Title

Review and performance evaluation of trait-based between-community dissimilarity measures

Author details

Attila Lengyel[1*] & Zoltán Botta-Dukát[2*]

*Centre for Ecological Research, Institute of Ecology and Botany, Alkotmány u. 2-4., H-2163 Vácrátót, Hungary

[1] corresponding author, lengyel.attila@ecolres.hu

[2] botta-dukat.zoltan@ecolres.hu

Abstract

1. In the recent years a variety of indices have been proposed with the aim of quantifying functional dissimilarity between communities. These indices follow different approaches to account for between-species similarities in the calculation of community dissimilarity, yet they all have been proposed as straightforward tools.
2. In this paper we reviewed the trait-based dissimilarity indices available in the literature, contrasted the approaches they follow, and evaluated their performance in terms of correlation with an underlying environmental gradient using individual-based community simulations with different gradient lengths. We tested how strongly dissimilarities calculated by different indices correlate with environmental distances. Using random forest models we tested the importance of gradient length, the choice of data type (abundance vs. presence/absence), the transformation of between-species similarities (linear vs. exponential), and the dissimilarity index in the predicting correlation value.
3. We found that many indices behave very similarly and reach high correlation with environmental distances. There were only a few indices (e.g. Rao's D_Q, and representatives of the nearest neighbour approach) which performed regularly poorer than the others. By far the strongest determinant of correlation with environmental distance was the gradient length, followed by the data type. The dissimilarity index and the transformation method seemed not crucial decisions when correlation with an underlying gradient is to be maximized.
4. *Synthesis*: We provide a framework of functional dissimilarity indices and discuss the approaches they follow. Although, these indices are formulated in different ways and follow different approaches, most of them perform similarly well. At the same time,

sample properties (e.g. gradient length) determine the correlation between trait-based dissimilarity and environmental distance more fundamentally.

Keywords

beta diversity, dissimilarity index, distance metric, community ecology, functional traits

Abbreviations

CDF = cumulative distribution function, CWM = community-weighted mean, FDissim = functional dissimilarity, VIS = variable importance score

Introduction

Understanding and explaining the variation of living communities along dimensions of space and time have been in the focus of ecological research ever since. The widely applied scheme by Whittaker (1960, 1972) to tackle questions of different aspects of community variation divides community diversity into alpha (within-community), beta (between-community) and gamma (across-community) components. It is no exaggeration to say that among these three, beta diversity sparked the most controversy due to the multitude of ways how it can be formulated (Tuomisto 2010a,b, Anderson et al. 2011, Podani & Schmera 2011, Baselga & Leprieur 2015). One of the most popular approaches to beta diversity builds upon quantification of variation between pairs of communities using dissimilarity indices (Anderson et al. 2006, Legendre & De Cáceres 2013, Ricotta 2017). A broad spectrum of such dissimilarity indices are available for many specific purposes providing elementary tools for different fields of ecology and beyond (see reviews by Legendre & Legendre 1998, Podani

2000). Nevertheless, choosing from such many options requires a more or less subjective decision from the researcher which may affect the final result of the analysis. Comparative reviews of dissimilarity indices (Faith et al. 1987, Koleff et al. 2003) and evaluations of effects of methodological decisions (Lengyel & Podani 2015) are inevitably helpful in making these decisions.

The most popular, yet not exclusive, interpretations of diversity for long time considered species as variables which are unrelated with each other. In the last two decades, however, the functional approach to ecological questions gained unprecedented attention (Díaz & Cabido 2001, McGill et al. 2006). This approach relies on the fact that species are not all maximally different from each other, rather they can be considered related with respect to similarities in their traits thought to represent their roles in ecosystems (Violle et al. 2007). The need for explicitly accounting for between-species relatedness generated a wave of methodological improvements that introduced new methods in the calculation of diversity. Next to a lively scientific discussion on how functional alpha diversity can be appropriately quantified (Mason et al. 2005, Petchey & Gaston 2006, Villéger et al. 2008, Mouchet et al. 2010), suggestions were made also for the expression of functional beta diversity (Swenson 2011, Botta-Dukát 2018, Chao et al. 2019). Among them, a large variety of indices for calculating dissimilarity between pairs of communities on the basis of the traits of their species have been proposed (e.g. Ricotta & Burrascano 2008, Cardoso et al. 2014, Ricotta & Pavoine 2015). Although these indices have been introduced as straightforward measures for revealing between-community dissimilarity on the basis of traits, they have very different concepts behind, and we still lack a comparative review of them.

In this paper we aim to provide an overview and a conceptual framework for the pairwise functional dissimilarity (hereafter called FDissim) measures available in the literature to our best knowledge. We start with a (1) short overview of the concept and indices of ecological

(dis-)similarity without accounting for relatedness of species, then (2) we review and classify FDissim indices according to their conceptual basis, and (3) we test the performance of FDissim indices.

Short overview of taxon-based (dis-)similarity methods

Most FDissim measures are generalizations of simple indices which were originally designed for expressing dissimilarity based on species composition (that is, omitting similarities between species). We start the review of trait-based (dis-)similarity measures with a brief summary of these species-based indices. Then, we present a framework of approaches including several families of trait-based dissimilarity indices.

Species-based indices

Most indices can be written in either similarity (s) or dissimilarity (d=1-s) form but when we do not see necessary to specify the form, we call them 'resemblances'. In the case of presence/absence data, these indices are based on the well-known 2×2 contingency table whose cells represent the number of species shared (denoted by a), as well as the number of species occurring only in one of the communities (b and c). The fourth cell of the contingency table quantifying the number of shared absences is disregarded by these indices and rarely used in ecological analyses (but see Tamás et al. 2001). All these indices agree that they express similarity as the proportion of shared diversity to total diversity. Hence, all of them range between 0 and 1. In the case of presence/absence data the number of shared species, a, in the numerator stands for shared diversity for all indices, while the denominators are different. In the Sørensen index (s_S) the denominator is the arithmetic mean of the species numbers of the two communities, in Ochiai index (s_O) it is their geometric mean, in Kulczynski (s_K) it is their harmonic mean, while in Simpson index (s_{Si}) it is the richness of the

species poorer community. If the two communities are equally species-rich, then these indices are equal, otherwise $s_S < s_O < s_K < s_{Si}$. In the Jaccard index (s_J), the denominator is the total number of species in the two communities, while in Sokal & Sneath index (s_{SS}) species occurring in a single community are taken into account with double weight. There is a direct and monotonic relationship between Jaccard, Sørensen, and Sokal & Sneath indices (see Appendix S1). Table 1 summarizes the similarity and dissimilarity forms of the above indices.

For abundance data, the resemblance of two communities is derived from the summation of species-wise differences, with the simplest interpretation being the Euclidean and the Manhattan distances, respectively:

Eq. 1. $$d_{Euclidean} = \sqrt{\sum_{i=1}^{S_{jk}} \left(x_{ij} - x_{ik}\right)^2}$$

Eq. 2. $$d_{Manhattan} = \sum_{i=1}^{S_{jk}} \left|x_{ij} - x_{ik}\right|$$

where x_{ij} and x_{ik} are the abundance of species i in communities j and k, S_{jk} is the total number of species in j and k. For both indices, the minimum is 0 but the maximum of Euclidean distance is the square-root of the sum of squared abundances, while for Manhattan distance the maximum is the sum of abundances. Obviously, their dependence on total abundance makes these index values difficult to compare across samples; therefore, indices including a standardization have become more popular in ecological studies. The standardization is possible in several ways. The first option is to standardize raw species contributions to between-community dissimilarity (x_{ij}-x_{ik}), and then to sum them. Therefore, each species-level difference in abundance should be divided by a scaling factor in a way that maximal species-level difference is 1 and this difference is maximal if species present only one of the compared communities. Summing x_{ij} and x_{ik} in the denominator satisfies this requirement and gives a well-known distance measure, the Canberra index:

Eq. 3. $$d_{Canberra} = \sum_{i=1}^{S_{jk}} \frac{|x_{ij}-x_{ik}|}{(x_{ij}+x_{ik})}$$

However, Canberra index still ranges between 0 and S_{jk}. According to Ricotta & Podani (2017), the normalized Canberra index can be derived by unweighted averaging of species contributions:

Eq. 4. $$d_{NCanberra} = \frac{1}{S_{jk}} \sum_{i=1}^{S_{jk}} \frac{|x_{ij}-x_{ik}|}{(x_{ij}+x_{ik})}$$

Alternatively, species-level differences can be divided by max(x_{ij}, x_{ik}). It also results unity, if species occur only either of the plots. Ricotta & Podani (2017) called this modified Canberra index, whose normalized version follows:

Eq. 5. $$d_{NMCanberra} = \frac{1}{S_{jk}} \sum_{i=1}^{S_{jk}} \frac{|x_{ij}-x_{ik}|}{\max\ (x_{ij},x_{ik})}$$

Calculating from binary data, both normalized Canberra and normalized modified Canberra result in Jaccard dissimilarity.

A different way of standardization is possible if raw species-level differences are summed and divided by the sum of their theoretical maxima. In this case, the denominator can follow the logic of Canberra index, thus leading to the Bray-Curtis index:

Eq. 6. $$d_{BC} = \frac{\sum_{i=1}^{S_{jk}} |x_{ij}-x_{ik}|}{\sum_{i=1}^{S_{jk}} (x_{ij}+x_{ik})}$$

Analogously with the normalized modified Canberra index, instead of the sum, the denominator may contain the maximum of abundance, resulting in the formula known as Marczewski-Steinhaus index:

Eq. 7. $$d_{MS} = \frac{\sum_{i=1}^{S_{jk}} |x_{ij}-x_{ik}|}{\sum_{i=1}^{S_{jk}} \max\ (x_{ij},x_{ik})}$$

Worth to note that Bray-Curtis and Marczewski-Steinhaus indices calculated on presence/absence data return the values of Sørensen index and Jaccard index in dissimilarity form, respectively. Moreover, several abundance-based indices can be expressed if we generalize *a*, *b*, and *c* quantities used during the definition of indices for presence/absence data (Tamás et al. 2001).

Eq. 8. $$a' = \sum_{i=1}^{S_{jk}} \min \left(x_{ij}, x_{ik}\right)$$

Eq. 9. $$b' = \sum_{i=1}^{S_{jk}} \left(\max \left(x_{ij}, x_{ik}\right) - x_{ij}\right)$$

Eq. 10. $$c' = \sum_{i=1}^{S_{jk}} \left(\max \left(x_{ij}, x_{ik}\right) - x_{ik}\right)$$

Substituting *a*, *b* and *c* with *a'*, *b'* and *c'* into the formula of Sørensen index gives Bray-Curtis, and doing so with Jaccard index results in the Marczewski-Steinhaus. Abundance versions of all other presence/absence indices can be created in the same manner.

A classification of FDissim indices

FDissim indices incorporate trait information into the calculation of dissimilarity in different ways. The simplest solution is when summary statistics or distributions are calculated for the two communities and a measure of distance or segregation is calculated between them. We call this the *summary-based class*, and in our review, we include two approaches within this, the *typical value approach* and the *distribution-based approach*. In the second class we include indices which utilize a symmetrical species by species (dis-)similarity matrix and link it directly through matrix operations with the compositional matrix. We call this the *dissimilarity-based class* which includes the *probabilistic*, the *ordinariness-based*, the *diversity partitioning*, and the *nearest neighbour approaches*. The third class includes methods which make use of between-species (dis-)similarities for classification of species;

therefore, we call it the *classification-based class*. The classification either transforms the original structure of the dissimilarity matrix into discrete groups of species which can be used as functional types, or expresses dissimilarities in a form of a tree-graph where between-species dissimilarities are organized in an inclusive hierarchy. This is a widespread approach for accounting for phylogenetic relatedness, since phylogenies are commonly summarized in the form of cladograms. Such methods heavily rely on the algorithm chosen for the classification, including the decisions about the number of clusters and the method for breaking tied values. Examples are provided by Hérault & Honnay (2007), Nipperess et al. (2010), and Cardoso et al. (2014), while a review is available by Pavoine (2016). As there is no general recommendation for the classification method, we omit this class from the framework detailed below and the comparative test. The classification of trait-based dissimilarity indices and their main properties are summarized on Table 2.

Typical value approach

Indices following this approach represent each community with a typical trait value, and calculate a distance metric between them. The most commonly applied typical trait value is the community weighted mean (CWM; Garnier et al. 2004). The rationale behind the CWM can be linked with the mass ratio hypothesis (Grime 1998) stating that the effect of species on ecosystem functioning is proportional to their relative abundances. Although, several issues emerged regarding its limited applicability in statistical inference (Hawkins et al. 2017, Peres-Neto et al. 2017, Zeleny 2018) and its negligence of within-community variation (Muscarella & Uriarte 2016), difference in CWM is still considered a reliable indicator of robust changes in trait composition induced by selective forces like environmental matching or succession (De Bello et al. 2007, 2013, Kleyer et al. 2012). Ricotta et al. (2015) investigated the relatedness of the distance between CWMs with the probabilistic approach (see therein) and showed its applicability on phylogenetic data. Due to its tolerable requirements for

computational capacity, Lengyel et al. (2020) used the Euclidean distance between trait CWMs of phytosociological relevés for the trait-based numerical classification of grasslands of Poland with a sample size of 6985 sites and 885 species. Another advantage of this method is its Euclidean property. Besides the community-weighted mean, other typical values, e.g. the median or the mode, might be considered depending on the scaling of the trait variable and on specific research aims.

Distribution-based approach

Instead of typical values, the distribution of trait values is considered a more reliable representative of the trait composition and variability of a community. Continuous distributions can be defined by a density function, while discrete distributions by the probabilities of the possible values, while both types can be characterized by a cumulative distribution function (CDF). A useful analogue of the distance between typical values might be distance between discrete distributions, density functions or CDFs.

If data is available on intraspecific trait variation, trait values forms a continuous distribution. First, separate density functions have to be fitted within each species. Then, density function of this community-level distribution can be calculated as weighted sum of species level density functions (Carmona, de Bello, Mason, & Lepš, 2016). If such data is not available, we can use relative abundances as estimates of probabilities of the corresponding trait values. Pairs of trait values and their probability form a discrete distribution.

Similarity of density functions can be measured by their overlap (see Appendix S2 for overview of overlap measures). Overlap functions between within-species trait distributions has already been proved useful in the quantification of between-species niche segregation (MacArthur & Levins 1967, Mouillot et al. 2005) or trait-based dissimilarity of species (Lepš

et al. 2006, De Bello et al. 2013). Nevertheless, they are perfectly applicable to the community level as well.

Gregorius et al. (2003) proposed an index called *delta* for the quantification differences between discrete trait distributions. Delta is the minimal sum of frequencies shifted from one trait state to another trait state, weighted by the differences between the respective states. Minimizing the sum of shifted frequencies is known in linear programming as the transportation problem (Hitchcock 1941). Due to its relatively high computational demand, it is unfeasible for large compositional and trait data matrices typically used in ecological research, therefore, we exclude this index from our comparison.

Difference between two CDFs can be calculated at each possible trait values (i.e. not only the observed ones), then the sum of them can be used as a trait-based dissimilarity measure. In Appendix S3 we introduce the distance between CDFs in more detail.

Maximally distinct communities

Species-based dissimilarities, except Euclidean, Manhattan and (non-normalized) Canberra distances, equal unity, which is their maximum, when the two compared communities do not share any species. In this context, we could call such communities maximally distinct.

However, when traits are considered, two communities can be similar, even if they do not share any species. For example, if all species of community *A* is replaced by a similar species in community *B*, the two communities have no shared species, but from functional point of view, they are similar. In this context, two communities are maximally distinct, when similarity of any species from the first community is zero to any species in the other community. It is a desirable property for a functional similarity index to take the value 0 if and only if the two compared communities are maximally distinct.

Probabilistic approach

This approach can be traced back to the diversity framework proposed by Rao (1982), and recently extended by Pavoine & Ricotta (2014). Rao's within community diversity is defined as the expected dissimilarity between two randomly drawn individuals from a single community:

Eq. 11. $$Q(\mathbf{p}) = \sum_i \sum_j p_i p_j \delta_{ij}$$

where p_i is the relative abundance of the *i*th species in the community and δ_{ij} is the dissimilarity between species *i* and *j*. This has become a widely used index of functional alpha diversity (Botta-Dukát 2005). Likewise, a between-community component of diversity, $Q(\mathbf{p},\mathbf{q})$, can be defined as the dissimilarity between two random individuals, each selected from different communities:

Eq. 12. $$Q(\mathbf{p},\mathbf{q}) = \sum_i \sum_j p_i q_j \delta_{ij}$$

Between community diversity can be expressed using within community diversity of the two original communities (Q(**p**) and Q(**q**)) and the community with mean relative abundances; $Q\left(\frac{\mathbf{p}+\mathbf{q}}{2}\right)$.

Eq. 13. $$2Q\left(\frac{\mathbf{p}+\mathbf{q}}{2}\right) = 2\sum_i \sum_j \frac{p_i+q_i}{2}\frac{p_j+q_j}{2}\delta_{ij} = \frac{1}{2}\sum_i \sum_j \left[p_i p_j + q_i q_j + 2p_i q_j\right]\delta_{ij} = \frac{Q(\mathbf{p})+Q(\mathbf{q})}{2} + Q(\mathbf{p},\mathbf{q})$$

Subtracting mean within community diversity from the between community diversity leads to Rao's dissimilarity (also called *DISC*):

Eq. 14. $$D_Q = \sum_i \sum_j p_i q_j \delta_{ij} - \frac{\sum_i \sum_j p_i p_j \delta_{ij} + \sum_i \sum_j q_i q_j \delta_{ij}}{2} = \sum_i \sum_j p_i q_j \delta_{ij} - \frac{Q(\mathbf{p}) + Q(\mathbf{q})}{2} = 2Q\left(\frac{\mathbf{p}+\mathbf{q}}{2}\right) - Q(\mathbf{p}) - Q(\mathbf{q})$$

where p_i and q_i are the relative abundances of species i in the two communities. Champely and Chessel (2002) proved that if δ has squared Euclidean property, Rao quadratic entropy is concave function, i.e. $Q\left(\frac{\mathbf{p}+\mathbf{q}}{2}\right)$ is higher than or equal to mean of $Q(\mathbf{p})$ and $Q(\mathbf{q})$. Thus under this condition, $D_Q \geq 0$. If $0 \leq \delta_{ij} \leq 1$, $\sum_i \sum_j p_i q_j \delta_{ij}$, which is the weighted average of between-species distances, also has to be within this range. Therefore, $0 \leq D_Q \leq 1$. However, D_Q may be much less than 1, even if the two communities are completely distinct, when $Q(\mathbf{p})$ and $Q(\mathbf{q})$ are high. Therefore, Pavoine & Ricotta (2014) suggested dividing D_Q by its theoretical maximum (see equations 3 and 4 in Pavoine & Ricotta 2014). They recognized that the resulting indices are representatives of a broader family of indices, hereafter called *dsimcom*, which are actually the implementations of Rao's between-community and within-community components of diversity into the similarity formulae designed for presence/absence data. For this index, it is necessary to introduce the similarity between species, ε_{ij}=1- δ_{ij}. The expected similarity between individuals of different communities, $A = \sum_i \sum_j p_i q_j \varepsilon_{ij}$ is taken analogous with the shared diversity, a, according to the parameters of the similarity indices for presence/absence data disregarding species properties, while the expected similarities within communities ($A + B = \sum_i \sum_j p_i p_j \varepsilon_{ij}$ and $A + C = \sum_i \sum_j q_i q_j \varepsilon_{ij}$) are analogous with the species numbers (a+b, a+c). In this way, Pavoine & Ricotta (2014) presented formulae following the Sokal & Sneath, Jaccard, Sørensen, and Ochiai indices. Additionally, a formula analogous with Whittaker's effective species turnover (β=γ/α-1; Whittaker 1972, Tuomisto 2010a) is suggested for two communities, which in similarity form is shown to be identical with the overlap index of Chiu et al. (2014). In this formulation γ=A+B+C and α=(2A+B+ C)/2. Pavoine & Ricotta (2014) showed that members of the

dsimcom family provide meaningful values also if absolute abundances, percentage values or binary occurrences are used instead of relative abundances.

When ε_{ij} contains taxonomical similarities, its off-diagonal elements are 0, and $A=a$, $B=b$, and $C=c$.

Worth to note the inherent link between D_Q and *CWMdis* on the basis of the geometric interpretation by Pavoine (2012) and Ricotta et al. (2015). Pavoine (2012) showed that if between-species dissimilarities are in the form $\delta_{ij}=(d_{ij}^2)/2$ and d_{ij} is Euclidean embeddable, D_Q is half the squared Euclidean distance between the centroids of two communities – a function monotonically related with *CWMdis*, the simple Euclidean distance between centroids of communities. As Ricotta et al. (2015) argue, if species relatedness is only described by a dissimilarity matrix, which is the common case in phylogenetic analyses, species can be mapped into a principal coordinate analysis ordination using d_{ij}. Given the Euclidean embeddable property of d_{ij}, this ordination should produce S-1 or fewer ordination axes, all with positive eigenvalues. Ordination scores for species can be used as traits, and therefore, centroids of communities, and (squared) Euclidean distances between communities can be calculated. In the special case when between-species dissimilarities are Euclidean distances, D_Q must be equal with the Euclidean distance between the weighted averages of traits, that is, *CWMdis*.

It is also notable that Swenson et al. (2011) and Swenson (2011) use the quantity $Q(\mathbf{p}, \mathbf{q})$ as a standalone index of pairwise beta diversity and call it D_{pw} or "Rao's D". The latter name is misleading since Rao (1982) himself noted with D_{ij} the *DISC* (or D_Q) index. $Q(\mathbf{p}, \mathbf{q})$ measures dissimilarity between two communities but the dissimilarity of a community from itself is not zero. Swenson (2011) also presents a standardized version of $Q(\mathbf{p}, \mathbf{q})$ under the name Rao's H. With this formula the dissimilarity of a community to itself is scaled to 1, however, its

transformation to a meaningful scale where each community has dissimilarity value zero towards itself is not elaborated. Due to this drawback, we do not consider these indices in our review of functional dissimilarity measures.

Schmidt et al. (2017) proposed probabilistic indices with weighted and unweighted versions for expressing community similarity on the basis of taxa interaction networks (called *TINA*, taxa interaction-adjusted) and phylogenetic relatedness (*PINA*, phylogenetic interaction-adjusted). *TINA* and *PINA* differ only in what type of data the interaction matrix contains. Notably, the functional formula of weighted *TINA* is identical with the Ochiai version of *dsimcom*. However, the unweighted *TINA*, abbreviated *TU*, is not a special case of *TINA*, which we consider an inconsistency. Therefore, we did not include *TU* as a separate index.

Ordinariness-based approach

With respect to functional alpha diversity, Leinster & Cobbold (2012) introduced the concept of species ordinariness defined as the weighted sum of relative abundances of species similar to a focal species within the same community, or in other words, the expected similarity of an individual of the focal species and an individual chosen randomly from the same community.

According to Ricotta & Pavoine (2015) it is straightforward to replace abundances with ordinariness values in the species-based (dis-)similarity indices. Following this concept, Ricotta & Pavoine (2015) introduced a new family of trait-based similarity measures called *dissABC*. *dissABC* applies the schemes of Jaccard, Sørensen, Ochiai, Kulczynski, Sokal & Sneath, and Simpson indices. Either relative or absolute abundances can be chosen as input values. Species ordinariness values can be calculated either with respect to the pooled species list of the two communities under comparison, or to the total species list of the data matrix.

For species-based analyses, Ricotta & Podani (2017) suggested a general formula of distance measures in which community dissimilarity is calculated by the weighted averaging of species-level differences in abundance. From this formula, a normalized Canberra distance, Bray-Curtis distance, Marczewski-Steinhaus index, and an evenness-based dissimilarity index (Ricotta 2018) can be derived. According to Pavoine & Ricotta (2019), replacing species abundances with species ordinariness values, a meaningful dissimilarity index can be designed, which is called *generalized_Tradidiss*. Additionally, this index contains a factor which weights the contribution of each species to the overall dissimilarity between the two communities. This weight can be set to give even weight to all species or to weigh them proportionally to their relative abundance in the pooled communities.

Diversity partitioning approach

Following the work of Hill (1973), a community with *diversity of order q*, ${}^{q}D$, is as diverse as a theoretical community containing ${}^{q}D$ equally abundant species. The order of diversity, q, expresses the weight given to differences in species abundance, $q = 0$ representing the presence/absence case, $q = \infty$ considering only the relative abundance of the most abundant species in the community. Without accounting for interspecific similarities, there is emerging consensus that using effective numbers (also called number of equivalents) is a straightforward way for partitioning diversity into within-community (alpha), between-community (beta) and across-community (gamma) components (Jost 2007). Of these three, the between-community component, beta diversity, can be interpreted as a form of dissimilarity when applied for two communities (Ricotta 2017). Beta diversity can be derived from alpha and gamma diversity in a multiplicative (beta = gamma/alpha) or an additive way (beta = gamma – alpha). Jost (2007) and Chao et al. (2012) argued that multiplicative beta diversity is a useful way for quantifying community differentiation; however, due to its

scaling between 1 and N (N being the number of communities) it is not comparable across samples containing different numbers of communities. To remove this dependence, they offer three solutions with which the value of multiplicative beta can be normed. Although, for pairwise comparisons, N is always 2, it seems straightforward to follow these recommendations, since the scaling between 0 and 1 has several advantages, and most other indices also share this property. The rescaling formulae of Chao et al. (2012) embody different concepts of community (dis-)similarity, which together we call the family of *multiplicative beta indices*. The first formula is the relative turnover rate per community, which is a linear transformation of beta to the normed scale.

Eq. 15. $$\beta_{turnover}\langle q\rangle = ({}^{q}\beta - 1)/(N-1)$$

Here 0 means identical species composition, while 1 indicates totally distinct communities. In the pairwise comparison (N = 2), $\beta_{turnover}\langle q\rangle = {}^{q}\beta - 1$.

The second index measures homogeneity, and is a linear transformation of the inverse of beta. With respect to the fact that the complement term of homogeneity is heterogeneity, we call its dissimilarity form $\beta_{heterogeneity}$:

Eq. 16. $$\beta_{heterogeneity}\langle q\rangle = 1 - \left(\frac{1}{{}^{q}\beta} - \frac{1}{N}\right)\Big/\left(1 - \frac{1}{N}\right)$$

When N = 2, $\beta_{het}\langle q\rangle = 2 - 2/{}^{q}\beta$. With $q = 0$ (presence/absence case) the index is identical with Jaccard index, while with $q = \infty$ (abundance case) it is the Morisita & Horn index.

The third index measures overlap between communities, whose counterpart is segregation, thus we call it β_{segr}:

Eq. 17. $$\beta_{segregation}\langle q\rangle = 1 - \left[\left(\frac{1}{{}^{q}\beta}\right)^{q-1} - \left(\frac{1}{N}\right)^{q-1}\right]\Big/\left[1 - \left(\frac{1}{N}\right)^{q-1}\right]$$

With $q = 0$, $\beta_{segregation}\langle q\rangle = \beta_{turnover}\langle q\rangle$, and both gives the Sørensen index.

According to Leinster & Cobbold (2012), it is possible to implement species similarities in the calculation of effective numbers. This way, the meaning of ${}^{q}D^{Z}$, is the diversity of a theoretical community with ${}^{q}D^{Z}$ *equally abundant* and *maximally different* species. Hence, both unevenness in the abundance structure and the between-species similarities decrease the value of effective species number. Due to measuring diversity in effective numbers, it is possible to partition diversity into alpha, beta, and gamma fractions (Leinster & Cobbold 2012; Botta-Dukát 2018) in the multiplicative way. Then, this multiplicative beta can be rescaled using the formulae proposed by Chao et al. (2012). These indices behave consistently only if abundances are taken into account as relative abundances.

Nearest neighbour approach

The earliest representatives of this family were shown by Clarke & Warwick (1998) and Izsák & Prince (2001), then Ricotta & Burrascano (2008), and Ricotta & Bacaro (2010; see D_{CW} and D_{IP} indices). Later Ricotta et al. (2016) introduced a new, general family called *PADDis*. All these indices were primarily defined for presence-absence data type. The approach is based on a re-definition of the *b* and *c* quantities of the 2×2 contingency table. Looking at species as maximally different, and taking X and Y the two communities under comparison, *b* can be viewed as the total uniqueness of community X. The uniqueness of a single species in *X* is 1 if it is absent in *Y*, otherwise it is 0. Therefore, *b* is the sum of species uniqueness values. However, from a functional perspective, the uniqueness of a species present only in *X* should be between 0 and 1 if it is absent in *Y* but a similar species present there. Therefore, it is possible to define the analogue of *b* which accounts for similarities between species:

Eq. 18. $$B = \sum_{i \in X}\left(1 - \max_{j \in Y} \varepsilon_{ij}\right) = S_x - \sum_{i \in X} \max_{j \in Y} \varepsilon_{ij}$$

The same logic applies for *c*, which is the uniqueness of community *Y*, where *C* expresses the degree of uniqueness:

Eq. 19. $$C = \sum_{i \in Y}\left(1 - \max_{j \in X} \varepsilon_{ij}\right) = S_y - \sum_{i \in Y} \max_{j \in X} \varepsilon_{ij}$$

Ricotta et al. (2016) define the *A* quantify as follows:

Eq. 20. $$A = a + (b - B) + (c - C)$$

Having *A*, *B*, and *C* defined as analogues of *a*, *b*, and *c*, it is now possible to design trait-based similarity measures following the logics of Jaccard, Sørensen, Sokal & Sneath, Kulczynski, Ochiai and Simpson indices. It is notable that Ricotta et al. (2016) define *A* as a quantity that ensures the components *B* and *C* to add up to *a* + *b* + *c* but with no explicit biological interpretation. Notably, D_{IP} and D_{CW} are identical with the Sørensen and Kulczynski forms of *PADDis*. The generalization of D_{IP} and D_{CW} to relative abundances, $D_{CW}(Q)$, was also derived by Ricotta & Bacaro (2010). For these two versions, it is not necessary to explicitly define the *A* component. Using the relationships between Jaccard, Sørensen, Kulczynski, Ochiai and Sokal & Sneath indices, from $D_{CW}(Q)$ it is theoretically possible to derive the extension of *PADDis* to relative abundances; however, the biological interpretation of *A* remains dubious in this framework.

Methods

The performance of FDissim indices can be reliably tested on data sets with known background processes driving community assembly which is hardly possible to satisfy with real data. Therefore, we compared the performance of FDissim indices using simulated data sets. The data sets were generated using the *comm.simul* function of the comsimitv R package (Botta-Dukát & Czúcz 2016, Botta-Dukát 2020). This function follows an individual-based

model for a meta-community comprising *N* communities and a regional pool of *S* species. Local communities include *J* individuals, and are distributed equidistantly along a continuous environmental gradient (with gradient values between 0 and 1). Each individual possesses three traits: an 'environmental', a 'competitive' trait, and a neutral trait, all ranging on [0; 1]. Intraspecific variation in trait values is neglected in the simulation, that is, individuals belonging to the same species are identical. The environmental trait defines the optimum of the species along the environmental gradient. The closer the position of a community along the environmental gradient to the environmental trait value of a species, the more suitable it is for that species:

Eq. 21. $$suitability = \frac{(environment - environmental\ trait)^2}{\sigma}.$$

where σ (sigma) is adjustable so as to change the niche width of the species, and hence, the length of the gradient (see later). The competitive trait represents the resource acquisition strategy of the individual. The more similar the latter value between two individuals, the higher the competition is between them, which means that intraspecific competition is the strongest. The neutral trait has no effect on community assembly, thus it is not considered in our study. The simulation starts with the random assignment of all individuals of all communities to species. The second step is a 'disturbance' event, when one individual 'dies' in each community. This individual is to be replaced by an offspring of other individuals within the same community or those of other communities. Each individual produces one offspring or does not reproduce. Probability of reproduction depends on the strength of competition. The offspring remains in the same community or randomly disperses into any of the other communities. Finally, the dead individual is replaced by one new individual from the seeds produced and dispersed. The probability that an individual of a certain species replaces the dead individual is defined by the number of seeds of that species and the

suitability of the habitat. Steps between the disturbance event and the establishment of a new individual constitute a single ‘generation’. Community composition is evaluated after lot of generations. The strength of the environmental filtering can be adjusted by the sigma parameter, respectively. When sigma is 0, all species are maximally specialist, which means that they can occur only at the optimum point of the gradient (that is, at the exact value for the environmental trait). If sigma is infinity, species are maximally generalist and all points along the environmental gradient are equally suitable for them. Therefore, sigma is the parameter which defines the suitability of each point of the gradient for each species based on its distance from the respective optima. We generated data sets with sigma values 0.01, 0.1, 0.25, 0.5, 1 and 5 in order to simulate situations with different strength of environmental filtering. The number of communities was 30, each community comprised 200 individuals, the number of species in the species pool was 300, the simulation iterated for 100 generations, and we allowed no intraspecific trait variation. For all the other parameters, we used the default options.

However, it needed further explanation what real situations the six simulated levels of environmental filtering represent. To provide a reference and assist interpretation, we calculated two species-based beta-diversity measures, the multiplicative beta (Whittaker 1960) and the gradient length of the first axis of a detrended correspondence analysis (DCA) ordination (Hill & Gauch 1980; Appendix S5, Fig. S5.1). The former gives the number of distinct communities present in the total species pool of the gradient, while the latter is minimal number of average niche breadths (also called turnover units) necessary for covering the total gradient length. Moreover, we plotted the abundance of species in the sample units along the gradient as a visual tool for assessing gradient length (Appendix S5, Fig. S5.2). All these methods indicated that with sigma = 0.01 the gradient is extremely long: there are more than 10 distinct communities and near 20 turnover units along the gradient. Samples with such

high beta diversity are very rare and special in real ecological research; therefore, findings from simulations with sigma = 0.01 are mostly of theoretical importance. Beta diversity values from sigma = 0.1 to sigma = 1 are more similar to real study situations, hence they should be more relevant for practice. At sigma = 5, environmental filtering is practically not operating, between-community variation is driven by interspecific relations and chance.

We calculated between-species dissimilarities as the Gower distance between their environmental trait values which in this case equals the Euclidean distance scaled to [0; 1]. These distances had to be transformed to similarities according to the requirements of the FDissim indices. Several formulae are available with which it is possible; however, they may assume different functional relationships between similarity and distance. One of such formulae we used is the linear transformation according to *Similarity = 1-Distance*. Besides this, we also used $Similarity = e^{-u \times Distance}$ which supposes a curvilinear function between similarity and distance (Leinster & Cobbold 2012). With this exponential formula, it is possible to weight the importance of small Gower distances between species relative to large distances. With changing the parameter u it is possible to adjust how steeply similarity decreases with increasing distance. We set $u = 10$ which leads to a relatively steep decline. Although, after this transformation the minimal value for similarity is higher than zero, we considered it negligibly low ($e^{-10} \approx 0.000045$) so we did not apply the transformation proposed by Botta-Dukát (2018). For all FDissim indices where it was necessary we used the similarity matrix or a dissimilarity matrix calculated as *Dissimilarity = 1-Similarity* as input. The dissimilarity matrix is identical with the Gower distance matrix if the similarities were calculated in a linear way, but in the other case, it keeps the exponential relationship between distance and (dis-)similarity.

Dissimilarity matrices were calculated for the four community data sets with different sigma values, with the two functions transforming Gower distances, and across a broad range of

available FDissim indices. For indices where absolute or relative abundances could have been taken into account, we opted for relative abundance for the sake of better comparability. With *generalized_Tradidiss*, we calculated the ‘even’ and the ‘uneven’ weighting versions. The entire analysis was run with abundance and presence/absence data. Some FDissim indices are only suitable for binary data, thus the number of indices applied for relative abundance and binary data were 25 and 31, respectively. In cases of indices handling both data types, we used exactly the same version of the index as with abundance data, hence communities with different numbers of species were given equal weight due to division by community totals. Additionally, dissimilarity matrices were also calculated using the Bray-Curtis index (for binary data: Sørensen index in dissimilarity form) to provide a contrast against the case disregarding between-species dissimilarities.

Then for each dissimilarity matrices, we conducted two types of analyses. Firstly, we compared how strongly the dissimilarity indices correlate with the environmental distance using Kendall tau rank correlation. This gives an estimate of how well a dissimilarity index reveals the monotonic relationship between trait composition of local communities and the environmental gradient. We visually assessed the shape of relationship between dissimilarity and environmental distance in the case of lowest sigma (i.e., longest gradient) when the distortion of linear relationship between the two is supposed to be the strongest. Then, to disentangle the effects of different methodological decisions and the sigma parameter on the correlation between FDissim indices and environmental distance we calculated a random forest model. In this model the dependent variable was the Kendall tau correlation coefficient, while the independent variables were the sigma, the data type (abundance vs. presence/absence), the transformation method for Gower distances (linear vs. exponential), and the FDissim method. Within approaches FDissim methods often strongly correlated that resulted in very similar Kendall’s tau values. Therefore, only the Sørensen/Bray-Curtis

versions of *dsimcom*, *dissABC*, *PADDis*/D_{CW}, *generalized_Tradidiss* with uneven weights, as well as $\beta_{turnover}$, *CWMdis*, and the *CDFdis* were included into this analysis. Variable importance scores (VIS) in the random forest were estimated by the permutation approach based on mean decrease in log-likelihood using the *varimp* function of the *partykit* package. The effects of the model terms were also illustrated by heat-maps.

All statistical analyses were done in R (R Core Team 2019) using the *FD* (Laliberté & Legendre 2010, Laliberté et al. 2014), *adiv* (Pavoine 2020a,b), *comsimitv* (Botta-Dukát 2020,) *vegan* (Oksanen et al. 2019), *DescTools* (Signorell et al. 2020), *partykit* (Hothorn et al. 2006, Strobl et al. 2007, Strobl et al. 2008, Hothorn & Zeileis 2015) packages.

Results

Kendall tau correlation coefficients decreased as the strength of environmental filtering decreased (that is, with increasing sigma) in all examined cases. For FDissim indices which handled both data types, presence/absence data resulted in lower correlations than abundance data for all indices. For most indices, this difference was highest at intermediate values for sigma. These trends were consistent between the linear and the exponential transformations. Correlations for all indices at all sigma values with linear transformation are shown in Table 3 for abundances data and in Table 4 for presence/absence data.

In most simulation scenarios, the FDissim indices correlated more strongly with the environmental gradient than the species-based Bray-Curtis index. However, in several occasions, indices belonging to the nearest neighbour family performed poorer than the species-based dissimilarity. Notably, at the highest sigma and with presence/absence data, all indices showed correlation near to zero but among them the Bray-Curtis index had the highest correlation with environmental distance.

As expected, we found perfect rank correlations among Jaccard, Sørensen, Sokal-Sneath and Whittaker's beta versions of *dsimcom*, among Jaccard, Sørensen and Sokal-Sneath forms of *dissABC*, between D_{IP} and Sørensen form of *PADDis* (only for presence-absence data), between D_{CW} and Kulczynski form of *PADDis* (only for presence-absence data), and between D_{IP} and D_{CW} (for abundance data type).

Dissimilarity indices showed various shapes of relationship with environmental distance (Appendix S4). At strongest environmental filtering, all FDissim indices had dissimilarity values near zero at minimal environmental distance, only the species-based Bray-Curtis which had dissimilarity was near 0.4 at the smallest environmental distances. In case of linear transformation of Gower distances and presence/absence data, approximately linear relationship was found for *CWMdis*, *CDFdis*, D_Q, Sørensen and Ochiai forms of *dsimcom*, Jaccard form of *dissABC*, Marczewski-Steinhaus form of *generalized_Tradidiss* with both weighting versions, $\beta_{heterogeneity}$ and $\beta_{segregation}$; although, most other indices showed only a small degree of distortion of linear function (Figure S4.1). Exponential relationship was found for the evenness-based (PE) form of *generalized_Tradidiss*. Notably, the taxon-based Bray-Curtis index had the steepest asymptotic function among all. In case of exponential transformation all other indices relying on between-species dissimilarities showed an asymptotic curve (Figure S4.2).

In the random forest, niche width (that is, sigma) acquired by far the highest variable importance score (VIS=0.114). The less important variables were the data type (VIS=0.0176), the dissimilarity method (VIS=0.0037) and the transformation (VIS=-0.00001). The heat map (Figure 1) also revealed a strong decrease in correlation along increasing sigma. It is also clearly shown that in most cases abundance data resulted in significantly higher correlation than presence/absence. The difference between linear and exponential transformation methods was not always visible. Regarding variation between dissimilarity indices, the most striking

patterns were the relatively poor performance of the *PADDis*/D_{CW} indices. All but the latter index combined with abundance data and linear transformation of dissimilarities lead to the highest correlation with environmental distance.

Discussion

General patterns in the correlation with environmental distance

We ran different simulation scenarios with varying strength of environmental filtering. We expected that the correlation between FDissim indices and environmental distance to be the highest when the environmental filtering is the strongest, and the correlation to become neutral when environmental filtering is not effective. When environmental filtering was strongest (that is, minimal overlap of species niches along the environmental gradient), all FDissim indices correlated highly with the environmental gradient. As expected, correlation between trait dissimilarity and environmental distance decreased as filtering weakened, moreover, differences between families of indices became more apparent. This result suggests that all tested methods are able to reveal the strong environmental filtering processes.

As the contribution of competitive exclusion and stochastic processes approach or override environmental filtering, the correlation between FDissim indices and the background gradient becomes weaker. This decrease itself is not a drawback of the FDissim methods, rather it is a consequence of our study design, since we applied a series of scenarios where the effect of niche filtering was decaying. However, we think that the degree of the decrease reflects the sensitivity of the FDissim indices to the underlying trait-environmental relationship. Indices, which showed high correlation with environmental distance, could be capable of revealing the environmental signal even when it is weak. Actually, in our tests, most indices reached similarly high correlation, and there were only a few combinations of simulation parameters

which resulted in a decreased correlation with environmental distance for some dissimilarity indices.

Determinants of the correlation based on the random forest model

The random forest model revealed that the effect of gradient length is the most important determinant of the correlation between dissimilarity and environmental distance, while methodological decisions had much lower variable importance. These observations suggest that the absolute value of the correlation between dissimilarity and environmental distance is primarily dependent on the sample in hand, and can be influenced by methodological decisions to a limited extent.

Correlations were stronger with abundance than with presence/absence data. This finding is at least partly attributable to our simulation design where community composition was driven by individual-based processes: birth, fitness difference, reproduction, and death. As a result, species relative abundances had to be proportional with their environmental suitability in the local community. Transforming such data to binary scale loses meaningful information and weakens the correlation between dissimilarity in trait composition and environmental background. In cases when presences and absences of species respond more robustly to the main environmental gradient, while relative abundances change stochastically, or abundance estimations are inaccurate, the binary data type might be more straightforward.

Transforming between-species dissimilarities has a potential to conform distributional requirements, to approximate expert intuitions about relatedness of species or to customize sensitivity to functional difference with respect to specific research aims. For most indices across the tested range of gradient length and data type, the exponential transformation resulted a somewhat lower correlation than with linear transformation. More insight is provided by examining the shape of the relationships besides the pure correlation value. After

linear transformation of Gower distances, most dissimilarity indices showed a linear or slightly curved function along environmental distance; although the scatter of the evenness-based *generalized_Tradidiss* differed considerably from the straight line towards an exponentially increasing one. After exponential transformation of between-species trait dissimilarities, all indices in the direct dissimilarity-based class showed a rather steeply increasing asymptotic function. This result suggests that with the exponential transformation of between-species dissimilarities, it is possible to make FDissim indices more sensitive to smaller differences in functional composition. Certainly, summary-based indices (*CWMdis*, *CDFdis*) are not affected by this transformation, since they are not based on between-species dissimilarities.

Comparison of taxon-based vs. trait-based dissimilarity

The basic assumption of functional ecology is that the traits of individuals should be in closer relationship with ecological properties than their taxonomical status. Following this argument, we expected that trait-based dissimilarity measures correlate more strongly with the environmental background than species-based indices. In contrast, higher correlation of species-based dissimilarity than trait-based dissimilarity indicates loss of information with the introduction of between-species similarity – which is non-sensual since our data was simulated in a way to possess a strong pattern in trait-environment relationship. We used the Sørensen/Bray-Curtis index in a dissimilarity form as a reference method representing species-based dissimilarity calculations disregarding traits. Our expectation was fulfilled by all indices with the exception of the members of the nearest neighbour family (D_{IP}, D_{CW} and *PADDis*). We suspect two potential reasons behind the low performance of these latter groups of indices. The first one is the improper scaling factor used for standardizing the ‘operational part’ of the indices (see the description in of the *PADDis* family and the discussion about it under the paragraph “Within-family variation of indices”). Second, these indices rely on the

quantities of minimally different species in the two communities under comparison. However, the minimum is a less robust descriptor of any sample distribution because of its dependency on sampling error; therefore, it might provide a poor representation of total community dissimilarity.

Although, we did not include dissimilarity values at exactly zero distance, the *y*-intercept (also called 'nugget') of the dissimilarity vs. environmental distance functions can be extrapolated with negligible error (Fortin & Dale 2005). Brownstein et al. (2012) argued that the nugget of the distance decay relationship is a direct estimate of the amount of chance in the variation between local communities. In this respect worth noting is that the nugget with species-based Bray-Curtis index was near 0.4, while with all trait-based indices the nugget was near zero. This suggests that without accounting for species similarities, environmental distance between communities can be overestimated due to similar species replacing each other.

Within-family variation of indices

The perfect correlation between Jaccard, Sørensen and Sokal-Sneath forms of *dsimcom* and *dissABC* families was expected, since the original, taxon-based Jaccard, Sørensen and Sokal-Sneath indices are algebraically related, too (Janson & Vegelius 1981). However, for *PADDis* Jaccard, Sørensen and Sokal-Sneath forms showed correlation below 1. At this family, the B and C components of the 2×2 contingency table are defined as measurable quantities with clear interpretation: the sum of species uniqueness values within each community. The total diversity ($A+B+C$) is defined to be equal with the species richness of the pooled pair of communities ($a+b+c$), and the quantity A is derived by subtracting ($B+C$) from it. With this definition, A remains a virtual quantity with no biological interpretation. In PADDis indices, trait-based quantities B and C appear in the numerator (the 'operational part' sensu Ricotta et al. 2016) of the indices, while in the denominators (i.e., in the 'scaling factor') the taxon-based quantities, a, b and c are used. We argue that the inconsistent behaviour of *PADDis* is

due to the application of taxon-based quantities for scaling factors of trait-based operational parts. At the same time, we acknowledge that we either see no obvious solution to define total diversity or shared diversity according to the uniqueness-based idea behind *PADDis* in a more realistic way. In the *generalized_Tradidiss* family, the trait-based analogue of Bray-Curtis index can be achieved by calculating generalized Canberra distance with uneven weighting of species. We expected this to be perfectly correlated with Marczewski-Steinhaus form of *generalized_Tradidiss* index with uneven weighting, since Bray-Curtis and Marczewski-Steinhaus indices are the abundance forms of Sørensen and Jaccard indices, respectively. However, the correlation between them was lower. In the *generalized_Tradidiss* family, between-community dissimilarity is calculated as weighted sum a standardized differences in species ordinariness values. Species ordinariness is calculated on the basis of species abundance and trait values; however, weights used for adjusting species-level contributions are derived solely from abundances. Therefore, *generalized_Tradidiss* also follows a 'hybrid' approach in accounting for taxon-based vs. trait-based information. We argue that this is the reason why the algebraic relationships between the original Sørensen and Jaccard indices does not apply to its Sørensen/Bray-Curtis-type and Jaccard/Marczewski-Steinhaus-type forms. To sum up, we point to our observation that Jaccard, Sørensen and Sokal-Sneath forms of certain families of indices do not satisfy the algebraic relationships they supposed to, opening space for potential confusion. These algebraic relations hold only if *A*, *B* and *C* quantities are explicitly and consistently defined.

Families of FDissim indices combine abundance difference of species between plots and interspecific trait differences in a unique way, while indices belonging to the same family differ in how they relate this amount of 'unshared' variation (summarized as the *b* and *c* portions of the contingency table) to the shared (*a*) variation. Some indices are able to handle abundances either as absolute or relative abundance (e.g. *dsimcom*, *generalized_Tradidiss*,

dissABC), while others divide absolute abundances by their sum over the respective community, thus they work only with relative abundances. When indices in the former group are set to consider absolute abundances, they become sensitive to variation in the summed abundances of the communities under comparison. To place our tests on a common ground, we simulated communities with equal total number of individuals, and set all indices, where relevant, to work with relative abundances. Hence, we removed the effect of differences in total abundance. The constant number of individuals might have increased the similarity between FDissim indices belonging to the same family and the correlation with the environmental gradient. The sum of abundances, let them be measured on any quantitative scale, may vary considerably in real study situations due to aggregated distribution of individuals or uneven sampling effort. Therefore, our findings are more likely valid for settings when the sum of abundances are relatively stable, e.g. when sampling effort is controlled and individuals are dispersed evenly, or when abundances are recorded on percentage scale.

Limitations of our study

In our study, we simulated a research situation in a simplistic way. We applied only one environmental gradient which operated as an environmental filter driving convergence on a single trait. Besides this, we applied another trait which was constantly affected by a low level of competitive exclusion. These two traits were uncorrelated. Nevertheless, there was some effect of random drift on community composition due to the probabilistic components of the simulation algorithm. We varied the strength of environmental filtering thus it had different relative contribution compared with competitive exclusion and stochasticity. In real research situations local trait composition is influenced by a wide range of processes, including several abiotic and biotic filters acting simultaneously. Unless they are manipulated as parts of an experimental system, the full set of such filters are usually unknown for the researchers. The

multiplicity of filters may reduce the ability of FDissim indices in recovering trait-environment relationships. Further research should clarify how increasing complexity of the sample affects the behaviour of FDissim indices.

Conclusions

Considering the diversity of concepts they are built upon, FDissim indices showed unexpectedly low variation in performance. CWMdis, dsimcom, generalized_Tradidiss acquired the highest correlation with environmental distance in all simulation scenarios, therefore they seem to be equally suitable for quantifying pairwise beta diversity based on traits. Nevertheless, the most important determinant of the matching between trait-based dissimilarity and environmental distance is the length of the trait gradient. Besides this, the data type (presence/absence vs. abundance) also affected the correlation more strongly than the choice of FDissim method. Extending the comparative tests of FDissim measure to more complex gradients and real data sets could offer further insight into their behaviour.

Data availability

Simulated data was generated using the comsimitv R package. Own functions for functional dissimilarity indices are made available through the Zenodo public repository: 10.5281/zenodo.4323590.

Author contributions

A.L. designed and carried out the analysis, lead writing, Z.B.D. discussed the concept and the results, wrote parts of and commented on the manuscript.

References

Anderson, M. J., Crist, T. O., Chase, J. M., Vellend, M., Inouye, B. D., Freestone, A. L., Sanders, N. J., Cornell, H. V., Comita, L. S., Davies, K. F., Harrison, S. P., Kraft, N. J. B., Stegen, J. C. & Swenson, N. G. (2011). Navigating the multiple meanings of β diversity: a roadmap for the practicing ecologist. Ecology Letters, 14(1), 19-28. doi:10.1111/j.1461-0248.2010.01552.x

Anderson, M. J., Ellingsen, K. E. & McArdle, B. H. (2006). Multivariate dispersion as a measure of beta diversity. Ecology Letters, 9(6), 683-693. doi:10.1111/j.1461-0248.2006.00926.x

Baselga, A. & Leprieur, F. (2015). Comparing methods to separate components of beta diversity. Methods in Ecology and Evolution, 6: 1069-1079. doi:10.1111/2041-210X.12388

Botta□Dukát, Z. & Czúcz, B. (2016). Testing the ability of functional diversity indices to detect trait convergence and divergence using individual□based simulation. Methods in Ecology and Evolution, 7, 114-126. https://doi.org/10.1111/2041-210X.12450

Botta□Dukát, Z. (2005). Rao's quadratic entropy as a measure of functional diversity based on multiple traits. Journal of Vegetation Science, 16, 533-540. https://doi.org/10.1111/j.1654-1103.2005.tb02393.x

Botta□Dukát, Z. (2018). The generalized replication principle and the partitioning of functional diversity into independent alpha and beta components. Ecography, 41: 40-50. doi:10.1111/ecog.02009

Botta-Dukat, Z. (2020). comsimitv: Flexible Framework for Simulating Community Assembly. R package version 0.1.4. https://CRAN.R-project.org/package=comsimitv

Brownstein, G., Steel, J.B., Porter, S., Gray, A., Wilson, C., Wilson, P.G. & Wilson, J. B. (2012). Chance in plant communities: a new approach to its measurement using the nugget from spatial autocorrelation. Journal of Ecology, 100, 987-996. https://doi.org/10.1111/j.1365-2745.2012.01973.x

Cardoso, P., Rigal, F., Carvalho, J.C., Fortelius, M., Borges, P.A.V., Podani, J. & Schmera, D. (2014). Partitioning taxon, phylogenetic and functional beta diversity into replacement and richness difference components. Journal of Biogeography, 41, 749-761. doi:10.1111/jbi.12239

Carmona, C. P., de Bello, F., Mason, N. W. H., Lepš, J. (2016). Traits without borders: Integrating functional diversity across scales. Trends in Ecology and Evolution 31(5), 382-394. doi: 10.1016/j.tree.2016.02.003

Champely, S., Chessel, D. (2002). Measuring biological diversity using Euclidean metrics. Environmental and Ecological Statistics 9, 167–177. https://doi.org/10.1023/A:1015170104476

Chao, A., Chiu, C. and Hsieh, T.C. (2012). Proposing a resolution to debates on diversity partitioning. Ecology, 93, 2037-2051. https://doi.org/10.1890/11-1817.1

Chao, A., Chiu, C.□H., Villéger, S., Sun, I□F., Thorn, S., Lin, Y.□C., Chiang, J.□M., & Sherwin, W. B. (2019). An attribute□diversity approach to functional diversity, functional beta diversity, and related (dis)similarity measures. Ecological Monographs, 89(2), e01343. 10.1002/ecm.1343

Chiu, C.-H., Jost, L. & Chao, A. (2014). Phylogenetic beta diversity, similarity, and differentiation measures based on Hill numbers. Ecological Monographs, 84(1), 21-44.

Clarke, K.R. & Warwick, R.M. (1993). Quantifying structural redundancy in ecological communities. Oecologia, 113(2), 278-289.

De Bello, F., Carmona, C.P., Mason, N.W.H., Sebastià, M.□T. and Lepš, J. (2013). Which trait dissimilarity for functional diversity: trait means or trait overlap? Journal of Vegetation Science, 24, 807-819. doi:10.1111/jvs.12008

De Bello, F., Lepš, J., Lavorel, S., & Moretti, M. (2007). Importance of species abundance for assessment of trait composition: an example based on pollinator communities. Community Ecology, 8(2), 163–170. https://doi.org/10.1556/ComEc.8.2007.2.3

Díaz, S., & Cabido, M. (2001). Vive la différence: plant functional diversity matters to ecosystem processes. Trends in Ecology and Evolution, 16(11), 646–655. https://doi.org/10.1016/S0169-5347(01)02283-2

Faith, D. P., Minchin, P. R. & Belbin, L. (1987). Compositional dissimilarity as a robust measure of ecological distance. Vegetatio 69, 57-68.

Fortin, M.□J. & Dale, M.R.T. (2005). Spatial Data Analysis: a Guide for Ecologists. Cambridge University Press, Cambridge.

Garnier, E., Cortez, J., Billès, G., Navas, M., Roumet, C., Debussche, M., Laurent, G., Blanchard, A., Aubry, D., Bellmann, A., Neill, C. & Toussaint, J. (2004). Plant functional markers capture ecosystem properties during secondary succession. Ecology, 85, 2630-2637. doi:10.1890/03-0799

Gregorius, H.□R., Gillet, E.M. & Ziehe, M. (2003). Measuring Differences of Trait Distributions Between Populations. Biometrical Journal, 45, 959-973. https://doi.org/10.1002/bimj.200390063

Grime, J. P. (1998). Benefits of plant diversity to ecosystems: immediate, filter and founder effects. Journal of Ecology, 86, 902–910.

Hawkins, B.A., Leroy, B., Rodríguez, M.Á., Singer, A., Vilela, B., Villalobos, F., Wang, X. & Zelený, D. (2017). Structural bias in aggregated species level variables driven by repeated species co occurrences: a pervasive problem in community and assemblage data. Journal of Biogeography, 44, 1199-1211.

Hérault, B., & Honnay, O. (2007). Using life-history traits to achieve a functional classification of habitats. Applied Vegetation Science, 10(1), 73–80. https://doi.org/10.1111/j.1654-109X.2007.tb00505.x

Hill, M. O. & Gauch, H. G. (1980). Detrended Correspondence Analysis: An Improved Ordination Technique. Vegetatio, 42, 47–58.

Hill, M. O. (1973). Diversity and evenness: a unifying notation and its consequences. Ecology, 54(2), 427–432.

Hitchcock, F.L. (1941). Distribution of a product from several sources to numerous localities. Journal of Mathematical Physics, 20: 224-230.

Hothorn, T., Hornik, K., Van de Wiel, M. A. & Zeileis, A. (2006). A Lego System for Conditional Inference. The American Statistician, 60(3), 257–263.

Hothorn, T., Zeileis, A. (2015). partykit: A Modular Toolkit for Recursive Partytioning in R. Journal of Machine Learning Research, 16, 3905-3909. URL http://jmlr.org/papers/v16/hothorn15a.html

Izsák, C., & Price. R. G. (2001). Measuring b-diversity using a taxonomic similarity index, and its relation to spatial scale. Marine Ecology Progress Series 215, 69–77.

Janson, S. & J. Vegelius (1981). Measures of ecological association. Oecologia, 49(3), 371-376.

Jost, L. (2007). Partitioning diversity into independent alpha and beta components. Ecology, 88, 2427–2439.

Kleyer, M., Dray, S., Bello, F., Lepš, J., Pakeman, R.J., Strauss, B., Thuiller, W. & Lavorel, S. (2012). Assessing species and community functional responses to environmental gradients: which multivariate methods? Journal of Vegetation Science, 23, 805-821. doi:10.1111/j.1654-1103.2012.01402.x: 1199–1211.

Koleff, P., Gaston, K. J. & Lennon, J. J. (2003). Measuring beta diversity for presence–absence data. Journal of Animal Ecology, 72, 367-382. doi:10.1046/j.1365-2656.2003.00710.x

Laliberté, E. & P. Legendre (2010). A distance-based framework for measuring functional diversity from multiple traits. Ecology, 91, 299-305.

Laliberté, E., Legendre, P., & Shipley, B. (2014). FD: measuring functional diversity from multiple traits, and other tools for functional ecology. R package version 1.0-12.

Legendre, P. & Legendre, L. (1998) Numerical ecology. Elsevier, Amsterdam, NL

Legendre, P., De Cáceres, M. (2013). Beta diversity as the variance of community data: dissimilarity coefficients and partitioning. Ecology Letters 16, 951–963

Leinster, T. & Cobbold, C.A. (2012). Measuring diversity: the importance of species similarity. Ecology, 93, 477-489. doi:10.1890/10-2402.1

Lengyel, A. & Podani, J. (2015). Assessing the relative importance of methodological decisions in classifications of vegetation data. Journal of Vegetation Science, 26, 804-815. doi:10.1111/jvs.12268

Lengyel, A., Swacha, G., Botta-Dukát, Z. & Kacki, Z. (2020). Trait-based numerical classification of mesic and wet grasslands in Poland. Journal of Vegetation Science, 31, 319–330. https://doi.org/10.1111/jvs.12850

Lepš, J., de Bello, F., Lavorel, S. & Berman, S. (2006). Quantifying and interpreting functional diversity of natural communities: practical considerations matter. Preslia, 78, 481–501.

MacArthur, R., Levins, R. (1967). Limiting similarity convergence and divergence of coexisting species. American Naturalist, 101, 377–387.

Mason, N. W. H., Mouillot, D., Lee, W. G. & Wilson, J. B. (2005). Functional richness, functional evenness and functional divergence: the primary components of functional diversity. Oikos, 111, 112-118. doi:10.1111/j.0030-1299.2005.13886.x

McGill, B., Enquist, B. J., Weiher, E., Westoby, M. (2006). Rebuilding community ecology from functional traits. Trends in Ecology and Evolution 21(4), 178-185.

Mouchet, M.A., 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. Functional Ecology, 24, 867-876. doi:10.1111/j.1365-2435.2010.01695.x

Mouillot, D., Stubbs, W., Faure, M., Dumay, O., Tomasini, J.A., Wilson, J.B. & Chi, T.D. (2005). Niche overlap estimates based on quantitative functional traits: a new family of non□parametric indices. Oecologia, 145, 345–353.

Muscarella, R. & Uriarte, M. (2016). Do community-weighted mean functional traits reflect optimal strategies? Proceedings of the Royal Society B, 283, 20152434. https://doi.org/10.1098/rspb.2015.2434

Nipperess, D.A., Faith, D.P. & Barton, K. (2010), Resemblance in phylogenetic diversity among ecological assemblages. Journal of Vegetation Science, 21, 809-820. doi:10.1111/j.1654-1103.2010.01192.x

Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Peter R. Minchin, P. R., O'Hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. M., Szoecs, E. & Wagner, H. (2019). vegan: Community Ecology Package. R package version 2.5-6. https://CRAN.R-project.org/package=vegan

Pavoine, S. & Ricotta, C. (2014). Functional and phylogenetic similarity among communities. Methods in Ecology and Evolution, 5, 666--675.

Pavoine, S. & Ricotta, C. (2019). Measuring functional dissimilarity among plots: Adapting old methods to new questions. Ecological Indicators, 97, 67-72.

Pavoine, S. (2012). Clarifying and developing analyses of biodiversity: towards a generalisation of current approaches. Methods in Ecology and Evolution, 3, 509-518. doi:10.1111/j.2041-210X.2011.00181.x

Pavoine, S. (2016). A guide through a family of phylogenetic dissimilarity measures among sites. Oikos, 125, 1719-1732. doi:10.1111/oik.03262

Pavoine, S. (2020). adiv: An R package to analyse biodiversity in ecology. Methods in Ecology and Evolution, 11, 1106– 1112. https://doi.org/10.1111/2041-210X.

Peres-Neto, P.R., Dray, S. & ter Braak, C.J.F. (2017). Linking trait variation to the environment: critical issues with community-weighted mean correlation resolved by the fourth-corner approach. Ecography, 40, 806-816.

Petchey, O. L. & Gaston, K. J. (2006). Functional diversity: back to basics and looking forward. Ecology Letters, 9, 741-758. doi:10.1111/j.1461-0248.2006.00924.x

Podani, J. & Schmera, D. (2011). A new conceptual and methodological framework for exploring and explaining pattern in presence – absence data. Oikos, 120, 1625-1638. doi:10.1111/j.1600-0706.2011.19451.x

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

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

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

Ricotta C. & Burrascano S. (2008). Beta diversity for functional ecology. Preslia, 80, 61–71.

Ricotta, C. & G. Bacaro. (2010). On plot-to-plot dissimilarity measures based on species functional traits. Community Ecology, 11, 113–119.

Ricotta, C. & J. Podani. (2017). On some properties of the Bray-Curtis dissimilarity and their ecological meaning. Ecological Complexity, 31, 201–205.

Ricotta, C. & Pavoine, S. (2015). Measuring similarity among plots including similarity among species: an extension of traditional approaches. Journal of Vegetation Science, 26, 1061-1067. doi:10.1111/jvs.12329

Ricotta, C. (2017). Of beta diversity, variance, evenness, and dissimilarity. Ecology and Evolution 7, 4835– 4843. https://doi.org/10.1002/ece3.2980

Ricotta, C. (2018). A family of (dis)similarity measures based on evenness and its relationship with beta diversity. Ecological Complexity, 34, 69-73. DOI: 10.1016/j.ecocom.2018.03.002

Ricotta, C., Bacaro, G., Caccianiga, M., Cerabolini, B.E.L. & Moretti, M. (2015). A classical measure of phylogenetic dissimilarity and its relationship with beta diversity. Basic and Applied Ecology 16(1), 10-18. https://doi.org/10.1016/j.baae.2014.10.003

Ricotta, C., Podani, J., Pavoine, S. (2016). A family of functional dissimilarity measures for presence and absence data. Ecology and Evolution, 6, 5383–5389. DOI: 10.1002/ece3.2214

Schmidt, T., Matias Rodrigues, J. & von Mering, C. (2017). A family of interaction-adjusted indices of community similarity. ISME Journal 11, 791–807. https://doi.org/10.1038/ismej.2016.139

Signorell, A. et mult. al. (2020). DescTools: Tools for descriptive statistics. R package version 0.99.38.

Strobl, C., Boulesteix, A.L., Kneib, T., Augustin, T. & Zeileis, A. (2008). Conditional Variable Importance for Random Forests. BMC Bioinformatics, 9(307). http://www.biomedcentral.com/1471-2105/9/307

Strobl, C., Boulesteix, A.L., Zeileis, A. & Hothorn, T. (2007). Bias in Random Forest Variable Importance Measures: Illustrations, Sources and a Solution. BMC Bioinformatics, 8, 25. http://www.biomedcentral.com/1471-2105/8/25

Swenson N. G., Anglada-Cordero P. & Barone J. A. (2011). Deterministic tropical tree community turnover: evidence from patterns of functional beta diversity along an elevational gradient. Proceedings of the Royal Society B, 278, 877–884.

Swenson, N. G. (2011). Phylogenetic Beta Diversity Metrics, Trait Evolution and Inferring the Functional Beta Diversity of Communities. PLoS ONE 6(6), e21264. https://doi.org/10.1371/journal.pone.0021264

Tamás, J., Podani, J. & Csontos, P. (2001). An extension of presence/absence coefficients to abundance data: a new look at absence. Journal of Vegetation Science, 12, 401-410. doi:10.2307/3236854

Tuomisto, H. (2010a). A diversity of beta diversities: straightening up a concept gone awry. Part 1. Defining beta diversity as a function of alpha and gamma diversity. Ecography, 33, 2-22. doi:10.1111/j.1600-0587.2009.05880.x

Tuomisto, H. (2010b). A diversity of beta diversities: straightening up a concept gone awry. Part 2. Quantifying beta diversity and related phenomena. Ecography, 33, 23-45. doi:10.1111/j.1600-0587.2009.06148.x

Villéger, S., Mason, N.W.H. & Mouillot, D. (2008). New multidimensional functional diversity indices for a multifaceted framework in functional ecology. Ecology, 89, 2290-2301. doi:10.1890/07-1206.1

Violle, C., Navas, M.□L., Vile, D., Kazakou, E., Fortunel, C., Hummel, I. & Garnier, E. (2007). Let the concept of trait be functional! Oikos, 116, 882-892. doi:10.1111/j.0030-1299.2007.15559.x

Whittaker, R. H. (1960). Vegetation of the Siskiyou Mountains, Oregon and California. Ecological Monographs, 30, 280–338.

Whittaker, R. H. (1972). Evolution and measurement of species diversity. Taxon, 21, 213-251.doi:10.2307/1218190

Zelený, D. (2018). Which results of the standard test for community weighted mean approach are too optimistic? Journal of Vegetation Science 29, 953-966.

Tables and Figures

Table 1. Similarity and dissimilarity forms of resemblance indices for presence-absence data

Name of the index	Similarity version	Dissimilarity version
Sørensen	$s_S = \frac{2a}{2a+b+c} = \frac{a}{(S_j+S_k)/2}$	$d_S = \frac{b+c}{2a+b+c} = \frac{b+c}{S_j+S_k}$
Ochiai	$s_O = \frac{a}{\sqrt{(a+b)(a+c)}} = \frac{a}{\sqrt{S_j S_k}}$	$d_O = \frac{b+c}{\sqrt{S_j S_k}}$
Kulczynski	$s_K = \frac{1}{2}\left(\frac{a}{a+b} + \frac{a}{a+c}\right) = \frac{a}{2/(1/S_j + 1/S_k)}$	$d_K = \frac{1}{2}\left(\frac{b}{a+b} + \frac{c}{a+c}\right) = \frac{1}{2}\left(\frac{b}{S_j} + \frac{c}{S_k}\right)$
Simpson	$s_{Si} = \frac{a}{a+\min(b,c)} = \frac{a}{\min(S_j, S_k)}$	$d_{Si} = \frac{b+c}{min(S_j, S_k)}$
Jaccard	$s_J = \frac{a}{a+b+c} = \frac{a}{S_{jk}}$	$d_J = \frac{b+c}{a+b+c} = \frac{b+c}{S_{jk}}$
Sokal & Sneath	$s_{SS} = \frac{a}{a+2(b+c)}$	$d_{SS} = \frac{2(b+c)}{a+2(b+c)}$

Table 2. Classification of trait-based dissimilarity indices. In columns of input data type X-es indicate, if abundance (A), relative abundance (R), and presence-absence data can be used as input.

Class	**Approach**	**Family**	**References**	**Input Data tpye**			**R function**
				A	R	P/A	
Summary-based	*Typical value*	*CWM-based*	Ricotta et al. (2015)	X	X	X	FD:::functcomp
	Distribution-based	*CDF-based*	Appendix S3	X	X	X	our new functions, see Data availability
Direct dissimilarity	*Probabilistic*	*DISC/D_Q*	Rao 1982, Pavoine & Ricotta (2014)	X	X	X	adiv::SQ
		dsimcom	Pavoine & Ricotta (2014)	X	X	X	adiv:::dsimcom
	Ordinariness-based	*dissABC*	Pavoine & Ricotta (2015)	X	X	X	adiv:::dissABC
		generalized_Tradidiss	Pavoine & Ricotta (2019)	X	X		adiv:::generalized_Tradidiss
	Diversity	*multiplicative beta*	Chao et al. (2012)		X		our new functions, see Data

	partitioning						availability
	Nearest neighbour	D_{CW}, $D_{CW}(Q)$	Clarke & Warwick (1998), Ricotta & Bacaro (2010)		X	X	our new functions, see Data availability
		D_{IP}	Izsák & Prince (2001), Ricotta & Bacaro (2010)		X	X	our new functions, see Data availability
		PADDis	Ricotta et al. (2016)			X	adiv:::PADDis
Classification-based	not discussed	not discussed	Hérault & Honnay (2007), Nipperess et al. (2010), Cardoso et al. (2014), Pavoine (2016)				

Table 3. Kendall tau correlations between environmental distance and the functional dissimilarity measures at different values of sigma and with abundance data type

	Sigma=0.01	**Sigma=0.1**	**Sigma=0.25**	**Sigma=0.5**	**Sigma=1**	**Sigma=5**
CWMdis	0.974	0.905	0.846	0.828	0.649	0.251
CDFdis	0.974	0.904	0.845	0.83	0.646	0.255
D(Q)	0.974	0.912	0.832	0.828	0.637	0.243
dsimcom.SS	0.974	0.911	0.832	0.829	0.638	0.243
dsimcom.Jac	0.974	0.911	0.832	0.829	0.638	0.243
dsimcom.Sor	0.974	0.911	0.832	0.829	0.638	0.243
dsimcom.Och	0.974	0.911	0.832	0.829	0.639	0.243
dsimcom.Beta	0.974	0.911	0.832	0.829	0.638	0.243
dissABC.Jac	0.967	0.899	0.82	0.813	0.617	0.243
dissABC.Sor	0.967	0.899	0.82	0.813	0.617	0.243
dissABC.SS	0.967	0.899	0.82	0.813	0.617	0.243
dissABC.Och	0.968	0.899	0.819	0.814	0.618	0.243
dissABC.Kul	0.968	0.898	0.819	0.814	0.619	0.243
dissABC.Si	0.954	0.867	0.789	0.791	0.616	0.243
Tradidiss.GC.even	0.974	0.908	0.816	0.829	0.626	0.245
Tradidiss.MS.even	0.974	0.908	0.814	0.828	0.623	0.243
Tradidiss.PE.even	0.974	0.907	0.828	0.831	0.639	0.25
Tradidiss.GC.uneven	0.967	0.901	0.827	0.815	0.622	0.244
Tradidiss.MS.uneven	0.966	0.899	0.823	0.813	0.618	0.243
Tradidiss.PE.uneven	0.969	0.905	0.837	0.821	0.637	0.249
$\beta_{turnover}$	0.974	0.911	0.837	0.829	0.641	0.251
$\beta_{heterogeneity}$	0.974	0.911	0.837	0.829	0.641	0.251
$\beta_{segregation}$	0.974	0.911	0.837	0.829	0.641	0.251

D_{IP}	0.923	0.778	0.68	0.565	0.338	0.034
D_{CW}	0.923	0.778	0.68	0.565	0.338	0.034
Bray-Curtis (species-based)	0.711	0.832	0.778	0.678	0.455	0.086

Table 4. Kendall tau correlations between environmental distance and the functional dissimilarity measures at different values of sigma and with presence/absence data type

	Sigma=0.01	**Sigma=0.1**	**Sigma=0.25**	**Sigma=0.5**	**Sigma=1**	**Sigma=5**
CWMdis	0.944	0.818	0.691	0.568	0.353	0.014
CDFdis	0.943	0.820	0.700	0.594	0.306	-0.001
D(Q)	0.941	0.818	0.701	0.59	0.313	-0.003
dsimcom.SS	0.944	0.821	0.705	0.593	0.316	-0.003
dsimcom.Jac	0.944	0.821	0.705	0.593	0.316	-0.003
dsimcom.Sor	0.944	0.821	0.705	0.593	0.316	-0.003
dsimcom.Och	0.944	0.822	0.707	0.592	0.323	0.000
dsimcom.Beta	0.944	0.821	0.705	0.593	0.316	-0.003
dissABC.Jac	0.946	0.819	0.704	0.592	0.292	-0.006
dissABC.Sor	0.946	0.819	0.704	0.592	0.292	-0.006
dissABC.SS	0.946	0.819	0.704	0.592	0.292	-0.006
dissABC.Och	0.946	0.819	0.704	0.591	0.293	-0.006
dissABC.Kul	0.946	0.819	0.704	0.591	0.294	-0.006
dissABC.Si	0.939	0.835	0.701	0.556	0.324	0.019
Tradidiss.GC.even	0.945	0.820	0.707	0.593	0.305	-0.005
Tradidiss.MS.even	0.945	0.819	0.707	0.592	0.304	-0.005
Tradidiss.PE.even	0.947	0.821	0.704	0.595	0.323	-0.002
Tradidiss.GC.uneven	0.946	0.819	0.702	0.592	0.308	-0.006
Tradidiss.MS.uneven	0.946	0.819	0.703	0.592	0.307	-0.006
Tradidiss.PE.uneven	0.947	0.820	0.698	0.593	0.326	-0.003
$\beta_{turnover}$	0.943	0.817	0.699	0.585	0.331	-0.003
$\beta_{heterogeneity}$	0.943	0.817	0.699	0.585	0.331	-0.003
$\beta_{segregation}$	0.943	0.817	0.699	0.585	0.331	-0.003

D_{IP}	0.905	0.696	0.597	0.435	0.158	-0.017
D_{CW}	0.904	0.694	0.593	0.431	0.160	-0.017
PADDis.Jac	0.904	0.679	0.575	0.418	0.154	-0.025
PADDis.Sor	0.905	0.696	0.597	0.435	0.158	-0.017
PADDis.SS	0.902	0.662	0.546	0.396	0.144	-0.034
PADDis.Och	0.904	0.697	0.596	0.436	0.159	-0.017
PADDis.Simp	0.881	0.620	0.474	0.343	0.158	0.030
PADDis.Kul	0.904	0.694	0.593	0.431	0.160	-0.017
Sørensen (species-based)	0.698	0.724	0.606	0.415	0.127	0.048
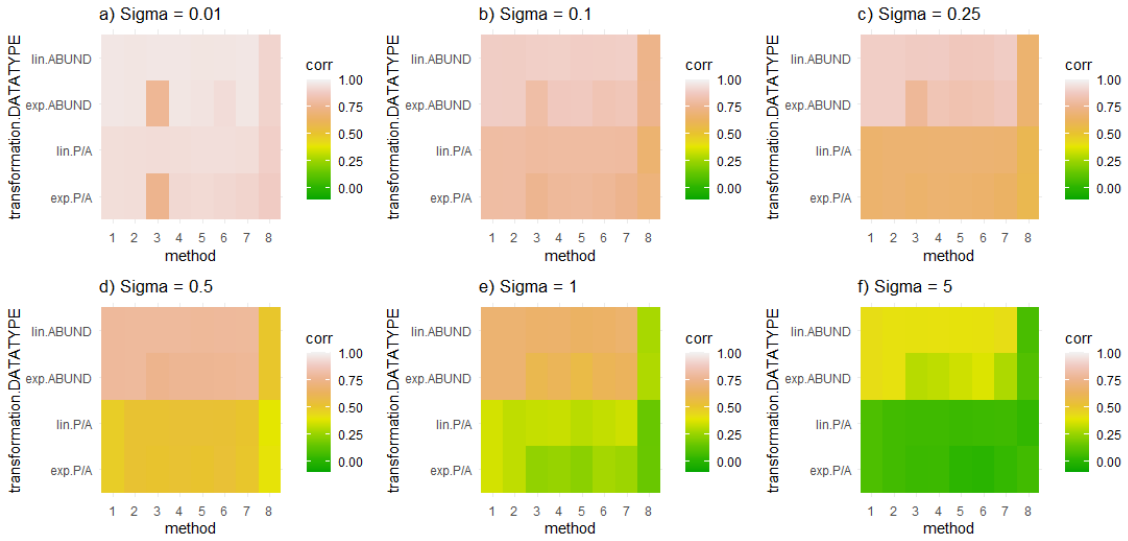
Figure 1. Heat maps showing the interactive effects of niche width (sigma), transformation of between-species dissimilarities (lin = linear, exp = exponential), data type (ABUND = abundance, P/A = presence/absence), and dissimilarity index (1 – *CWMdis*, 2 – *CDFdis*, 3 – D_Q, 4 – *dsimcom*/Sørensen, 5 – *dissABC*/Sørensen, 6 – *generalized_Tradidiss*/generalized Canberra, uneven weighting, 7 – $\beta_{turnover}$, 8 – D_{CW}) on the correlation with environmental distance