An ordination approach to explore similarities among communities

Sandrine Pavoine

To cite this version:

Sandrine Pavoine. An ordination approach to explore similarities among communities. Journal of Theoretical Biology, 2019, 462, pp.85-96. 10.1016/j.jtbi.2018.11.002 . hal-02291830

An ordination approach to explore similarities among communities

Sandrine Pavoine*

Centre d'Ecologie et des Sciences de la Conservation (CESCO), Muséum national d'Histoire naturelle, Centre National de la Recherche Scientifique, Sorbonne Université, CP 135, 57 rue Cuvier, 75005 Paris, France

* Corresponding author. E-mail address: sandrine.pavoine@mnhn.fr

Correspondence address: Sandrine Pavoine, UMR 7204 CESCO, Muséum National d’Histoire Naturelle, CP 135, 61 rue Buffon, 75005 Paris, France

Running title: Depicting similarities among communities

Number of tables: 0; number of figures: 7

ABSTRACT

Analysis of similarities among communities can help to decipher the biogeographical, evolutionary, and ecological factors that drive local diversity. Recent indices of similarity among communities incorporate not only information on species presence and abundance but also information on how similar species are in their traits and how closely related they are in terms of taxonomy or phylogeny. Towards this aim, trait-based, taxonomic or phylogenetic similarities among species have been defined and bounded between 0 (species are maximally distinct) and 1 (species are similar). A required property for an index of similarity between two communities is that it must provide minimum similarity (0) where communities have maximally distinct species, as well as maximum similarity (1) where communities are equivalent in their trait, taxonomic or phylogenetic compositions. Here, I developed a new ordination methodology that conforms to the requirement: double similarity principal component analysis (DSPCA). DSPCA summarizes multidimensional trait-based, taxonomic or phylogenetic similarities among communities into orthogonal axes. The species that drive each similarity pattern can be identified together with their traits or with their taxonomic or phylogenetic positions. I applied this methodology to theoretical examples and to empirical data sets on bird and bat communities to illustrate key properties of DSPCA. I compared the results obtained with DSPCA with those provided by related approaches. Theoretical and empirical case studies highlight the following additional properties of DSPCA: (*i*) axes are orthogonal and identify independent (dis)similarity patterns between communities; (*ii*) the more functionally, taxonomically or phylogenetically similar communities are, the closer they are on an axis; (*iii*) the coordinate of a species on an axis expresses how representative the species is of the pattern identified by the axis; and (*iv*) a species is representative of *x* communities if the functional, taxonomic or phylogenetic characteristics of this species are very common within each of these *x* communities. DSPCA is an efficient approach to visualize functional, taxonomic and phylogenetic similarities between communities. It is also a useful alternative to recent methods dedicated to phylogenetic diversity patterns. It will be an asset for all studies that aim to compare functional, taxonomic, genetic and phylogenetic diversity.

Keywords:

Beta diversity

Biodiversity

Functional traits

Phylogeny

Taxonomy

1. Introduction

In ecology, similarities among communities are considered to pinpoint in space and time where and when patterns of community structure change. These changes might be driven, for example, by abiotic and biotic environments, geographic barriers, and dispersal limitations. Similarities among communities depend on which species they contain and potentially on the relative abundances of these species. Recent developments of similarity coefficients also include taxonomic, phylogenetic or trait-based similarities among the species that compose the communities (e.g., Pavoine et al., 2004; Ferrier et al., 2007; Bryant et al., 2008, Graham and Fine, 2008; Webb et al., 2008; Ricotta and Szeidl, 2009; Pavoine and Ricotta, 2014; Ricotta et al., 2016). In species characterization, the traits selected for a given study may be qualified as functional when they are associated with the ability of species to gain resources, disperse, reproduce, respond to loss and generally persist (Weiher et al., 2011) or when they influence ecosystem properties or species responses to environmental conditions (Lavorel and Garnier, 2002; Hooper et al., 2005). Functional traits lead to measures of functional similarity between species and between communities. Two levels of similarities are thus nested: one among the species and one among the communities. Estimating trait-based similarities among communities can reveal, for example, that some species are filtered out from an environment because of their traits, while others can expand, being adapted or tolerant to the environmental conditions (environmental filtering). This approach can also reveal that species with differences in fitness but similarities in niches rarely co-exist within the same community (competitive exclusion) (Mayfield and Levine, 2010). Estimation of the phylogenetic similarities among communities − especially when the lineages driving these similarities are clearly identified − can provide insights into historical and evolutionary mechanisms, including the potential for allopatric and ecological speciation (Graham and Fine, 2008).

Referring to Jost's (2006) observations on more traditional indices, Ricotta and Szeidl (2009) observed that two communities should be completely distinct (similarity=zero) if they have no species in common and if their species have no (trait-based, taxonomic or phylogenetic) similarities. The absence of trait-based similarities among species can be observed if these species have maximally distinct trait states. The absence of phylogenetic similarity would be obtained relative to a given delimited clade if the species of the first community diverged from the species of the second community at the root of the clade without any subsequent shared history. This point of view assumes that previously shared history outside the clade is discarded. In all cases, Ricotta and Szeidl's viewpoint assumes that the differences between species have a maximum that cannot be exceeded.

Pavoine and Ricotta (2014) responded to this definition of completely distinct communities by developing a new family of indices for measuring the trait-based, taxonomic and phylogenetic similarity between two communities. Let $\mathbf{S}^{\text{spe}}=(s_{kl}^{\text{spe}})$ be a matrix where s_{kl}^{spe} is the similarity between species *k* and species *l*; $s_{kk}^{\text{spe}}=1$ for all *k*, and $0\leq s_{kl}^{\text{spe}}\leq 1$ for all *k* and *l*. The matrix is non-negative definite (Seber, 2008), so that for any real vector $\mathbf{x}=\left(x_1\ \ldots\ x_n\right)^t$, $\sum_{k,l}x_k x_l s_{kl}^{\text{spe}}\geq 0$ (*n* is the number of species; and $\sum_{k,l}$ is the double summation $\sum_{k=1}^{n}\sum_{l=1}^{n}$). Let $\mathbf{p}_i=(p_{i1}\ \ldots\ p_{in})^t$ be the vector of species' proportions (e.g., relative abundances in terms of number of individuals or biomass) in community *i* with $p_{ik}\geq 0$ and $\sum_k p_{ik}=1$. Pavoine and Ricotta (2014) introduced, among others, the following index of similarity between two communities *i* and *j*:

$$S_{Ochiai}(\mathbf{p}_i,\mathbf{p}_j)=\frac{\sum_{k,l}p_{ik}p_{jl}s_{kl}^{\text{spe}}}{\sqrt{\sum_{k,l}p_{ik}p_{il}s_{kl}^{\text{spe}}}\sqrt{\sum_{k,l}p_{jk}p_{jl}s_{kl}^{\text{spe}}}} \tag{1.1}$$

When $s_{kl}^{\text{spe}}=0$ for all $k\neq l$, $\sqrt{2\left(1-S_{Ochiai}\right)}$ is a generalization of the Chord distance applied to species' abundance, an index first introduced in ecology by Orloci (1967): $\sqrt{2\left(1-\sum_k p_{ik}p_{jk}\Big/\sqrt{\sum_k p_{ik}^2\sum_k p_{jk}^2}\right)}$. In addition, when $p_{ik}=1/n_i$, where n_i is the number of species in community *i*, then S_{Ochiai} is equivalent to Ochiai's (1957) index of similarity that uses species presence and absence in communities: $a_{ij}\Big/\sqrt{n_i n_j}$, where a_{ij} is the number of species shared by communities *i* and *j*. The problem raised by Jost (2006), concerning completely distinct communities, was known by quantitative ecologists: with certain dissimilarity indices centered on species' identity only, two sites without any species in common may be attributed a smaller dissimilarity than another pair of sites sharing species (Orloci, 1967; Legendre and Legendre, 1998). Orloci (1967) therefore developed an index derived from the chord distance to circumvent this paradox. This issue was extended to phylogenetic and functional diversity by Ricotta and Szeidl (2009).

Let $\mathbf{S}^{\text{com}}=(s_{ij}^{\text{com}})$ be the matrix of similarities between communities obtained from eqn. 1.1 (i.e., $s_{ij}^{\text{com}}=S_{Ochiai}(\mathbf{p}_i,\mathbf{p}_j)$). The objective of this study is to develop a new ordination method that analyzes and summarizes the information driven by matrix $\mathbf{S}^{\text{com}}$ of similarity among communities into independent one-dimensional axes that can be directly explained by the composition of species communities, by species' trait, taxonomic or phylogenetic positions. These methodological advances are illustrated with: 1) theoretical examples; 2) a case study where the taxonomic and trait-based (dis)similarities between bird communities are depicted along environmental gradients under Mediterranean and temperate bioclimates; and 3) a case study on the phylogenetic dissimilarities between bat communities along a disturbance gradient in Selva Lacandona of Chiapas, Mexico.

2. Materials and Methods

2.1. DSPCA

As highlighted above, for the matrix $\mathbf{S}^{\mathrm{spe}}$ to be used in index S_{Ochiai}, it needs to have a special mathematical property, i.e., non-negative definite. Pavoine and Ricotta (2014) described various ways of obtaining a non-negative definite matrix $\mathbf{S}^{\mathrm{spe}}$ from trait-based, taxonomic and phylogenetic data and demonstrated that, in that case, the matrix $\mathbf{S}^{\mathrm{com}}$ has values bounded between 0 and 1. I show in Appendix A that if $\mathbf{S}^{\mathrm{spe}}$ is non-negative definite, $\mathbf{S}^{\mathrm{com}}$ is also non-negative definite. These mathematical properties common to $\mathbf{S}^{\mathrm{spe}}$ and $\mathbf{S}^{\mathrm{com}}$ are exploited in DSPCA.

DSPCA can be related to the analysis of correlation matrices in normed principal component analysis (Corsten and Gabriel, 1976; Seber, 2004). The approach can be described in four main steps: (1) obtaining a space in which species are positioned according to their similarities, (2) positioning the communities in this space according to the species they contain and the abundances of these species, (3) obtaining new axes which successively optimize the representation in few dimensions of the similarities among the communities, and (4) projecting species and communities on these new axes.

The details of the approach are as follows. For the first step, similarities among species are described on a series of independent axes obtained from the eigen-decomposition of $\mathbf{S}^{\mathrm{spe}}$: $\mathbf{S}^{\mathrm{spe}} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^t$, where the columns of $\mathbf{U}$ contain eigenvectors and the diagonal values of $\mathbf{\Lambda}$ contain eigenvalues. The rows of $\mathbf{X} = \mathbf{U}\mathbf{\Lambda}^{1/2}$ provide coordinates for the species. The axes on which these coordinates are defined are called principal components in the context of multivariate analyses of correlation matrices. The expression "principal component" is also retained here although similarities replace correlations. Let $\mathbf{P} = \left(\mathbf{p}_1 \,|\, \mathbf{p}_2 \,|\, ... \,|\, \mathbf{p}_m\right)$ be the $n \times m$

matrix with the proportions of n species in m communities ($\mathbf{P}^t\mathbf{1}_n = \mathbf{1}_m$, with $\mathbf{1}_n$ and $\mathbf{1}_m$ the n×1 and m×1 vectors of units, respectively). For the second step, communities are positioned at the center of their species; the rows of $\mathbf{Y} = \mathbf{P}^t\mathbf{X}$ thus provide coordinates for the communities. These coordinates are normalized as follows: $\tilde{\mathbf{Y}} = \mathbf{Q}^{-1}\mathbf{P}^t\mathbf{X}$, where $\mathbf{Q}$ is a squared, diagonal matrix with $\sqrt{\mathbf{p}_i^t\mathbf{S}^{\text{spe}}\mathbf{p}_i} = \sqrt{\sum_{k,l} p_{ik}\,p_{il}\,s_{kl}^{\text{spe}}}$ at line i and column i and 0 out of the diagonal. The diagonal values of $\mathbf{Q}$ are the square root of the diagonal values of $\mathbf{YY}^t$. If presences/absences are used, the proportion of a species present within a community i that contains n_i species is set to $1/n_i$ (S_{Ochiai} is not impacted by considering relative rather than absolute abundances, see Appendix A). The third step is determined by the eigen-decomposition of $\tilde{\mathbf{Y}}^t\tilde{\mathbf{Y}}$: $\tilde{\mathbf{Y}}^t\tilde{\mathbf{Y}} = \mathbf{B\Psi B}^t$, with eigenvectors in $\mathbf{B}$, and positive eigenvalues in $\mathbf{\Psi}$. This third step allows switching from a space where the axes successively describe similarities among species to a space where the axes successively best describe similarities among communities in light of their species composition. In the fourth step, the final coordinates of the species are presented in the rows of $\mathbf{X}_{\text{final}} = \mathbf{XB}$, and those of the communities in the rows of $\mathbf{Y}_{\text{final}} = \tilde{\mathbf{Y}}\mathbf{B} = \mathbf{Q}^{-1}\mathbf{P}^t\mathbf{X}_{\text{final}}$. The columns of matrices $\mathbf{X}_{\text{final}}$ and $\mathbf{Y}_{\text{final}}$ are principal components and the rows within each matrix represent the species and the communities, respectively. A community point is located on the axes in the direction of the (abundance-weighted) center of its species; its exact position satisfies the requirement that the norm of the community coordinates is 1 (community and species are located in a ball of radius 1 such as variables in a normed principal component analysis). In the final multidimensional space, entities (species and communities) can be displayed by arrows starting from the origin of the space to the vertices defined by the rows of $\mathbf{Y}_{\text{final}}$ and $\mathbf{X}_{\text{final}}$, respectively. A community arrow is thus unit length and points to a direction defined by a weighted mean of species' arrows; weights are the proportions (e.g., relative abundance) of the species in the community. It can be shown (Appendix A) that $\mathbf{Y}_{\text{final}}\mathbf{Y}_{\text{final}}^{t} = \mathbf{S}^{\text{com}}$ (with similarities among communities calculated with index S_{Ochiai}), so that the similarities among communities are preserved in the final space.

2D-graphics can be displayed using any two principal components of the communities. The first principal component contains the largest part of the similarities among communities, the second is orthogonal to the first and contains the second largest part, and so on. These 2D-graphics optimize the visualization of the similarities among communities while explaining these similarities with their species. In the multidimensional space, the arrows of any two communities *i* and *j* form an angle. The cosine of this angle is s_{ij}^{com}. This means that, in this graphical approach, two communities are similar if their arrows form a very acute angle. The larger the angle, the more dissimilar they are. Community and species coordinates are bounded between -1 and 1. In 2D-graphics, they can thus be represented within a circle of unit radius. The coordinate of a species in a principal component expresses how representative the species is of the similarity pattern identified by the principal component (see Appendix B in the Supplementary material and the case studies below).

The sum of all eigenvalues in $\mathbf{\Psi}$ is equal to the number of communities. The number of axes examined in an analysis depends on these eigenvalues. Several coefficients can be used to evaluate the quality of the graphical representation of the similarities obtained by retaining the first *k* out of *K* axes, including

$$\alpha_k = \left(\sum_{i=1}^{k}\psi_i \,/\, \sum_{i=1}^{K}\psi_i\right)\times 100\%$$

(see Seber (2004) for indices developed in other contexts). The first eigenvalue, λ_1, reflects the amount of overall similarity among all communities. Its value is approximately equal to 1 + (m-1) $\bar{s}$ (Friedman and Weisberg, 1981), where $\bar{s}$ is the mean similarity between any two communities and *m* the number of communities. If communities are not completely distinct, the last eigenvalue expresses the full dissimilarities between the communities (what is left when all similarities have been described). Intermediate eigenvalues detail multivariate similarity patterns, that is to say the fact that some similarities concern only part of the compared communities. In the extreme case where the similarities between communities are equal, say to s, then $\lambda_1 = 1 + (m\text{-}1)\, s$ (Morrison, 1978, p. 289). For example, if communities are completely distinct, then $s = 0$ and $\lambda_1=1$, which is the lowest possible value for λ_1. In that case, all m eigenvalues are equal to 1. When the similarity between any two communities is positive, then at least λ_1 is higher than 1 and at least λ_m lower than 1. If there are only two communities compared, then s is the similarity between these two communities, $\lambda_1 = 1 + s$ and λ_2 the second and last eigenvalue equals $1 - s$, expressing thus the dissimilarity between the two communities.

2.2. Case studies

Calculations were performed with R (R Core Team, 2018) as described in Appendices C and D of the Supplementary material.

2.2.1. Theoretical data set #1

Within-community diversity influences the length of the species arrows; for example, if the functional diversity of a community is high, then the constitutive species have low similarity in terms of their functional traits. Each species of the community is thus unlikely to be representative of others. More generally, if PCi, the ith axis of DSPCA, represents a certain similarity between x communities, then the contribution of a species shared by the x communities to the identified similarity pattern is high if the functional, taxonomic or phylogenetic characteristics of this species are very common within each of these x communities. To illustrate this point, I use three simple examples as described in Fig. 1.

2.2.2. Theoretical data set #2

The second theoretical data set aims to highlight the main discrepancies between DSPCA and another ordination approach: double principal coordinate analysis (DPCoA) developed by Pavoine et al. (2004). First, DSPCA uses similarities among species and communities whereas DPCoA focuses on dissimilarities. Second, DPCoA and DSPCA differ in their treatment of completely distinct communities. DPCoA was not defined to be restricted to bounded dissimilarities between communities. In this particular case, however, the distance between completely distinct communities in DPCoA maps depends on the diversity within each community. By contrast, DSPCA always provides zero similarity between completely dissimilar communities. To highlight these main differences between DPCoA and DSPCA, I applied both approaches to the following theoretical data set: 110 species, named s1 to s110, have no similarities with each other. $\mathbf{S}^{spe}$ is thus a diagonal matrix with 110 rows and 110 columns, with unit values on the diagonal and 0s elsewhere. Four communities have no species in common. The first community c1 has species s1 to s50; the second, c2, has species s51 to s100; the third, c3, has species s101 to s105; and the fourth, c4, has species s106 to s110. Species' proportions within communities are even.

2.2.3. Theoretical data set #3.

A common practice when analyzing pair-wise dissimilarities between communities is to use non-metric (nMDS) or metric (MDS) multidimensional scaling depending on the Euclidean properties of the dissimilarity matrix of interest. For example, MDS can be applied to a matrix of dissimilarities obtained with $\sqrt{1-S_{Ochiai}}$. When MDS and nMDS are used, however, information about species is lost, and it may not be possible to identify which species, trait, or phylogenetic position contributed to the dissimilarities among communities a posteriori. Placing species a posteriori at the barycenter of their communities in MDS or nMDS maps may be misleading. In doing so, the position of the species will reflect their abundance within communities, but not their functional, taxonomic or phylogenetic dissimilarities. To illustrate this fact, I used the theoretical data set described in Fig. 2a. It contains 36 species distributed among 4 communities and is described by two quantitative traits. Application of the Gower (1971) distance to the trait data led to a matrix $\mathbf{S}^{\mathrm{spe}}$ of similarity between species; then, coefficient S_{Ochiai} of similarity between sites led to $\mathbf{S}^{\mathrm{com}}$=(s_{ij}^{com}), where the similarity, s_{ij}^{com}, between any two sites i and j $i \neq j$ was 0.79 ($s_{ii}^{\mathrm{com}}=1\ \forall i$). I applied MDS to $\mathbf{D}^{\mathrm{com}}=\left(\sqrt{1-s_{ij}^{com}}\right)_{i=1,\ldots,4;\,j=1,\ldots,4}$ and DSPCA to $\mathbf{S}^{\mathrm{spe}}$ and the matrix of species presence/absence in communities.

2.2.4. Bird data set

I applied DSPCA to the same data set as that used to illustrate DPCoA in Pavoine et al. (2004). The data set (Blondel et al., 1984) contains bird communities living in different parts of the world under Mediterranean bioclimates: central Chile, California (United States), and Provence (France). These regions were compared to a control region under a temperate bioclimate: Burgundy (France). Blondel et al. (1984) determined equivalent habitats among the four regions in terms of structure, height and physiognomy of vegetation. Overall, the habitats form a gradient of vegetation complexity from habitat#1 (the least complex) to habitat#4 (the most complex). The data set contains data on species' foraging substrate (multichoice nominal variable), morphometry (quantitative variable) and taxonomy. The effects of species abundance and species-to-species similarities on the results of DSPCA can be analyzed by considering both presence-absence data and abundance data, and by considering species as maximally dissimilar in addition to analyzing trait and phylogenetic information on species (see Appendix E in the Supplementary material for a pedagogic illustration). Here I explored the effect of species-to-species similarities by considering four matrices of species similarity: 1) $\mathbf{S}_{\mathrm{MAX}}^{\mathrm{spe}}$ contains 1 on the diagonal and 0 elsewhere, which means that species are maximally dissimilar; 2) $\mathbf{S}_{\mathrm{FOR}}^{\mathrm{spe}}$ was defined as a function of the substrates where species forage using the Ochiai index of similarity; 3) $\mathbf{S}_{\mathrm{MOR}}^{\mathrm{spe}}$ was obtained by applying Gower's (1971) similarity to species morphometric traits; and 4) $\mathbf{S}_{\mathrm{TAX}}^{\mathrm{spe}}$ has 1 on the diagonal, 3/4 between species of the same genus, 1/2 between species of the same family but distinct genera, 1/4 between species of similar order but distinct families, and 0 between species of different orders, families and genera. The method used to calculate taxonomic similarities is also related to the Ochiai coefficient. Indeed the taxonomic similarity between two species can be expressed as $t_{kl} \big/ \sqrt{t_{kk} t_{ll}}$, where t_{kl} is the number of taxonomic levels shared by the two species and t_{kk} is the total number of taxonomic levels that describe any species k (here 4 levels: species, genus, family, and order). This leads to t_{kk} being equal to 4 for all k. The taxonomic similarity between two species k and l is thus t_{kl} / 4. The calculation of all similarity matrices is detailed in this Appendix C of the Supplementary material.

2.2.5. Bat data set

I also applied DSPCA to data from Medellín et al. (2000) on bats in four habitats in the Selva Lacandona of Chiapas, Mexico, with Fritz et al. (2009) phylogeny pruned for retaining only the species present in the Medellín et al. data set. The four compared habitats were distributed on a disturbance gradient from an active cornfield (the most disturbed), through old fields and cacao plantations, to rainforests (the least disturbed). The phylogenetic similarity between two species k and l was defined as $c_{kl} \big/ \sqrt{c_{kk} c_{ll}}$: c_{kl} is the sum of branch lengths on the shortest path that connects the most recent common ancestor of the two species to the root of the tree, and c_{kk} is the sum of branch lengths on the shortest path that connects species k to the root of the tree (Pavoine and Ricotta, 2014). This coefficient is thus also related to the Ochiai index. Because the phylogenetic tree is ultrametric, $c_{kk} = H$, the height of the tree, for all species k, and the phylogenetic similarity between two species k and l reduces thus to c_{kl}/H. I compared the results obtained with DSPCA with those produced by evoPCA$_{\text{Chord}}$, an ordination approach I developed in Pavoine (2016) to specifically analyze phylogenetic tree data.

3. Results

3.1. Theoretical data set #1

When communities are maximally dissimilar (Fig. 1a), the species within a community are linked only to this community in DSPCA. Their arrows superimpose that of the community. The lengths of species arrows, however, depend on how representative each species is of the community. The more numerous species are within the community and the more distinct they are (from a functional, taxonomic or phylogenetic perspective), the less representative each species is of the community composition. When a community is nested within another, the similarity between these two communities depends on the number of species shared and on the number of similarities between these species and between unshared species (Fig. 1b). The lengths of species arrows also depend on these two factors. When communities do not share species, they can still be similar if the most representative species of each community are similar (Fig. 1c). In any case, the species arrows tend towards the communities where they occur and their length depends on how well they represent the composition of each community.

3.2. Theoretical data set #2

DSPCA identifies the absence of similarity between communities, placing them on orthogonal axes, with unit eigenvalues (Fig. 3). The arrows for species point to the direction of the communities in which they occur. However, their sizes change depending on the diversity within the associated community. As observed above, the size of a species arrow expresses how representative a species is of the similarity pattern. The example in Fig. 3 is extreme, so that each axis represents a community, and species are all maximally dissimilar. In that case, the size of a species arrow associated with community *i* is $1/\sqrt{n_i}$, where n_i is the number of species in community *i*. The size of a species arrow is thus inversely linked with the number of species within the community. By contrast, DPCoA identifies higher similarity between the most diverse communities.

3.3. Theoretical data set #3

I analyzed the data set presented in Fig. 2a using DSPCA (Fig. 2b) and MDS (Fig. 2c). MDS places the communities at the vertices of a regular tetrahedron (Fig. 2c). As communities do not share species, positioning species on the map of MDS due to their distribution in communities places them on the point of their community as shown in Figure 2c and thus independently of their traits. With DSPCA, the directions of species arrows indicate which community(ies) each species belongs to, and the size of a species arrow indicates how representative the species is of the(se) community(ies) compared to other communities (Fig. 2b). For example, species s1 with a low value for trait t1 and a medium value for trait t2 is the most characteristic of community c1 compared to other communities. Species s9, s10, s27 and s28, with medium values for the two traits, are the least original species and have close-to-zero coordinates on the axes. They are the four species that discriminate the least among the four communities.

3.4. Bird data set

When bird species were considered maximally dissimilar, DSPCA identified four main principal components (axes) (Fig. 4): the first one for the similarities between Burgundy and Provence; the second for the similarities between habitats in Chile; the third for similarities between habitats in California; and the fourth for the distinction between habitats in Provence and those in Burgundy. The fifth and sixth principal components then highlight the gradient of vegetation complexity in Chile and California, respectively. The length of species arrows on these six axes increases with the number of habitats in which they were observed (from 1 to 4 per region) and decreases with the number of species in the region and each of its habitats. The orthogonal patterns highlight that California, Chile and France do not share species.

When applied to foraging substrate, the first principal component of DSPCA highlighted high similarities between all communities (Fig. 5). Species coordinates reveal that the species most representative of the study area forage on the ground solely or in addition to other substrates. The second and third principal components highlight the environmental gradient within each region, from species foraging on the ground in open habitats, to a large diversity of foraging substrates in closed habitats. These principal components are close, but not equal, to the first and second axes of DPCoA applied to the same data set (Pavoine et al., 2004).

When applied to morphometric data, DSPCA identified the most common morphological shapes for a bird species in the data set and, inversely, the most original shapes (Fig. 6a). The species with the highest coordinates on the first principal component, *Sylvia hortensis*, is the most representative of bird morphology in the study area (considering that a species that occurs in many places also increases similarities among these places). The five species with the lowest coordinates and thus the most morphometrically original species are *Ammodramus sandwichensis* with a relatively short tail, *Sylviorthorhynchus desmursii* with a relatively very long tail, and the three hummingbirds, notably with their unique beak shape, *Archilochus alexandri*, *Calypte costae*, *Calypte anna*. The eigenvalues of other axes were very low, which indicates low morphometric differences between communities within and across regions.

With taxonomic information, DSPCA underlined on the first principal component the dominance, in terms of species occurrences, of Passeriformes in all habitats of all regions (Fig. 6b). The second and third principal components highlighted minor differences discriminating the four regions from each others: e.g., the more frequent presence of Emberizidae species in open habitats of California and Chile, Piciformes in close habitats in California, Chile, and Burgundy, Paridae species in close habitats of Burgundy, species of the genus *Sylvia* in open habitats of Provence and more generally, Sylviidae and Turdidae in Provence and Burgundy.

3.5. Bat case study

I applied DSPCA to the phylogenetic similarities between bat communities in Selva Lacandona of Chiapas, Mexico. The first principal component highlighted high similarities between all habitats (high eigenvalue and close-to-1 scores for all habitats) (Fig. 7a). The sets of the most abundant species in each habitat are closely related. The least representative species in the study area (*Thyroptera tricolor*, *Bauerus dubiaquercus* and *Myotis keaysi*, with close-to-zero scores) are the most isolated on the phylogenetic tree. They are also among the least abundant. The results obtained on the second and third principal components are close

(Fig. 7c,d), but not equal, to those obtained with evoPCA$_{\text{Chord}}$ (Pavoine, 2016). Compared with evoPCA$_{\text{Chord}}$, DSPCA does not directly position the nodes of the phylogenetic tree on the factorial maps. DSPCA distinguishes cornfields with high abundance of *Sturnira lilium* from old fields with high abundance of *Carollia brevicauda* and *C. perspicillata* and the rainforest, which is distinguished by the higher relative abundance of 10 species including *Artibeus jamaicensis, A. lituratus, Dermanura watsoni* and *D. phaeotis* (Fig. 7c,d). On the third principal component, *Glossophaga commissarisi* and *G. soricina* characterize both cornfields and old fields compared to other habitats (Fig. 7c,d). This pattern was not revealed by evoPCA$_{\text{Chord}}$.

4. Discussion

Connections exist between ordination analyses and diversity measurements (e.g., Pélissier et al., 2003). While measures value biodiversity, ordination analyses use these values to depict structures in the diversity of communities. They identify, for instance, which communities are similar. Some can also identify which species, taxa, clades or traits are responsible for these similarities (e.g., Pavoine et al., 2004). Recent approaches have focused on describing the phylogenetic patterns of communities (e.g., Duarte, 2011; Pavoine, 2016). DSPCA can describe how functionally or phylogenetically similar communities are. It is flexible in the type of similarities measured between species. DSPCA orders communities along axes, the number of which depends on the complexity of the similarity matrix among communities. The axes are orthogonal, provide independent information and are organized from the main to the most residual pattern of similarity. The strength of the similarity pattern provided by an axis is represented by a numerical value, which is an eigenvalue. It is thus possible to describe a pattern of similarity and to provide a value of its importance compared with the pattern of similarity expressed by all other available axes. If patterns are not presented per axis but for a set of axes, a coefficient is provided to evaluate the amount of information extracted by these axes (e.g., Seber, 2004). The methodology offers direct solutions for explaining the pattern of similarities among communities with their compositions in species and the functional, taxonomic or phylogenetic links specified between them.

DSPCA analyzes both similarities and dissimilarities between communities. For example, in the bat dataset, DSPCA revealed low effects of habitat disturbance on the phylogenetic structure of bat communities: the measured phylogenetic similarity between the four compared habitats was high and the only identified differences between habitats concerned young clades and terminal branches of the phylogenetic tree. The bird data set showed that in all regions, the species composition changed along the gradient of vegetation complexity, from species foraging on the ground to species using a large diversity of foraging substrates in close habitats. Despite identified changes in species identity, in particular, despite the absence of species shared between California, Chile and the two French regions, DSPCA revealed high similarities between all regions and habitats in terms of species taxonomy and morphometry. DSPCA can thus be usefully applied to communities that share no species, because different species may have similarities due to their traits, phylogenetic or taxonomic positions. This shows that DSPCA could also be applied to entities that are systematically unshared by communities such as individuals and populations. DSPCA could thus be applied in the future to explore within-species variation by focusing on individuals or populations considering that species trait may vary from community to community.

Here, I analyzed trait-based (dis)similarities and phylogenetic (dis)similarities separately. Further applications of the approach could explore new ways of measuring the similarities among species to analyze trait-based diversity in light of phylogeny. For example, new approaches could be considered to apportion a matrix of species traits into a matrix of phylogenetically explained variations in traits among species, and inversely, a matrix of trait-based information independent of phylogeny (see, e.g., Diniz-Filho et al., 1998; Desdevises et al., 2003; Giannini, 2003). Using the latter matrix to calculate similarities among species in DSPCA could reveal trait-based similarities among communities not driven by phylogeny. An alternative would be to follow Cadotte et al. (2013) by developing similarities between species that are nonlinear combinations of trait-based similarities and phylogenetic similarities.

DSPCA also allows identification of the most representative species of one or several communities compared to other communities. In the bird data set, for example, DSPCA identified the species *S. hortensis* (the western Orphean warbler) as the most representative of the morphometric aspects of birds in the whole data set. DSPCA also allows the identification of the species with the rarest characteristics, such as the hummingbirds in the bird data set, with their unique beak shape. In the bat data set, DSPCA identified the most phylogenetically isolated species with the lowest abundance as the least representative species in the study area. Compared with other ordination approaches, DSPCA is thus able to identify not only original species in an original, species-poor community but also original species within a diverse and otherwise common community. The identification of original species may be important if these species are keystone, being rare while having important functions in the ecosystem (Mouillot et al., 2013; Power et al., 1996). Inversely, the most representative species may represent the species most adapted to their biotic and abiotic environments. The amount of functional redundancy in an assemblage, for instance, may enhance the resilience of the assemblage after a disturbance if functionally similar species differ in their response to disturbance (Walker, 1992). DSPCA thus allows a complete evaluation of the trait-based, taxonomic or phylogenetic diversity within and between communities due to its description of (dis)similarities between species and communities and as a result of the identification of original and redundant species.

DSPCA ensures, via index S_{Ochiai}, that two completely distinct communities always have zero similarity, as recommended by Ricotta and Szeidl (2009). By contrast, in DPCoA, the similarity between two communities is considered high whenever the average similarity between an individual from the first community and an individual from the second community is approximately the same as the average similarity between two individuals drawn from the same community. DPCoA should be preferred over DSPCA when dissimilarities among species do not have to be bounded between 0 and 1. In that case, maximally dissimilar species cannot exist and neither can maximally dissimilar communities. The use of DSPCA or DPCoA relates to how the biological dissimilarities and similarities among communities have to be defined considering the objective of the study at hand. DSPCA uses the S_{Ochiai} index (eqn. 1.1) while DPCoA relies on Rao's (1982) *DISC* index, defined as follows:

$$DISC\left(\mathbf{p}_i,\mathbf{p}_j\right)=\sum\nolimits_{k,l} p_{ik}p_{jl}d_{kl}^{\mathrm{spe}}-\tfrac{1}{2}\sum\nolimits_{k,l} p_{ik}p_{il}d_{kl}^{\mathrm{spe}}-\tfrac{1}{2}\sum\nolimits_{k,l} p_{jk}p_{jl}d_{kl}^{\mathrm{spe}}$$

where d_{kl}^{spe} depicts the (trait-based, taxonomic or phylogenetic) dissimilarities between two species *k* and *l*. An advantage of DPCoA over DSPCA is that it has been extended, for instance, to evaluate how two interacting factors (e.g., habitat and geography) affect the compositions of communities in terms of the functions or lineages they contain (Pavoine et al., 2013). Such developments for DSPCA are directions for future research.

Compared with DSPCA and DPCoA, evoPCA$_{\mathrm{Chord}}$ is dedicated to phylogenetic data expressed by a hierarchical tree describing the evolutionary relationships between species. Both DSPCA and DPCoA can handle a variety of data including functional, taxonomic and phylogenetic data. A common feature of all three approaches, however, is that the (dis)similarity indices they use are rooted in traditional literature on biodiversity. Indeed, if species have no similarities with each other, the index used by DPCoA is the Euclidean distance between the vectors of species proportions of the two compared communities, which corresponds to the index of β diversity developed for Gini-Simpson diversity (Lande, 1996; see also Appendix A). If species have no similarities with each other, the indices used by DSPCA and evoPCA$_{\text{Chord}}$ are both related to the Orloci (1967) index. If these species also have equal proportions within communities, these indices reduce to the Ochiai (1957) index. Compared with simply applying MDS or nMDS to a matrix of dissimilarity between communities, DSPCA, DPCoA and evoPCA$_{\text{Chord}}$ all permit the identification of the species, taxonomic groups, clades or traits responsible for the identified patterns of (dis)similarity between communities.

5. Conclusion

DSPCA summarizes multidimensional similarities into individual similarity patterns represented by orthogonal axes. These individual similarity patterns are ordered and their relative strength evaluated. Applied to the phylogenetic distribution of a group, DSPCA has the potential to raise hypotheses about historical processes such as colonization processes and dispersal limitation. Applied to (morphological, behavioral or life-history) traits of the species, DSPCA could also reveal the influence of the environment on the evolution of labile species functional traits or on the impact of conserved functional traits on the dispersal abilities of these species. If no information on the phylogeny and functional traits is given, this approach is still valid. In that case, it evaluates similarities in species abundances between sites. A comparison of the results obtained with DSPCA applied to species, functional, and phylogenetic data could increase the chance of identifying key ecological and evolutionary mechanisms that shape community assembly (e.g., Pavoine and Bonsall, 2011; Stegen and Hurlbert, 2011; see also Swenson, 2013).

Acknowledgements

I thank reviewers for their useful comments on the paper.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Appendix A. Mathematical proofs

The notations here are the same as in the main text.

A.1. If the matrix of similarities among species is non-negative definite, then the matrices of similarity among communities obtained with coefficient S_{Ochiai} is also non-negative definite

For any matrix **A**, the matrixes $\mathbf{A}^t\mathbf{A}$ and $\mathbf{A}\mathbf{A}^t$ are non-negative definite (e.g., Albert, 1969). By definition,

$$\mathbf{S}^{\text{com}} = \left(\mathbf{Q}^{-1}\mathbf{P}^t\right)\mathbf{S}^{\text{spe}}\left(\mathbf{Q}^{-1}\mathbf{P}^t\right)^t$$

Because $\mathbf{S}^{\text{spe}}$ is non-negative definite, there is a matrix **R** so that $\mathbf{S}^{\text{spe}} = \mathbf{R}\mathbf{R}^t$ (e.g., Seber, 2008). Then,

$$\mathbf{S}^{\text{com}} = \left(\mathbf{Q}^{-1}\mathbf{P}^t\mathbf{R}\right)\left(\mathbf{Q}^{-1}\mathbf{P}^t\mathbf{R}\right)^t$$

Let $\mathbf{A}=\left(\mathbf{Q}^{-1}\mathbf{P}^{t}\mathbf{R}\right)^{t}$, $\mathbf{S}^{\mathrm{com}}=\mathbf{A}^{t}\mathbf{A}$. The matrix $\mathbf{S}^{\mathrm{com}}$ that contains $S_{Ochiai}(\mathbf{p},\mathbf{q})$ for any number of communities is thus non-negative definite.

A.2. Conservation of the similarities among species and among communities in DSPCA

$\tilde{\mathbf{Y}}$ is the matrix with m rows and r columns defined in the main text. Matrices $\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}^{t}$ and $\tilde{\mathbf{Y}}^{t}\tilde{\mathbf{Y}}$ have the same s non-zero eigenvalues, where s = rank($\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}^{t}$) = rank($\tilde{\mathbf{Y}}^{t}\tilde{\mathbf{Y}}$), $s \leq$ min(r,m).

Consider the following eigenvalue decompositions:

$$\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}^{t}=\mathbf{A}_{m}\mathbf{\Psi}_{m}\mathbf{A}_{m}^{t}$$

where $\mathbf{A}_m$ is a matrix with eigenvectors (in columns) associated with all eigenvalues of $\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}^{t}$, and $\mathbf{\Psi}_{m}$ is the diagonal matrix with all eigenvalues on the diagonal including potential zero eigenvalues. Because $\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}^{t}$ is real symmetric, $\mathbf{A}_m$ is an $m \times m$ orthogonal matrix satisfying $\mathbf{A}_{m}^{t}\mathbf{A}_{m}=\mathbf{A}_{m}\mathbf{A}_{m}^{t}=\mathbf{I}_{m}$, where $\mathbf{I}_{m}$ is the $m \times m$ identity matrix (spectral decomposition theorem).

$$\tilde{\mathbf{Y}}^{t}\tilde{\mathbf{Y}}=\mathbf{B}_{r}\mathbf{\Psi}_{r}\mathbf{B}_{r}^{t}$$

where $\mathbf{B}_{r}$ is a matrix with eigenvectors (in columns) associated with all eigenvalues of $\tilde{\mathbf{Y}}^{t}\tilde{\mathbf{Y}}$, and $\mathbf{\Psi}_{r}$ is the diagonal matrix with all eigenvalues including potential zero eigenvalues on the diagonal. Because $\tilde{\mathbf{Y}}^{t}\tilde{\mathbf{Y}}$ is real symmetric, $\mathbf{B}_{r}$ is an $r \times r$ orthogonal matrix satisfying $\mathbf{B}_{r}^{t}\mathbf{B}_{r}=\mathbf{B}_{r}\mathbf{B}_{r}^{t}=\mathbf{I}_{r}$, where $\mathbf{I}_{r}$ is the $r \times r$ identity matrix (spectral decomposition theorem).

The following equalities also hold:

$$\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}^t = \mathbf{A}_s\boldsymbol{\Psi}_s\mathbf{A}_s^t$$

where $\mathbf{A}_s$ is a matrix with eigenvectors (in columns) associated with positive (non-zero) eigenvalues (in columns) for $\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}^t$, and $\boldsymbol{\Psi}_s$ is the diagonal matrix with positive eigenvalues on the diagonal. In addition, $\mathbf{A}_s^t\mathbf{A}_s = \mathbf{I}_s$.

$$\tilde{\mathbf{Y}}^t\tilde{\mathbf{Y}} = \mathbf{B}_s\boldsymbol{\Psi}_s\mathbf{B}_s^t$$

where $\mathbf{B}_s$ is a matrix with eigenvectors (in columns) associated with positive (non-zero) eigenvalues (in columns) for $\tilde{\mathbf{Y}}^t\tilde{\mathbf{Y}}$, and $\boldsymbol{\Psi}_s$ is the diagonal matrix with all positive eigenvalues on the diagonal. In addition, $\mathbf{B}_s^t\mathbf{B}_s = \mathbf{I}_s$.

Matrix $\mathbf{A}_s$ can be chosen to be equal to $\tilde{\mathbf{Y}}\mathbf{B}_s\boldsymbol{\Psi}_s^{-1/2}$. Indeed,

$$\left(\tilde{\mathbf{Y}}\mathbf{B}_s\boldsymbol{\Psi}_s^{-1/2}\right)^t\tilde{\mathbf{Y}}\mathbf{B}_s\boldsymbol{\Psi}_s^{-1/2} = \boldsymbol{\Psi}_s^{-1/2}\mathbf{B}_s^t\tilde{\mathbf{Y}}^t\tilde{\mathbf{Y}}\mathbf{B}_s\boldsymbol{\Psi}_s^{-1/2}$$

$$\left(\tilde{\mathbf{Y}}\mathbf{B}_s\boldsymbol{\Psi}_s^{-1/2}\right)^t\tilde{\mathbf{Y}}\mathbf{B}_s\boldsymbol{\Psi}_s^{-1/2} = \boldsymbol{\Psi}_s^{-1/2}\mathbf{B}_s^t\mathbf{B}_s\boldsymbol{\Psi}_s\mathbf{B}_s^t\mathbf{B}_s\boldsymbol{\Psi}_s^{-1/2} = \mathbf{I}_s$$

and

$$\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}^t\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}^t = \left(\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}^t\right)\left(\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}^t\right) = \mathbf{A}_s\boldsymbol{\Psi}_s^2\mathbf{A}_s^t$$

$$\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}^t\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}^t = \tilde{\mathbf{Y}}\left(\tilde{\mathbf{Y}}^t\tilde{\mathbf{Y}}\right)\tilde{\mathbf{Y}}^t = \tilde{\mathbf{Y}}\mathbf{B}_s\boldsymbol{\Psi}_s\mathbf{B}_s^t\tilde{\mathbf{Y}}^t$$

$$\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}^t\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}^t = \tilde{\mathbf{Y}}\mathbf{B}_s\boldsymbol{\Psi}_s^{-1/2}\boldsymbol{\Psi}_s^2\boldsymbol{\Psi}_s^{-1/2}\mathbf{B}_s^t\tilde{\mathbf{Y}}^t$$

Because $\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}^t$ is by definition a non-negative definite matrix, the previous equations (Seber, 2008, theorem 10.8, p. 220) imply that

$$\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}^t = \mathbf{A}_s \mathbf{\Psi}_s \mathbf{A}_s^t$$

and

$$\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}^t = \tilde{\mathbf{Y}}\mathbf{B}_s \mathbf{\Psi}_s^{-1/2} \mathbf{\Psi}_s \mathbf{\Psi}_s^{-1/2} \mathbf{B}_s^t \tilde{\mathbf{Y}}^t$$

and thus that matrix $\mathbf{A}_s$ can be chosen to be equal to $\tilde{\mathbf{Y}}\mathbf{B}_s \mathbf{\Psi}_s^{-1/2}$.

The final coordinates of the communities in DSPCA are thus given by

$$\mathbf{Y}_{\text{final}} = \tilde{\mathbf{Y}}\mathbf{B}_s = \mathbf{A}_s \mathbf{\Psi}_s^{1/2}$$

The similarities among communities are contained in $\mathbf{S}^{\text{com}} = \tilde{\mathbf{Y}}\tilde{\mathbf{Y}}^t$.

Given that

$$\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}^t = \tilde{\mathbf{Y}}\mathbf{B}_s \mathbf{\Psi}_s^{-1/2} \mathbf{\Psi}_s \mathbf{\Psi}_s^{-1/2} \mathbf{B}_s^t \tilde{\mathbf{Y}}^t$$

and thus

$$\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}^t = \tilde{\mathbf{Y}}\mathbf{B}_s \mathbf{B}_s^t \tilde{\mathbf{Y}}^t$$

then,

$$\mathbf{S}^{\text{com}} = \mathbf{Y}_{\text{final}} \mathbf{Y}_{\text{final}}^t .$$

The similarities among communities are conserved in the final space of DSPCA.

A.3. S_{Ochiai} treats relative and absolute abundances equally

Consider that for any i and k, $p_{ik} = n_{ik} / n_{i+}$, where n_{ik} is the absolute abundance of species k at site i (e.g., number of individuals from species k at site i), and n_{i+} is the total abundance at site i ($n_{i+} = \sum_k n_{ik}$).

$$S_{Ochiai} = \frac{\sum_{k,l} \frac{n_{ik}}{n_{i+}} \frac{n_{jl}}{n_{j+}} s_{kl}^{\text{spe}}}{\sqrt{\sum_{k,l} \frac{n_{ik}}{n_{i+}} \frac{n_{il}}{n_{i+}} s_{kl}^{\text{spe}}} \sqrt{\sum_{k,l} \frac{n_{jk}}{n_{j+}} \frac{n_{jl}}{n_{j+}} s_{kl}^{\text{spe}}}}$$

which yields

$$S_{Ochiai} = \frac{\sum_{k,l} n_{ik} n_{jl} s_{kl}^{\text{spe}}}{\sqrt{\sum_{k,l} n_{ik} n_{il} s_{kl}^{\text{spe}}} \sqrt{\sum_{k,l} n_{jk} n_{jl} s_{kl}^{\text{spe}}}}$$

A.4. On the dissimilarity index used by DPCoA

When species have no similarity and the dissimilarity (d_{kl}^{spe}) between any two species k and l is set equal to 1, then

$$DISC\left(\mathbf{p}_i, \mathbf{p}_j\right) = \frac{1}{2} \sum_{k,l} \left(p_{ik} - p_{il} \right)^2$$

If these species also have equal proportions in each of the compared communities and if a is the number of species shared by communities i and j, b is the number of species found in community i only, and c is the number of species found in community j but not i, then

$$DISC = \frac{1}{2}\left[a\left(\frac{1}{a+b} - \frac{1}{a+c} \right)^2 + b\frac{1}{\left(a+b\right)^2} + c\frac{1}{\left(a+c\right)^2} \right]$$

From this equation, it can easily be noted that *DISC* depends on the diversity within communities even if the two communities have no species in common. For instance, if $a = 0$, then if $b = 1$ and $c = 1$, $DISC = 1$; if $b = 1$ and $c = 10$, $DISC = 0.55$; if $b = 10$ and $c = 10$, $DISC = 0.10$. By contrast, the Ochiai index used by DSPCA when species have no similarity is

$$Ochiai = \frac{a}{\sqrt{a+b}\sqrt{a+c}}$$

If $a = 0$, the Ochiai index equals 0 and it does not depend on b and c.

Supplementary material

Supplementary material associated with this article can be found in the online version, at ---.

Appendixes B to E. Supplementary materials

References

Albert, A., 1969. Conditions for positive and nonnegative definiteness in terms of pseudoinverses. SIAM Journal on Applied Mathematics, 17, 434−440. DOI: 10.1137/0117041

Blondel, J., Vuilleumier, F. Marcus, L.F., Terouanne, E., 1984. Is there ecomorphological convergence among mediterranean bird communities of Chile, California, and France? In M.K. Hecht, B. Wallace, R.J. MacIntyre (Eds.), Evolutionary Biology (pp. 141−213). New York, NY: Plenum Press.

Bryant, J.A., Lamanna, C., Morlon, H., Kerkhoff, A.J., Enquist, A.J., Green, J.L., 2008. Microbes on mountainsides: contrasting elevational patterns of bacterial and plant diversity. Proceedings of the National Academy of Sciences of the United States of America, 105, 11505−11511. doi: 10.1073/pnas.0801920105

Cadotte, M.W., Albert, C., Walker, S., 2013. The ecology of differences: integrating evolutionary and functional distances. Ecology Letters, 16, 1234−1244. DOI: 10.1111/ele.12161

Corsten, L.C., Gabriel, K.R., 1976. Graphical exploration in comparing variance matrices. Biometrics, 32, 851−863. DOI: 10.2307/2529269

Desdevises, Y., Legendre, P., Azouzi, L., Morand S., 2003. Quantifying phylogenetically structured environmental variation. Evolution, 57, 2647−2652. DOI: 10.1111/j.0014-3820.2003.tb01508.x

Diniz-Filho, J.A.F., de Sant'Ana, C.E.R., Bini, L.M. 1998. An eigenvector method for estimating phylogenetic inertia. Evolution, 52, 1247−1262. DOI: 10.1111/j.1558-5646.1998.tb02006.x

Duarte, L.D.S, 2011. Phylogenetic habitat filtering influences forest nucleation in grasslands. Oikos, 120, 208–215. DOI: 10.1111/j.1600-0706.2010.18898.x

Ferrier, S., Manion, G., Elith, J., Richardson, K., 2007. Using generalized dissimilarity modelling to analyse and predict patterns of beta diversity in regional biodiversity assessment. Diversity and Distributions, 13, 252–264. DOI: 10.1111/j.1472-4642.2007.00341.x

Friedman, S., Weisberg, H.F., 1981. Interpreting the first eigenvalue of a correlation matrix. Education and Psychological Measurement, 41, 11–21. DOI: 10.1177/001316448104100102

Fritz, S.A., Bininda-Emonds, O.R.P., Purvis, A., 2009. Geographic variation in predictors of mammalian extinction risk: big is bad, but only in the tropics. Ecology Letters, 12, 538–549. DOI: 10.1111/j.1461-0248.2009.01307.x

Giannini, N.P., 2003. Canonical phylogenetic ordination. Systematic Biology, 52, 684–695. DOI: 10.1080/10635150390238888

Gower, J.C., 1971. A general coefficient of similarity and some of its properties. Biometrics, 27, 857–871. DOI: 10.2307/2528823

Graham, C.H., Fine, P.V.A., 2008. Phylogenetic beta diversity: linking ecological and evolutionary processes across space and time. Ecology Letters, 11, 1265–1277. DOI: 10.1111/j.1461-0248.2008.01256.x

Hooper, D.U., Chapin, F.S., Ewel, J.J., Hector, A., Inchausti, P., Lavorel, S., Lawton, J.H., Lodge, D.M., Loreau, M., Naeem, S., Schmid, B., Setälä, H., Symstad, A.J., Vandermeer, J., Wardle, D.A., 2005. Effects of biodiversity on ecosystem functioning: a consensus of current knowledge. Ecological Monographs, 75, 3–35. DOI: 10.1890/04-0922

Jost, L., 2006. Entropy and diversity. Oikos, 113, 363–375. DOI: 10.1111/j.2006.0030-1299.14714.x

Lande, R., 1996. Statistics and partitioning of species diversity, and similarity among multiple communities. Oikos, 76, 5–13. DOI: 10.2307/3545743

Lavorel, S., Garnier, E., 2002. Predicting changes in community composition and ecosystem functioning from plant traits: revisiting the Holy Grail. Functional ecology, 16, 545–556. DOI: 10.1046/j.1365-2435.2002.00664.x

Legendre, P., Legendre, L., 1998. Numerical ecology. Amsterdam, The Netherlands: Elsevier Science BV.

Mayfield, M.M., Levine, J.M., 2010. Opposing effects of competitive exclusion on the phylogenetic structure of communities. Ecology Letters, 13, 1085–1093. DOI: 10.1111/j.1461-0248.2010.01509.x

Medellín, R., Equihua, M., Amin, M.A., 2000. Bat diversity and abundance as indicators of disturbance in Neotropical rainforest. Conservation Biology, 14, 1666–1675. DOI: 10.1111/j.1523-1739.2000.99068.x

Morrison, D.F., 1978. Multivariate statistical methods. Singapore: McGraw-Hill.

Mouillot, D., Bellwood, D.R., Baraloto, C., Chave, J., Galzin, R., Harmelin-Vivien, M., Kulbicki, M., Lavergne, S., Lavorel, S., Mouquet, N., Paine, C.E.T., Renaud, J., Thuiller, W., 2013. Rare species support vulnerable functions in high-diversity ecosystems. PLoS Biology, 11, e1001569. DOI: 10.1371/journal.pbio.1001569

Ochiai, A., 1957. Zoogeographic studies on the soleoid fishes found in Japan and its neighbouring regions. Bulletin of the Japanese Society of Scientific Fisheries, 22, 526–530. DOI: 10.2331/suisan.22.526

Orloci, L., 1967. An agglomerative method for classification of plant communities. Journal of Ecology, 55, 193–206. DOI: 10.2307/2257725

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., Bonsall, M., 2011. Measuring biodiversity to explain community assembly: a unified approach. Biological Reviews, 86, 792–812. DOI: 10.1111/j.1469-185X.2010.00171.x

Pavoine, S., Dufour, A.B., Chessel, D., 2004. From dissimilarities among species to dissimilarities among communities: a double principal coordinate analysis. Journal of Theoretical Biology, 228, 523–537. DOI: 10.1016/j.jtbi.2004.02.014

Pavoine, S., Blondel, J., Dufour, A.B., Gasc, A., Bonsall, M.B., 2013. A new technique for analysing interacting factors affecting biodiversity patterns: crossed-DPCoA. PloS ONE, 8, e54530. DOI: 10.1371/journal.pone.0054530

Pavoine, S., Ricotta, C., 2014. Functional and phylogenetic similarity among communities. Methods in Ecology and Evolution, 5, 666−675. DOI: 10.1111/2041-210X.12193

Pélissier, R., Couteron, P., Dray, S., Sabatier, D., 2003. Consistency between ordination techniques and diversity measurements: two strategies for species occurrence data. Ecology, 84, 242–251. DOI: 10.1890/0012-9658(2003)084[0242:CBOTAD]2.0.CO;2

Power, M.E., Tilman, D., Estes, J.A., Menge, B.A., Bond, W.J., Mills, L.S., Daily, G., Castilla, J.C., Lubchenco, J., Paine, R.T., 1996. Challenges in the quest for keystones. BioScience, 46, 609–620. DOI: 10.2307/1312990

R Core Team, 2018. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Rao, C.R., 1982. Diversity and dissimilarity coefficients: a unified approach. Theoretical Population Biology, 21, 24–43. DOI: 10.1016/0040-5809(82)90004-1

Ricotta, C., Szeidl, L., 2009. Diversity partitioning of Rao's quadratic entropy. Theoretical Population Biology, 76, 299–302. DOI: 10.1016/j.tpb.2009.10.001

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

Seber, G.A.F., 2004. Multivariate observations. Hoboken, NJ: Wiley.

Seber, G.A.F., 2008. A matrix handbook for statisticians. Hoboken, NJ: Wiley.

Stegen, J., Hurlbert, A.H., 2011. Inferring ecological processes from taxonomic, phylogenetic and functional trait β-diversity. PloS ONE, 6, e20906. DOI: 10.1371/journal.pone.0020906

Swenson, N.G., 2013. The assembly of tropical tree community – the advances and shortcomings of phylogenetic and functional trait analyses. Ecography, 36, 264–276. DOI: 10.1111/j.1600-0587.2012.00121.x

Walker, B.H., 1992. Biodiversity and ecological redundancy. Conservation Biology, 6, 18–23. DOI: 10.1046/j.1523-1739.1992.610018.x

Webb, C.O., Ackerly, D.D., Kembel, S.W., 2008. Phylocom: software for the analyses of phylogenetic community structure and trait evolution. Bioinformatics, 24, 2098–2100. DOI: 10.1093/bioinformatics/btn358

Weiher, E., Freund, D., Bunton, T., Stefanski, A., Lee, T., Bentivenga, S., 2011. Advances, challenges and a developing synthesis of ecological community assembly theory. Philosophical Transactions of the Royal Society of London B: Biological Sciences, 366, 2403–2413. DOI: 10.1098/rstb.2011.0056

Figure legends

Fig. 1. Results of DSPCA applied to theoretical data set #1. The data set is described in the figure based on the matrix of species abundance in communities and the matrix of inter-specific similarities. Species are numbered from s1 to s12, and communities are numbered c1 and c2. The data set is split into three examples: (a) both species and communities are maximally distinct, but communities have different levels of species richness; (b) species are maximally distinct and community c2 is nested in c1; and (c) c1 and c2 share no species, but the most abundant species in c1 is similar to species in c2. Similarities here are theoretical in that they could represent functional, taxonomic or phylogenetic proximities. Principal components (PCs) 1 and 2 in the graphs are unit length and designate the first and second axes, respectively, of the DSPCA. The associated eigenvalues are shown in parentheses. Species arrows are frequently superimposed. For example, s1-8 means that the arrows associated with these species, from s1 to s8, are identical.

Single-column figure

Fig. 2. Analysis of theoretical data set #3: (a) the data set with the matrix of species' incidence in communities (species in blue, communities in red) and the table of trait values per species (I considered two traits t1 and t2), (b) results of DSPCA applied to the data set, (c) results of multidimensional scaling (MDS) applied to the matrix of distances between communities associated with DSPCA. I used grey levels (in b) and colors (in c) to better reveal the 3D regular tetrahedron formed by community points. In (b), species are distributed as follows from the center of the space to the periphery: from s9 to s1 along the arrow of community c1; from s11 to s18 along the c2 arrow, from s27 to s19 along the c3 arrow, and from s28 to s36 along the c4 arrow. The species that best particularizes each community is

indicated on the map. The PCi's in the graphs are unit length principal components of DSPCA. The scale of MDS axes is indicated in (c). The eigenvalues associated with DSPCA principal components and with MDS axes are shown in parentheses. The violet color is used each time species and community arrows and points are superimposed.

Single-column figure

Fig. 3. Results of (a-b) DSPCA and (c-e) DPCoA applied to theoretical data set #2. The four eigenvalues of DSPCA are all equal to 1. Those of DPCoA, three in number, are all equal to 0.009. The order of the axes is thus random in each of these analyses. With DSPCA, I provide a factorial map with principal components (PCs) 1 and 2 (c) and then with PC3 and PC4 (d) because the dissimilarity between any two communities is 1 (zero similarity) according to S_{Ochiai}. The dissimilarities between communities calculated by DPCoA are 0.2 between c1 and c2; 0.47 between c1 or c2 and c3 or c4 and 0.63 between c3 and c4. I display the associated factorial maps using all combinations of the three axes: (c) axes 1 and 2, (d) 1 and 3, and (e) 2 and 3. In the graphs produced by DPCoA, community and species points are superimposed.

Single-column figure

Fig. 4. Result of DSPCA applied to the bird data set considering species as maximally dissimilar: (a) Principal component (PC) 1 and 4; (b) PC2 and 6; (c) PC3 and 5. The eigenvalues associated with each PC are shown in parentheses. I provide the arrows of species on the factorial map of each panel together with community arrows. Labels for communities are defined as follows: Bu = Burgundy, PR = Provence (Pr), Ca = California, Ch = Chile; numbers 1 to 4 associated with the code of the region indicate the position on the gradient of vegetation complexity. Communities not positioned on a map actually have zero

coordinates on this map. For example, California and Chile have zero coordinates on map (a) because they do not share species with Provence and Burgundy. I also zoom in on species arrows on the left of the factorial maps. Arrows of species with similar distributions across the regions are superimposed. The number of superimposed arrows is indicated: for example, "3x" means 3 arrows for 3 species with the indicated distribution profile. Next to each species arrow, I provide the incidence of the species in each habitat of each region: in (a), eight squares indicate whether the species was (closed square) or was not (open square) observed in habitats 1 to 4 (from left to right), first in Burgundy and then in Provence; similarly in (b) and (c), four squares indicate whether the species was (closed square) or was not (open square) observed in habitats 1 to 4 (from left to right) in California (b) or Chile (c).

2-column figure

Fig. 5. Result of DSPCA applied to the bird data set considering similarities between species according to their foraging habits: (a) Principal component (PC) 1 with species arrows; (b) PC1 with community arrows; (b) PC2 and PC3 with species arrows; and (c) PC2 and PC3 with community arrows. Bar plots are presented above each species to indicate its affinity with the ground, trunk, bush, twig, foliage and aerial strata for foraging activities. The white color indicates that the species does not use the strata. In panels (a) and (b), strata are shown using the order indicated in (a). The arrows of species that use similar foraging strata are superimposed. See Fig. 4 for codes associated with communities.

Single-column figure

Fig. 6. Result of DSPCA applied to the bird data set considering similarities between species according to (a) their morphometry and (b) their taxonomy. See Fig. 4 for codes associated with communities. In each panel (a) and (b), I provide factorial maps separately for species

and communities to ease the visualization of arrows and labels. In both analyses (with morphometry and taxonomy), the first principal component (PC)1 was largely dominant with a very high eigenvalue compared to other axes. I thus provide PC1 first and then 2-dimensional plots with PC2 (abscissa) and PC3 (ordinates). Eigenvalues associated with each PC are shown in parentheses. On PC1, the community coordinates were so clustered that I have not indicated their labels. For both panels (a) and (b), PC1 is unit length; the scale for PC2 and PC3 is indicated separately for species coordinates and for community coordinates. In (a), I indicate the names of species with the five lowest and five highest coordinates on PC1. I also indicate the names of the species with the largest coordinates on either PC2 or PC3. Codes for species are (in alphabetical order): Aale = *Archilochus alexandri*, Asan = *Ammodramus sandwichensis*, Cann = *Calypte anna*, Ccos = *Calypte costae*, Csor = *Contopus sordidulus*, Ecit = *Emberiza citrinella*, Igal = *Icterus galbula*, Lexc = *Lanius excubitor*, Pcae = *Passerina caerulea*, Pery = *Pipilo erythrophthalmus*, Pfus = *Pipilo fuscus*, Pnit = *Phainopepla nitens*, Psib = *Phylloscopus sibilatrix*, Ptro = *Phylloscopus trochilus*, Salb = *Scelorchilus albicollis*, Scom = *Sylvia communis*, Sdes = *Sylviorthorhynchus desmursii*, Seur = *Sitta europaea*, Shor = *Sylvia hortensis*, Sloy = *Sturnella loyca*, Sneg = *Sturnella neglecta*, Svul = *Sturnus vulgaris*, Tmer = *Turdus merula*, Vhut = *Vireo huttoni*, Zmel = *Zonotrichia melodia*. In b), I group species by taxonomic group (genus, family or order depending on how close species from these groups were on the map). PC1 simply distinguishes Passeriformes with medium coordinates from species of other orders with low coordinates.

single-column figure

Fig. 7. Result of DSPCA applied to the abundance of bat species and their phylogenetic similarities along the disturbance gradient in Selva Lacandona of Chiapas. (a) Species scores on the first principal component (PC)1; (b) community scores on PC1; (c) species scores on

PC2 and PC3; (d) community scores on PC2 and PC3. Codes for communities are: F = rainforest; P = cacao plantation; O = old field; C = cornfield. Codes for species: Ajam= *Artibeus jamaicensis*; Alit= *A. lituratus*; Bdub= *Bauerus dubiaquercus*; Cbre= *Carollia brevicauda*; Cper= *C. perspicillata*; Dpha= *Dermanura phaeotis*; Dwat= *D. watsoni*; Gcom= *Glossophaga commissarisi*; Gsor= *G. soricina*; Mkea= *Myotis keaysi*; Mmeg= *Mormoops megalophylla*; Ppar= *Pteronotus parnellii*; Slil= *Sturnira lilium*; Ttri= *Thyroptera tricolor*; n27= all species descending from node named n27 in the phylogenetic tree (see Appendix C in Supplementary material); these include *Chiroderma villosum*, *Platyrrhinus helleri; Vampyressa pusilla*, *Vampyrodes major*, and *Uroderma bilobatum*.

single-column figure

Figure 1

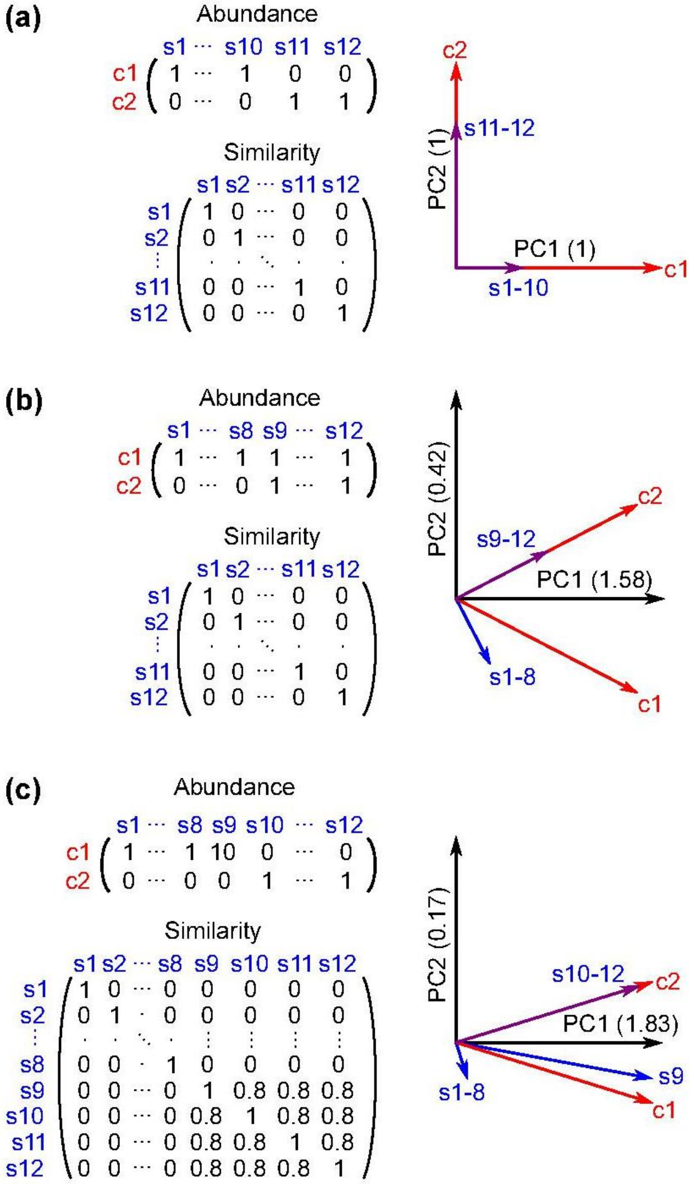
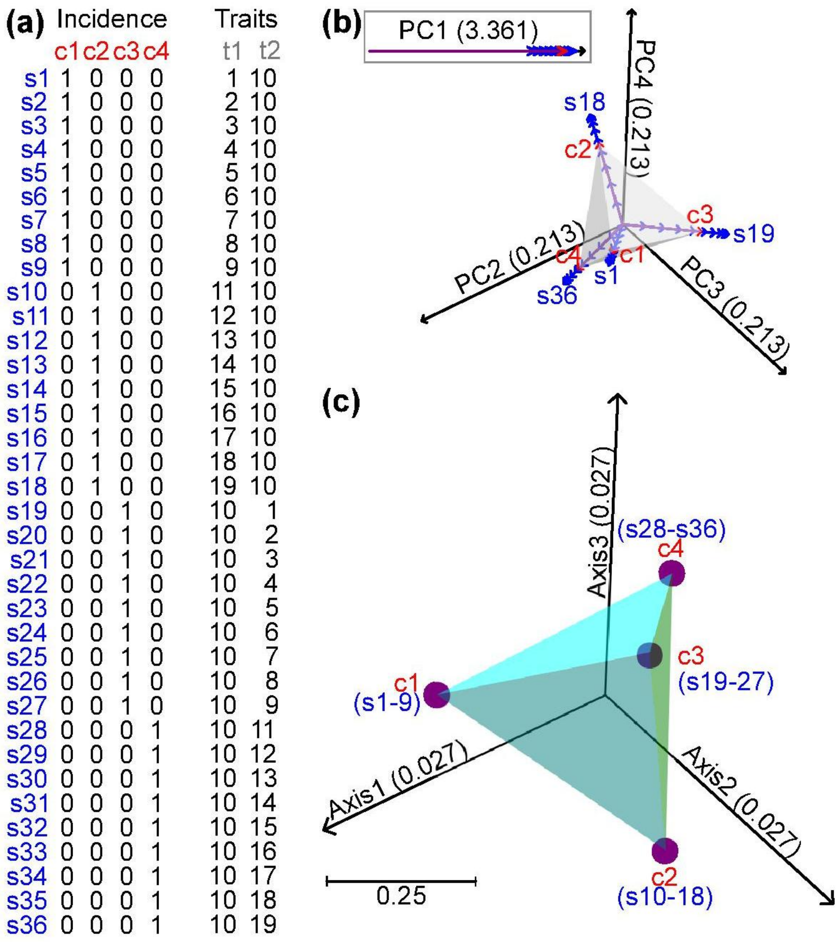

(a)
Abundance
s1 ··· s10 s11 s12
c1 1 ··· 1 0 0
c2 0 ··· 0 1 1
Similarity
s1 s2 ··· s11 s12
s1 1 0 ··· 0 0
s2 0 1 ··· 0 0
s11 0 0 ··· 1 0
s12 0 0 ··· 0 1
c2
s11-12
PC2 (1)
PC1 (1)
c1
s1-10
(b)
Abundance
s1 ··· s8 s9 ··· s12
c1 1 ··· 1 1 ··· 1
c2 0 ··· 0 1 ··· 1
Similarity
s1 s2 ··· s11 s12
s1 1 0 ··· 0 0
s2 0 1 ··· 0 0
s11 0 0 ··· 1 0
s12 0 0 ··· 0 1
PC2 (0.42)
c2
s9-12
PC1 (1.58)
s1-8
c1
(c)
Abundance
s1 ··· s8 s9 s10 ··· s12
c1 1 ··· 1 10 0 ··· 0
c2 0 ··· 0 0 1 ··· 1
Similarity
s1 s2 ··· s8 s9 s10 s11 s12
s1 1 0 ··· 0 0 0 0 0
s2 0 1 · 0 0 0 0 0
s8 0 0 · 1 0 0 0 0
s9 0 0 ··· 0 1 0.8 0.8 0.8
s10 0 0 ··· 0 0.8 1 0.8 0.8
s11 0 0 ··· 0 0.8 0.8 1 0.8
s12 0 0 ··· 0 0.8 0.8 0.8 1
PC2 (0.17)
s10-12
c2
PC1 (1.83)
s9
s1-8
c1

Figure 2



(a)

	Incidence				Traits	
	c1	c2	c3	c4	t1	t2
s1	1	0	0	0	1	10
s2	1	0	0	0	2	10
s3	1	0	0	0	3	10
s4	1	0	0	0	4	10
s5	1	0	0	0	5	10
s6	1	0	0	0	6	10
s7	1	0	0	0	7	10
s8	1	0	0	0	8	10
s9	1	0	0	0	9	10
s10	0	1	0	0	11	10
s11	0	1	0	0	12	10
s12	0	1	0	0	13	10
s13	0	1	0	0	14	10
s14	0	1	0	0	15	10
s15	0	1	0	0	16	10
s16	0	1	0	0	17	10
s17	0	1	0	0	18	10
s18	0	1	0	0	19	10
s19	0	0	1	0	10	1
s20	0	0	1	0	10	2
s21	0	0	1	0	10	3
s22	0	0	1	0	10	4
s23	0	0	1	0	10	5
s24	0	0	1	0	10	6
s25	0	0	1	0	10	7
s26	0	0	1	0	10	8
s27	0	0	1	0	10	9
s28	0	0	0	1	10	11
s29	0	0	0	1	10	12
s30	0	0	0	1	10	13
s31	0	0	0	1	10	14
s32	0	0	0	1	10	15
s33	0	0	0	1	10	16
s34	0	0	0	1	10	17
s35	0	0	0	1	10	18
s36	0	0	0	1	10	19

Figure 3

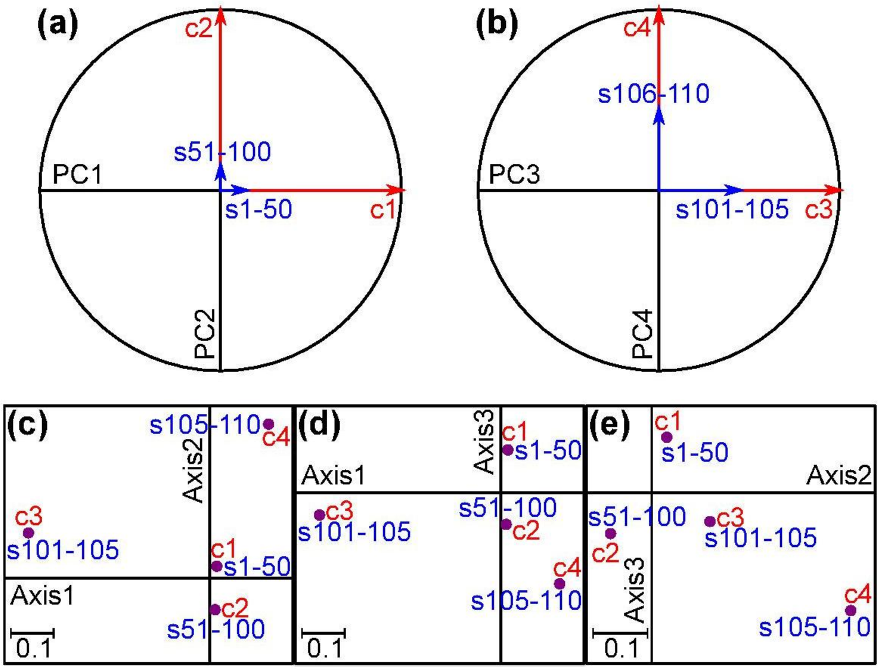

(a)
c2
s51-100
PC1
s1-50
c1
PC2
(b)
c4
s106-110
PC3
s101-105
c3
PC4
(c)
s105-110
c4
Axis2
c3
s101-105
c1
s1-50
Axis1
c2
s51-100
0.1
(d)
Axis3
c1
s1-50
Axis1
c3
s101-105
s51-100
c2
c4
s105-110
0.1
(e)
c1
s1-50
Axis2
s51-100
c2
c3
s101-105
Axis3
c4
s105-110
0.1

Figure 4

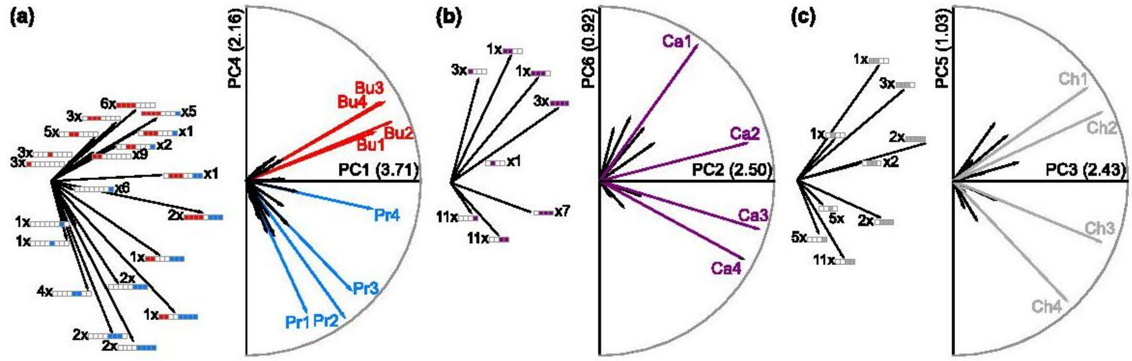

(a)
6x
3x
x5
5x
x1
3x
x2
3x
x9
x1
x6
2x
1x
1x
1x
2x
4x
1x
2x
2x
PC4 (2.16)
Bu3
Bu4
Bu2
Bu1
PC1 (3.71)
Pr4
Pr3
Pr1
Pr2
(b)
1x
3x
1x
3x
x1
11x
x7
11x
PC6 (0.92)
Ca1
Ca2
PC2 (2.50)
Ca3
Ca4
(c)
1x
3x
1x
2x
x2
5x
2x
5x
11x
PC5 (1.03)
Ch1
Ch2
PC3 (2.43)
Ch3
Ch4

Figure 5

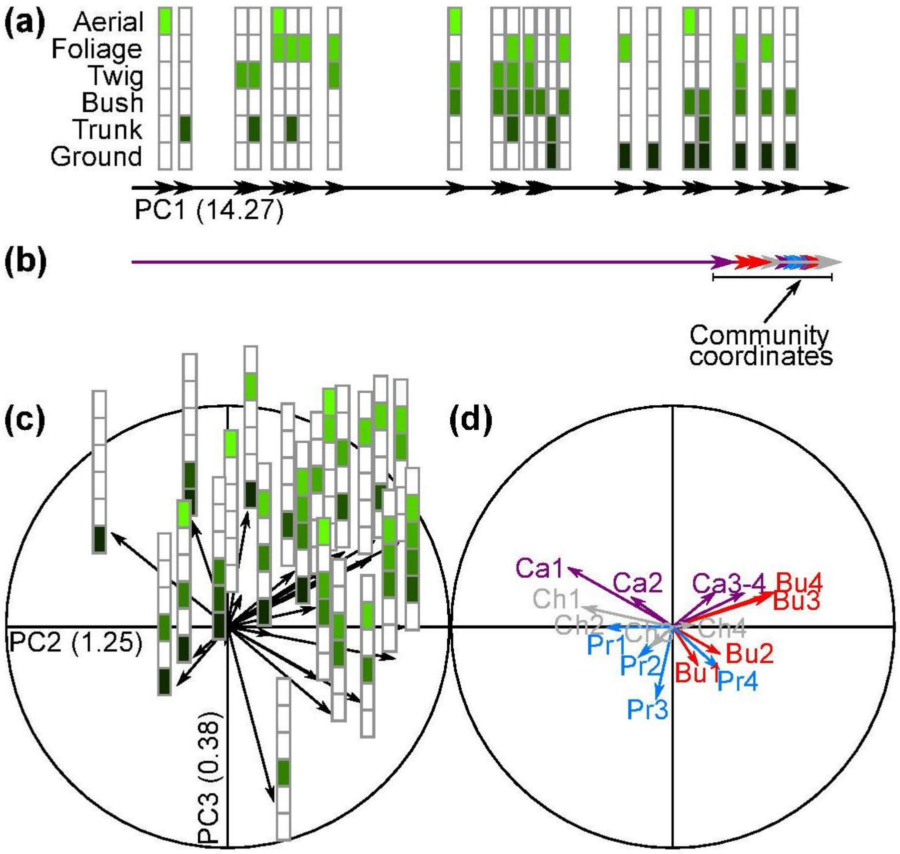

(a)
Aerial
Foliage
Twig
Bush
Trunk
Ground
PC1 (14.27)
(b)
Community coordinates
(c)
PC2 (1.25)
PC3 (0.38)
(d)
Ca1
Ca2
Ca3-4
Bu4
Bu3
Ch1
Ch2
Ch3
Ch4
Pr1
Pr2
Pr3
Pr4
Bu1
Bu2

Figure 6

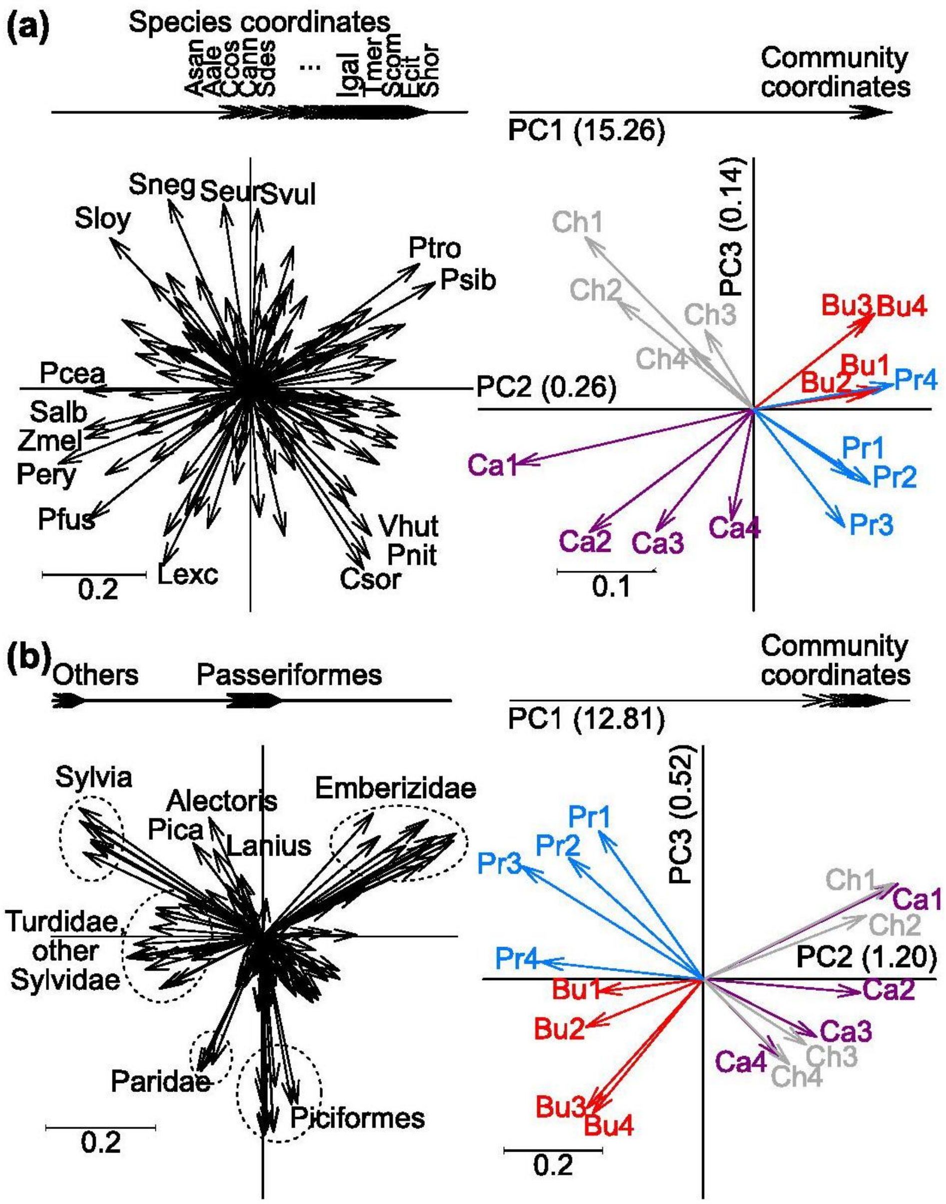
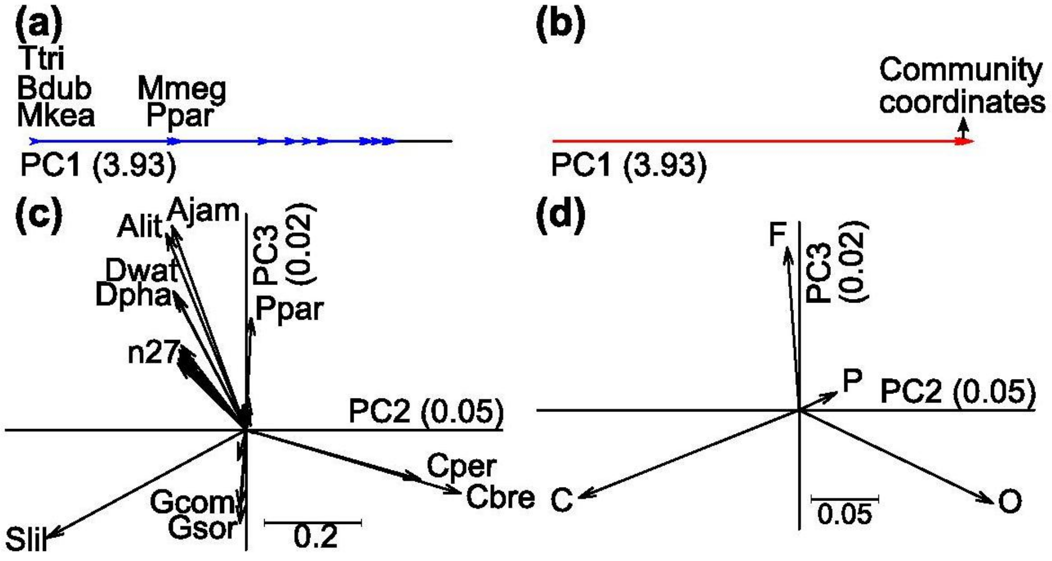

(a)
Species coordinates
Asan
Aale
Ccos
Cann
Sdes
...
Igal
Tmer
Scom
Ecit
Shor
Community coordinates
PC1 (15.26)
Sneg
Seur
Svul
Sloy
Ptro
Psib
Pcea
PC2 (0.26)
Salb
Zmel
Pery
Pfus
Lexc
Vhut
Pnit
Csor
0.2
PC3 (0.14)
Ch1
Ch2
Ch3
Ch4
Bu3
Bu4
Bu1
Bu2
Pr4
Pr1
Pr2
Pr3
Ca1
Ca2
Ca3
Ca4
0.1
(b)
Others
Passeriformes
Community coordinates
PC1 (12.81)
Sylvia
Alectoris
Pica
Lanius
Emberizidae
Turdidae, other Sylvidae
Paridae
Piciformes
0.2
PC3 (0.52)
Pr1
Pr2
Pr3
Pr4
Ch1
Ca1
Ch2
PC2 (1.20)
Bu1
Bu2
Ca2
Ca3
Ch3
Ca4
Ch4
Bu3
Bu4
0.2

Figure 7


(a)
Ttri
Bdub
Mkea
Mmeg
Ppar
PC1 (3.93)
(b)
Community
coordinates
PC1 (3.93)
(c)
Alit
Ajam
Dwat
Dpha
n27
PC3
(0.02)
Ppar
PC2 (0.05)
Cper
Cbre
Gcom
Gsor
Slil
0.2
(d)
F
PC3
(0.02)
P
PC2 (0.05)
C
O
0.05

Appendix B. An ordination approach to explore similarities among communities

S. Pavoine

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

The coordinate of a species in a principal component expresses how representative the species is of a similarity pattern

The coordinate of a species in a principal component expresses how representative the species is of the similarity pattern identified by the principal component. To illustrate this point, I used a series of short examples where only two communities (named c1 and c2) were compared for a total of two species only (named s1 and s2). I considered four matrices of species abundance within communities and two matrices of species similarities as described in Figure B.1. I applied DSPCA to each combination of the abundance and similarity matrices.

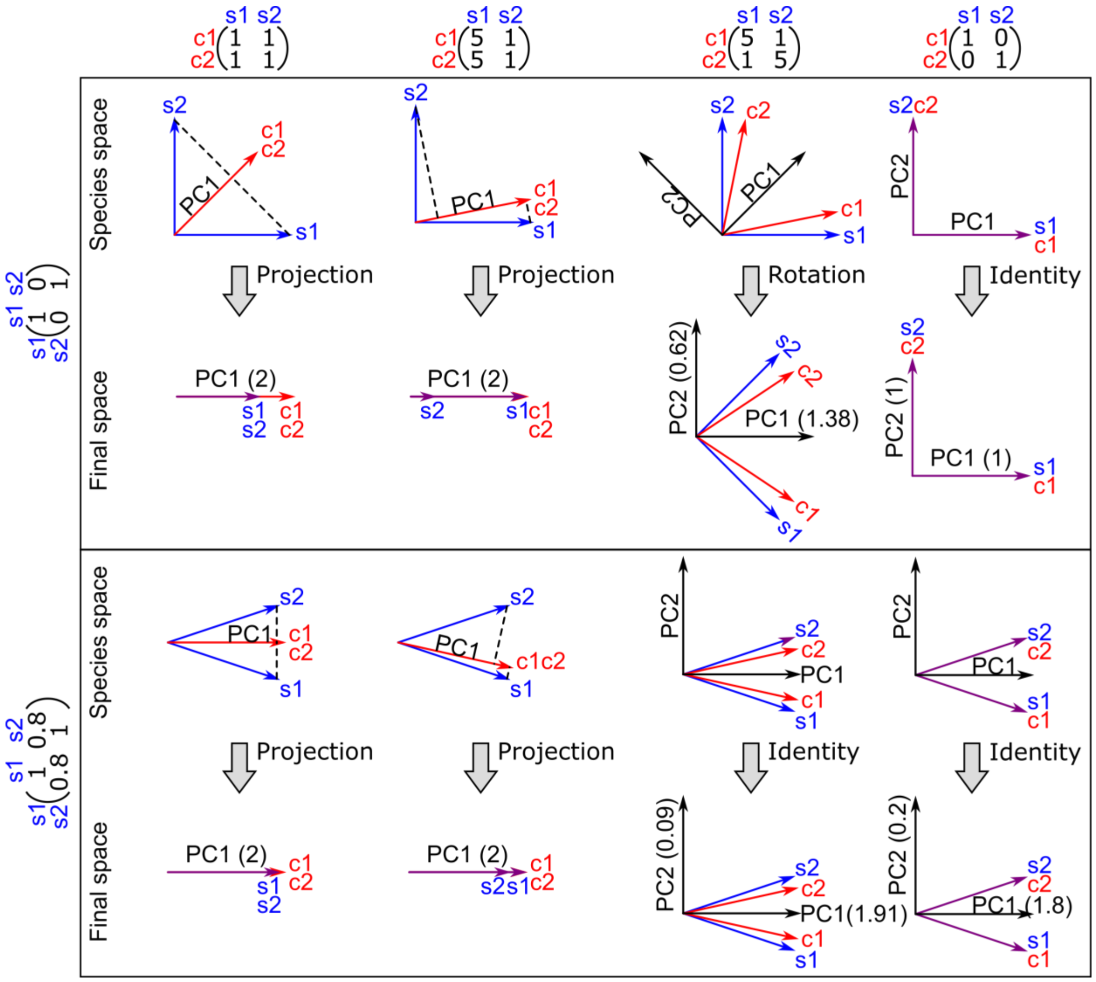


Fig. B.1. Results of DSPCA applied to theoretical data set #1. The results are presented for each matrix of species abundance within communities (columns) and for each matrix of similarities between species (rows). I first provide the intermediate *species space* where the representation of the similarities among species is optimized: species coordinates are shown by the rows of matrix **X** and community coordinates by the rows of $\mathbf{Y}_1$ (see Materials and Methods). Then, I provide the *final space* of the DSPCA where the representation of the similarities of the communities is optimized. PC1 and PC2 in the graphs are unit length and represent the first and second axes, respectively, of DSPCA. The associated eigenvalues are shown in parentheses in the final space. When DSPCA leads to a single axis because communities are identical (PC1), broken lines indicate directions of projection on this axis for the species arrows. Communities labeled c1 and c2 are displayed in red, and species labeled s1 and s2 are displayed in blue. The violet color is used each time species and community arrows are superimposed.

When species are maximally distinct, their arrows are orthogonal in the species space, the dimension of which is thus equal to the number of species. The arrow of a community is defined as the mean of the species arrows weighted by the species proportions in this community. All community arrows are then transformed to be unique lengths. The principal components of the community arrows are determined, and all arrows are projected in the space formed by these principal components. The simple examples provided in Figure B.1 show that the higher the redundancy between the species that drive the similarity structure between communities, the longer the species arrows.

An ordination approach to explore similarities among communities by S. Pavoine

Appendix C. Manual for R scripts

1 Functions

1.1 Function dspca

The R function dspca performs the ordination approach DSPCA. It will become part of package adiv of R. The reader can refer to the package for updated versions of the function. dspca has the following usage:

```
> dspca(com, S, tol=1e-8)
```

The parameters are defined as follows:

Parameter	**Explanation**
com	Data frame or matrix with communities as rows, species as columns and abundances, proportions or presences/absences (1/0) as entries.
S	Matrix of similarities among species (species as rows and columns in the same order as in df).
tol	a tolerance threshold: an absolute value is zero if it is lower than tol.

The result is a list of the following objects:

Parameter	**Explanation**
eig	Final eigenvalues (diagonal values of Ψ in the main text): positive eigenvalues of the matrix of similarities among communities.
X	Final coordinates of the species ($\mathbf{X}_{\text{final}}$): matrix with the coordinates of the species on the principal components associated with the matrix of similarities among communities. The names of the matrix start with "CPC" indicating "communities' principal component".
Y	Final coordinates of the communities ($\mathbf{Y}_{\text{final}}$): matrix with the coordinates of the communities on the principal components associated with the matrix of similarities among communities. The names of the matrix start with "CPC" indicating "communities' principal component".
Scom	Matrix of similarities among communities (obtained with coefficient S_{Ochiai}).

Coordinates can be visualized with graphic tools available in R. Examples are provided in the next section entitled "Applications".

2 Applications

Load the R function contained in Appendix D. For that, you can use:

```
> source(file.choose())
```

Install packages ade4, adiv, cluster, phylobase, adephylo and ape of R

```
> install.packages("ade4")
> install.packages("adiv")
> install.packages("cluster")
> install.packages("phylobase")
> install.packages("adephylo")
> install.packages("ape")
```

Load the packages:

```
> library(ade4)
> library(adiv)
> library(cluster)
> library(phylobase)
> library(adephylo)
> library(ape)
```

2.1 bird case study

Load the data set on bird communities:

```
> data(ecomor)
```

Species are coded in this data set. Latin names associated with codes are available in object labels of the list ecomor:

```
> head(ecomor$labels)
```

```
     latin                      abbr
E033 "Archilochus alexandri"    "Arc|ale"
E034 "Calypte anna"             "Cal|ann"
E035 "Calypte costae"           "Cal|cos"
E070 "Patagona gigas"           "Pat|gig"
E071 "Sephaniodes sephaniodes" "Sep|sep"
E001 "Columba palumbus"         "Col|pal"
```

Here are the instructions needed to reproduce the analyses done in the main text:

```
> com <- t(ecomor$habitat)
```

Species are maximally dissimilar

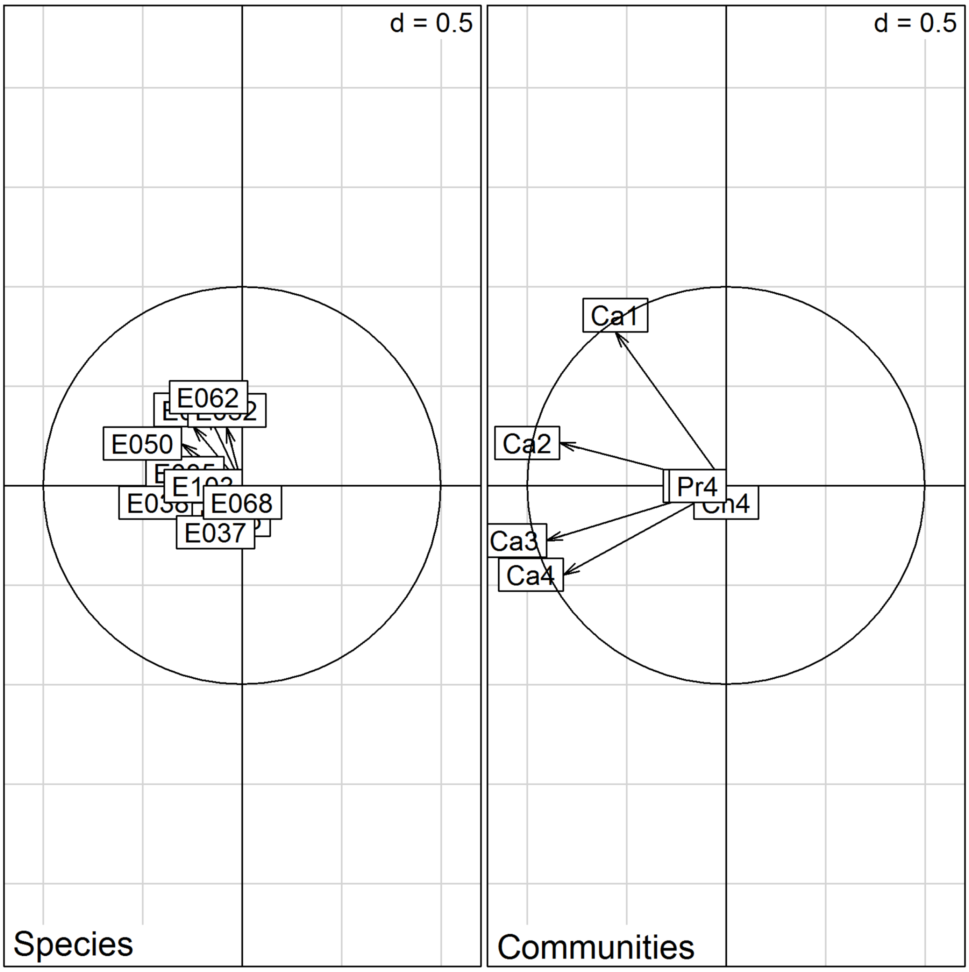
```
> Stax <- diag(rep(1,129))
> # DSPCA
> pcatax <- dspca(com, Stax)
> # Eigenvalues
> pcatax$eig
```

```
 [1] 3.71372822 2.50486238 2.43484097 2.16247525 1.02929799 0.92234935
 [7] 0.80009896 0.67074332 0.39492251 0.36252798 0.30807938 0.22778166
[13] 0.17786576 0.13431687 0.10314198 0.05296741
```

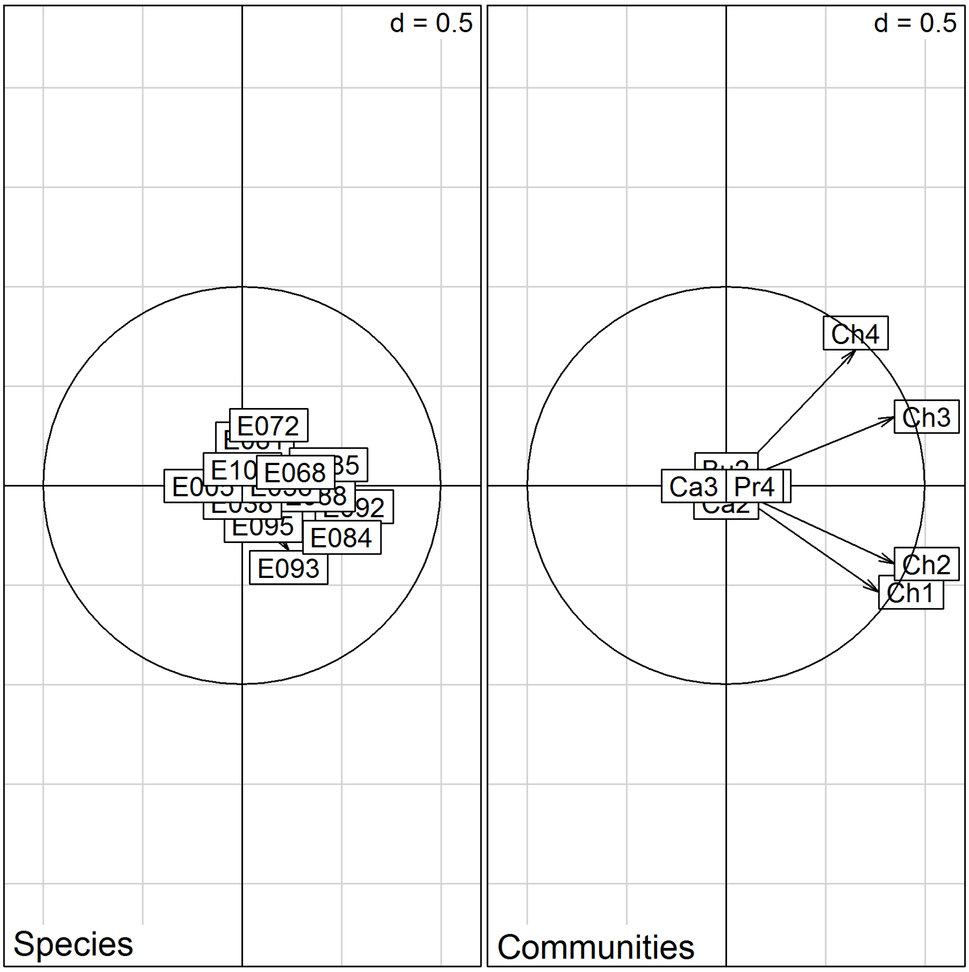
```
> # Axes 1 and 4:
> par.mar <- par()$mar
```

```
> par(mar=rep(0.1,4))
> par(mfrow=c(1,2))
> # Species
> ade4::s.arrow(pcatax$X, xax=1, yax=4, ylim=c(-1.2,1.2), xlim=c(-1.2,1.2),sub="Species")
> symbols(0,0,1, inch=F, add=TRUE)
> # Communities
> ade4::s.arrow(pcatax$Y, xax=1, yax=4, ylim=c(-1.2,1.2), xlim=c(-1.2,1.2),sub="Communities")
> symbols(0,0,1, inch=F, add=TRUE)
> par(mar=par.mar)
```

```
> # Axes 2 and 6:
> par.mar <- par()$mar
> par(mar=rep(0.1,4))
> par(mfrow=c(1,2))
> # Species
> ade4::s.arrow(pcatax$X, xax=2, yax=6, ylim=c(-1.2,1.2), xlim=c(-1.2,1.2),sub="Species")
> symbols(0,0,1, inch=F, add=TRUE)
> # Communities
> ade4::s.arrow(pcatax$Y, xax=2, yax=6, ylim=c(-1.2,1.2), xlim=c(-1.2,1.2),sub="Communities")
> symbols(0,0,1, inch=F, add=TRUE)
> par(mar=par.mar)
```



```
> # Axes 3 and 5:
> par.mar <- par()$mar
> par(mar=rep(0.1,4))
> par(mfrow=c(1,2))
> # Species
> ade4::s.arrow(pcatax$X, xax=3, yax=5, ylim=c(-1.2,1.2), xlim=c(-1.2,1.2),sub="Species")
> symbols(0,0,1, inch=F, add=TRUE)
> # Communities
> ade4::s.arrow(pcatax$Y, xax=3, yax=5, ylim=c(-1.2,1.2), xlim=c(-1.2,1.2),sub="Communities")
> symbols(0,0,1, inch=F, add=TRUE)
> par(mar=par.mar)
```

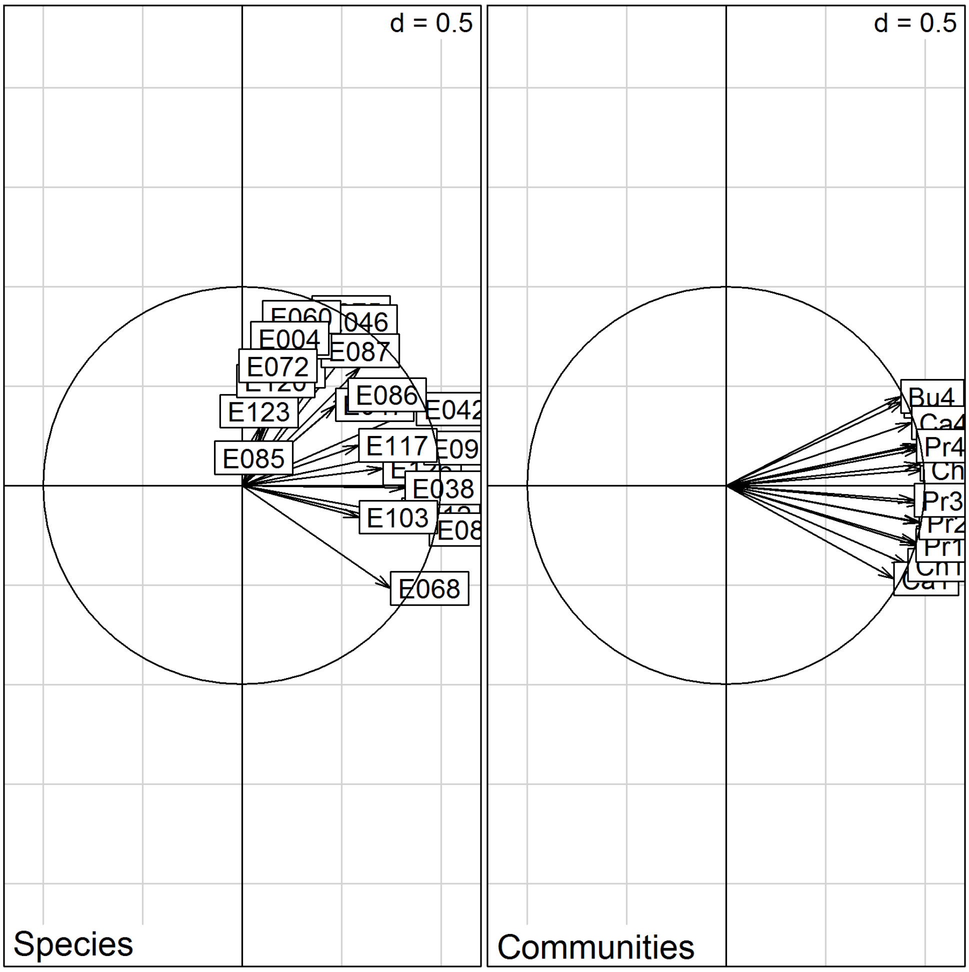


```
> # Similarities between communities (sample):
> (pcatax$Y%*%t(pcatax$Y))[1:5,1:5]

             Bu1          Bu2          Bu3          Bu4          Ca1
Bu1 1.000000e+00 8.576900e-01 5.003702e-01 4.728054e-01 3.412386e-18
Bu2 8.576900e-01 1.000000e+00 6.685032e-01 6.316762e-01 9.412505e-19
Bu3 5.003702e-01 6.685032e-01 1.000000e+00 9.449112e-01 1.250755e-18
Bu4 4.728054e-01 6.316762e-01 9.449112e-01 1.000000e+00 7.037151e-19
Ca1 3.412386e-18 9.412505e-19 1.250755e-18 7.037151e-19 1.000000e+00

> pcatax$Scom[1:5,1:5]

          Bu1       Bu2       Bu3       Bu4 Ca1
Bu1 1.0000000 0.8576900 0.5003702 0.4728054   0
Bu2 0.8576900 1.0000000 0.6685032 0.6316762   0
Bu3 0.5003702 0.6685032 1.0000000 0.9449112   0
Bu4 0.4728054 0.6316762 0.9449112 1.0000000   0
Ca1 0.0000000 0.0000000 0.0000000 0.0000000   1
```

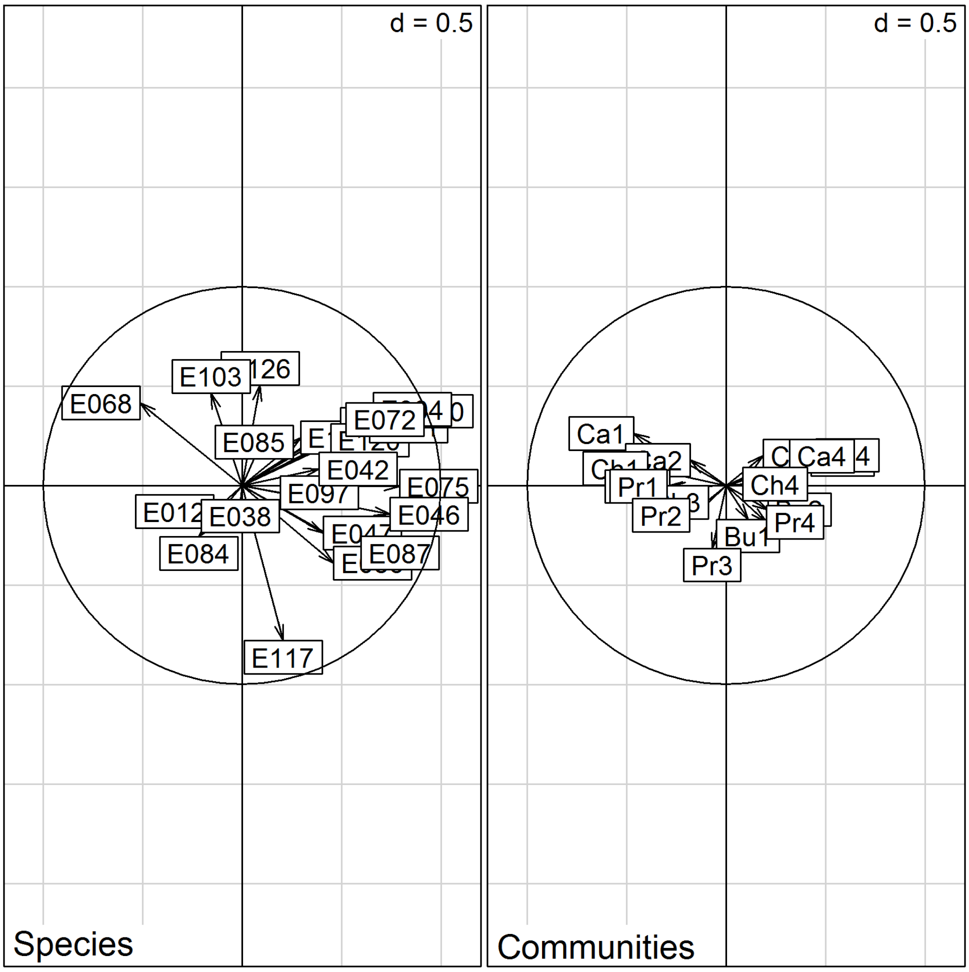
Similarities between species according to the place where they forage:

```
> Sfor <- dsimFun(ecomor$forsub, "M", method=4, type="similarity")
> # DSPCA
> pcafor <- dspca(com, Sfor)
> # Eigenvalues
> pcafor$eig
```

```
[1] 14.26843633  1.24601104  0.38472267  0.05064909  0.02574553  0.02443534

> # Axes 1 and 2:
> par.mar <- par()$mar
> par(mar=rep(0.1,4))
> par(mfrow=c(1,2))
> # Species
> ade4::s.arrow(pcafor$X, ylim=c(-1.2,1.2), xlim=c(-1.2,1.2),sub="Species")
> symbols(0,0,1, inch=F, add=TRUE)
> # Communities
> ade4::s.arrow(pcafor$Y, ylim=c(-1.2,1.2), xlim=c(-1.2,1.2),sub="Communities")
> symbols(0,0,1, inch=F, add=TRUE)
> par(mar=par.mar)
```



```
> # Axes 2 and 3:
> par.mar <- par()$mar
> par(mar=rep(0.1,4))
> par(mfrow=c(1,2))
> # Species
> ade4::s.arrow(pcafor$X, xax=2, yax=3, ylim=c(-1.2,1.2), xlim=c(-1.2,1.2),sub="Species")
> symbols(0,0,1, inch=F, add=TRUE)
> # Communities
> ade4::s.arrow(pcafor$Y, xax=2, yax=3, ylim=c(-1.2,1.2), xlim=c(-1.2,1.2),sub="Communities")
> symbols(0,0,1, inch=F, add=TRUE)
> par(mar=par.mar)
```



```
> # Similarities between communities (sample)
> (pcafor$Y%*%t(pcafor$Y))[1:5,1:5]
```

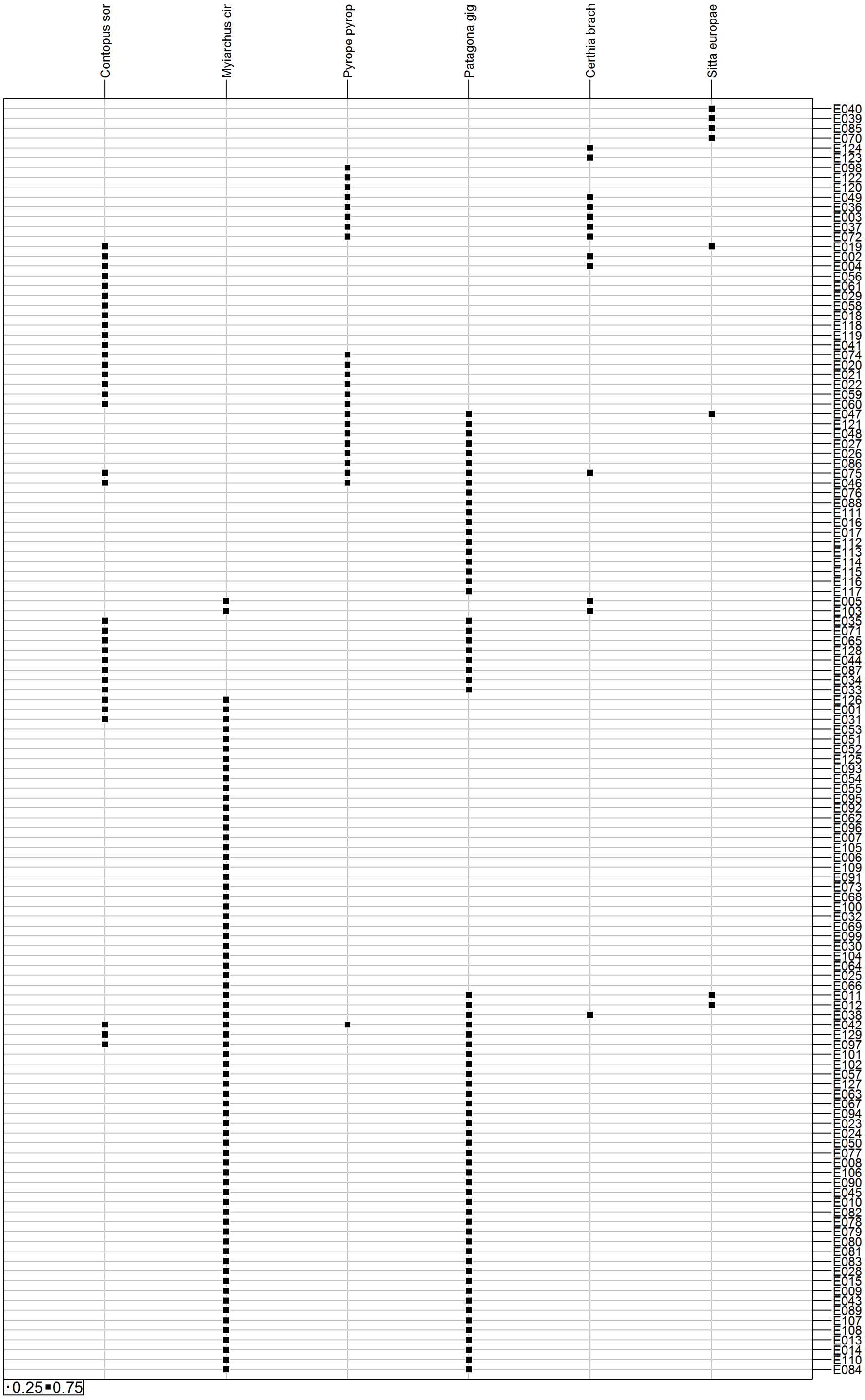
```
          Bu1       Bu2       Bu3       Bu4       Ca1
Bu1 1.0000000 0.9930174 0.8896489 0.8684885 0.7262468
Bu2 0.9930174 1.0000000 0.9318144 0.9159583 0.6864214
Bu3 0.8896489 0.9318144 1.0000000 0.9971959 0.5914888
Bu4 0.8684885 0.9159583 0.9971959 1.0000000 0.5692410
Ca1 0.7262468 0.6864214 0.5914888 0.5692410 1.0000000
```

```
> pcafor$Scom[1:5,1:5]
```

```
          Bu1       Bu2       Bu3       Bu4       Ca1
Bu1 1.0000000 0.9930174 0.8896489 0.8684885 0.7262468
Bu2 0.9930174 1.0000000 0.9318144 0.9159583 0.6864214
Bu3 0.8896489 0.9318144 1.0000000 0.9971959 0.5914888
Bu4 0.8684885 0.9159583 0.9971959 1.0000000 0.5692410
Ca1 0.7262468 0.6864214 0.5914888 0.5692410 1.0000000
```

Table of species foraging substrates where species are ordered according to the first axis of DSPCA:

```
> table.value(ecomor$forsub[order(pcafor$X[,1]), ], ppoints.cex = 0.2,
+     labelsx = ecomor$labels[rownames(ecomor$forsub[order(pcafor$X[,1]), ]), 1])
```



A close square in the graph means that the species forage on the specified substrate. Legends for substrates are available with the following instruction: ?ecomor.

Species are characterized according to morphometrical traits:

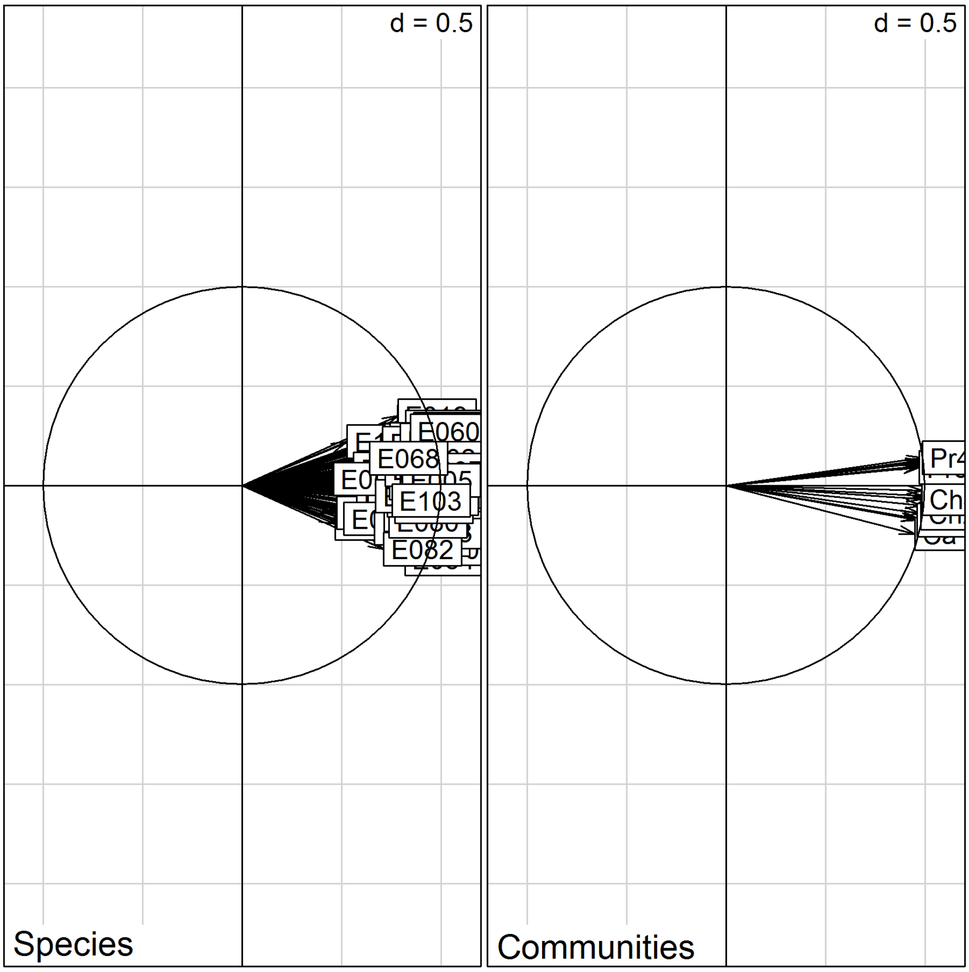
To remove redundancies between morphometric traits, I performed a principal component analysis (PCA) on the morphometric traits. Then, I applied Gower similarity index to the normed coordinates of the species in PCA:

```
> pcamorpho <- dudi.pca(log(ecomor$morpho), scann=FALSE, nf=8)
> Dmor <- dsimFun(pcamorpho$l1[colnames(com), ], "Q", type="dissimilarity")
> Smor <- 1- as.matrix(Dmor/max(Dmor))
```

```
> #DSPCA
> pcamor <- dspca(com, Smor)
> # eigenvalues
> pcamor$eig
```

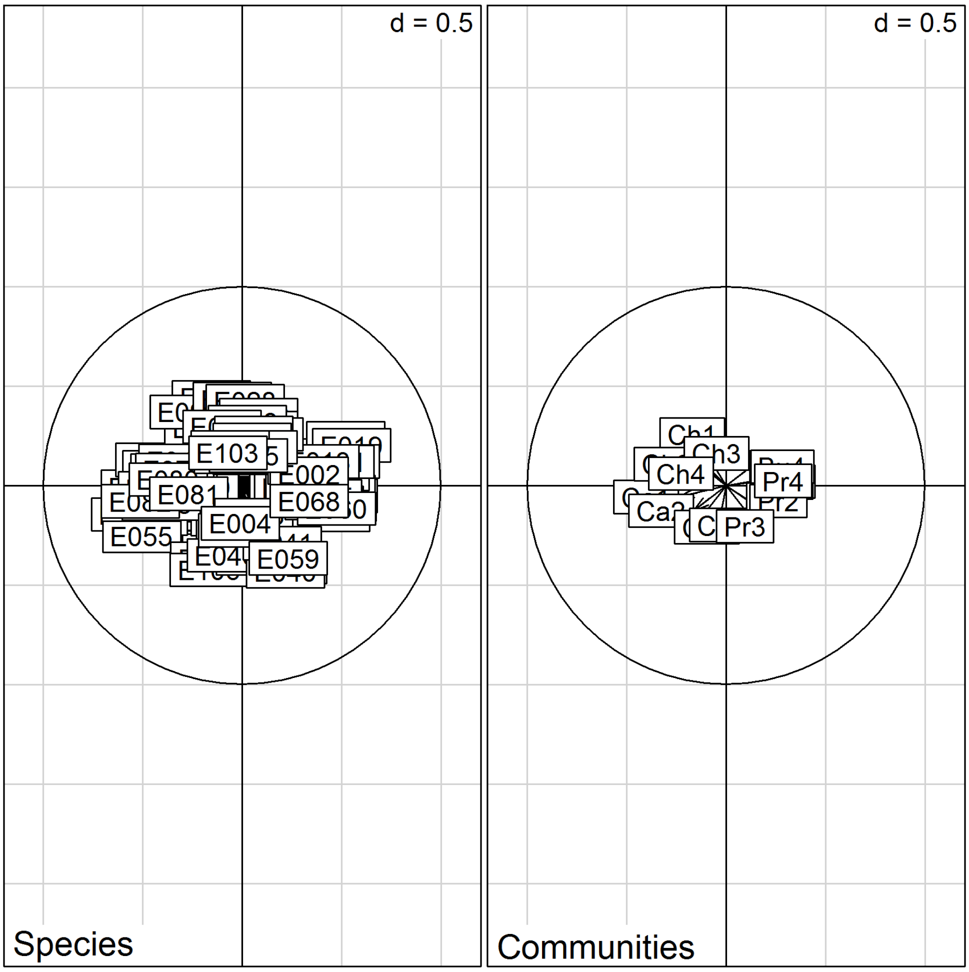
```
 [1] 1.525938e+01 2.631015e-01 1.438653e-01 1.176506e-01 7.967625e-02
 [6] 4.805947e-02 3.345960e-02 1.412404e-02 1.368605e-02 9.434050e-03
[11] 6.596879e-03 3.757320e-03 3.048406e-03 2.132844e-03 1.305462e-03
[16] 7.234662e-04
```

```
> # Axes 1 and 2
> par.mar <- par()$mar
> par(mar=rep(0.1,4))
> par(mfrow=c(1,2))
> # Species
> ade4::s.arrow(pcamor$X, ylim=c(-1.2,1.2), xlim=c(-1.2,1.2),sub="Species")
> symbols(0,0,1, inch=F, add=TRUE)
> # Communities
> ade4::s.arrow(pcamor$Y, ylim=c(-1.2,1.2), xlim=c(-1.2,1.2),sub="Communities")
> symbols(0,0,1, inch=F, add=TRUE)
> par(mar=par.mar)
```



```
> # Axes 2 and 3
> par.mar <- par()$mar
> par(mar=rep(0.1,4))
```

```
> par(mfrow=c(1,2))
> # Species
> ade4::s.arrow(pcamor$X, xax=2, yax=3, ylim=c(-1.2,1.2), xlim=c(-1.2,1.2),sub="Species")
> symbols(0,0,1, inch=F, add=TRUE)
> # Communities
> ade4::s.arrow(pcamor$Y, xax=2, yax=3, ylim=c(-1.2,1.2), xlim=c(-1.2,1.2),sub="Communities")
> symbols(0,0,1, inch=F, add=TRUE)
> par(mar=par.mar)
```



```
> # Similarities between communities (sample)
> (pcamor$Y%*%t(pcamor$Y))[1:5,1:5]

          Bu1       Bu2       Bu3       Bu4       Ca1
Bu1 1.0000000 0.9954334 0.9750488 0.9749757 0.9108033
Bu2 0.9954334 1.0000000 0.9852318 0.9855155 0.9178688
Bu3 0.9750488 0.9852318 1.0000000 0.9988637 0.8998454
Bu4 0.9749757 0.9855155 0.9988637 1.0000000 0.8979839
Ca1 0.9108033 0.9178688 0.8998454 0.8979839 1.0000000
```

```
> pcamor$Scom [1:5,1:5]

          Bu1       Bu2       Bu3       Bu4       Ca1
Bu1 1.0000000 0.9954334 0.9750488 0.9749757 0.9108033
Bu2 0.9954334 1.0000000 0.9852318 0.9855155 0.9178688
Bu3 0.9750488 0.9852318 1.0000000 0.9988637 0.8998454
```

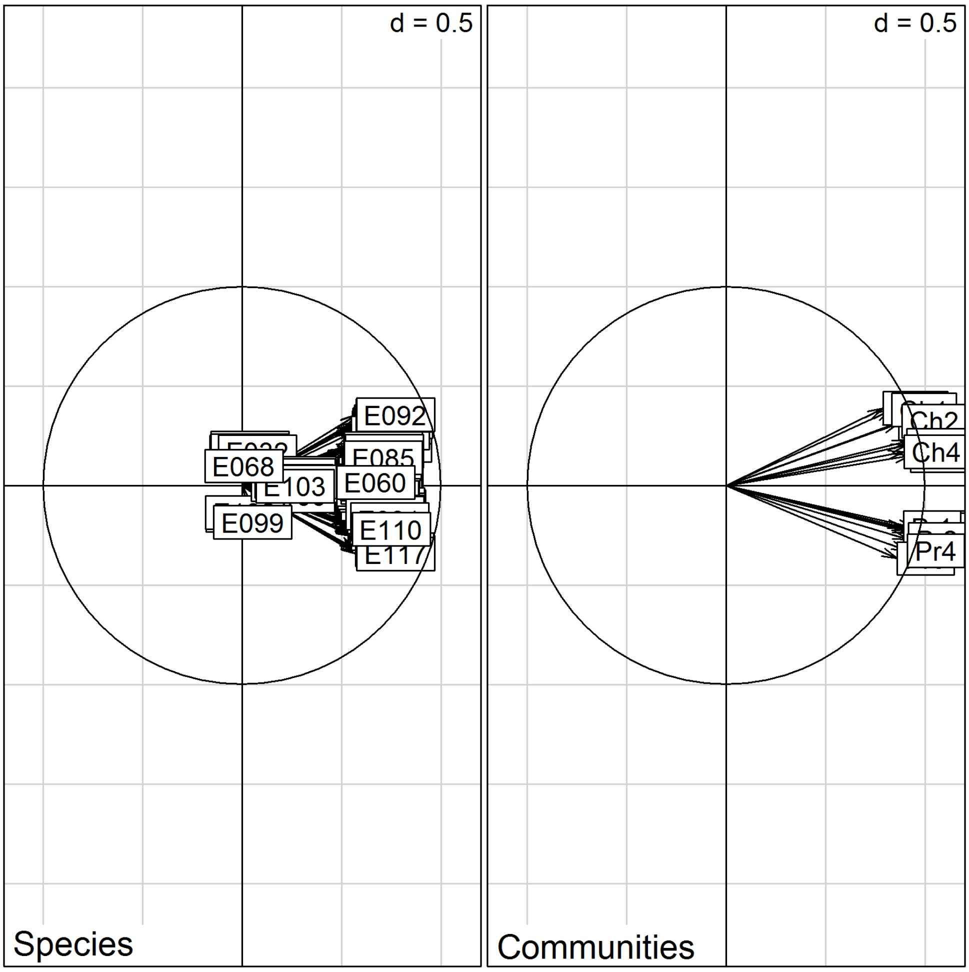
```
Bu4 0.9749757 0.9855155 0.9988637 1.0000000 0.8979839
Ca1 0.9108033 0.9178688 0.8998454 0.8979839 1.0000000
```

Taxonomic similarities between species

```
> Staxo <- dsimTaxo(ecomor$taxo[rownames(ecomor$habitat),], method=4)
> # DSPCA applied to taxonomic data:
> pcataxo <- dspca(com, Staxo)
> # eigenvalues
> pcataxo$eig
```

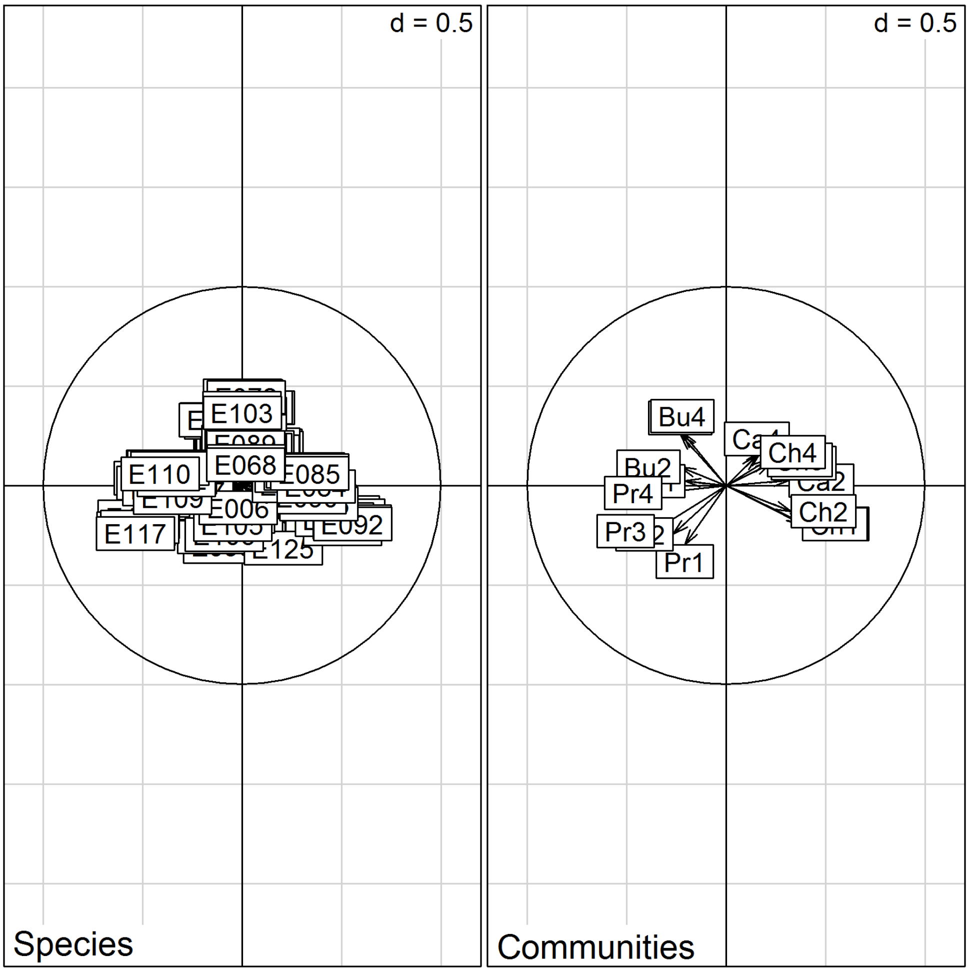
```
 [1] 12.806414191  1.204822868  0.521027183  0.417689369  0.333060229
 [6]  0.188640751  0.139448364  0.087736030  0.081941114  0.061053488
[11]  0.051925851  0.035700704  0.033457124  0.021674194  0.009519878
[16]  0.005888662
```

```
> # Axes 1 and 2
> par.mar <- par()$mar
> par(mar=rep(0.1,4))
> par(mfrow=c(1,2))
> # Species
> ade4::s.arrow(pcataxo$X, ylim=c(-1.2,1.2), xlim=c(-1.2,1.2),sub="Species")
> symbols(0,0,1, inch=F, add=TRUE)
> # Communities
> ade4::s.arrow(pcataxo$Y, ylim=c(-1.2,1.2), xlim=c(-1.2,1.2),sub="Communities")
> symbols(0,0,1, inch=F, add=TRUE)
> par(mar=par.mar)
```



```
> # Axes 2 and 3
> par.mar <- par()$mar
> par(mar=rep(0.1,4))
> par(mfrow=c(1,2))
> # Species
> ade4::s.arrow(pcataxo$X, xax=2, yax=3, ylim=c(-1.2,1.2), xlim=c(-1.2,1.2),sub="Species")
> symbols(0,0,1, inch=F, add=TRUE)
> # Communities
> ade4::s.arrow(pcataxo$Y, xax=2, yax=3, ylim=c(-1.2,1.2), xlim=c(-1.2,1.2),sub="Communities")
> symbols(0,0,1, inch=F, add=TRUE)
> par(mar=par.mar)
> # Similarities between communities (sample)
> (pcataxo$Y%*%t(pcataxo$Y))[1:5,1:5]

          Bu1       Bu2       Bu3       Bu4       Ca1
Bu1 1.0000000 0.9811047 0.9201980 0.9113432 0.6740227
Bu2 0.9811047 1.0000000 0.9512201 0.9425491 0.6616646
Bu3 0.9201980 0.9512201 1.0000000 0.9932061 0.6221710
Bu4 0.9113432 0.9425491 0.9932061 1.0000000 0.6202234
Ca1 0.6740227 0.6616646 0.6221710 0.6202234 1.0000000

> pcataxo$Scom [1:5,1:5]

          Bu1       Bu2       Bu3       Bu4       Ca1
Bu1 1.0000000 0.9811047 0.9201980 0.9113432 0.6740227
Bu2 0.9811047 1.0000000 0.9512201 0.9425491 0.6616646
```

```
Bu3 0.9201980 0.9512201 1.0000000 0.9932061 0.6221710
Bu4 0.9113432 0.9425491 0.9932061 1.0000000 0.6202234
Ca1 0.6740227 0.6616646 0.6221710 0.6202234 1.0000000
```



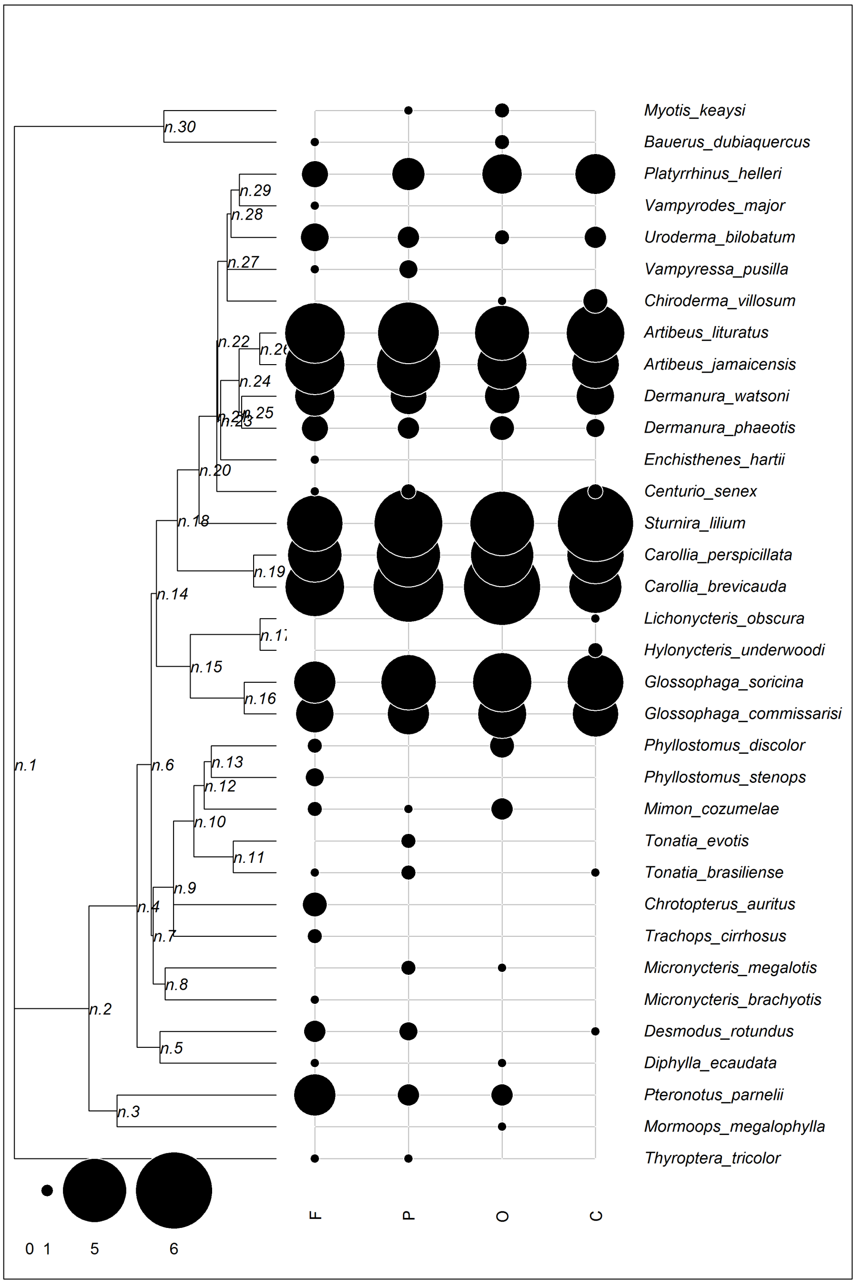
2.2 Bat data set

Load the data set on bat communities.

```
> data(batcomm)
> phy <- read.tree(text=batcomm$tre) # phylogenetic tree
> ab <- batcomm$ab # abundances of species within habitats
```

Species abundances in front of the phylogenetic tree (log-transformed abundance):

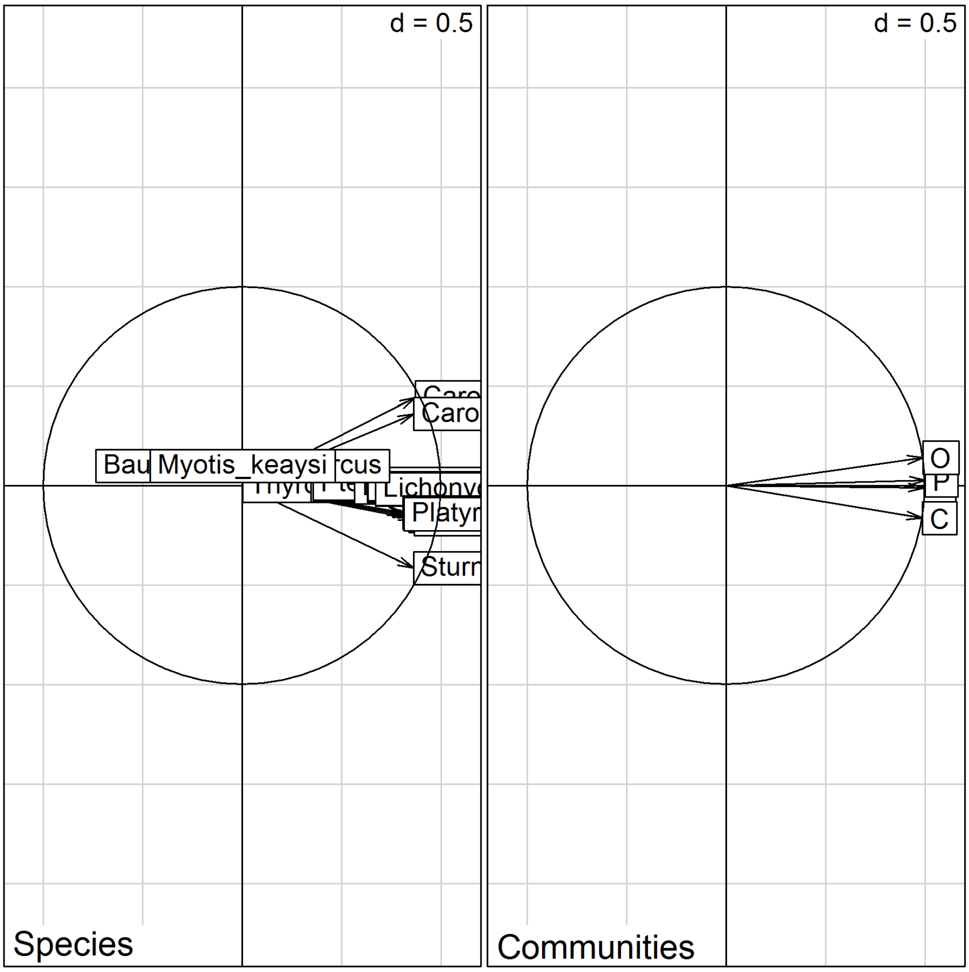
```
> # Axes 1 to 3
> bat.4d <- phylo4d(phy, log(t(ab[, phy$tip.label])+1))
> table.phylo4d(bat.4d, center = FALSE, scale = FALSE, cex.symbol=2)
```



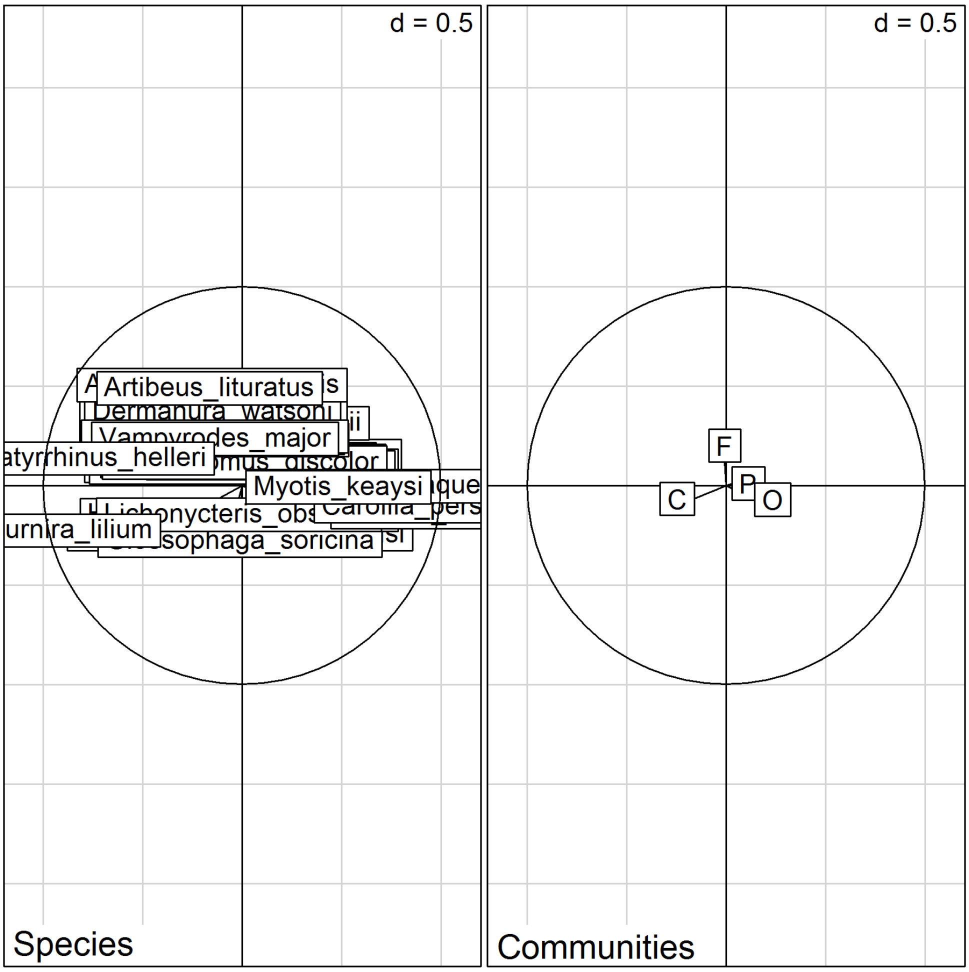
Legend: F=rainforest; P=cacao plantation; O=oldfields; C=cornfields

```
> # Phylogenetic similarities between species
> Sphy <- dsimTree(phy, method=4)
> # DSPCA
> pcaphy <- dspca(ab[, rownames(Sphy)], Sphy)

> # Axes 1 and 2
> par.mar <- par()$mar
> par(mar=rep(0.1,4))
> par(mfrow=c(1,2))
> # Species
```

```
> ade4::s.arrow(pcaphy$X, ylim=c(-1.2,1.2), xlim=c(-1.2,1.2),sub="Species")
> symbols(0,0,1, inch=F, add=TRUE)
> # Communities
> ade4::s.arrow(pcaphy$Y, ylim=c(-1.2,1.2), xlim=c(-1.2,1.2),sub="Communities")
> symbols(0,0,1, inch=F, add=TRUE)
> par(mar=par.mar)
```

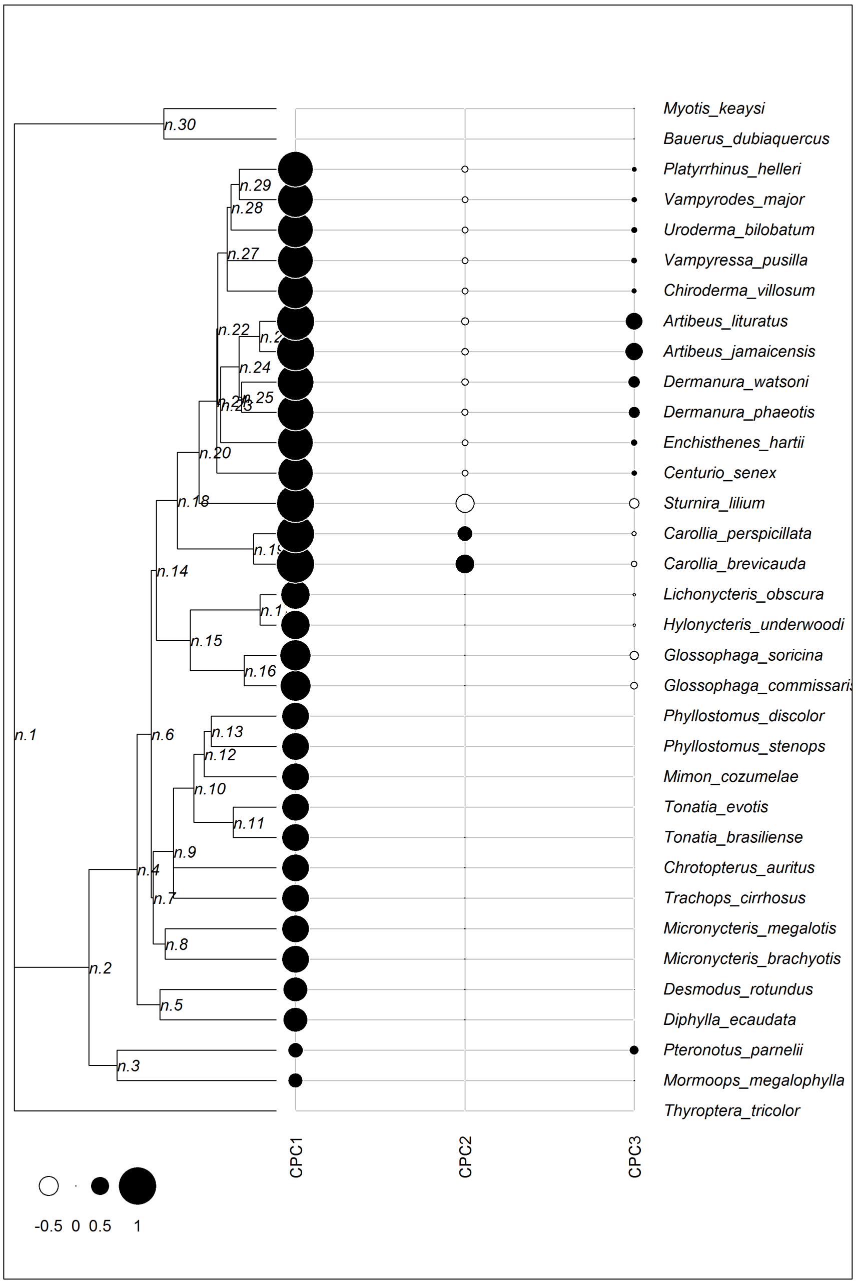


```
> # Axes 2 and 3
> par.mar <- par()$mar
> par(mar=rep(0.1,4))
> par(mfrow=c(1,2))
> # Species
> ade4::s.arrow(pcaphy$X, xax=2, yax=3, ylim=c(-1.2,1.2), xlim=c(-1.2,1.2),sub="Species")
> symbols(0,0,1, inch=F, add=TRUE)
> # Communities
> ade4::s.arrow(pcaphy$Y, xax=2, yax=3, ylim=c(-1.2,1.2), xlim=c(-1.2,1.2),sub="Communities")
> symbols(0,0,1, inch=F, add=TRUE)
> par(mar=par.mar)
```



Species coordinates in front of the phylogenetic tree:

```
> # Axes 1 to 3
> bat.4d <- phylo4d(phy, pcaphy$X[phy$tip.label, 1:3])
> table.phylo4d(bat.4d, center = FALSE, scale = FALSE)
```


Myotis_keaysi
Bauerus_dubiaquercus
Platyrrhinus_helleri
Vampyrodes_major
Uroderma_bilobatum
Vampyressa_pusilla
Chiroderma_villosum
Artibeus_lituratus
Artibeus_jamaicensis
Dermanura_watsoni
Dermanura_phaeotis
Enchisthenes_hartii
Centurio_senex
Sturnira_lilium
Carollia_perspicillata
Carollia_brevicauda
Lichonycteris_obscura
Hylonycteris_underwoodi
Glossophaga_soricina
Glossophaga_commissari
Phyllostomus_discolor
Phyllostomus_stenops
Mimon_cozumelae
Tonatia_evotis
Tonatia_brasiliense
Chrotopterus_auritus
Trachops_cirrhosus
Micronycteris_megalotis
Micronycteris_brachyotis
Desmodus_rotundus
Diphylla_ecaudata
Pteronotus_parnelii
Mormoops_megalophylla
Thyroptera_tricolor
n.30
n.29
n.28
n.27
n.22
n.24
n.25
n.20
n.18
n.19
n.14
n.15
n.16
n.1
n.6
n.13
n.12
n.10
n.11
n.9
n.4
n.7
n.8
n.2
n.5
n.3
CPC1
CPC2
CPC3
-0.5
0
0.5
1

```
# Appendix D - R scripts

dspca <- function(com, S=NULL, tol=1e-8){

    df <- t(com)
    if(is.null(S)){
        S <- matrix(diag(rep(1, nrow(df)))) # By default: minimum similarity between any two species
        colnames(S) <- rownames(S) <- rownames(df)
    }
    if(!inherits(df, "data.frame"))
        df <- as.data.frame(df)
    dfp <- t(t(df)/colSums(df))
    step1 <- S
    svd.step1 <- svd(step1)
    u <- svd.step1$u
    d <- svd.step1$d
    r1 <- sum(d > (d[1] * tol))
    dp <- d[1:r1]
    up <- u[, 1:r1]
    X <- up%*%diag(sqrt(dp))
    Y <- t(dfp)%*%X
    rownames(X) <- rownames(df)
    colnames(X) <- paste("SPC", 1:ncol(X), sep="")
    colnames(Y) <- paste("SPC", 1:ncol(Y), sep="")
    Scom <- Y%*%t(Y)
    Q <- diag(1/sqrt(diag(Scom)))
    Y1 <- Q%*%Y
    rownames(Y1) <- rownames(Y)
    colnames(Y1) <- paste("SPC", 1:ncol(Y1), sep="")
    Y1 <- Q%*%Y
    Scom <- Q%*%Scom%*%Q
    rownames(Scom) <- colnames(Scom) <- colnames(df)
    step2 <- t(Y1)%*%Y1
    svd.step2 <- svd(step2)
    d <- svd.step2$d
    r2 <- sum(d > (d[1] * tol))
    dp <- d[1:r2]
    u <- svd.step2$u
    up <- u[, 1:r2]
    step2.Y <- Y1%*%up
    rownames(step2.Y) <- colnames(df)
    step2.X <- X%*%up
    rownames(step2.X) <- rownames(df)
    colnames(step2.X) <- paste("CPC", 1:ncol(step2.X), sep="")
    colnames(step2.Y) <- paste("CPC", 1:ncol(step2.Y), sep="")
    res <- list()
    res$eig <- dp
    res$X <- step2.X
    res$Y <- step2.Y
    res$Scom <- Scom
    return(res)

}

dfunsimspe <- function(df, vartype=c("Q","N","M","P"), method=1:5, type=c("dissimilarity", "similarity")){

    meantype <- method[1]
```

```
if(!meantype%in%(1:5)) stop("Incorrect definition of method")
fun0 <- function(i){
df0 <- as.matrix(df[[i]])
type <- type[1]
vartype0 <- vartype[i]
if(vartype0=="Q" | vartype0=="N"){
   if(type=="dissimilarity")
      return(daisy(df0, metric = "gower")*ncol(df0))
   else
      return((1-as.matrix(daisy(df0, metric = "gower")))*ncol(df0))
}
if(vartype0=="P"){
   df0 <- sweep(df0, 1, rowSums(df0), "/")
}
if(vartype0=="P" | vartype0=="M"){
   A <- df0%*%t(df0)
   B <- diag(A)%*%t(rep(1, nrow(df0)))
   C <- rep(1, nrow(df0))%*%t(diag(A))
   if(meantype==4) S <- A/sqrt(B)/sqrt(C)
   else if(meantype==3){
      S <- 2*A/(B+C)
   }
   else if(meantype==1){
      S <- A/(2*B+2*C-3*A)
   }
   else if(meantype==2){
      S <- A/(B+C-A)
   }
   else S <- 4*A/(2*A+B+C)

   rownames(S)<-colnames(S)<-rownames(df0)
   if(type=="dissimilarity")
      return(as.dist(1-S))
   else
      return(S)
}
}
if(inherits(df, "ktab")){
   listdsim <- lapply(1:length(df$blo), fun0)
   res <- listdsim[[1]]
   if(length(listdsim)>1){
      for(i in 2:length(listdsim))
         res <- res + listdsim[[i]]
   }
   nk <- length(vartype[vartype!="Q" & vartype!="N"])
   nk <- nk + sum(df$blo[vartype=="Q" | vartype=="N"])
   return(res/nk)
}
else{
df <- as.matrix(df)
type <- type[1]
vartype <- vartype[1]
if(vartype=="Q" | vartype=="N"){
   if(type=="dissimilarity")
      return(daisy(df, metric = "gower"))
   else
      return(1-as.matrix(daisy(df, metric = "gower")))
```

```
    }
    if(vartype=="P"){
        df <- sweep(df, 1, rowSums(df), "/")
    }
    if(vartype=="P" | vartype=="M"){
        A <- df%*%t(df)
        B <- diag(A)%*%t(rep(1, nrow(df)))
        C <- rep(1, nrow(df))%*%t(diag(A))
        if(meantype==4) S <- A/sqrt(B)/sqrt(C)
        else if(meantype==3){
            S <- 2*A/(B+C)
        }
        else if(meantype==1){
            S <- A/(2*B+2*C-3*A)
        }
        else if(meantype==2){
            S <- A/(B+C-A)
        }
        else S <- 4*A/(2*A+B+C)

        rownames(S)<-colnames(S)<-rownames(df)
        if(type=="dissimilarity")
            return(as.dist(1-S))
        else
            return(S)
    }
    }
}
```

Appendix E. An ordination approach to explore similarities among communities

S. Pavoine

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

Effects of abundance and species-species similarities in DSPCA − A theoretical example

Here I consider a theoretical example to illustrate how one can evaluate the effects of abundance and similarity data on community-to-community similarities thanks to DSPCA.

R scripts used below are given in Appendix D; a manual is available in Appendix C. The scripts below also require that package adiv be loaded:

```
install.packages("adiv")
library(adiv)
```

I first define a matrix with the abundance of 10 species in five communities:

```
com <- matrix(c(10, 1, 0, 0, 0, 5, 2, 0, 0, 0, 2, 5, 0, 0, 0, 0, 0, 1, 0,
0, 1, 10, 0, 10, 1, 0, 0,10, 0, 0, 0, 0, 0, 5, 2, 0, 0, 1, 0, 0, 0, 0, 0,
2, 5, 0, 0, 0, 1, 10), 5, 10)
rownames(com) <- paste("c", 1:5, sep="")
colnames(com) <- paste("s", 1:10, sep="")
com
   s1 s2 s3 s4 s5 s6 s7 s8 s9 s10
c1 10  5  2  0  1  0  0  0  0   0
c2  1  2  5  0 10  0  0  0  0   0
c3  0  0  0  1  0 10  0  1  0   0
c4  0  0  0  0 10  0  5  0  2   1
c5  0  0  0  0  1  0  2  0  5  10
```

Then I define trait values for the 10 species:

```
trait <- c(-4,-2,-1,-0.8,0,0.2,1,1.2,2,4)
names(trait) <- colnames(com)
trait
  s1   s2   s3   s4   s5   s6   s7   s8   s9  s10 
-4.0 -2.0 -1.0 -0.8  0.0  0.2  1.0  1.2  2.0  4.0
```

The species traits are distributed on a segment from -4 to 4 with species s1 having the minimum value and species s10 the maximum value. s5 is in the middle of the segment. Species s3 and s4 have close trait values; same for s5 and s6 and s7 and s8.

I calculate similarities between species applying to the trait data Gower (1971) distance scaled between 0 and 1, as follows:

```
Strait <- dsimFun(trait, "Q", type="similarity")
```

The resulting matrix of species-species similarities has the following values:

```
Strait
       s1    s2    s3    s4    s5    s6    s7    s8    s9   s10
s1  1.000 0.750 0.625 0.600 0.500 0.475 0.375 0.350 0.250 0.000
s2  0.750 1.000 0.875 0.850 0.750 0.725 0.625 0.600 0.500 0.250
s3  0.625 0.875 1.000 0.975 0.875 0.850 0.750 0.725 0.625 0.375
s4  0.600 0.850 0.975 1.000 0.900 0.875 0.775 0.750 0.650 0.400
s5  0.500 0.750 0.875 0.900 1.000 0.975 0.875 0.850 0.750 0.500
s6  0.475 0.725 0.850 0.875 0.975 1.000 0.900 0.875 0.775 0.525
s7  0.375 0.625 0.750 0.775 0.875 0.900 1.000 0.975 0.875 0.625
s8  0.350 0.600 0.725 0.750 0.850 0.875 0.975 1.000 0.900 0.650
s9  0.250 0.500 0.625 0.650 0.750 0.775 0.875 0.900 1.000 0.750
s10 0.000 0.250 0.375 0.400 0.500 0.525 0.625 0.650 0.750 1.000
```

Now I run DSPCA on this dataset:

```
dspca1 <- dspca(com=com, S=Strait)
```

DSPCA leads to 5 orthogonal axes with the following eigenvalues:

```
dspca1$eig
[1] 4.06926031 0.71751617 0.17951337 0.02145132 0.01225883
```

The first eigenvalue indicates high average similarities between the 5 communities.

The similarities between communities can be obtained as follows:

```
dspca1$Scom
          c1        c2        c3        c4        c5
c1 1.0000000 0.8253012 0.6865797 0.6434091 0.3351955
c2 0.8253012 1.0000000 0.9530516 0.9340369 0.6434091
c3 0.6865797 0.9530516 1.0000000 0.9759833 0.7389547
c4 0.6434091 0.9340369 0.9759833 1.0000000 0.8253012
c5 0.3351955 0.6434091 0.7389547 0.8253012 1.0000000
```

The average similarity is:

```
mean(as.dist(dspca1$Scom))
[1] 0.7561222
```

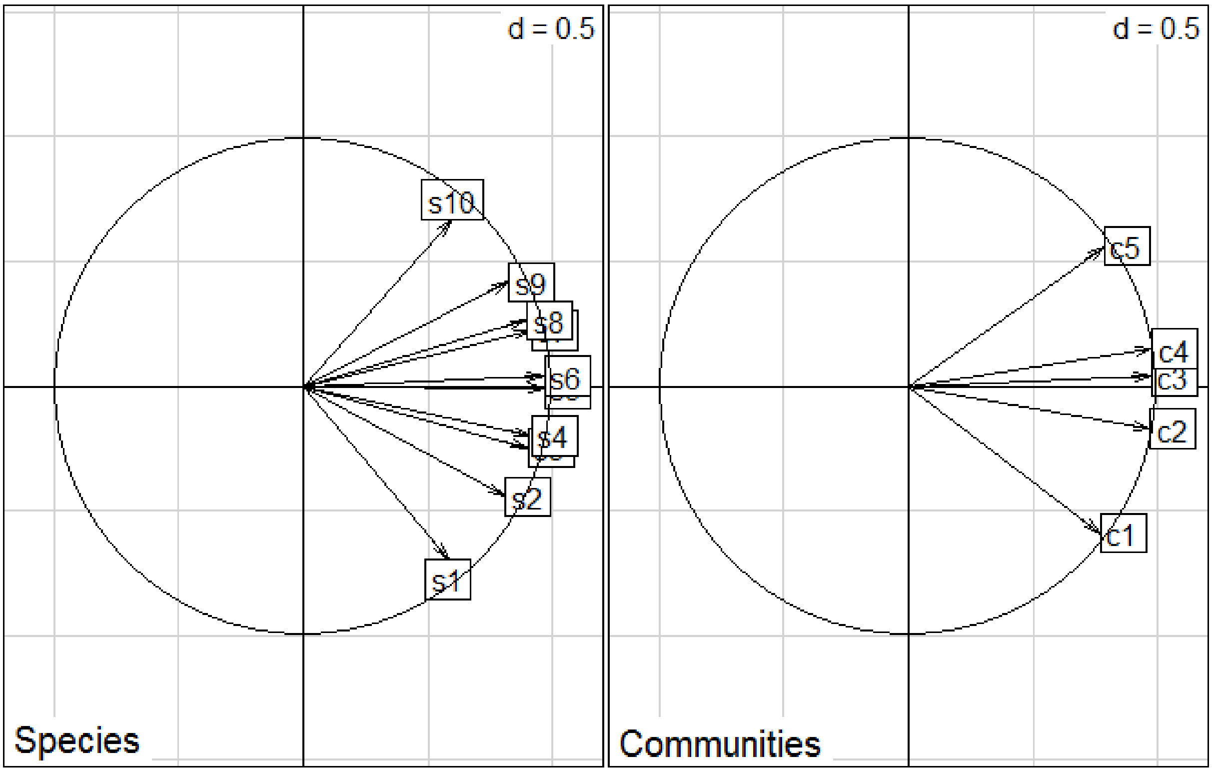
this value is close to:

```
 (dspca1$eig[1]-1)/4
[1] 0.7673151
```

$= (\lambda_1-1)/(m-1)$, where λ_1 is the first eigenvalue and m the number of communities.

The first two axes of DSPCA show that c3 is the community with the highest similarities with other species and that c1 and c5 are the most different. Indeed, although c3 do not share species with the other communities, its dominant species has close trait values with at least one of the species of the other communities. Although c1 an c2 have exactly the same species and c4 and c5 also have exactly the same species, the most abundant species of communities c2, c3, and c4 have trait values close or equal to zero. In contrast, the most abundant species of c1 has a trait value of -4 and that of c5 a trait value of 4.

```
par.mar <- par()$mar
par(mar=rep(0.1,4))
par(mfrow=c(1,2))
# Species
ade4::s.arrow(dspca1$X, ylim=c(-1.2,1.2), xlim=c(-1.2,1.2),sub="Species")
symbols(0,0,1, inch=F, add=TRUE)
# Communities
ade4::s.arrow(dspca1$Y, ylim=c(-1.2,1.2), xlim=c(-
1.2,1.2),sub="Communities")
symbols(0,0,1, inch=F, add=TRUE)
par(mar=par.mar)
```



To evaluate the effect of species-to-species similarities on community-to-community similarities, I run again DSPCA considering that the species are maximally dissimilar (i.e. ignoring trait data). The obtained results are quite different.

The new dissimilarities between species are defined as follows:

```
Stax <- diag(rep(1,10))
rownames(Stax) <- colnames(Stax) <- colnames(com)
Stax
    s1 s2 s3 s4 s5 s6 s7 s8 s9 s10
s1   1  0  0  0  0  0  0  0  0   0
s2   0  1  0  0  0  0  0  0  0   0
s3   0  0  1  0  0  0  0  0  0   0
s4   0  0  0  1  0  0  0  0  0   0
s5   0  0  0  0  1  0  0  0  0   0
s6   0  0  0  0  0  1  0  0  0   0
s7   0  0  0  0  0  0  1  0  0   0
s8   0  0  0  0  0  0  0  1  0   0
```

```
s9   0  0  0  0  0  0  0  0  1   0
s10  0  0  0  0  0  0  0  0   1
```

I now apply DSPCA to the community matrix and these new species-species similarities:

```
dspca2 <- dspca(com=com, Stax)
```

DSPCA leads to 5 orthogonal axes with the following eigenvalues:

```
dspca2$eig
[1] 1.9296770 1.0567796 1.0000000 0.8472461 0.1662974
```

The first eigenvalue of this new application of DSPCA indicates much more moderate similarities between the 5 communities than the previous application of DSPCA where species trait values were considered.

The similarities between communities can be obtained as follows:

```
dspca2$Scom
            c1         c2 c3         c4          c5
c1 1.000000000 0.30769231  0 0.07692308 0.007692308
c2 0.307692308 1.00000000  0 0.76923077 0.076923077
c3 0.000000000 0.00000000  1 0.00000000 0.000000000
c4 0.076923077 0.76923077  0 1.00000000 0.307692308
c5 0.007692308 0.07692308  0 0.30769231 1.000000000
```

The average similarity is:

```
mean(as.dist(dspca2$Scom))
[1] 0.1546154
```

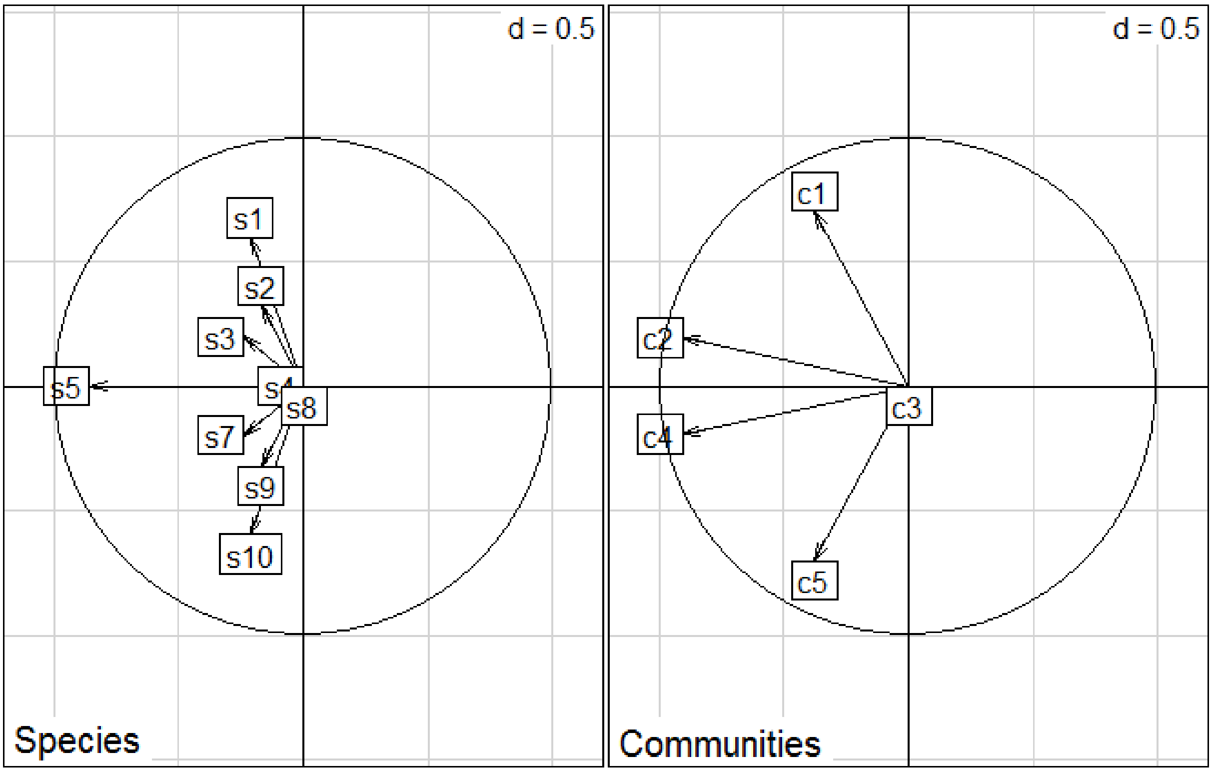
this value is lower than:

```
(dspca2$eig[1]-1)/4
[1] 0.2324193
```

= (λ_1-1)/(*m*-1), where λ_1 is the first eigenvalue and *m* the number of communities. This is consistent with Friedman and Weisberg (1981) statement that the estimate 1 + (n-1) $\bar{s}$ "deteriorates slightly" as the variance of the similarities increases. [Friedman and Weisberg (1981) actually analyzed correlation matrixes with positive values but their statement remain valid for similarity matrixes].

The first and second axes of DSPCA highlight similarity patterns between c1, c2, c4 and c5. Indeed these four communities share species s5. However s5 has the highest abundance in c2 and c4, whereas it has the lowest abundance in c1 and c5:

```
# Axes 1 and 2
par.mar <- par()$mar
par(mar=rep(0.1,4))
par(mfrow=c(1,2))
# Species
ade4::s.arrow(dspca2$X, ylim=c(-1.2,1.2), xlim=c(-1.2,1.2),sub="Species")
symbols(0,0,1, inch=F, add=TRUE)
# Communities
ade4::s.arrow(dspca2$Y, ylim=c(-1.2,1.2), xlim=c(-
1.2,1.2),sub="Communities")
symbols(0,0,1, inch=F, add=TRUE)
par(mar=par.mar)
```



These graphs also show that species s5 is the most representative species of communities c2 and c4, while s1 is the most characteristic species of c1 and s10 the most characteristic species of c5.

The coordinates of the communities on the five axes are as follows:

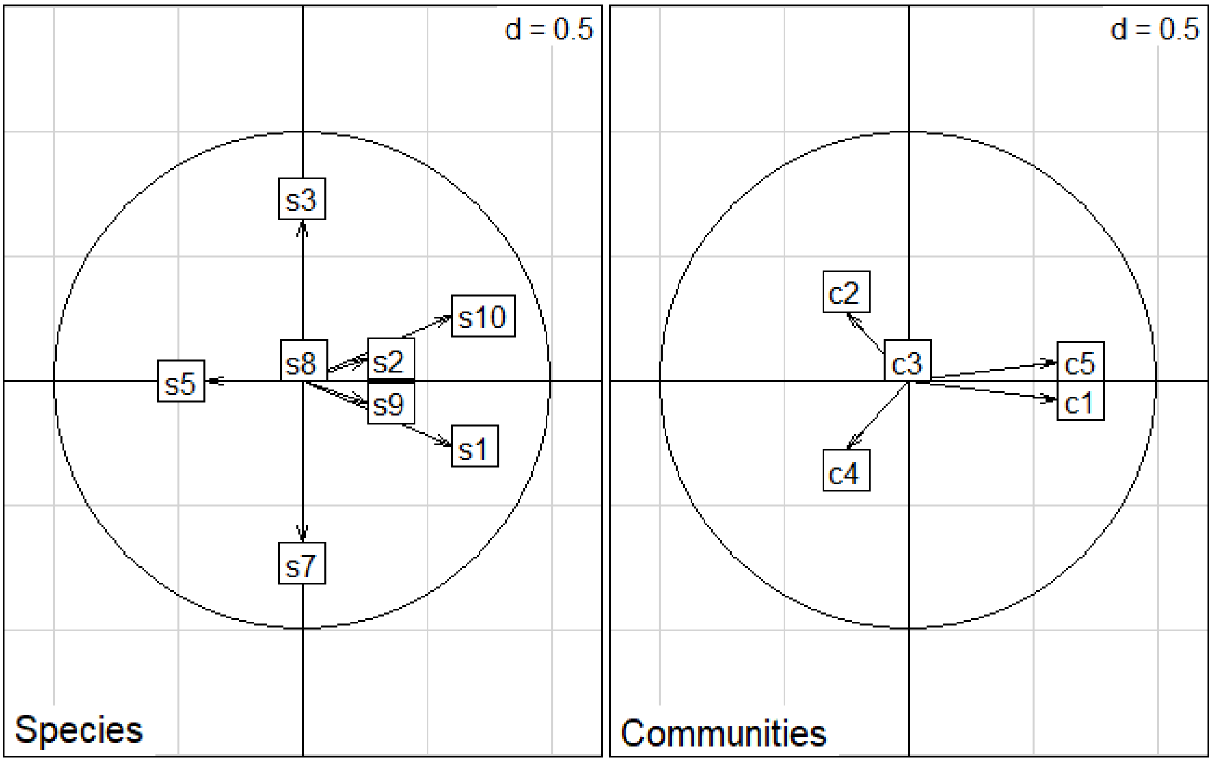
```
dspca2$Y
            CPC1          CPC2          CPC3          CPC4          CPC5
c1 -3.781743e-01  7.000956e-01  1.136607e-17  6.006916e-01 -7.758901e-02
c2 -9.065444e-01  1.955914e-01 -4.110979e-17 -2.505847e-01  2.777204e-01
c3 -1.126169e-20 -6.254382e-19  1.000000e+00  1.664908e-18  1.597530e-16
c4 -9.065444e-01 -1.955914e-01  4.070810e-17 -2.505847e-01 -2.777204e-01
c5 -3.781743e-01 -7.000956e-01 -9.948896e-18  6.006916e-01  7.758901e-02
```

This shows that community c3 has a coordinate equal to zero on all axes except axis 3. Axis 3 indicates the complete dissimilarity between c3 and the other communities, because c3 does not share species with the other communities and because information on species traits was ignored.

Then, axis 4 indicates the differences between c1-c5 and c2-c4, and axis 5 the differences between c2 and c4, which are the lowest differences between any two of the communities.

```
# Axes 4 and 5

par.mar <- par()$mar
par(mar=rep(0.1,4))
par(mfrow=c(1,2))
# Species
ade4::s.arrow(dspca2$X, ylim=c(-1.2,1.2), xlim=c(-1.2,1.2),sub="Species",
xax=4, yax=5)
symbols(0,0,1, inch=F, add=TRUE)
# Communities
```

```
ade4::s.arrow(dspca2$Y, ylim=c(-1.2,1.2), xlim=c(-
1.2,1.2),sub="Communities", xax=4, yax=5)
symbols(0,0,1, inch=F, add=TRUE)
par(mar=par.mar)
```



Now I evaluate the effect of species abundance on community-to-community similarities. By transforming abundance data into 0/1 data (0 for the absence, 1 for the presence of a species in a community).

The new matrix of species presence/absence in communities is obtained as follows:

```
comPA <- com
comPA[comPA>0] <- 1
comPA
   s1  s2  s3  s4  s5  s6  s7  s8  s9  s10
c1  1   1   1   0   1   0   0   0   0    0
c2  1   1   1   0   1   0   0   0   0    0
c3  0   0   0   1   0   1   0   1   0    0
c4  0   0   0   0   1   0   1   0   1    1
c5  0   0   0   0   1   0   1   0   1    1
```

I apply DSPCA to this matrix and the species-to-species trait similarities:

```
dspca3 <- dspca(com=comPA, S=Strait)
```

DSPCA leads to 3 orthogonal axes with the following eigenvalues:

```
dspca3$eig
[1] 4.37017950 0.59036930 0.03945121
```

The first eigenvalue of this new application of DSPCA indicates high similarities between the 5 communities.

The similarities between communities can be obtained as follows:

```
dspca3$Scom
          c1        c2        c3        c4        c5
c1 1.0000000 1.0000000 0.8787559 0.7058824 0.7058824
c2 1.0000000 1.0000000 0.8787559 0.7058824 0.7058824
c3 0.8787559 0.8787559 1.0000000 0.9183618 0.9183618
c4 0.7058824 0.7058824 0.9183618 1.0000000 1.0000000
c5 0.7058824 0.7058824 0.9183618 1.0000000 1.0000000
```

The average similarity is:

```
mean(as.dist(dspca3$Scom))
[1] 0.8417765
```

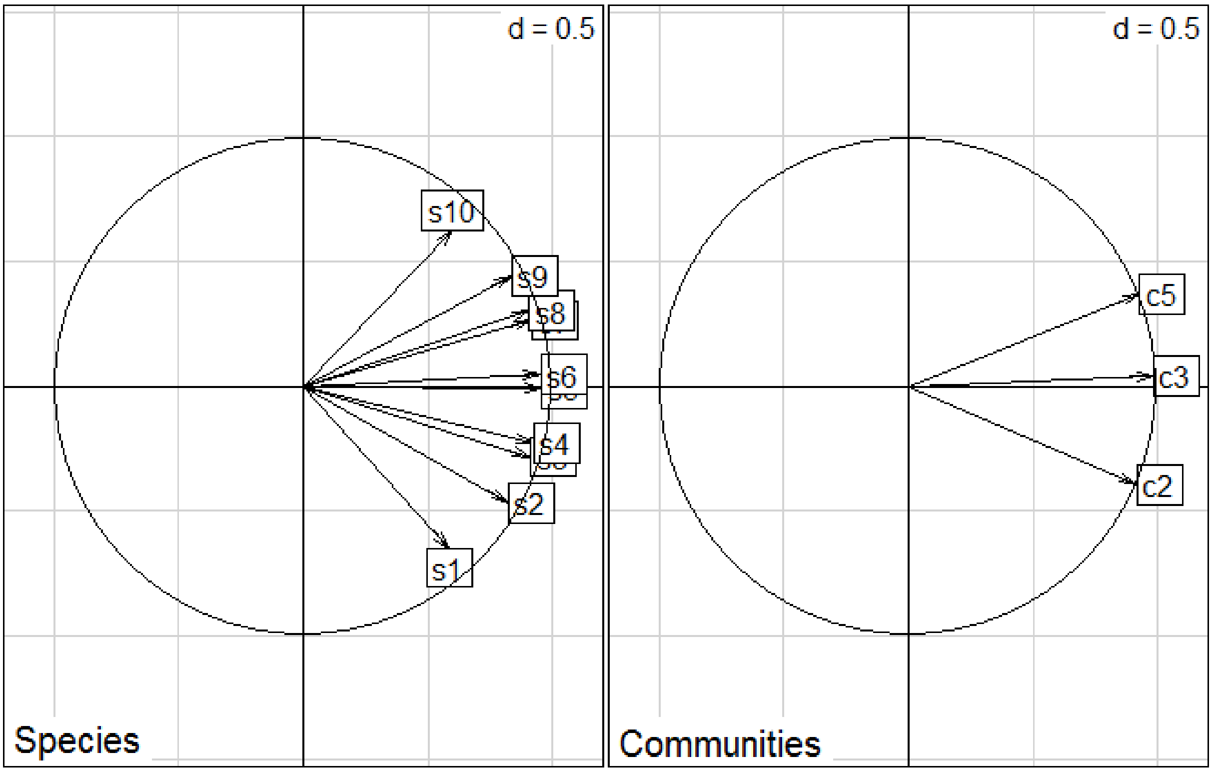
this value is close to:

```
(dspca3$eig[1]-1)/4
[1] 0.8425449
```

= (λ_1-1)/(m-1), where λ_1 is the first eigenvalue and m the number of communities.

With presence/absence data, communities c1 and c2 become similar to each other; c4 and c5 are also similar to each other. The overall similarities between the five communities (evaluated by the first eigenvalue) increases compared to the DSPCA applied to abundance data. Indeed, considering presence/absence data increases the similarities between c1 and c5 and the other communities. This pattern of similarity is shown on the first two axes of DSPCA where the points of communities c1 and c2 are superimposed and the points of communities c4 and c5 are also superimposed:

```
dspca3$Y
        CPC1        CPC2        CPC3
c1 0.9171264 -0.39655411 -0.04029864
c2 0.9171264 -0.39655411 -0.04029864
c3 0.9837132  0.04131551  0.17493265
c4 0.9274282  0.37023776 -0.05292358
c5 0.9274282  0.37023776 -0.05292358

# Axes 1 and 2
par.mar <- par()$mar
par(mar=rep(0.1,4))
par(mfrow=c(1,2))
# Species
ade4::s.arrow(dspca3$X, ylim=c(-1.2,1.2), xlim=c(-1.2,1.2),sub="Species")
symbols(0,0,1, inch=F, add=TRUE)
# Communities
ade4::s.arrow(dspca3$Y, ylim=c(-1.2,1.2), xlim=c(-
1.2,1.2),sub="Communities")
symbols(0,0,1, inch=F, add=TRUE)
par(mar=par.mar)
```


d = 0.5
s10
s9
s8
s6
s4
s2
s1
Species
d = 0.5
c5
c3
c2
Communities