Measuring functional dissimilarity among plots: Adapting old methods to new questions

Sandrine Pavoine[a,1], Carlo Ricotta[b,1,*]

[a] *Centre d'Ecologie et des Sciences de la Conservation (CESCO), Muséum National d'Histoire Naturelle, CNRS, Sorbonne Université, 43 Rue Cuvier, CP 135 75005 Paris, France*
[b] *Department of Environmental Biology, University of Rome, 'La Sapienza', Piazzale Aldo Moro 5, 00185 Rome, Italy*

ABSTRACT

Ecologists routinely use dissimilarity measures between pairs of plots to explore the complex mechanisms that drive community assembly. Traditional dissimilarity measures usually quantify plot-to-plot dissimilarity based either on species presences and absences within plots or on species abundances, thus assuming that all species are equally and maximally distinct from one another. However, the value of dissimilarity measures that incorporate information on functional differences among species is becoming increasingly recognized. Since these 'functional dissimilarity measures' have been developed for capturing new aspects of plot-to-plot dissimilarity, they have usually little to do with more traditional measures based solely on species incidence or abundance data. In this paper we introduce a general method for adapting a large family of traditional dissimilarity coefficients to the measurement of functional differences among plots. The behavior of the proposed method, for which we provide a simple R function, was evaluated with published data on plant communities in a coastal marsh plain in Algeria. As shown by the worked example, our proposal produces a coherent framework for summarizing functional dissimilarity among plots. Being based on a generalization of classical dissimilarity measures with well-known properties, this new family of functional dissimilarity indices also has a great potential for future theoretical and applied developments in this field of research.

1. Introduction

Ecologists have developed a multitude of (dis)similarity measures between pairs of plots (or communities, assemblages, relevés, sites, quadrats, etc.) for exploring various aspects of the complex mechanisms that drive community assembly (see e.g. Orlóci, 1978; Podani, 2000; Legendre and Legendre, 2012). These measures usually quantify plot-to-plot dissimilarity based either on species incidence or abundance within plots, thus assuming that all species are equally and maximally distinct from one another, while neglecting information on functional differences among species.

More recently, a number of 'functional dissimilarity measures' have been proposed for summarizing different facets of functional differences among plots (Rao, 1982; Clarke and Warwick, 1998; Izsák and Price, 2001; Champely and Chessel, 2002; Pavoine et al., 2004; Chiu et al., 2014; Pavoine and Ricotta, 2014; Ricotta et al., 2016). Such new dissimilarity measures incorporate information on the species functional traits. Therefore, they are expected to correlate more strongly

with ecosystem-level processes, as species influence these processes via their traits (Mason and de Bello, 2013). Since functional dissimilarity measures have been developed for capturing new aspects of ecological differences among plots, they are usually not directly related to more traditional measures of species turnover which are based solely on species incidence or abundance data.

The aim of this paper is thus to introduce a methodological framework for adapting a large family of traditional dissimilarity coefficients to the measurement of functional differences among plots. The main advantage of this approach is that, being based on a generalization of well-known dissimilarity measures which have been extensively used in ecology for a long time, these new functional measures benefit from decades of research on multivariate resemblance.

---

* Correspondence author.
  E-mail address: carlo.ricotta@uniroma1.it (C. Ricotta).
[1] Both authors equally contributed to this paper.

2. Methods

2.1. A general family of traditional dissimilarity measures

Let U and V be two plots where x_{Uj} and x_{Vj} are the abundance values of species j in plots U and V, respectively, and N is the total number of species sampled in both plots (i.e. the species for which $\min\{x_{Uj}, x_{Vj}\} > 0$). A desirable property for a dissimilarity coefficient is the so-called 'sum property' (Ricotta and Podani, 2017). That is, its ability to be additively partitioned into species-level contributions, thus enabling to emphasize the relevance of single species to plot-to-plot dissimilarity.

Among the many dissimilarity measures that conform to the sum property, the Canberra distance (Lance and Williams, 1967) is obtained by standardizing separately for each species the absolute difference in species abundances by the sum of the abundances in both plots:

$$CD = \sum_{j=1}^{N} \frac{|x_{Uj} - x_{Vj}|}{(x_{Uj} + x_{Vj})} \qquad (1)$$

The Canberra distance is bounded between 0 and N. Therefore, a normalized measure in the range [0,1] is obtained by division by N:

$$NC = \frac{1}{N} \sum_{j=1}^{N} \frac{|x_{Uj} - x_{Vj}|}{(x_{Uj} + x_{Vj})} \qquad (2)$$

On the other hand, the dissimilarity coefficient proposed by Bray and Curtis (1957), one of the most popular measures of compositional dissimilarity among ecologists, implies normalization of the sum of species-wise differences by the total abundance of species in both plots:

$$BC = \frac{\sum_{j=1}^{N} |x_{Uj} - x_{Vj}|}{\sum_{j=1}^{N} (x_{Uj} + x_{Vj})} \qquad (3)$$

To emphasize the relationship between the Bray-Curtis dissimilarity and the Canberra distance, we can rewrite BC as:

$$BC = \sum_{j=1}^{N} \frac{|x_{Uj} - x_{Vj}|}{\sum_{k=1}^{N} (x_{Uk} + x_{Vk})} = \sum_{j=1}^{N} w_j \frac{|x_{Uj} - x_{Vj}|}{(x_{Uj} + x_{Vj})} \qquad (4)$$

where

$$w_j = (x_{Uj} + x_{Vj}) / \sum_{k=1}^{N} (x_{Uk} + x_{Vk}) \qquad (5)$$

with $0 \leq w_j \leq 1$ and $\sum_{j=1}^{N} w_j = 1$.

As shown by Eq. (4), the Bray-Curtis dissimilarity can be expressed as a normalized form of the Canberra distance in which the contribution of species j to overall dissimilarity $|x_{Uj} - x_{Vj}|/(x_{Uj} + x_{Vj})$ is weighted by the relative abundance of j in plots U and V. Based on the observed relationship between the Bray-Curtis dissimilarity and the normalized Canberra distance, Ricotta and Podani (2017) defined a 'generalized Canberra dissimilarity' which includes BC and NC as special cases:

$$GC = \sum_{j=1}^{N} \pi_j \frac{|x_{Uj} - x_{Vj}|}{(x_{Uj} + x_{Vj})} \qquad (6)$$

with $0 \leq \pi_j \leq 1$ and $\sum_{j=1}^{N} \pi_j = 1$. The weights π_j may be related to any species-specific ecological variable that is assumed to influence ecosystem functioning, such as the species phylogenetic and/or functional originality, or their conservation value. For $\pi_j = 1/N$, we obtain the normalized Canberra distance, while setting $\pi_j = w_j$, we obtain the Bray-Curtis dissimilarity.

Going a step further, we can define an even larger family of dissimilarity coefficients as:

$$D_\pi = \sum_{j=1}^{N} \pi_j d(x_{Uj}, x_{Vj}) \qquad (7)$$

where the term $d(x_{Uj}, x_{Vj})$ represents the single-species dissimilarity of U and V for species j in the range [0, 1].

Members of this family are the generalized version of the Marczewski-Steinhaus coefficient (Ricotta and Podani, 2017):

$$D_{MS} = \sum_{j=1}^{N} \pi_j \frac{|x_{Uj} - x_{Vj}|}{\max\{x_{Uj}, x_{Vj}\}} \qquad (8)$$

or the evenness-based dissimilarity coefficients proposed by Ricotta (2018):

$$D_{EVE} = \sum_{j=1}^{N} \pi_j (1 - EVE_j) \qquad (9)$$

where EVE_j is a measure of the evenness of species j in plots U and V. For example, EVE_j can be obtained with Pielou's evenness:

$$EVE_j = H_j / \log 2 \qquad (10)$$

where H_j is the Shannon entropy of species j: $H_j = -\frac{x_{Uj}}{x_{Uj} + x_{Vj}} \log(\frac{x_{Uj}}{x_{Uj} + x_{Vj}}) - \frac{x_{Vj}}{x_{Uj} + x_{Vj}} \log(\frac{x_{Vj}}{x_{Uj} + x_{Vj}})$.

In the next sections, we will show how to apply this family of dissimilarity coefficients to the measurement of functional dissimilarity among plots.

2.2. A new family of functional dissimilarity measures

Regardless of how pairwise functional differences among species i and j are calculated, they are usually represented by symmetric dissimilarity coefficients δ_{ij} with $\delta_{ij} = \delta_{ji}$ and $\delta_{ii} = 0$. If δ_{ij} is bounded in the range [0, 1], we can derive a corresponding similarity coefficient $\sigma_{ij} = 1 - \delta_{ij}$ as the complement of δ_{ij}. Note that dissimilarity measures with an upper bound $\delta_{\max} > 1$ can be normalized by dividing each term by $\delta_{\max}$, while for dissimilarity measures without an upper bound we can still get a locally normalized dissimilarity in the range [0, 1] by dividing each term by the maximum value in the data set.

Combining species abundances x_{Uj} and between-species similarities σ_{ij}, Leinster and Cobbold (2012) defined the 'ordinariness' of species j as the abundance of all species in plot U that are functionally similar to j such that:

$$z_{Uj} = \sum_{i=1}^{N} x_{Ui} \sigma_{ij} \qquad (11)$$

where z_{Uj} is the abundance of all species that are functionally similar to j (including j itself). For species j, z_{Uj} thus measures the commonness of all individuals in plot U that support the functions associated with j. z_{Uj} ranges from x_{Uj} if all species $i \neq j$ are maximally dissimilar from j, such that $\sigma_{ij} = 0$, to $\sum_{j=1}^{N} x_{Uj}$ (i.e. the total species abundance in plot U) if all species $i \neq j$ are functionally identical to j such that $\sigma_{ij} = 1$. Hence, the abundance of species similar to j is at least as great as the abundance of j itself.

Given two functionally identical plots U and V, we have that for each species in U and V, the ordinariness $z_{Uj} = \sum_{i=1}^{N} x_{Ui} \sigma_{ij}$ in plot U is equal to the corresponding value $z_{Vj} = \sum_{i=1}^{N} x_{Vi} \sigma_{ij}$ in plot V. In other words, for two functionally identical plots, the abundance of the species similar to j in plot U is equal to the abundance of the species similar to j in plot V (for definitions and proofs see Appendix 1). In contrast, for two maximally distinct assemblages, either z_{Uj} or z_{Vj} is equal to zero, meaning that the species in plot U do not have any functional analogue in plot V.

Hence, in principle, we could calculate a measure of functional dissimilarity among plots with any of the many traditional dissimilarity measures developed for species abundance data by simply replacing the species abundances x_{Uj} and x_{Vj} in both plots with their corresponding species ordinariness z_{Uj} and z_{Vj}. The resulting measures meet the foremost requirements for a dissimilarity coefficient in the range [0, 1]: for two maximally dissimilar plots (i.e. two plots with no species in

common and $\sigma_{ij} = 0$ for species i belonging to the first plot and species j to the second plot), the measure takes the value one, whereas for two functionally identical plots the measure takes the value zero (see Pavoine and Ricotta, 2014).

2.3. An additional requirement for functional dissimilarity

A potential drawback of this approach is that the dissimilarity indices obtained by substituting the species abundances x_{Uj} and x_{Vj} in U and V with the species ordinariness z_{Uj} and z_{Vj} tend to overemphasize the relevance of the most functionally redundant species. For example, consider two plots U and V and two maximally dissimilar species i and j with $\sigma_{ij} = 0$. Plot U contains species i with abundance $x_{Ui} = 5$ and species j also with abundance $x_{Uj} = 5$, while plot V contains only species j with abundance $x_{Vj} = 10$:

	i	j
Plot U	5	5
Plot V	0	10

Since the two species are maximally dissimilar, their ordinariness is identical to their abundances: $x_{Ui} = z_{Ui} = 5$, $x_{Uj} = z_{Uj} = 5$, and $x_{Vj} = z_{Vj} = 10$. Therefore, using the species ordinariness for calculating the functional dissimilarity between U and V with the Bray-Curtis coefficient (see Eq. (3)), we obtain $BC_z = 0.5$.

However, if we split species i into two functionally identical species i' and i'', each with abundance $x_{Ui'} = x_{Ui''} = 2.5$ and $\sigma_{i'i''} = 1$:

	i'	i''	j
Plot U	2.5	2.5	5
Plot V	0	0	10

we have $z_{Ui'} = z_{Ui''} = 5$, $z_{Uj} = 5$ and $z_{Vj} = 10$. Accordingly, $BC_z = 0.6$, whereas, intuitively, the functional dissimilarity between U and V should not change if one species is replaced by two identical species with the same total abundance.

More generally, irrespective of the values of the pairwise similarities among species pairs, the species ordinariness overemphasizes the role of the functionally redundant species in both plots. If the species in U are functionally similar to the species in V, the similarity between U and V will be exaggerated. To the contrary, if the species in U are functionally different from the species in V, the similarity between U and V will be excessively reduced.

Therefore, a meaningful measure of functional dissimilarity between pairs of plots needs to conform to the following branching requirement: the value of functional dissimilarity among a pair of plots U and V should not change if a given species j is replaced by two functionally identical species with the same total abundance of j in U and V.

In Appendix 1, we show that a special case of the general family of dissimilarity coefficients D_π conforms to this additional requirement. If the single-species dissimilarities d in Eq. (7) are calculated from the species ordinariness z_{Uj}, while the weighting factors w_j are calculated from the original species abundances $w_j = (x_{Uj} + x_{Vj})/\sum_{k=1}^{N}(x_{Uk} + x_{Vk})$, we get a family of measures:

$$D_w = \sum_{j=1}^{N} w_j d(z_{Uj}, z_{Vj}) \tag{12}$$

which conforms to the branching requirement for functional dissimilarity coefficients (proof in Appendix 1):

Take for example the generalized version of the Bray-Curtis coefficient. If this measure is calculated as $BC_w = \sum_{j=1}^{N} w_j |z_{Uj} - z_{Vj}|/(z_{Uj} + z_{Vj})$, it is easily shown that BC_w conforms to the branching requirement. The same holds for all functional dissimilarity coefficients that can be expressed according to Eq. (12).

3. Worked example

We analyzed the functional composition of plant communities in the coastal marsh plain of La Mafragh (36°48′ N – 008°00′ E) in Algeria. Detailed information on the data set can be found in de Bélair (1981) and Pavoine et al. (2011). The area is a basin filled by alluvial and colluvial deposits and furrowed by rivers. The lowest parts are composed of large and small marshes. The study area is bounded by dunes with a narrow connection (Oued Mafragh) to the Mediterranean Sea in the north , by Numidian clay-sandstone mountains in the south, by a river (Oued El Kebir) in the east, and by an irrigated agricultural zone in the west. It is about 15 km large from east to west and 8 km from north to south.

97 study sites were regularly distributed on the plain. de Bélair (1981) collected species abundances and environmental data (de Bélair 1981), and Pavoine et al. (2011) collected species traits from the literature. As in Pavoine et al. (2011), we focused our analyses on four traits: life cycle (categorical multichoice variable with four levels [perennial; annual; biennial; seasonal]), pollination (categorical multichoice variable with three levels [autogamous, entomogamous, anemogamous]), presence of spiky structures (ordinal variable with three levels [0 = no; 1 = sometimes; 2 = yes]), and hairy leaves (ordinal variables with three levels [0 = no; 1 = sometimes; 2 = yes]). We calculated the functional dissimilarities between species as the average dissimilarity over all traits (see Pavoine et al., 2009). We used the Sørensen index for categorical multichoice traits and the Euclidean distance on rank-transformed ordinal variables.

To obtain species-based dissimilarities among plots, we applied the traditional Bray-Curtis dissimilarity (Eq. (4)) to the species relative abundances. Then, to obtain functional dissimilarities among plots, we applied the functional version of the Bray-Curtis dissimilarity to the species ordinariness z_{Uj} (see Eq. (12)). We then analyzed these plot-to-plot dissimilarities with Principal Coordinate Analysis (PCoA) after Cailliez (1983) transformation. This operation consists in adding the smallest constant to all dissimilarities to make the resulting dissimilarity matrix Euclidean. We finally compared the results obtained from the PCoA with the spatial distribution of eight environmental variables (Fig. 1): elevation [m], clay [%], silt [%], sand [%], K₂O [‰], Mg²⁺ [mEq/100 g], Na⁺ [mEq/100 g], K⁺ [mEq/100 g]. The R scripts developed for this analysis are provided in Appendix 2.

4. Results

We retained the first axis of the PCoA applied to plot-to-plot functional dissimilarities (Fig. 2) as its eigenvalue is about five-fold that of the second axis (Fig. 3). The high number of eigenvalues close to 0.006 in this analysis (Fig. 3A) is due to the Cailliez transformation used to render the dissimilarities Euclidean. Similar patterns were also obtained using non-metric Multidimensional Scaling (Appendix 2), meaning that the results were not affected by the ordination method used for analyzing the data. According to Figs. 1 and 2, the negative part of the PCoA axis corresponded mainly to plots at lower elevation in the middle of the study area where clay-rich soils had high concentrations of K+, Mg++, Na+ and K₂O, indicating high salinity. On the other hand, plots with positive coordinates tended to be located at higher elevations on soils with comparatively higher proportions of silt and sand and with lower salinity levels, mostly on the West side of the plain.

Looking only at species-based dissimilarities, without considering trait data, the first eigenvalue of the PCoA is about twice as large as the second eigenvalue (Fig. 3B). In that case, the sites were roughly divided into two groups: those in the central part of the plain and those on the East and West side (Fig. 4A). Compared to the smoother gradient obtained with trait data, this pattern is more sensitive to changes in the elevation of the study area and less to the salinity gradient.
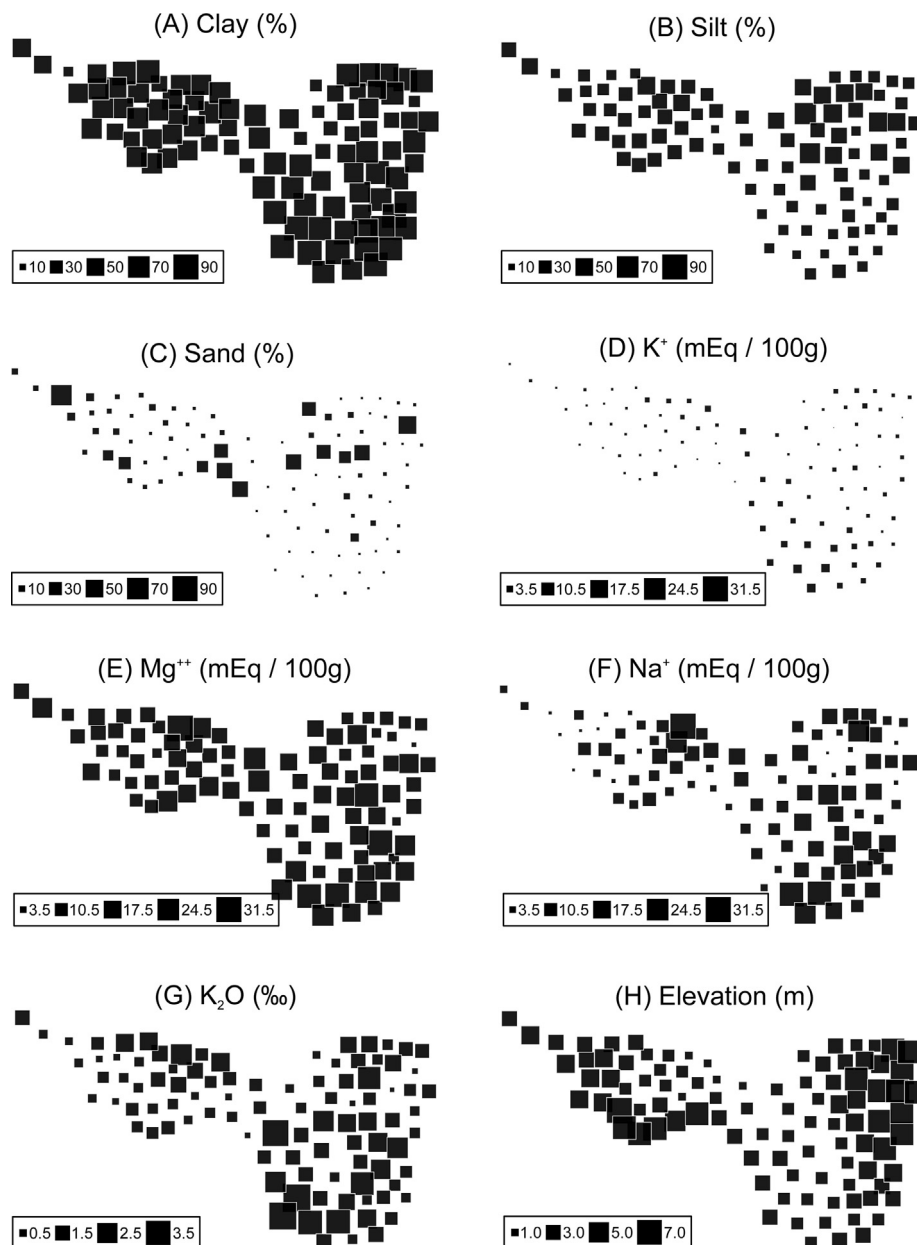
Fig. 1. Environmental patterns in the coastal marsh plain of La Mafragh. The locations of the squares correspond to the geographic positions of the sample plots. The square size provides the level of the environmental variables. North is at the top of the panels.

We finally compared the plot-to-plot dissimilarities obtained solely from species abundances with the corresponding functional dissimilarities. As expected, relaxing the constraint that all species are equally and maximally distinct from one another, trait-based dissimilarities were always lower than species-based dissimilarities (Fig. 4B). The functional dissimilarities among sites with no species in common (species-based dissimilarity = 1) varied between about 0 and 0.5 (Fig. 4B).

5. Discussion

Ecological data are often multivariate of high dimensionality. Therefore, no single measure adequately summarizes all aspects of (functional) dissimilarity among plots. In order to assess differences in the functional organization of species assemblages, a multifaceted approach is needed. In this paper, we thus proposed a new method for adapting traditional dissimilarity coefficients to the measurement of

functional differences among plots. The resulting measures generalize species dissimilarity to include functional information. As such, they allow plot-to-plot comparisons at different levels of ecological complexity (i.e. compositional vs. functional dissimilarity), while avoiding confounding effects which may be due to differences in the calculation of the indices (Ricotta and Pavoine, 2015).

The basic unit for the calculation of functional dissimilarity is the species ordinariness z_{Uj}. In Appendix 3 we show that, in the special case where the species ordinariness are calculated from the relative abundances $p_{Uj} = x_{Uj}/\sum_{j=1}^{N} x_{Uj}$ of each species, the resulting values of z_{Uj} are tightly related to the notion of species rarity used by Patil and Taillie (1982) for providing a generalized definition of diversity measures. As such, our proposal may provide a starting point for developing a unified framework for the summarization of functional diversity and dissimilarity.

Note that species ordinariness was already used by Ricotta et al. (2015) for developing a single Bray and Curtis-like index of functional
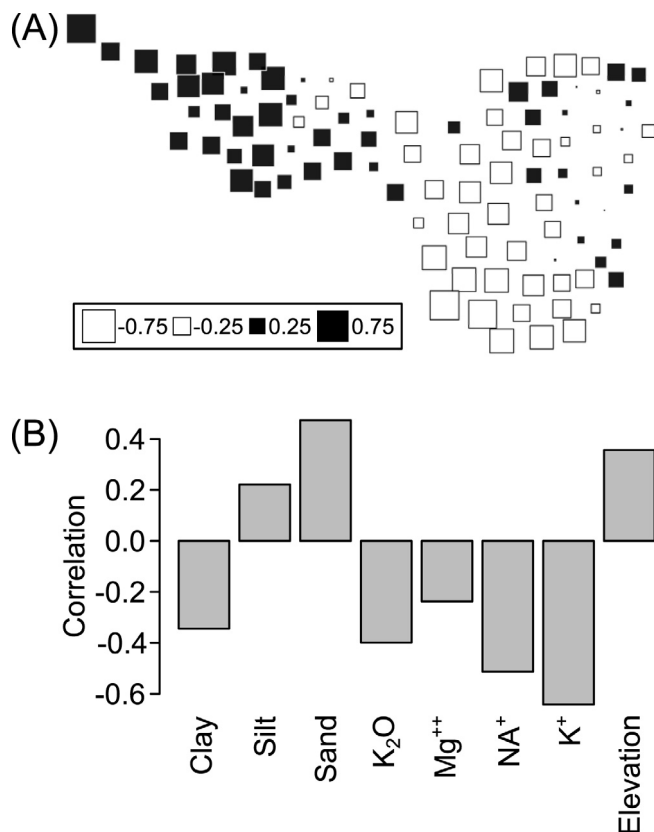
Fig. 2. Results of the principal coordinate analysis applied to plot-to-plot functional dissimilarities after Cailliez transformation. (A) First axis of the PCoA applied to plot-to-plot functional dissimilarities. (B) Bar plot of the Spearman correlations between the first axis of the PcoA and the environmental variables.
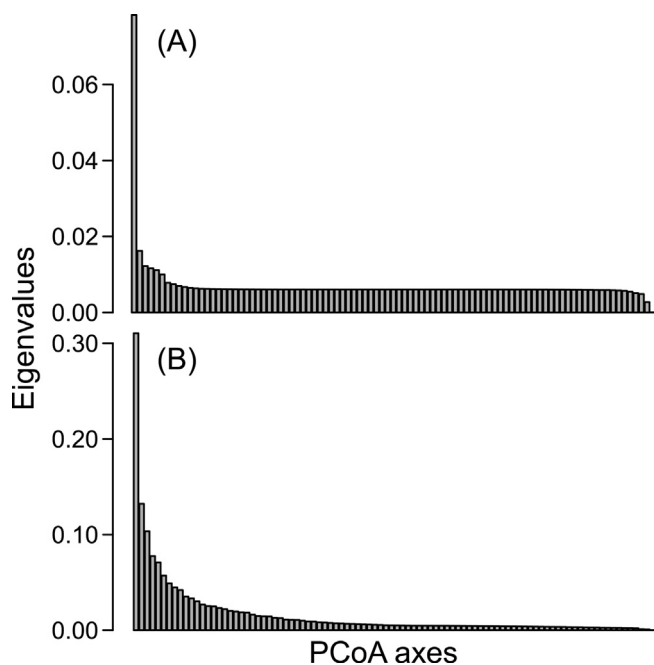

Fig. 3. Eigenvalue bar plot of the principal coordinate analysis applied to (A) plot-to-plot functional dissimilarities and (B) plot-to-plot species-based dissimilarities, after Cailliez transformation.

dissimilarity and by Ricotta and Pavoine (2015) for constructing a family of functional dissimilarity measures based on a generalization of the classical 2 × 2 contingency table to include species abundances and
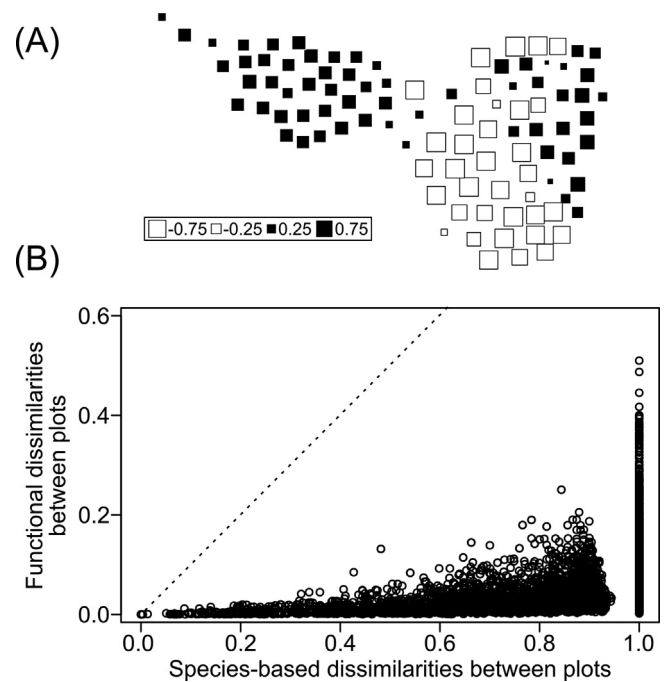

Fig. 4. Species-based dissimilarities among plant communities in the coastal marsh plain of La Mafragh. (A) First axis of the principal coordinate analysis applied to species-based dissimilarities among plots. (B) Relationship between functional and species-based dissimilarities among plots. The dashed line corresponds to the theoretical scenario where species-based and functional dissimilarities would have equal values.

interspecies resemblances. However, these measures do not conform to the newly proposed branching requirement for functional dissimilarity coefficients.

Given two assemblages U and V, the major difference between classical species abundance values x_{Uj} and the corresponding species ordinariness z_{Uj} is that the former are stand-alone quantities, which are independent of the abundances of the other species in U and V, whereas the ordinariness of species j in plot U depends not only on its abundance x_{Uj}, but also on its functional resemblance with the other species in U and V. Hence, using species ordinariness for the calculation of plot-to-plot functional dissimilarity, we implicitly assume that species with similar traits are likely to support similar functions, such that the selection of the most appropriate traits needs to be explicitly targeted on the ecological process of interest. Also, although the formulation of species uniqueness does not consider within-species differentiation, it is commonly agreed that species differences in trait values are related to different mechanisms of resource use, such that the higher the species overlap in functional space, the higher the species are expected to support similar functions. Therefore, using dissimilarity measures based on functional overlap among species (Lepš et al., 2006; Blonder et al., 2014), will allow us to include within-species differentiation in the calculation of plot-to-plot functional dissimilarity.

To conclude, we hope our proposal will prove fruitful for summarizing plot-to-plot functional dissimilarity in a meaningful way and for getting insights into the complex ecological mechanisms that drive community structure.

6. Author contribution statement

CR formulated the idea. SP and CR developed the methodology. SP wrote the R code and analyzed the data. SP and CR wrote the manuscript.

7. Compliance with ethical standards

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ecolind.2018.09.048.

References

Blonder, B., Lamanna, C., Violle, C., Enquist, B.J., 2014. The n-dimensional hypervolume. Glob. Ecol. Biogeogr. 23, 595–609.

Bray, J.R., Curtis, J.T., 1957. An ordination of the upland forest communities of southern Wisconsin. Ecol. Monogr. 27, 325–349.

Cailliez, F., 1983. The analytical solution to the additive constant problem. Psychometrika 48, 305–308.

Champely, S., Chessel, D., 2002. Measuring biological diversity using Euclidean metrics. Environ. Ecol. Stat. 9, 167–177.

Chiu, C.-H., Jost, L., Chao, A., 2014. Phylogenetic beta diversity, similarity, and differentiation measures based on Hill numbers. Ecol. Monogr. 84, 21–44.

Clarke, K.R., Warwick, R.M., 1998. A taxonomic distinctness index and its statistical properties. J. Appl. Ecol. 35, 523–531.

de Bélair, G., 1981. Biogéographie et aménagement: la plaine de La Mafragh (Annaba, Algérie). PhD Thesis. Université Paul Valéry, Montpellier.

Izsák, C., Price, R.G., 2001. Measuring β-diversity using a taxonomic similarity index, and its relation to spatial scale. Mar. Ecol. Prog. Ser. 215, 69–77.

Lance, G.N., Williams, W.T., 1967. Mixed data classificatory programs. I. Agglomerative systems. Aust. Comput. J. 1, 15–20.

Legendre, P., Legendre, L., 2012. Numerical Ecology. Elsevier, Amsterdam.

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

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.

Mason, N.W.H., de Bello, F., 2013. Functional diversity: a tool for answering challenging ecological questions. J. Veg. Sci. 24, 777–780.

Orlóci, L., 1978. Multivariate Analysis in Vegetation Research. Junk, The Hague, NL.

Patil, G.P., Taillie, C., 1982. Diversity as a concept and its measurement. J. Am. Statist. Assoc. 77, 548–561.

Pavoine, S., Dufour, A.-B., Chessel, D., 2004. From dissimilarities among species to dissimilarities among communities: a double principal coordinate analysis. J. Theor. Biol. 228, 523–537.

Pavoine, S., Ricotta, C., 2014. Functional and phylogenetic similarity among communities. Meth. Ecol. Evol. 5, 666–675.

Pavoine, S., Vallet, J., Dufour, A.-B., Gachet, S., Daniel, H., 2009. On the challenge of treating various types of variables: application for improving the measurement of functional diversity. Oikos 118, 391–402.

Pavoine, S., Véla, E., Gachet, S., de Bélair, G., Bonsall, M.B., 2011. Linking patterns in phylogeny, traits, abiotic variables and space: a novel approach to linking environmental filtering and plant community assembly. J. Ecol. 99, 165–175.

Podani, J., 2000. Introduction to the Exploration of Multivariate Biological Data. Backhuys, Leiden, NL.

Rao, C.R., 1982. Diversity and dissimilarity coefficients: a unified approach. Theor. Popul. Biol. 21, 24–43.

Ricotta, C., 2018. A family of (dis)similarity measures based on evenness and its relationship with beta diversity. Ecol. Complex 34, 69–73.

Ricotta, C., Bacaro, G., Pavoine, S., 2015. A cautionary note on some phylogenetic dissimilarity measures. J. Plant Ecol. 8, 12–16.

Ricotta, C., Pavoine, S., 2015. Measuring similarity among plots including similarity among species: an extension of traditional approaches. J. Veg. Sci. 26, 1061–1067.

Ricotta, C., Podani, J., 2017. On some properties of the Bray-Curtis dissimilarity and their ecological meaning. Ecol. Complex 31, 201–205.

Ricotta, C., Podani, J., Pavoine, S., 2016. A family of functional dissimilarity measures for presence and absence data. Ecol. Evol. 6, 5383–5389.

Appendix 1. Definitions and Proofs

Hereafter, we consider two plots U and V, where x_{Uj} and x_{Vj} are the abundance values of species j in plots U and V, respectively, N is the number of species of the pooled pair of plots (i.e. all species with non-zero abundance in at least one plot), and σ_{ij} is a symmetric similarity coefficient between species i and j bounded in the range $[0,1]$ with $\sigma_{ij} = \sigma_{ji}$ and $\sigma_{ii} = 0$.

Two species i and j will be said functionally identical if their functional similarity is maximal: $\sigma_{ij} = 1$.

Two plots U and V with no shared species are said functionally identical if for any species j (or any set of functionally identical species j, j', j'', ...) in plot U, a species k (or a set of functionally identical species k, k', k'', ...) exist in plot V such that $x_{Uj} = x_{Vk}$ and $\sigma_{jk} = 1$ (or $x_{Uj} + x_{Uj'} + x_{Uj''} + ... = x_{Vk} + x_{Vk'} + x_{Vk''} + ...$ and $\sigma_{jk} = 1$, in case of functionally identical species in U and/or in V), and if similarly for any species k (or a set of functionally identical species k, k', k'' , ...) in plot V, a species j (or any set of functionally identical species j, j', j'', ...) exist in plot U such that $x_{Uj} = x_{Vk}$ and $\sigma_{jk} = 1$ (or $x_{Uj} + x_{Uj'} + x_{Uj''} + ... = x_{Vk} + x_{Vk'} + x_{Vk''} + ...$ and $\sigma_{jk} = 1$, in case of functionally identical species in U and/or in V).

1. For two identical plots U and V, the ordinariness z_{Uj} of species j in plot U is equal to the ordinariness z_{Vj} of species j in plot V.

For two identical plots, $x_{Uj} = x_{Vj}$ for all N species in U and V. Therefore, since $z_{Uj} = \sum_{i=1}^{N} x_{Ui} \sigma_{ij}$ (see Eq. 11 in the main text):

$$z_{Uj} = \sum_{i=1}^{N} x_{Ui} \sigma_{ij} = \sum_{i=1}^{N} x_{Vi} \sigma_{ij} = z_{Vj} \qquad \qquad (A1)$$

2. For two functionally identical plots U and V with no shared species, the ordinariness of species j in plot U is equal to the ordinariness of species j in plot V.

Consider the simplest case where U does not contain functionally identical species such that for all i and j in U, $\sigma_{ij} < 1$, and similarly V does not contain functionally identical species such that for all m and n in V, $\sigma_{mn} < 1$. In that case, functionally identical plots necessarily have the same number of species. In addition, for each pair of species i and j in U, we have a functionally identical pair of species m and n in V such that $x_{Ui} = x_{Vm}$, $\sigma_{im} = 1$, $x_{Uj} = x_{Vn}$, $\sigma_{jn} = 1$ and for all k in U and in V, $\sigma_{ik} = \sigma_{mk}$ and $\sigma_{kj} = \sigma_{kn}$. Therefore, we have:

$$z_{Uj} = \sum_{i \in U} x_{Ui} \sigma_{ij} \qquad \qquad (A2)$$

and

$$z_{Vj} = z_{Vn} = \sum\nolimits_{m \in V} x_{Vm} \sigma_{mn} = \sum\nolimits_{i \in U} x_{Ui} \sigma_{in} = \sum\nolimits_{i \in U} x_{Ui} \sigma_{ij} = z_{Uj} \qquad (A3)$$

The demonstration for the more general case (where some species are allowed to be functionally identical within U and/or within V) can simply be obtained by pooling a set of functionally identical species within a plot into a single super-species with its abundance equal to the summed abundance of the functionally identical species. Indeed, the ordinariness values of functionally identical species in a plot are equal to each other, and are also equal to the ordinariness of the super-species (see proof at point 4 below, Eq. A8 and A9). For example, let $J=\{j, j', j'', ...\}$ be such a super-species in U, such that $x_{UJ} = x_{Uj} + x_{Uj'} + x_{Uj''} + ...$. If U and V are functionally identical, then there exist a species or a set of functionally identical species K in V such that $x_{UJ} = x_{VK}$ and $\sigma_{JK} = 1$.

3. For two maximally distinct plots with no species in common and with $\sigma_{jm} = 0$ for species j belonging to plot U and species m to plot V, z_{Vj} is always equal to zero, meaning that the species in U do not have any functional analogue in V, and z_{Um} is always equal to zero, meaning that the species in V do not have any functional analogue in U

Observing that

$$z_{Vj} = \sum\nolimits_{m \in V} x_{Vm} \sigma_{jm} \qquad (A4)$$

and

$$z_{Um} = \sum\nolimits_{j \in U} x_{Uj} \sigma_{jm} \qquad (A5)$$

Since $\sigma_{jm} = 0$ by definition, we have $z_{Vj} = z_{Um} = 0$.

4. The general family of functional dissimilarity measures $D_w = \sum\nolimits_{j=1}^{N} w_j d\left(z_{Uj}, z_{Vj}\right)$, with $w_j = \left(x_{Uj} + x_{Vj}\right) \Big/ \sum\nolimits_{k=1}^{N} \left(x_{Uk} + x_{Vk}\right)$ (see Eq. 5 and 12 in the main text) conforms to the following branching requirement: the value of functional dissimilarity D_w among a pair of plots U and V does not change if a given species j is replaced by two functionally identical species with the same total abundance of j in U and V.

If species j is split into two species j' and j'' which are functionally identical to j such that $x_{Uj'} + x_{Uj''} = x_{Uj}$, $x_{Vj'} + x_{Vj''} = x_{Vj}$. Because j, j' and j'' are functionally identical $\sigma_{j'j''} = \sigma_{j'j} = \sigma_{j''j} = 1$, and for any species i in U and/or V, $\sigma_{ij} = \sigma_{ij'} = \sigma_{ij''}$. In addition, we have:

$$w_{j'} + w_{j''} = w_j \qquad (A6)$$

and

$$z_{Uj} = \sum\nolimits_{i}^{N} x_{Ui} \sigma_{ij} = \sum\nolimits_{i}^{N} x_{Ui} \sigma_{ij'} = \sum\nolimits_{i}^{N} x_{Ui} \sigma_{ij''} \qquad (A7)$$

2

It follows that

$$z_{Uj'} = z_{Uj''} = z_{Uj} \tag{A8}$$

and similarly

$$z_{Vj'} = z_{Vj''} = z_{Vj} \tag{A9}$$

This yields

$$d\left(z_{Uj'}, z_{Vj'}\right) = d\left(z_{Uj''}, z_{Vj''}\right) = d\left(z_{Uj}, z_{Vj}\right) \tag{A10}$$

Therefore, since $d\left(z_{Uj'}, z_{Vj'}\right)w_{j'} + d\left(z_{Uj''}, z_{Vj''}\right)w_{j''} = d\left(z_{Uj}, z_{Vj}\right)w_j$, it follows that D_w does not change if a given species j is split into two functionally identical species.

Appendix 2. R scripts

The R function "generalized_Tradidiss" calculates plot-to-plot dissimilarity taking account of functional dissimilarities between species. This program is free software: you can redistribute it and/or modify it under the terms of the GNU General Public License http://www.gnu.org/licenses/. It will be integrated in a version of the adiv package after the publication of the associated paper.

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

Description: given a matrix of S species' relative or absolute abundance values $\times$ N plots, together with an $S \times S$ (functional) dissimilarity matrix, the function "generalized_Tradidiss" calculates a semimatrix with the values of the dissimilarity index for each pair of plots, as proposed in the main text.

Dependencies: none.

Usage:
generalized_Tradidiss <-
function(comm, dis, method = c("GC", "MS", "PE"), abundance = c("relative", "absolute", "none"), weights = c("uneven", "even"), tol = 1e-8)

Arguments:
comm: a matrix of N plots $\times$ S species containing the relative or absolute abundance of all species. Columns are species and plots are rows. Column labels (species names) should be assigned as in the *dis* matrix.

dis: an $S \times S$ matrix of dissimilarities between species rescaled in the range [0, 1] or an object of class "dist" [obtained by functions like vegdist in package vegan (Oksanen *et al.* 2013), gowdis in package FD (Laliberté & Shipley 2011), or dist.ktab in package ade4 for functional dissimilarities (Dray *et al.* 2007), or functions like cophenetic.phylo in package ape (Paradis *et al.* 2004) or distTips in package adephylo (Jombart and Dray 2010) for phylogenetic dissimilarities]. If the dissimilarities are outside the range 0-1, a warning message is displayed and each dissimilarity is divided by the maximum over all pair-wise dissimilarities.

method: a character string giving the name of the coefficient of plot-to-plot dissimilarity to be applied to species ordinariness (Z vectors defined in the main text, Eq. 11). It can be "GC", "MS", or "PE". See details.

abundance: a character string with three possible values: "relative" for the use of relative species abundance, "absolute" for the use of absolute species abundance, and "none" for the use of presence/absence data (1/0).

weights: Two types of weights are available in the function: "uneven" (Eq. 5 in the main text) or "even" (1/S, where S is the number of species in the two compared plots)

tol: a numeric tolerance threshold: a value between -tol and tol is considered as null.

Details:

The plot-to-plot dissimilarity coefficients used in this function are those listed in the main text. Codes are as follows:

"GC": Equation 6 in the main text

"MS": Equation 8 in the main text

"PE": Equations 9 and 10 in the main text

Value:

The function returns a semi-matrix of class "dist" with the values of the proposed dissimilarities for each pair of plots.

Function Syntax

```
generalized_Tradidiss <-
function(comm, dis, method = c("GC", "MS", "PE"), abundance = c("relative",
"absolute", "none"), weights = c("uneven", "even"), tol = 1e-8)
{
    if(inherits(dis, "dist")) dis <- as.matrix(dis)
    if(any(dis>1)){
        warning("Phylogenetic dissimilarities are not in the range 0-1. They have
been normalized by the maximum")
        dis <- dis/max(dis)
    }
    if(any(!colnames(comm) %in%rownames(dis))) stop("At least one species in the
matrix of abundances is missing in the matrix of dissimilarities")
    if(any(!colnames(comm) %in%colnames(dis))) stop("At least one species in the
matrix of abundances is missing in the matrix of dissimilarities")
    method <- method[1]
    if(!method%in%c("GC", "MS", "PE")) stop("Incorrect definition for method")
    abundance <- abundance[1]
    if(!abundance%in%c("relative", "absolute", "none"))
stop("Incorrect definition for abundance")
    weights <- weights[1]
    if(!weights%in%c("uneven", "even")) stop("Incorrect definition for weights")
    dis <- dis[colnames(comm), colnames(comm)]
    dataset <- t(comm)
    similarities <- 1-as.matrix(dis)
    total <- colSums(dataset)
    if(abundance == "relative")
        abu <- sweep(dataset, 2, total, "/")
    else if(abundance == "none"){
        abu <- dataset
        abu[abu>tol] <- 1
        abu[abu<=tol] <- 0
    }
    else abu <- dataset
    num.plot<-dim(dataset)[2]
    num.sp <- dim(dataset)[1]
    names<-list(colnames(dataset), colnames(dataset))
    dis.matrix<-matrix(0, nrow=num.plot, ncol=num.plot, dimnames=names)
    for (i in 2:num.plot) {
    for (j in 1:(i-1)) {
        mat_folk <- similarities*abu[, j]
        mat_folk2 <- similarities*abu[, i]
        Zik <- colSums(mat_folk)
        Zih <- colSums(mat_folk2)
        tabZ <- rbind.data.frame(Zik, Zih)
        garde <- apply(tabZ, 2, sum)>tol & apply(abu[,c(i,j)], 1, sum)>tol
        Zik <- Zik[garde]
        Zih <- Zih[garde]
        tabZ <- tabZ[, garde]
        if(weights == "even") wk <- rep(1/length(Zih), length(Zih))
        else {
        wk <- (abu[, j]+abu[, i])/sum(abu[, c(i,j)])
        wk <- wk[garde]
        }
```

```
        if(method == "CG") index <- sum((wk*abs(Zik-Zih)/apply(tabZ, 2, sum)))
        else if(method == "MS") index <- sum((wk*abs(Zik-Zih)/apply(tabZ, 2, max)))
        else {
            zik <- Zik / (Zik + Zih)
            zih <- Zih / (Zik + Zih)
            tabz <- rbind.data.frame(zik, zih)
            enlev <- as.vector(apply(tabz, 2,
                function(x) any(x<=tol)))
            zik <- zik[!enlev]
            zih <- zih[!enlev]
            wkN <- wk[!enlev]
            wkY <- wk[enlev]
            index <- sum(wkY) + sum(wkN*(1
                +(zik * log(zik, base = 2)
                + zih * log(zih, base = 2))))
        }
        dis.matrix[i, j] <- index
    }
    }
    dis.matrix <- dis.matrix + t(dis.matrix)
    dis.matrix <- dis.matrix + diag(rep(0, num.plot))
    return(as.dist(dis.matrix))
}
```

Case study

Loading of package ade4 of R:
```
library(ade4)
```

Loading of the data set:
```
data(mafragh)
```

Species names:
```
namspe <- rownames(mafragh$traits[[1]])
```

Matrix with the abundance of plant species (columns) in plots (rows):
```
M <- mafragh$flo
colnames(M) <- namspe
```

Calculation of the functional dissimilarities between species (bounded between 0 and 1):
```
Bin <- prep.binary(mafragh$traits$tabBinary, c(3, 4))
distraits <- dist.ktab(ktab.list.df(list(mafragh$traits$tabOrdinal[,2:3],
Bin)), c("O","B"), scan=TRUE)
Choose your metric for binary variables
1 = JACCARD index (1901) S3 coefficient of GOWER &
LEGENDRE
s1 = a/(a+b+c) --> d = sqrt(1 - s)
2 = SOCKAL & MICHENER index (1958) S4 coefficient of GOWER & LEGENDRE
s2 = (a+d)/(a+b+c+d) --> d = sqrt(1 - s)
3 = SOCKAL & SNEATH(1963) S5 coefficient of GOWER &
LEGENDRE
s3 = a/(a+2(b+c)) --> d = sqrt(1 - s)
4 = ROGERS & TANIMOTO (1960) S6 coefficient of GOWER &
LEGENDRE
s4 = (a+d)/(a+2(b+c)+d) --> d = sqrt(1 - s)
5 = CZEKANOWSKI (1913) or SORENSEN (1948) S7 coefficient of GOWER &
LEGENDRE
s5 = 2*a/(2*a+b+c) --> d = sqrt(1 - s)
6 = S9 index of GOWER & LEGENDRE (1986)
s6 = (a-(b+c)+d)/(a+b+c+d) --> d = sqrt(1 - s)
7 = OCHIAI (1957) S12 coefficient of GOWER & LEGENDRE
s7 = a/sqrt((a+b)(a+c)) --> d = sqrt(1 - s)
```

```
8 = SOKAL & SNEATH (1963) S13 coefficient of GOWER &
LEGENDRE
s8 = ad/sqrt((a+b)(a+c)(d+b)(d+c)) --> d = sqrt(1 - s)
9 = Phi of PEARSON = S14 coefficient of GOWER & LEGENDRE
s9 = (ad-bc)/sqrt((a+b)(a+c)(b+d)(d+c)) --> d = sqrt(1 -
s)
10 = S2 coefficient of GOWER & LEGENDRE
s10 =  a/(a+b+c+d) --> d = sqrt(1 - s) and unit
self-similarity
Select an integer (1-10): 5
Choose your metric for ordinal variables
1 = ranked variables treated as quantitative variables
2 = Podani (1999)'s formula
Select an integer (1-2): 1
Choose your metric for quantitative variables
1 = Euclidean
d1 = Sum((x(i)-y(i))^2)/n
2 = Manhattan
d2= Sum(|x(i)-y(i)|)/n
Select an integer (1-2): 1
```

Application of the Bray-Curtis index (Eq. 4 of the main text) or equivalently index GC (Eq. 6) with uneven weights for species (Eq. 5) to the data set:
```
disGC <- generalized_Tradidiss(M, distraits, method="GC")
```

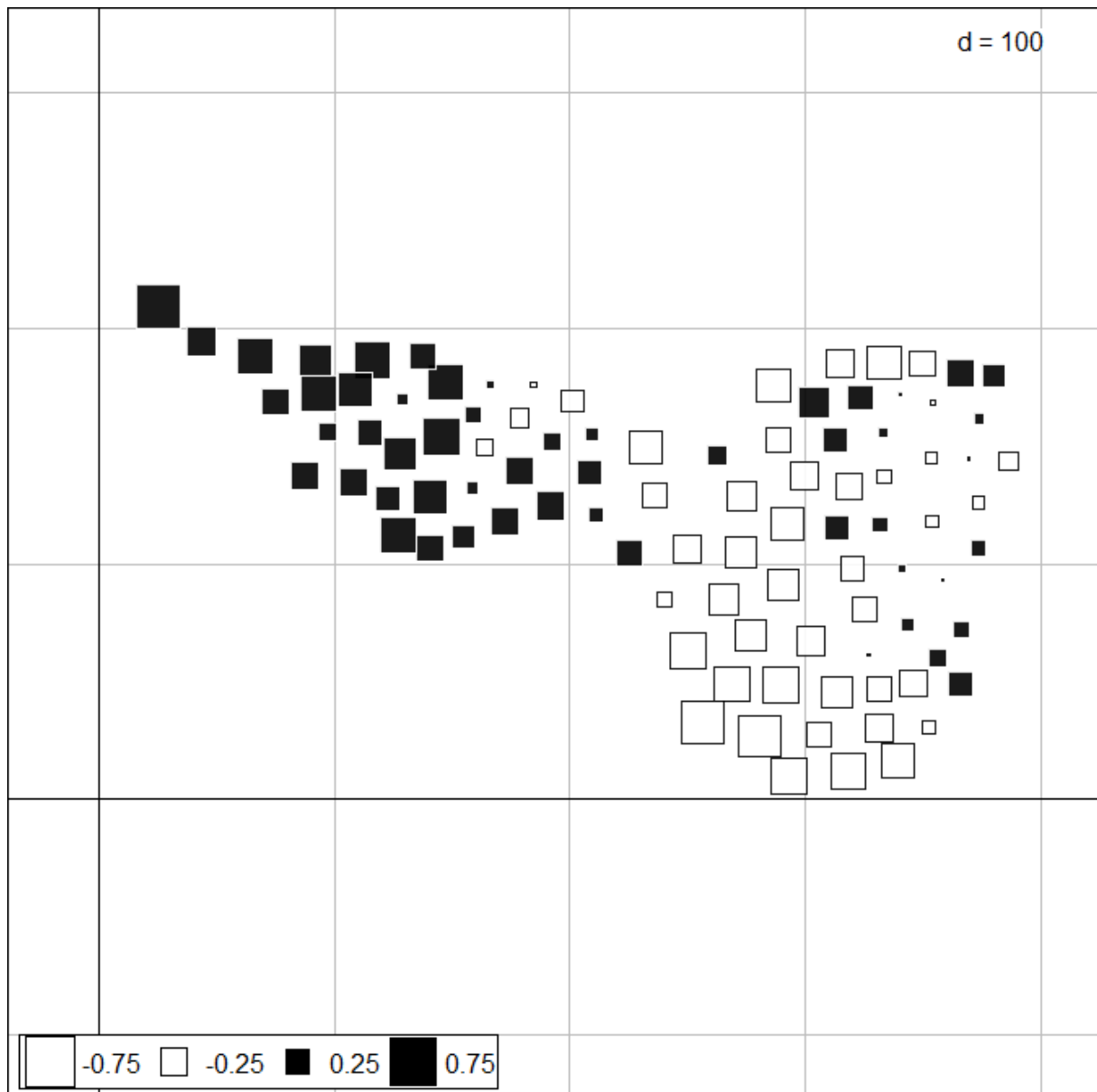The functional dissimilarities between plots do not have the Euclidean properties
```
is.euclid(as.dist(disGC))
[1] FALSE
```

First approach: Principal coordinate analysis (PCoA) on the functional dissimilarities transformed by Cailliez approach (addition to the dissimilarities of the smallest constant that renders the dissimilarities Euclidean)
```
pcoGC <- dudi.pco(as.dist(cailliez(disGC)))
Select the number of axes: 2
```

Coordinates of the species on the first axis of the PCoA (given by the color and size of a square) at their spatial location in the plain La Mafragh (a square is at the geographic position of its associated plot):
```
s.value(mafragh$xy, pcoGC$li[,1])
```

Second approach: non-metric Multidimensional Scaling analysis (nMDS) on the functional dissimilarities

```
CMD <- cmdscale(disGC)
summary(CMD)
            V1                      V2
 Min.   :-0.258573   Min.    :-0.077484
 1st Qu.:-0.060668   1st Qu.:-0.014220
 Median : 0.005213   Median :-0.001209
 Mean   : 0.000000   Mean    : 0.000000
 3rd Qu.: 0.053966   3rd Qu.: 0.010996
 Max.   : 0.291525   Max.    : 0.099926
```

The coordinates on the second axis of the nMDS are low indicating that most of the functional variation between plots is expressed by the first axis.

Spatial pattern of the first axis of the nMDS:
`s.value(mafragh$xy, CMD[,1])`
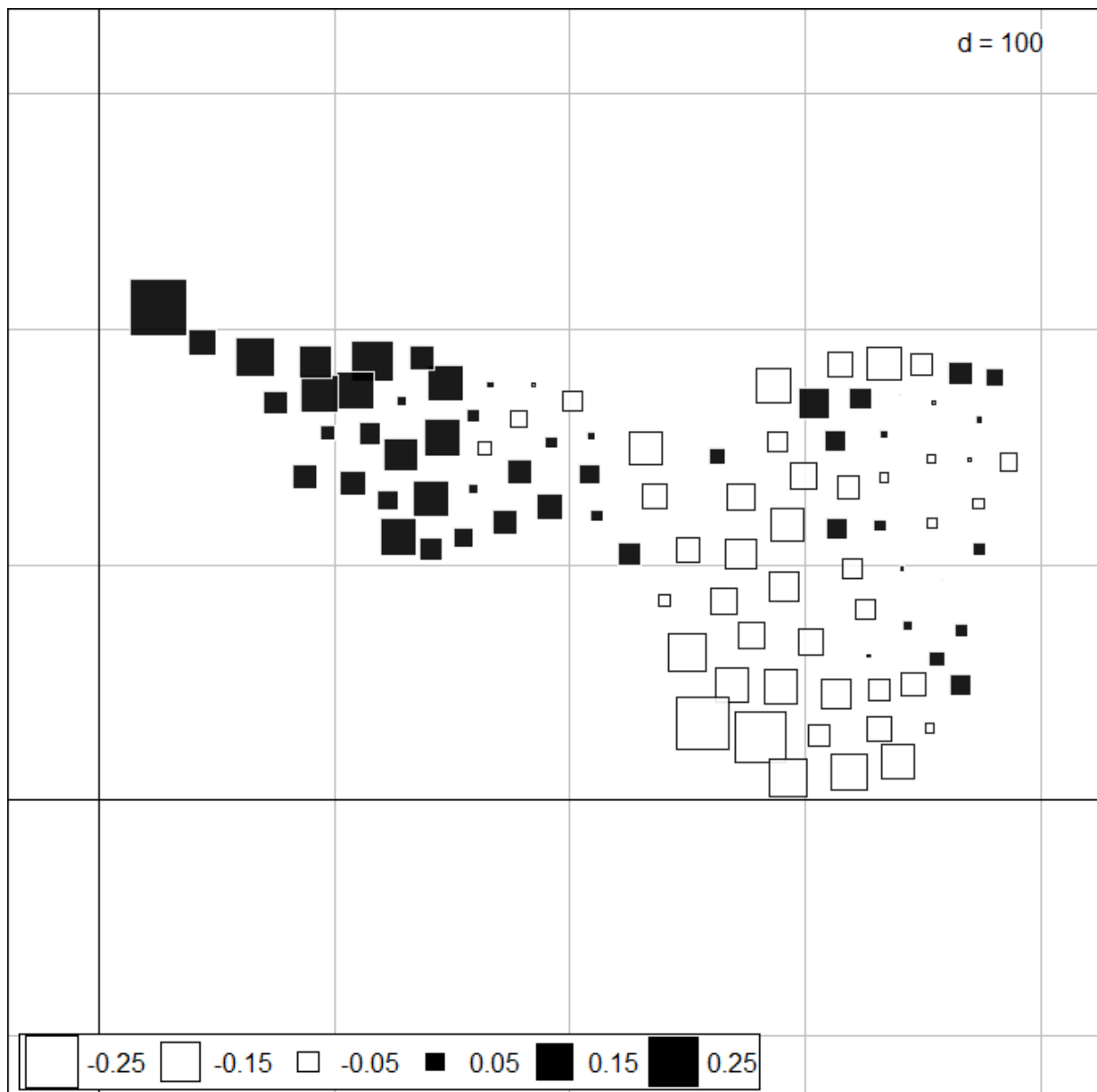
Link between the coordinates of plots in PCoA and those in the nMDS:

`plot(pcoGC$li[,1], CMD[,1], xlab="First axis of PCoA", ylab="First axis of nMDS")`
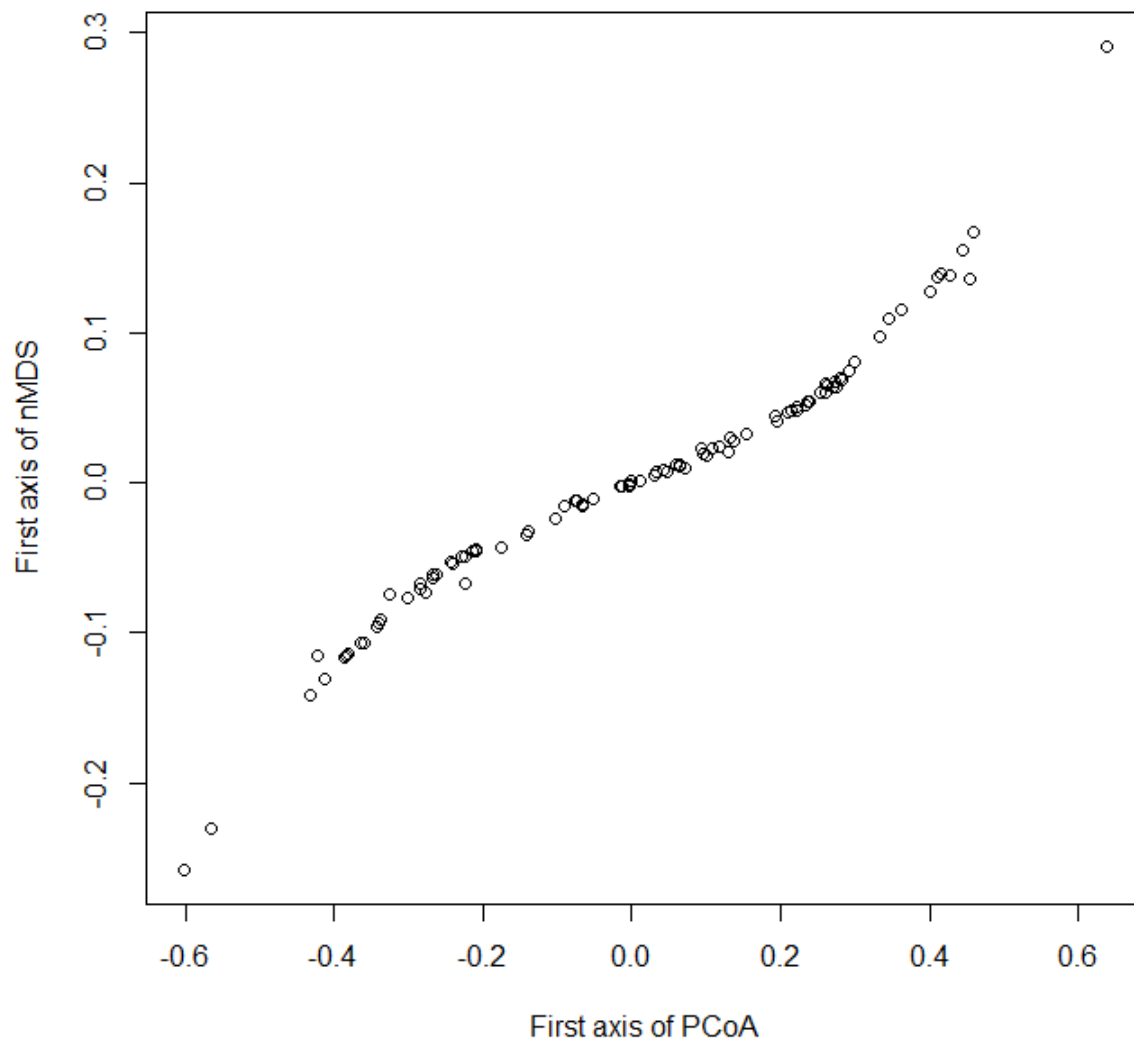
Analysis of the species-based dissimilarities between plots:

Species are defined as maximally dissimilar:
```
DM1 <- as.dist(matrix(1, 56,56))
attributes(DM1)$Labels <- attributes(distraits)$Labels
```

Application of the Bray-Curtis index (Eq. 4 of the main text) or equivalently index GC (Eq. 6) with uneven weights for species (Eq. 5) to the data set with species defined as maximally dissimilar:
```
disM1 <- generalized_Tradidiss(M, DM1)
```

The species-based dissimilarities between plots do not have the Euclidean properties
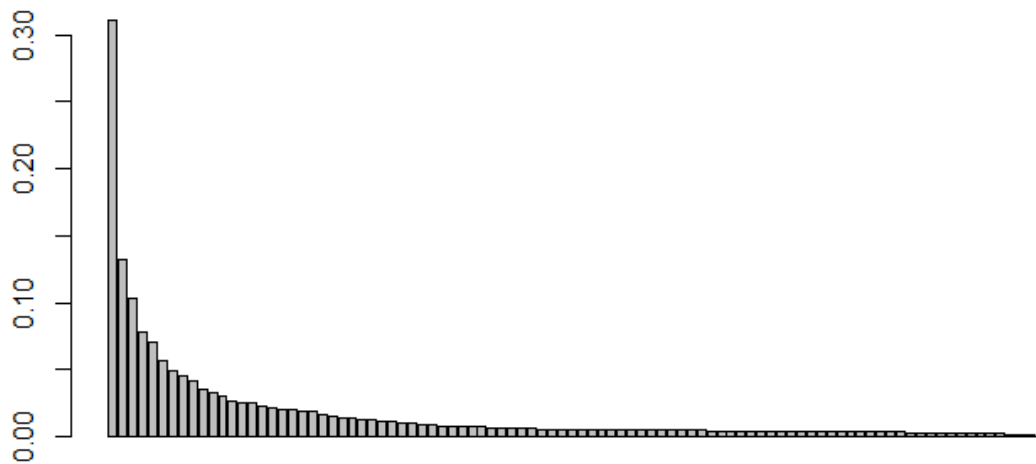```
is.euclid(disM1)
[1] FALSE
```

Principal Coordinate Analysis of the species-based dissimilarities between plots
```
pcoDM1 <- dudi.pco(as.dist(cailliez(disM1)))
```
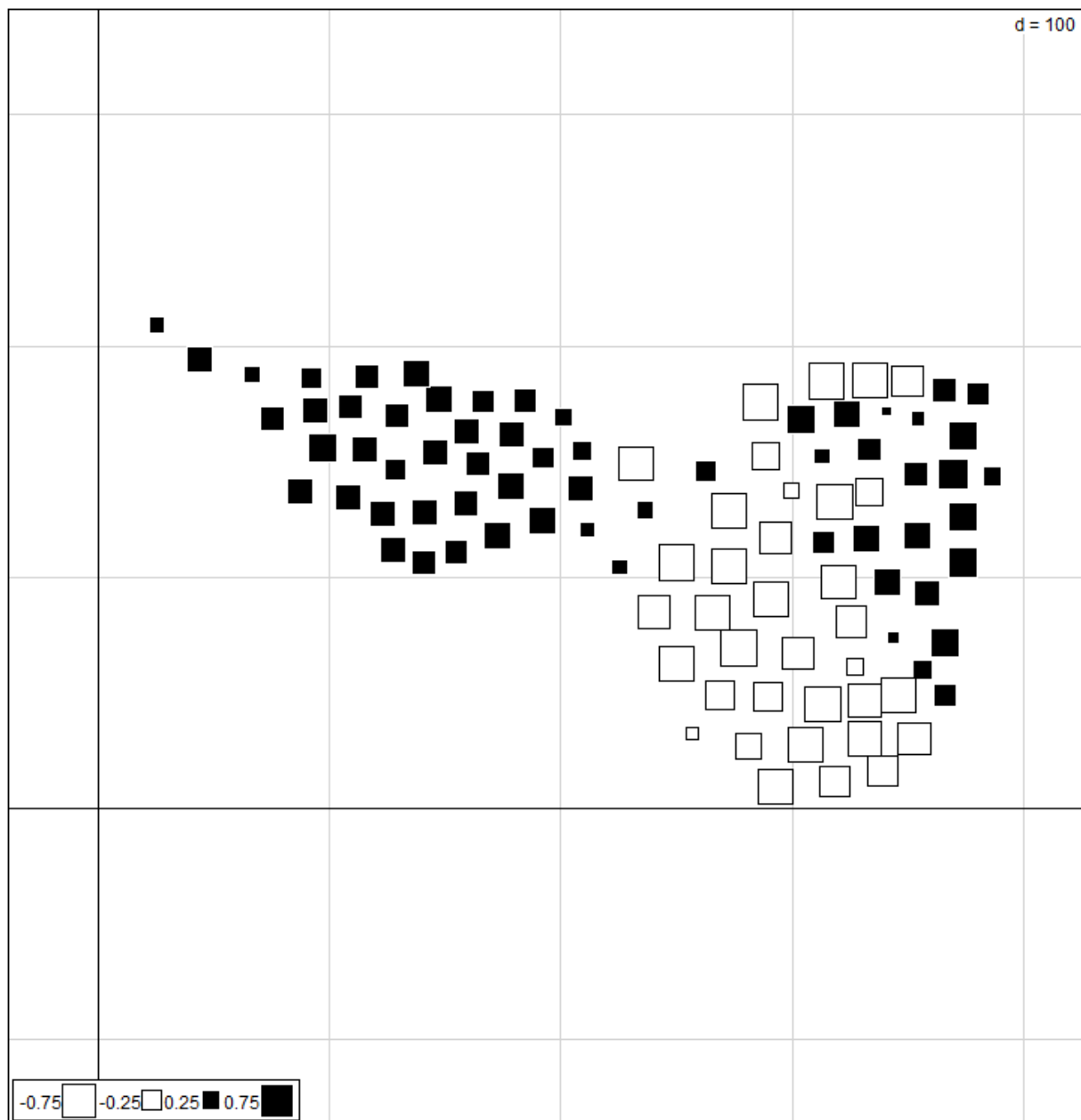
Eigenvalue bar plot:
`barplot(pcoDM1$eig)`



Coordinates of the species on the first axis of the PCoA (given by the color and size of a square) at their spatial location in the plain La Mafragh (a square is at the geographic position of its associated plot):
`s.value(mafragh$xy, pcoDM1$li[,1], csize=0.5)`

**Link between the species-based and the functional dissimilarities between plots:**

```
plot(disM1, disGC, asp=1, ylim=c(0,1), xlab="species-based dissimilarities
between plots", ylab="functional dissimilarities between plots")
lines(c(0,1), c(0,1))
```
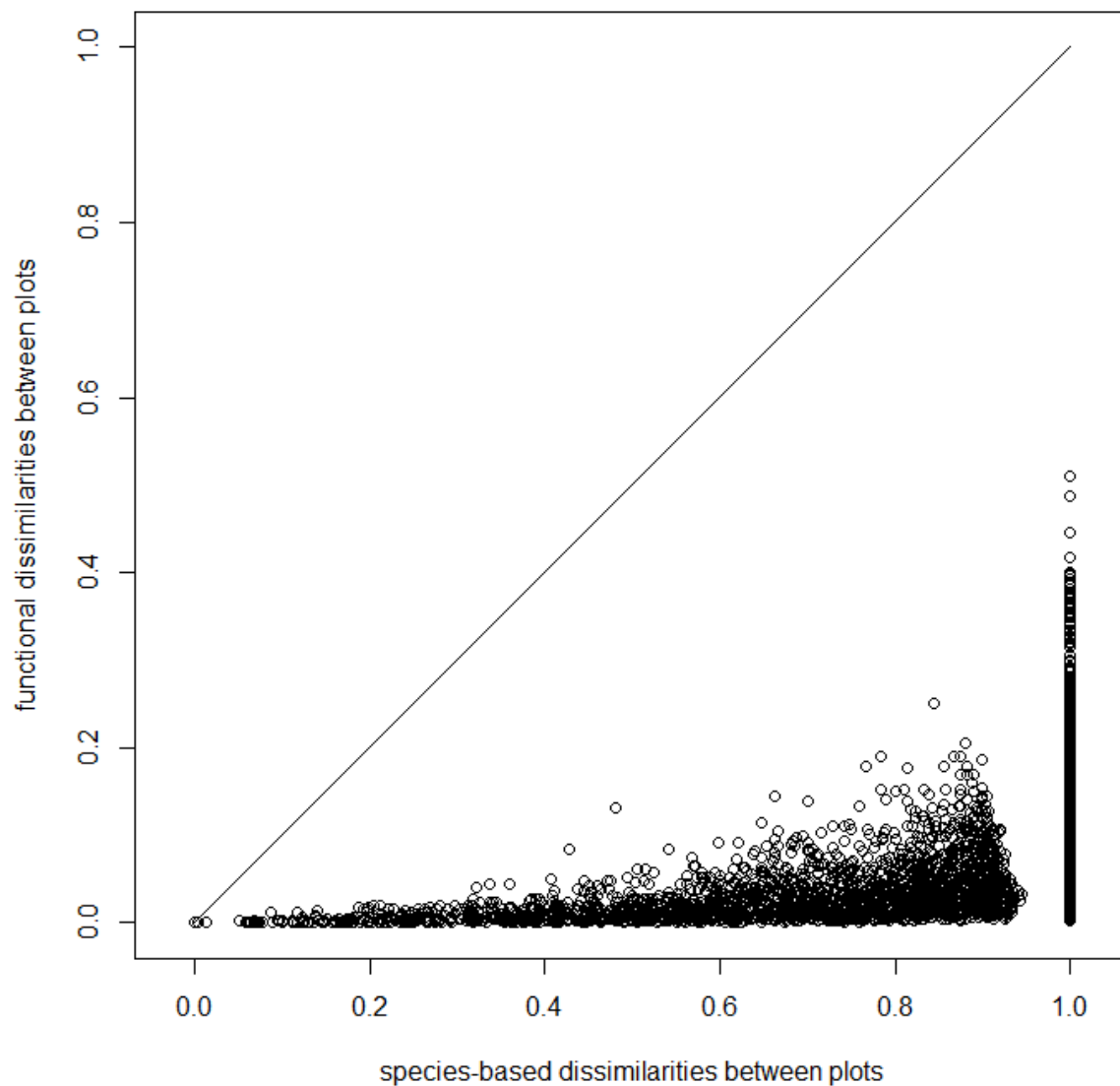
Analysis of environmental data on the factorial map of the functional PCoA:
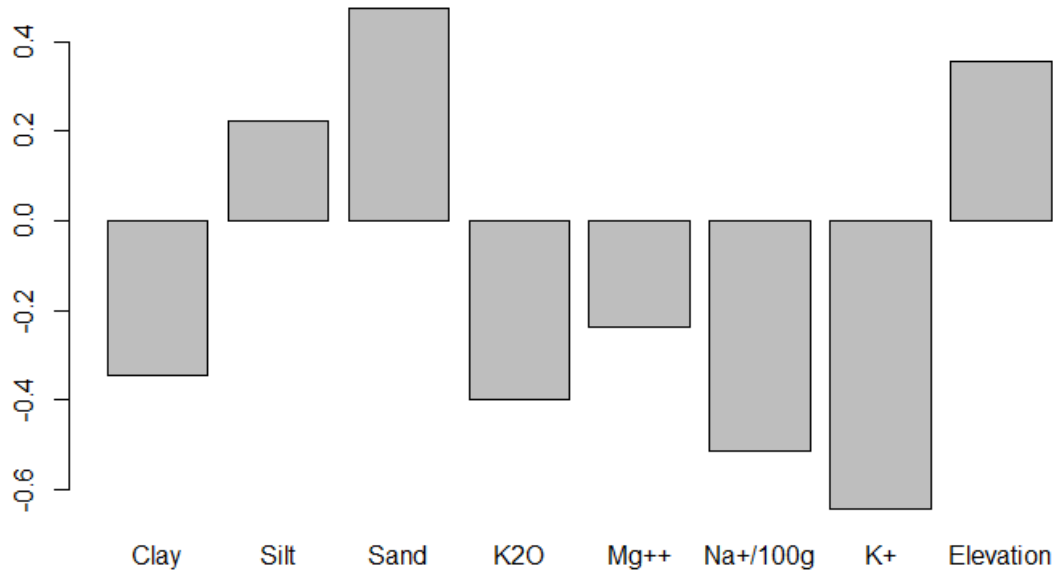
Environmental data:
```
env <- mafragh$env[-(8:10)]
```

Correlations between environmental data and the coordinates of plots on the first axis of the functional PCoA:
```
cor(env, pcoGC$li[,1], method= "spearman")
              [,1]
Clay     -0.3441044
Silt      0.2214038
Sand      0.4734112
K2O      -0.3989277
Mg++     -0.2372428
Na+/100g -0.5128574
K+       -0.6414609
```

```
Elevation  0.3561348
```

```
barplot(cor(env, pcoGC$li[,1], method="spearman"))
```

References

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

Jombart, T., Dray S. (2008) adephylo: exploratory analyses for the phylogenetic comparative method. Bioinformatics 26: 1907–1909.

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

Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Henry, M., Stevens, H., Szoecs, E., Wagner, H. (2018) vegan: Community Ecology Package. R package version 2.4-6. https://CRAN.R-project.org/package=vegan.

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

Appendix 3. On the relationship between species ordinariness and species rarity

Let U and V be two plots (or communities, assemblages, relevés, sites, quadrats, etc.) where x_{Uj} and x_{Vj} are the abundance values of species j in plots U and V, respectively, $p_{Uj} = x_{Uj} / \sum_{j=1}^{N} x_{Uj}$ is the relative abundance of species j in plot U (with $0 \leq p_{Uj} \leq 1$ and $\sum_{j=1}^{N} p_{Uj} = 1$), and N is the total number of species sampled in both plots (i.e. the species for which $\min\{x_{Uj}, x_{Vj}\} > 0$). Leinster and Cobbold (2012) defined the ordinariness of species j in plot U as:

$$z_{Uj} = \sum_{i=1}^{N} x_{Ui} \sigma_{ij} \tag{A1}$$

where z_{Uj} is the abundance of all species that are functionally similar to j (including j itself), and σ_{ij} is a symmetric similarity coefficient between species i and j such that $\sigma_{ij} = \sigma_{ji}$ and $\sigma_{jj} = 0$. For species j, z_{Uj} thus measures the commonness of all individuals in plot U that support the functions associated with j.

In the main text, we showed that the quantity z_{Uj} can be used for deriving a general family of functional dissimilarity coefficients among plots U and V (see Eq. 10 of the main text). Here, we show that if species ordinariness z_{Uj} is calculated from the species relative abundances p_{Uj} such that

$$z_{Uj} = \sum_{i=1}^{N} p_{Ui} \sigma_{ij} \tag{A2}$$

the resulting measure is tightly related to the notion of species rarity used by Patil and Taillie (1982) for providing a general definition of diversity coefficients.

Rao (1982) introduced a diversity index, termed Quadratic diversity, for summarizing mean interspecies differences within a given plot. The quadratic diversity of plot U is defined as the expected dissimilarity between two individuals drawn randomly with replacement from the plot:

$$Q_U = \sum_{j=1}^{N} \sum_{i=1}^{N} p_{Ui} p_{Uj} \delta_{ij} \tag{A3}$$

where δ_{ij} is the dissimilarity among species i and j.

For a review of some relevant mathematical properties of Rao's Q see Pavoine (2012). Here, it is sufficient to observe that if δ_{ij} is bounded between 0 and 1, assuming that all species are maximally distinct from one another (i.e. if $\delta_{ij} = 1$ for all $i \neq j$), Q_U reduces to the classical Simpson index of diversity $S_U = 1 - \sum_{j=1}^{N} p_{Uj}^2$. Also, because $0 \leq \delta_{ij} \leq 1$, an associated measure of concentration or dominance (i.e. the opposite of diversity) can be defined as the complement of Rao's Q (Leinster and Cobbold 2012):

$$D_U = 1 - Q_U = \sum_{j=1}^{N} \sum_{i=1}^{N} p_{Ui} p_{Uj} \sigma_{ij} \tag{A4}$$

where D_U is the average similarity between all species in plot U, and $\sigma_{ij} = 1 - \delta_{ij}$.

To gain insight into the type of diversity measured by the Simpson and the Rao coefficients, it is convenient to express Simpson in the form $S_U = \sum_{j=1}^{N} p_{Uj}\left(1 - p_{Uj}\right)$. In the context of interspecific encounter theory, Patil and Taillie (1982) defined the Simpson diversity as the probability that two individuals drawn randomly from a given plot do not belong to the same species. Therefore, the Simpson diversity is essentially a measure of uncertainty in predicting the relative abundance of the species in the plot. In information theory and statistics, the quantity $(1 - p_{Uj})$ is usually referred to as the surprise or rarity associated to species j (Ricotta and Szeidl 2006; Leinster and Cobbold 2012). If p_{Uj} is small, we would be quite surprised in finding species j. To the contrary, if p_{Uj} is large, the surprise in finding j is small. Patil and Taillie (1982) thus proposed a general definition of diversity (Δ) as the average rarity $\rho(p_{Uj})$ of a given assemblage

$$\Delta = \sum_{j=1}^{N} p_{Uj} \rho(p_{Uj}) \tag{A5}$$

where $\rho(p_{Uj})$ is some decreasing function of the species relative abundance p_{Uj}.

Klir and Wierman (1999) and Ricotta and Szeidl (2006) noted that the quantity $(1 - p_{Uj}) = \sum_{i \neq j} p_{Ui}$ measures the relative abundance of all species that differ from species j. That is, the probabilities of all events that conflict with p_{Uj}. Accordingly, assuming that all species are maximally distinct from each other (i.e. if $\delta_{ij} = 1$ for all $i \neq j$), the Simpson index can be interpreted as a measure of the expected conflict $C(p_{Uj})$ of all species in the assemblage (Ricotta and Szeidl 2006):

$$S_U = \sum_{j=1}^{N} p_{Uj} \sum_{i \neq j}^{N} p_{Ui} = \sum_{j=1}^{N} p_{Uj} C(p_{Uj}) \tag{A6}$$

Relaxing the constraint of maximal dissimilarity among species, it seems natural to summarize the conflict of species j with the other species in the assemblage taking into account their pairwise distances δ_{ij} such that:

$$C(p_{Uj}) = \sum_{j=1}^{N} p_{Uj} \delta_{ij} \tag{A7}$$

with $\delta_{jj} = 0$. Accordingly, incorporating a measure of pairwise dissimilarity among species, the Simpson expression of expected conflict becomes Rao's quadratic diversity $\sum_{j=1}^{N} p_{Uj} \sum_{i=1}^{N} p_{Ui} \delta_{ij}$. Rao's Q can be thus interpreted as a general measure of expected rarity or conflict in a given assemblage if the species are not treated as maximally distinct from each other (Ricotta and Szeidl 2006).

Likewise, we can define the ordinariness of species j as the complement of its conflict:

$$z_{Uj} = 1 - C(p_{Uj}) = \sum_{j=1}^{N} p_{Uj} \sigma_{ij} \tag{A8}$$

such that the expected species concentration in Eq. A4 can be rewritten as $D_U = \sum_{j=1}^{N} p_{Uj} \sum_{i=1}^{N} p_{Ui} \sigma_{ij} = \sum_{j=1}^{N} p_{Uj} z_{Uj}$, while Rao's diversity becomes $Q_U = 1 - \sum_{j=1}^{N} p_{Uj} z_{Uj}$, where the ordinariness z_{Uj} is the expected similarity between an individual of species j and an individual drawn randomly from the assemblage.

Therefore, according to the above derivation, if the values of z_{Uj} are calculated from the species relative abundances p_{Uj}, species ordinariness may represent a common currency for expressing both the species functional diversity in single plots and their functional dissimilarity among pair of plots.

References

Klir, G.J., Wierman, M.J. (1999) Uncertainty-Based Information. Physica-Verlag, Heidelberg.

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

Patil, G.P., Taillie, C. (1982) Diversity as a concept and its measurement. Journal of the American Statistical Association 77: 548–561.

Pavoine, S. (2012) Clarifying and developing analyses of biodiversity: towards a generalisation of current approaches. Methods in Ecology and Evolution 3: 509–518.

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

Ricotta, C., Szeidl, L. (2006) Towards a unifying approach to diversity measures: bridging the gap between the Shannon entropy and Rao's quadratic index. Theoretical Population Biology 70: 237–243.