is an obvious symmetry between trait states 0 and 1. Trait 2 exhibits maximum 'variation' because half the species take state 0, the others take 1. The information theoretical expression summarizing trait heterogeneity in the above matrix is as follows:

$$FH = t \log s - \frac{1}{s} \sum_{i=1}^{t} \{ x_{i.} \log x_{i.} + (s - x_{i.}) \log(s - x_{i.}) \} \qquad (2)$$

a quantity called „information content" by Sneath and Sokal (1973) and used infrequently as a clustering criterion in hierarchical classifications of s objects (see also Orlóci 1969). Mathematically, formula (2) is the pooled entropy of t variables in matrix **X**. In the present context, we shall refer to it as *functional heterogeneity*. Since each trait has two states, the most appropriate base of logarithm is 2 as used in this paper without further notice. As seen, formula (2) is calculated from the marginal row totals of **X**, which are at the same time marginal totals in the $2 \times 2 \times ... \times 2$ t-way contingency table as well. For the above example, we have FH = $3 \times \log 8 - [2 \times 7$ $\log 7 + 2 \times 4 \log 4]/8 = 1.4$. The contributions of individual traits to heterogeneity are 0.2, 1, and 0.2, respectively. Note the far analogy of FH to the pooled variance of t variables measured on the continuous scale. FH takes the value of zero if all variables are constant (zero heterogeneity) and is the maximum (t) when all traits have maximum entropy, i.e., $x_{i.} = s - x_{i.}$ for all i.

To our knowledge, formula 2 has only been used in the context of functional diversity by Trousselier and Legendre (1981) to express 'total information' in tests by strains data matrices in microbial ecology. Those authors did not use it directly, but suggested to divide FH by the maximum possible value to obtain their 'functional evenness index'. They argued that whenever 'all the strains in the sample are described by different combinations of the characters…' the value of their formula 'should be close to maximum'. However, as we show below, the situation is not that simple because, being a pooled measure, FH is absolutely insensitive to the relationships among the t traits – which will not be so with the following two quantities.

Combinatorial functional diversity – CFD

The intuitive meaning of combinatorial functional diversity is best understood based on the following two examples. Assume that we have three binary traits for eight species in a hypothetical community A and for 8 other species in community B, as summarized by the data matrices given below:

Community A

Trait 1: 1 0 1 0 1 1 0 0
Trait 2: 1 0 1 0 1 1 1 1
Trait 3: 1 1 1 1 0 0 1 1

Community B

Trait 1: 1 0 1 0 0 0 0 0
Trait 2: 1 0 1 1 1 1 1 1
Trait 3: 1 1 0 1 1 1 1 1

These assemblages have identical CFR because the same four combinations of traits appear in both. However, in Community A all realized combinations occur twice, while community B is dominated by combination {0,1,1} appearing five times (i.e., five species are identical in all traits), while the other combinations appear only once. Thus, Community A is more diverse than community B in a sense that the realized combinations are dispersed more evenly among species. This can be measured by the well-known entropy function as suggested by Juhász-Nagy (1984) in which the values refer to estimated *probabilities of trait combinations* rather than of single traits:

$$CFD = - \sum_{k=1}^{w} \hat{p}_k \log \hat{p}_k , \qquad (3)$$

where $\hat{p}_k$ is the relative frequency of trait combination k in the data matrix and w is the number of all possible combinations of t binary traits, i.e., $w = 2^t$. In terms of frequencies, f_k, we can also write that

$$CFD = \log s - \frac{1}{s} \sum_{k=1}^{w} f_k \log f_k , \qquad (4)$$

which is called the joint entropy in the t-way contingency table (Juhász-Nagy 1984). The advantage of formula (4) is that it uses frequencies of trait combinations (f_k) so that there is no need to calculate relative frequencies (estimated probabilities in eq. 3). The minimum of CFD is 0, when all species have the same trait combination, while the maximum is log s obtained in the case with $f_k = 1$ for all realized combinations. While CFD may be used as a measure of uncertainty in the assemblage, it is not comparable over different communities due to its obvious dependence on species richness. Therefore, one may wish to remove the effect of s, that is, the number of species detected in the community. Division of CFD by its potential maximum will lead to *combinatorial functional evenness*

$$CFE = CFD / \log s \qquad (5)$$

whose range is [0,1], 0 reflecting presence of a single realized combination and 1 corresponding to the situation when all realized combinations are different.

For the two assemblages above, we have the following results: CFD = log 8 – 4(2 log 2)/8 = 2 for A and log 8 – (5 log 5)/8 = 1.45 for B, while CFE = 2/log 8 = 0.66 for A and 1.45/log 8 = 0.48 for B.

Functional associatum – FA

Consider the following two assemblages of eight species described in terms of 2 traits:

Table 1. Species richness, combinatorial functional diversity and related functions as measured in Mediterranean vegetation

	s	CFR	CFD	CFE	FA	FH	Most frequent combination with number of occurrences	
H1	47	25	4.35	0.784	5.18	9.53	6	00111 10000 1
H2	77	37	4.72	0.752	4.72	9.44	9	00011 10000 0
H3	63	29	4.44	0.742	5.04	9.48	9	00011 10000 0
H4	81	37	4.7	0.742	4.55	9.25	10	00111 10000 1
H5	75	37	4.73	0.759	4.61	9.34	9	00111 10000 1
L1	33	22	4.18	0.829	5.21	9.39	5	00111 10000 1
L2	50	33	4.81	0.852	4.9	9.71	5	00111 10000 1
L3	44	27	4.5	0.824	5	9.5	6	00111 10000 1
L4	64	37	4.81	0.802	4.79	9.6	8	00111 10000 1

Community C

Trait 1: 1 1 1 1 0 0 0 0
Trait 2: 0 0 0 0 1 1 1 1

Community D

Trait 1: 1 1 1 1 0 0 0 0
Trait 2: 1 1 0 0 1 1 0 0

These two assemblages have the same FH and yet different CFD, the latter coming from the fact that the two traits are completely associated in Community C (trait 1 is completely predictable by knowing trait 2, and vice versa) and entirely independent in Community D. Association among binary variables is quantified in information theory as the difference between pooled entropy and joint entropy, giving the (mutual) information between variables. Thus, functional associatum is derived from FH and CFD after rearrangement of equation (1):

$$FA = (t-1)\log s - \frac{1}{s} \sum_{i=1}^{2} \{ x_{i.} \log x_{i.} + (s - x_{i.}) \log(s - x_{i.}) \} +$$

$$+ \frac{1}{s} \sum_{k=1}^{w} f_k \log f_k \qquad (6)$$

This is mathematically the contingency information in the $2 \times 2 \times \ldots \times 2$ t-way contingency table and measures overall association of functional traits.

Computer program

The calculations were made by the modified INPRO program (Podani 1993) developed originally for Juhász-Nagy's models. The application and a short user's guide can be downloaded from http://ramet.elte.hu/~podani.

Actual examples

1. Mediterranean communities affected by different fire regimes

The data include two sets of Mediterranean woody plant assemblages in the eastern Iberian Peninsula (from Verdú

and Pausas 2007, Pausas and Verdú 2008). The first set corresponds to communities living in environments with high fire frequency (H: stands H1 to H5), and the other to communities living in environments under low fire frequency (L: stands L1 to L4). All these communities were well separated from each other (independent) and their size varied largely (from 900 to 70,000 ha); only the most representative woody species (chamaephytes and phanerophytes) of the shrublands and woodlands were considered. The species presence/absence was obtained from local botanical studies (see location map and details in Verdú and Pausas 2007). The number of functional traits is 11, describing the species that vary greatly in number in the stands (from 33 to 81, Table 1). The question is whether functional diversity differs with stands and if there is a global difference between the two sets of assemblages defined by different fire regimes. The traits are as follows: eight traits are two state nominal [spinescence (yes vs no), leaf type (sclerophyllous vs other), leaf habit (deciduous vs evergreen), flower sexuality (hermaphroditic vs unisexual), perianth color (colored vs brownish-greenish), perianth reduction (complete vs reduced), pollinator type (insect vs wind) and seed dispersal mode (endozochorous vs other)] and three quantitative traits are binarized: plant height (= 0 below 1 m, = 1 otherwise), flower size (0 if perianth depth ´ width < 25 mm, 1 otherwise) and seed mass (= 0 below 2 mg, 1 otherwise). Combinatorial functional diversity indices were calculated for each stand. The resulting scores were subjected to standardized (i.e., correlation-based) Principal Components Analysis (Podani 2001) to illuminate relationships within the group of diversity measures. We are aware, of course, that these functions are inter-dependent mathematically (two values in Eq. 1 determine the third), therefore PCA is considered as a visualization tool, rather than a statistical technique of any sort.

Results. It is apparent from the results without any testing (Table 1) that there is no perfect distinction between H and L for all but one measures of functional diversity. Functional evenness, CFE, is the only function consistently higher in low fire frequency sites than in the high fire frequency stands. A potential reason is that while most trait states appear in both H and L communities, the frequency of different trait states differs between H and L. That is, in H stands species with large, hermaphrodite and colored flowers are more

common, while small unisexual and brownish or greenish flowers dominate in L plots, but the two types of species occur in both environments (H and L). The fact that CFE is lower in high fire frequency environments suggests that fire breaks the balance of trait combinations in the community: some combinations tolerate fire better than others. If fire is less frequent, then the community develops various combinations more evenly. As a result, volume in the trait hyperspace is also affected as shown by correlation analysis. Pausas and Verdú (2008) studied the same communities and computed the hyper-volume in the trait space of each community using the convex hull approach. Their standardized volume for each community (called the occupation index) is significantly and positively correlated with CFE (r = 0.813, p = 0.007). That is, low CFE is associated to a reduced trait volume (phenotypic clustering, in H) and high CFE to an expanded trait volume (phenotypic overdispersion, in L). It is also interesting to examine the most frequently detected trait combinations. There are only two different combinations in this list (Table 1), differing only in traits 3 and 11 (leaf habit and seed mass) which vary in the H communities. Considering these most common combinations, traits 1, 2, 6, 7, 8, 9, 10 take always the value of 0, whereas traits 4, 5, 6 always have 1. Thus, fire frequency seems to have little influence on which combination appears most often in the sites.

The Euclidean biplot of PCA (Figure 1) reveals relationships between the six variables considered, including the number of species. Three positive eigenvalues resulted, explaining 67%, 27% and 6% of variation on the first three components and suggesting that a two-dimensional scatter diagram sufficiently demonstrates data structure. The two groups of sites separate the best along the arrow pointing to CFE, which is most positively correlated with functional heterogeneity (FH) and less so with functional associatum (FA, Table 2). Component 1 is largely the contrast between these three functions and another group (CFD, CFR, s) within which correlations are very high. Notable is the high negative

correlation between s and CFE, and the high positive correlation between s and CFR.

2. Invertebrates in running waters

In 2004, the 'Ecological Survey of Surface Waters' program (ECOSURV) was established in Hungary, with the aim to monitor and assess the ecological status of surface waters based on biological (algae, macrophytes, aquatic invertebrates and fish) and hydromorphological and physico-chemical quality descriptors. The aquatic invertebrate data from the ECOSURV project were used in this article. We decided to use presence/absence data, and considered only running-water sites and taxa identified to species level. Records with missing data were omitted. A total of 317 sites with presence–absence data from 444 species fulfilled these criteria (see further details in Schmera and Baur 2011). Sites featuring less than 10 species were omitted from the analysis because these were considered to give insufficient information on the frequency of trait combinations. Finally, 286 sites were retained with the number of species ranging from 10 to 62. The fauna in each site was described in terms of a traits-by-species presence-absence matrix. For all species, we used the following five traits (or, more precisely, feeding groups): shredders, grazers, collectors, predators and others. Trait definitions followed Moog (1995, see Electronic Appendix 1). Note that a given species can be characterized by the presence of several traits simultaneously. Each site was described in terms of two environmental variables: stream/river width (W) and the degree of human impact (I) on stream morphology. This second environmental variable expresses the number of different types of human influence (e.g., bed fixation, bank fixation, stagnation, torrent modification) visually assessed at the sampling sites. Combinatorial functional diversity indices were calculated for the species assemblages of all study sites. The resulting scores, together with the environmental descriptors were subjected to standardized (i.e., correlation-based) Principal Components Analysis to reveal
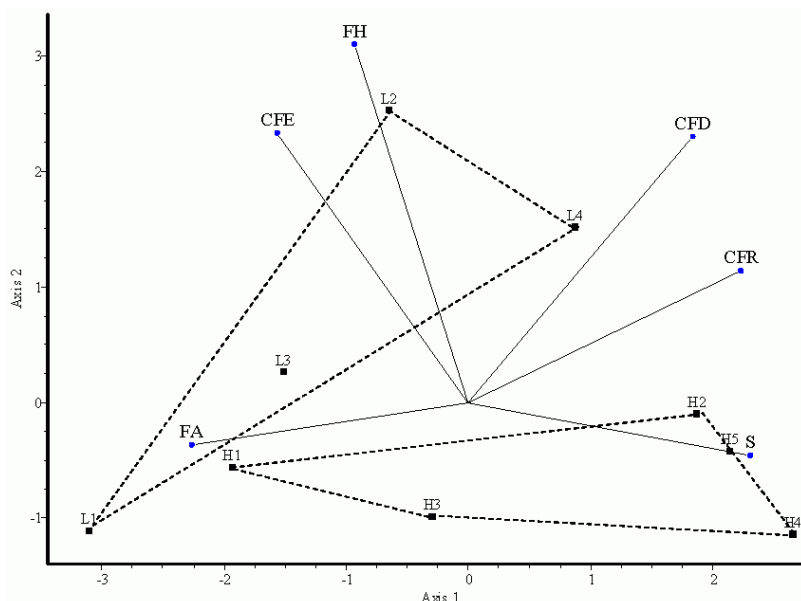

Figure 1. Euclidean biplot from a standardized PCA of 9 vegetation stands from eastern Spain characterized by species richness (*s*) and various indices of combinatorial functional diversity (FH, FA, CFD, CFR, and CFE) using 11 binary functional characters. Convex hulls indicate groups of low fire frequency (L1-L4) and high fire frequency (H1-

Table 2. Correlation matrix between species richness, combinatorial functional diversity measures and two habitat decriptors used in the second illustrative study. Lower semimatrix: Mediterranean plant communities, upper semimatrix: invertebrate stream communities. See text, for abbreviations.

	s	CFD	FH	FA	CFR	CFE	W	I
s		0.10	0.25	0.18	0.63	−0.56	−0.06	0.10
CFD	0.68		0.48	−0.36	0.79	0.72	−0.11	−0.18
FH	−0.43	0.20		0.63	0.49	0.23	−0.19	0.00
FA	−0.89	−0.82	0.39		−0.11	−0.38	−0.11	0.16
CFR	0.88	0.93	−0.12	−0.94		0.15	−0.18	−0.07
CFE	−0.79	−0.12	0.64	0.48	−0.43		−0.02	−0.22
W								0.21

relationships within the group of diversity measures (as in the previous case study) and between these and the two habitat descriptors.

Results. The correlation matrix between the 8 variables is given in Table 2. The Euclidean PCA biplot (Fig. 2) shows the arrangement of the 286 sites for dimensions 1 vs 2 (the associated percentages are 32% and 27%, while the third axis explained 16%), with variable positions overlaid and emphasized by arrows. Since the number of points is relatively high and the number of variables is also larger than in the previous case, the total variance explained on the plane of components 1-2 is 'only' 59%. The sites form a fairly homogeneous cluster, without any observable structure. On the basis of this large set of sites, the interrelationships between functional diversity measures differ from those revealed for the much smaller set of vegetation stands. Whereas FA was highly negatively correlated with *s* for the vegetation stands, their correlation is now positive. On the other hand, the high negative correlation between *s* and CFE, and the high positive correlation between *s* and CFR are apparent in both case studies. Of the habitat descriptors, the degree of human impact (I) is negatively correlated with CFE (human impact decreases the number of realized combinations) and positively with FA (human impact increases trait associations, thereby increasing predictability of trait pattern), while W has the highest negative correlation with CFR. It may be somewhat surprising that in general wider streams support fewer combinations than narrower ones. CFE and W are apparently orthogonal, suggesting that combinatorial functional evenness is not influenced by the width of streams.

Discussion

This paper presents a new conceptual and methodological scheme for the analysis of functional diversity in communities. It is entirely different from existing procedures in that the pooled entropy, mutual information (associatum) and joint entropy are used to express several aspects of diversity simultaneously. In essence, the approach is the extension of the florula diversity concept developed by Juhász-Nagy in the 1970's for analyzing species presence-absence community data. There is also an evident relationship to the taxon-free character-based sequential scheme as described by Orlóci (1991, see also Pillar and Orlóci 2004). The method is applicable to two-state nominal functional characters and examines the number and frequency distribution of character combinations realized in the sample. In this way, we obtain estimates on the multiple association among the variables and on the diversity and evenness of character combinations. The main difference between Juhász-Nagy's original approach and ours is that for presence-absence species data the effects of spatial scale were evaluated (by calculating differ-
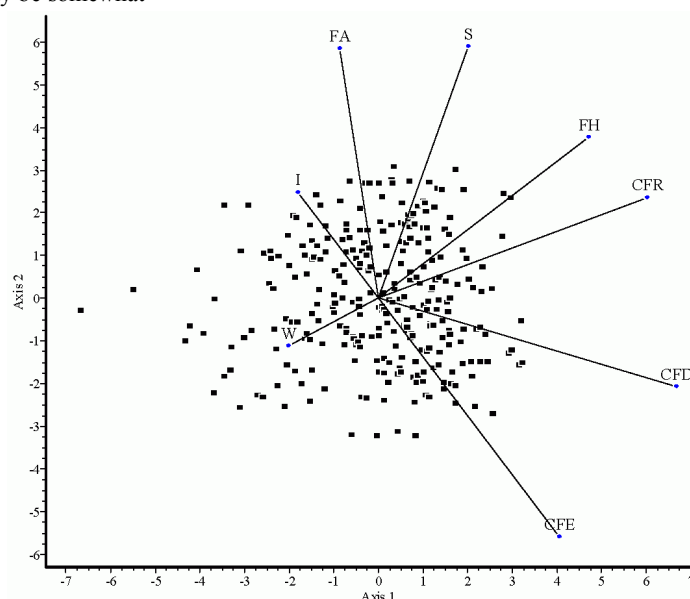

Figure 2. Euclidean biplot from standardized PCA of 286 running water sites from Hungary characterized by species richness (s) and various indices of combinatorial functional diversity (FH, FA, CFD, CFR, and CFE) using 5 binary functional characters and two habitat descriptors (W–stream width, I–degree of human impact).

ent functions for a series of plot sizes), while in the present case external environmental information is contrasted with the results.

Two actual case studies illustrated these contrasts. In Mediterranean woody communities from eastern Spain, we found that the effect of fire frequency is manifested most clearly in functional combinatorial evenness: assemblages prone to high fire frequency exhibited consistently lower functional evenness, and thus lower functional diversity, than assemblages less frequently affected by fire. This observation correlates with other findings derived from convex hull analyses of functional diversity (Pausas and Verdú 2008). Thus, fire seems to have an equalizing effect on functional combinations. In case of the aquatic invertebrates example, combinatorial functional evenness was shown to be correlated mostly with the degree of human impact, expressed in terms of the number of artificial hydromechanical constructs present in a given site. The correlation was negative, showing clearly that habitat destruction equalizes the frequencies of functional combinations. On the other hand, this external variable was uncorrelated to the number of manifested functional combinations showing that habitat changes do not simply eliminate certain combinations. CFR was, however, negatively correlated to the width of streams, perhaps an unexpected result suggesting that narrower streams are often richer in functional combinations than wide ones. Overall these two examples show that disturbances (wildfire in plant communties, human impact on streams) reduce the possible combinations of functional traits, and point out the filtering effect of disturbance in the community assembly.

The interrelationships among the diversity functions used here are best revealed by standardized principal components analysis. Such a comparative study was made with success by Mouchet et al. (2010) on a different group of measures which did not consider trait combinations. The two PCA results we obtained differ from each other in several respects, probably because of the difference in the sample size (9 plots in the vegetation example and 286 plots in the stream example) and in the number of traits (11 and 5 for the vegetation and the stream examples, respectively). Nevertheless, there are agreements as well. We observed already that species richness, s, and CFR are strongly positively correlated in both situations. This may appear obvious: the more species are present the higher number of functional combinations may develop in the community. Increases in the number of species lead to lower functional evenness, however, as seen from the high negative correlations between s and CFE in both assemblage types examined. Thus, fewer species are more likely to produce an even distribution of functional combinations than many species. This observation is in contrast with the findings of Mouchet et al. (2010) who concluded from simulated experiments that their FD indices were strongly or weakly positively correlated with s, while some others were uncorrelated with species richness. No negative relationship was found under any model.

To explain our observations regarding the negative species richness-combinatorial functional evenness relationship, we can put forward the following hypothesis. High evenness means that the points in the abstract functional space are regularly spaced, which is explained by limiting (niche) similarity as the primary driving force in community assembly. That is, biotic interactions dominate in niche differentiation, leading to a community whose constituting elements differ in at least one functional trait. Lower evenness implies that points are more aggregated in the conceptual space, because certain combinations are more frequently represented than others. This result is often attributed to habitat filtering effects which are known to increase clustering of trait values by forcing ecologically similar species to co-occur (Pausas and Verdú 2010). From these findings and our observations it logically follows that the relative importance of limiting similarity and habitat filtering may be decisive not only at the level of species abundances (e.g., Maire et al. 2012), but for species presence-absences as well. Namely, species rich vegetation stands are more the result of habitat filtering effects produced by recurrent fires while the dominant assembly process in species poor stands was the limiting similarity. In case of the stream/river invertebrates, human impact is the filtering factor that decreases functional evenness while increasing species richness, and limiting similarity predominates in species poorer but intact stream sites. We admit, however, that the above hypothesis and interpretation require confirmation by involving more case studies of combinatorial functional evenness.

The discrepancy between the two case studies is most apparent for the multiple association among traits, i.e., functional associatum. For the relatively few sites in the woody vegetation, FA was in an inverse relationship to species richness suggesting that functional traits become less inter-dependent when species richness increases. It has to do probably with the relatively high number of traits. In the streams and rivers example FA is positively correlated with species richness, potentially explained by the fewer number of traits. We must conclude that, in general, results are case-dependent and note that the biological differences in study objects and traits used may also contribute to the discrepancies.

The relative magnitude of the number of species and the number of traits is crucial in the present approach. In the forest example, the s/t ratio at each site ranged from 3 to 7.5 whereas in the case of invertebrates varied from 2 to 12. We feel that sites with a lower ratio are less reliable for measuring functional diversity even at the descriptive level adopted in this study. In the future, it is to be examined how the – otherwise arbitrary – choice of the number of traits influences the results and conclusions in CFD analysis. Also, the distributional properties of these functions require attention, especially if some of them are used as estimators of community level phenomena and are tested for significance as well.

The current analytical models are applicable to two-state nominal traits, but extension to multistate ones is conceivable at least for Eq. 3-5. Abundance data may also be involved in

the calculations in such a way that the abundance data are pooled for each functional combination, and then normalized for the entire data set to unit sum. This produces relative frequencies for species combinations which can be evaluated by the Shannon diversity function in the traditional manner. Heino (2005) already used such an approach to generalize Shannon diversity and evenness to combinations („groups", see his Table 3.)

Our study demonstrated that functional diversity has components additional to those originally suggested (richness, evenness and divergence, Mason et al. 2005). These components involve quantification of trait combinations, rather than distances, volumes and lengths measured in the conceptual functional space. We introduced a new equations termed as combinatorial functional diversity, functional heterogeneity and functional associatum and showed that these equations are inter-related. Based on these, our functions supplement already existing approaches. Moreover, as these information theoretical equations are mathematically linked to each other, our approach presented here provides a rather complex framework for analyzing species by trait matrices in communities. The presented methodology may be used in in the future for examining the effect of trait selection before analysis, to mention only one example. Depending on associations, traits whose removal does not affect much the results are probably less influential in determining functionality of the assemblage than those characters which cannot be eliminated from the data without significant changes in the results. This hypothesis may also be tested in future studies.

References

Ackerly, D.D. 2004. Functional strategies of chaparral shrubs in relation to seasonal water deficit and disturbance. *Ecol. Monogr.* 75: 25-44.

Cornwell, W.K., Schwilk, D.W. and Ackerly, D.D. 2006. A trait-based test for habitat filtering: convex hull volume. *Ecology* 87: 1465-1471.

de Bello, F., Lavergne, S., Meynard, C.N., Lepš, J. and Thuiller, W. 2010. The partitioning of diversity: showing Theseus a way out of the labyrinth. *J. Veg. Sci.* 21: 992-1000.

Erős, T., Heino, J., Schmera, D. and Rask, M. 2009. Characterising functional trait diversity and trait-environment relationships in fish assemblages of boreal lakes. *Freshw. Biol.* 54: 1788-1803.

Gross, N., Suding, K.N. and Lavorel, S. 2007. Leaf dry matter content and lateral spread predict response to land use change for six subalpine grassland species. *J. Veg. Sci.* 18: 289-300.

Heino, J. 2005. Functional biodiversity of macroinvertebrate assemblages along major ecological gradients of boreal headwater streams. *Freshw. Biol.* 50: 1578-1587.

Juhász-Nagy, P. 1984. Spatial dependence of plant populations. Part 2. A family of new models. *Acta Bot. Acad. Sci. Hung.* 30: 363-402.

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

Juhász-Nagy, P. and Podani, J. 1983. Information theory methods for the study of spatial processes and succession. *Vegetatio* 51:129-140.

Kattge, J. et al. 2011. TRY – a global database of plant traits. *Global Change Biol.* 17: 2905-2935.

Kühner, A. and Kleyer, M. 2008. A parsimonious combination of functional traits predicting plant response to disturbance and soil fertility. *J. Veg. Sci.* 19: 681-692.

Lavorel, S., Grigulis, K., McIntyre, S., Williams, N.S.G., Garden, D., Dorrough, J., Berman, S., Quetier, F., Thebault, A. and Bonis, A. 2008. Assessing functional diversity in the field - methodology matters! *Funct. Ecol.* 22: 134-147.

Magurran, A.E. 1988. *Ecological Diversity and its Measurement.* Princeton University Press, Princeton.

Maire, V., Gross, N., Wirth, C., Pontes, L.D.S., Proulx, R., Börger, L., Soussana, J-F. and Loualt, F. 2012. Habitat-filtering and niche differentiation jointly determine species relative abundance within grassland communities along fertility and disturbance gradients. *New Phytol.* 196: 497-509.

Mason, N.W.H., Mouillot, D., Lee, W.G. and Wilson, J.B. 2005. Functional richness, functional evenness and functional divergence: the primary components of functional diversity. *Oikos* 111: 112-118.

McGill, B.J., Enquist, B.J., Weiher, E. and Westoby, M. 2006. Rebuilding community ecology from functional traits. *TREE* 21: 178-185.

Moog, O. 1995. *Fauna Aquatica Austriaca.* Lieferung Mai/95. Wasserwirtschaftskataster, Bundesministerium für Land- und Forstwirtschaft, Wien.

Mouchet, M., Villéger, S., Mason, N.W.H. and Mouillot, D. 2010. Functional diversity measures: an overview of their redundancy and their ability to discriminate community assembly rules. *Funct. Ecol.* 24: 867-876.

Mouillot, D., Culioli, J.M., Pelletier, D. and Tomasini, J.A. 2008. Do we protetct biological originality in protected areas? A new index and an application to the Bonifacio Strait Natural Reserve. *Biol. Cons.* 141: 1569-1580.

Orlóci, L. 1969. Information theory models for hierarchic and non-hierarchic classifications. In: Cole, A.D. (ed.), *Numerical Taxonomy.* Academic, London. pp. 148-164.

Orlóci, L. 1991. On character-based plant community analysis: choice, arrangement, comparison. In: Feoli, E. and Orlóci, L. (eds.), Computer Assisted Vegetation Analysis. Kluwer, Dordrecht. pp. 81- 86.

Pausas, J.G. and Verdú, M. 2008. Fire reduces morphospace occupation in plant communities. *Ecology* 89: 2181-2186.

Pausas, J.G. and Verdú, M. 2010. The jungle of methods for evaluating phenotypic and phylogenetic structure of communities. *BioScience* 60: 614-625.

Pavoine, S., Gasc, A., Bonsall, M.B. and Mason, N.W.M. 2013. Correlation between phylogenetic and functional diversity: mathematical artefacts or true ecological and evolutionary process? *J. Veg. Sci,* 24: 781-793.

Peet, R.K. 1974. The measurement of species diversity. *Ann. Rev. Ecol. Syst.* 5: 285-307.

Petchey, O.L. and Gaston, K.J. 2006. Functional diversity: back to basics and looking forward. *Ecol. Lett.* 9: 741-758.

Pillar, V.D. and Orlóci, L. 2004. *Character-Based Community Analysis: The Theory and an Application Program*. Electronic Edition available at http://ecoqua.ecologia.ufrgs.br.

Podani, J. 1993. SYN/TAX 5 User's Manual. Scientia, Budapest

Podani J. 2001. SYN-TAX 2000. *Computer programs for data analysis in ecology and systematics*. User's Manual. Scientia, Budapest.

Podani, J. 2009. Convex hulls, habitat filtering, and functional diversity: mathematical elegance versus ecological interpretability. *Community Ecol.* 10: 244-250.

Podani, J. and Schmera, D. 2006. On dendrogram-based measures of functional diversity. *Oikos* 115: 179-185.

Poff, N.L., Olden, J.D., Vieira, N.K.M., Finn, D.S., Simmons, M.P. and Kondratieff, B.C. 2006. Functional trait niches of North American lotic insects: traits-based ecological applications in light of phylogenetic relationships. *J. North Amer. Benthological Soc.* 25: 730-755.

Ricotta, C. 2003. On parametric evenness measures. *J. Theor. Biol.* 222: 189-197.

Ricotta, C. and Moretti, M. 2011. CWM and Rao's quadratic diversity: a unified framework for functional ecology. *Oecologia* 167: 181-188.

Roscher, C., Schumacher, J., Gubsch, M., Lipowsky, A., Weigelt, A., Buchmann, N., Schmid, B. and Schulze, E.D. 2012. Using plant functional traits to explain diversity–productivity relationships. *PLoS ONE* 7(5): e36760. doi:10.1371/journal.pone.0036760

Schmera, D. and Baur, B. 2011. Testing a typology system of running waters for conservation planning in Hungary. *Hydrobiologia* 665: 183-194.

Schmera, D., Erős, T. and Podani, J. 2009. A measure for assessing functional diversity in ecological communities. *Aquat. Ecol.* 43: 157-167.

Smith, B. and Wilson, J.B. 1996. A consumer's guide to evenness indices. *Oikos* 76: 70-82.

Taillie, C. 1979. Species equitability: a comparative approach. In: Grassle, J.F., Patil, G.P., Smith, W.K. and Taillie, C. (eds.), *Ecological Diversity in Theory and Practice*. International Cooperative Publishing House, Fairland, MD, pp. 51-62.

Troussellier, M. and Legendre, P. 1981. A functional diversity index for microbial ecology. *Microbial. Ecol.* 7: 283-292.

Tuomisto, A. 2012. An updated consumer's guide to evenness and related indices. *Oikos* 121: 1203-1218.

Verdú, M. and Pausas, J.G. 2007. Fire drives phylogenetic clustering in Mediterranean Basin woody plant communities. *J. Ecol.* 95: 1316-323

Villéger, S., Mason, N.W.H. and Mouillot, D. 2008. New multidimensional functional diversity indices for a multifaceted framework in functional ecology. *Ecology* 89: 2290-2301.

Whittaker, R.H. 1965. Dominance and diversity in land plant communities. *Science* 147: 250-260.

Electronic Appendix 1

The definitions of or functional feeding groups of the invertebrates used in this study. The file may be downloaded from the web site of the publisher at www.akademiai.com.

Electronic Appendix 1. The definitions of traits of invertebrates in running waters.

Trait	Verbal definition	Mathematical definition
shredder	Species feeding on fallen leaves, plant tissue, coarse particulate organic matter (CPOM), leaves of aquatic plants, algae and cells of aquatic plants or on woody debris.	The sum of SHR, MIN and XYL values in Moog (1995) is larger than zero
grazer	Species feeding on epilithic algal tissues, biofilm, partially particulate organic matter.	The GRA value in Moog (1995) is larger than zero.
collectors	Active filter feeders, passive filter-feeders and detritus feeders.	The sum of AFIL, PFIL and DET values in Moog (1995) is larger than zero.
predators	Predators and parasites.	The sum of PRE and PAR values in Moog (1995) is larger than zero.
others	Cannot be classified into the scheme.	The OTH value is larger than zero.

Reference

Moog, O. 1995. *Fauna Aquatica Austriaca*. Lieferung Mai/95. Wasserwirtschaftskataster, Bundesministerium für Land- und Fortwirtschaft, Wien.