



A new method for indicator species analysis in the framework of multivariate analysis of variance

Carlo Ricotta¹ | Sandrine Pavoine² | Bruno E. L. Cerabolini³ | Valério D. Pillar⁴

¹Department of Environmental Biology, University of Rome, 'La Sapienza', Rome, Italy

²Centre d'Ecologie et des Sciences de la Conservation (CESCO), Muséum National d'Histoire Naturelle, CNRS, Sorbonne Université, Paris, France

³Department of Biotechnology and Life Sciences, University of Insubria, Varese, Italy

⁴Departamento de Ecologia, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil

Correspondence

Carlo Ricotta, Department of Environmental Biology, University of Rome, 'La Sapienza', Piazzale Aldo Moro 5, 00185 Rome, Italy. Email: carlo.ricotta@uniroma1.it

Funding information

This research received no external funding.

Co-ordinating Editor: Zoltán Botta-Dukát

Abstract

Question: In vegetation science, the compositional dissimilarity among two or more groups of plots is usually tested with dissimilarity-based multivariate analysis of variance (db-MANOVA), whereas the compositional characterization of the different groups is performed by means of indicator species analysis. Although db-MANOVA and indicator species analysis are apparently very far from each other, the question we address here is: can we put both approaches under the same methodological umbrella?

Methods: We will show that for a specific class of dissimilarity measures, the partitioning of variation used in one-factor db-MANOVA can be additively decomposed into species-level values allowing us to identify the species that contribute most to the compositional differences among the groups.

Results: The proposed method, for which we provide a simple R function, is illustrated with one small data set on alpine vegetation sampled along a successional gradient.

Conclusion: The species that contribute most to the compositional differences among the groups are preferentially concentrated in particular groups of plots. Therefore, they can be appropriately called indicator species. This connects multivariate analysis of variance with indicator species analysis.

KEYWORDS

dissimilarity-based analysis of variance, Euclidean distance, randomization test, standardized effect size, within-group sum of squared dissimilarities

1 | INTRODUCTION

For a set of plots that are grouped according to some external criteria, such as selected environmental variables or different experimental treatments, two relevant questions usually asked by community ecologists are: "Are these groups compositionally different from each other?" and "Which species contribute most to the compositional differences among the groups?" These questions are generally

answered with different tools: the compositional dissimilarity among groups of plots is usually tested with dissimilarity-based multivariate analysis of variance (Pillar & Orlóci, 1996; Anderson, 2001), whereas the compositional characterization of the different groups is performed by means of indicator species analysis (Dufrêne & Legendre, 1997; Chytrý et al., 2002).

Dissimilarity-based multivariate analysis of variance (db-MANOVA) is applicable to any type of compositional data,

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *Journal of Vegetation Science* published by John Wiley & Sons Ltd on behalf of International Association for Vegetation Science

irrespective of the number of species sampled and the way they are sampled (i.e., presence/absence scores, number of individuals, species cover, etc.), provided that a meaningful measure is used to adequately represent the dissimilarity between pairs of plots (or relevés, communities, assemblages, quadrats, sites, etc.).

Given a community composition matrix containing the presence/absence or the abundance values x_{sj} of S species ($s = 1, 2, \dots, S$) in N plots ($j = 1, 2, \dots, N$), let the number of plots in group k ($k = 1, 2, \dots, K$) be N_k such that $\sum_{k=1}^K N_k = N$ and d_{ij} be the compositional dissimilarity between plot i and plot j . In the simplest situation of a single-factor analysis on a set of plots partitioned into K groups, the very essence of db-MANOVA is to compare the plot-to-plot dissimilarity within groups with the plot-to-plot dissimilarity among groups with an adequate test statistic (Figure 1). The larger the value of between-group dissimilarity compared to within-group dissimilarity, the more likely it is that the groups are compositionally different from each other (Anderson, 2001). For db-MANOVA, hypothesis testing is usually performed with randomization tests where the N plots are randomly permuted among the K groups, thus creating a reference distribution under the null hypothesis of no compositional difference between the different groups of plots.

Once the compositional differences among the groups of plots have been verified, the identification of diagnostic species for the different groups is a relevant step for their ecological characterization (Dufrêne & Legendre, 1997). In vegetation science, the concept of indicator or diagnostic species which we use here as synonyms has usually been associated with the species concentration in particular groups of plots. According to Chytrý et al. (2002), diagnostic

species “include species which preferably occur in a single vegetation unit (character species) or in a few vegetation units (differential species)”.

Several methods have been developed to identify indicator species, (Dufrêne & Legendre, 1997; Chytrý et al., 2002; Podani & Csányi, 2010; Ricotta et al., 2015). Indicator species are fundamentally defined as those species that are more common in a given group of plots than expected by chance alone. Therefore, to determine if a given species is significantly associated with a target group of plots, most of these methods compare the actual abundance of that species in the target group of plots with a reference distribution obtained from a random null model in which the species occurrences or abundances are permuted among the N plots. The null hypothesis is that all plots have equal probability to host each species, irrespective of the species' environmental preferences (De Cáceres & Legendre, 2009).

Although dissimilarity-based MANOVA and indicator species analysis are apparently unrelated, in this paper we will show that for a particular group of dissimilarity measures the classical partitioning of variation used for hypothesis testing in one-factor db-MANOVA can be additively decomposed into species-level values, thus connecting indicator species analysis with multivariate analysis of variance.

2 | METHODS

Several test statistics have been proposed for dissimilarity-based MANOVA. For review, see Pillar (2013) and references therein. For dissimilarity matrices with Euclidean properties (see Gower & Legendre, 1986), Pillar and Orlóci (1996) proposed as test criterion the sum of squared dissimilarities between groups. Given a symmetric $N \times N$ dissimilarity matrix whose elements d_{ij} represent the pair-wise dissimilarities between plot i and j such that $d_{ij} = d_{ji}$ and $d_{ii} = 0$, the between-group sum of squared dissimilarities Q_B is:

$$Q_B = Q_T - Q_W \quad (1)$$

where

$$Q_T = \frac{1}{N} \sum_{i>j}^N d_{ij}^2 \quad (2)$$

is the total sum of squared dissimilarities. Hence, to calculate Q_T we have to add the $N(N-1)/2$ plot-to-plot squared dissimilarities d_{ij}^2 in the sub-diagonal half of the dissimilarity matrix and then divide by the total number of plots N . The term Q_W is the sum of squares within the K groups, which is obtained as:

$$Q_W = \sum_{k=1}^K Q_{Wk} \quad (3)$$

where the sum of squares within group k is:

$$Q_{Wk} = \frac{1}{N_k} \sum_{i>j}^N d_{ij}^2 \epsilon_{ijk} \quad (4)$$

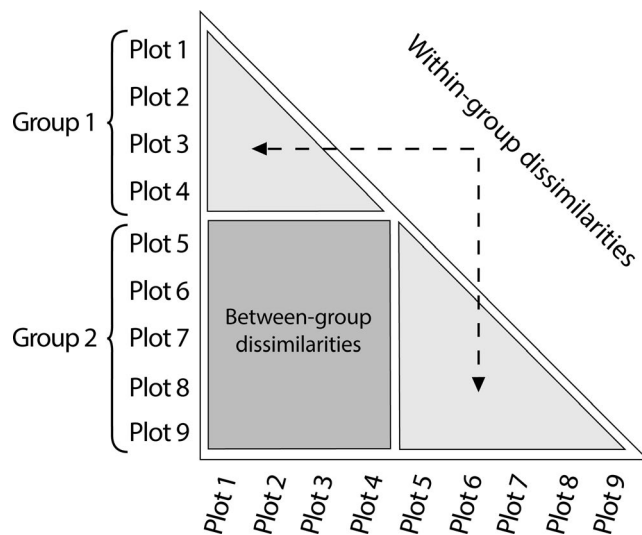


FIGURE 1 Schematic example of a semimatrix of pairwise dissimilarities among nine plots clustered into two groups. The very essence of dissimilarity-based MANOVA is to compare the within-group dissimilarity with the between-group dissimilarity with an adequate test statistic. The larger the value of between-group dissimilarity compared to within-group dissimilarity, the more likely it is that the groups are compositionally different from each other



The indicator variable ϵ_{ijk} takes the value 1 if plots i and j are both in group k . Otherwise it is 0. That is, Q_{Wk} is obtained by adding all squared dissimilarities d_{ij}^2 between all plots that occur in group k and then dividing the sum by the number of plots in that group N_k .

The sum of squares could be equally computed from deviations of single plots from group centroids, as described, among others, by Edgington (1987) or Manly (1991), but using plot-to-plot dissimilarities does not require the computation of group centroids. McArdle and Anderson (2001) further showed that this partition of the total sum of squares is also valid for dissimilarity coefficients that do not have Euclidean properties, thus making dissimilarity-based MANOVA a very flexible component of the ecologist's toolbox.

Since the distribution of Q_B under the null hypothesis of no compositional difference among the different groups of plots is generally unknown, a reference distribution of values of the test statistic is then obtained by Mantel randomization of the dissimilarity matrix. For dissimilarity-based one-way MANOVA, this is equivalent to randomly permuting the N plots among the K groups. Note that in one-factor MANOVA, since the total sum of squares Q_T is invariant over permutations, some test statistics, such as Q_B , Q_W and the Anderson (2001) pseudo- F ratio Q_B/Q_W are equivalent in randomization testing. That is, for the same data and permutations they will lead to identical p -values (Pillar, 2013).

A desirable property for a dissimilarity index is its ability to be decomposed into species-level values (Ricotta & Podani, 2017). In this way, it is possible to see which species contribute most to plot-to-plot dissimilarity. In this framework, the advantage of adopting Q_W as a test statistic over Q_B or Q_B/Q_W is that it can be decomposed into intuitively simple species-level values, provided that the squared dissimilarities d_{ij}^2 are also decomposable in the same way.

Take for example the Euclidean distance $\sqrt{\sum_{s=1}^S (x_{si} - x_{sj})^2}$. It is easily shown that the squared Euclidean distance $\sum_{s=1}^S (x_{si} - x_{sj})^2$ is additively decomposable into the species-level contributions $d_{ijs}^2 = (x_{si} - x_{sj})^2$ such that $d_{ij}^2 = \sum_{s=1}^S d_{ijs}^2$. Therefore, in a sum-of-squares-based test, if the plot-to-plot compositional dissimilarity is calculated with the Euclidean distance, the within-group sum of squares becomes:

$$Q_W = \sum_{s=1}^S Q_{Ws} = \sum_{s=1}^S \sum_{k=1}^K \frac{1}{N_k} \sum_{i>j}^N (x_{si} - x_{sj})^2 \epsilon_{ijk} \quad (5)$$

where the contribution of single species to overall Q_W is:

$$Q_{Ws} = \sum_{k=1}^K \frac{1}{N_k} \sum_{i>j}^N (x_{si} - x_{sj})^2 \epsilon_{ijk} \quad (6)$$

Geometrically, a smaller value of Q_W denotes higher within-group concentration and larger between-group differences (Cai, 2006). To test for significant compositional differences among the groups, the actual value of Q_W can be compared with a reference distribution of null values obtained by Mantel randomization of the dissimilarity matrix. p -Values can be computed as the proportion of null values

that are as low as or lower than the actual value of the test statistic. The quantities Q_{Ws} can then be used as in classical post-hoc analysis: if the dissimilarity-based MANOVA shows significant overall compositional difference among groups, the analysis can proceed by exploring differences in single-species occurrences (or abundances) among the groups. This can be done by using the same permutation procedures used for testing the significance of Q_W .

The contribution of a particular species to the compositional differences among the K groups can be summarized as standardized effect size (SES; Gotelli & McCabe, 2002):

$$SES = \frac{Obs(I) - Mean(I)}{SD(I)} \quad (7)$$

where $Obs(I)$ is the observed value of a given variable I , $Mean(I)$ is the mean of the null distribution of the variable in random assemblages, and $SD(I)$ is the standard deviation of the null distribution. For the pros and cons of SES, see Botta-Dukát (2018). Put simply, SES represents the departure of the observed variable from the mean of the null distribution, expressed in standard deviation units. However, since the observed Q_{Ws} values are generally lower than the corresponding null values, we typically have: $Mean(Q_{Ws}) > Obs(Q_{Ws})$. Therefore, in this paper, the standardized effect size of species s is formulated in such a way that the resulting values are mainly positive:

$$SES_s = \frac{Mean(Q_{Ws}) - Obs(Q_{Ws})}{SD(Q_{Ws})} \quad (8)$$

Note that in one-factor MANOVA, Equation 8 (i.e., the standardized effect size of Q_{Ws} changed in sign) is identical to the standardized effect size of the species-level contribution to the between-group sum of squared dissimilarities $Q_{Bs} = Q_{Ts} - Q_{Ws}$, with $Q_{Ts} = \frac{1}{N} \sum_{i>j}^N d_{ijs}^2$, thus reinforcing the generality of the proposed approach. For details, see Appendix S1.

3 | WORKED EXAMPLE

3.1 | Data

We used data on plant communities along a primary succession on the foreland of the Rutor Glacier (northern Italy). The data were sampled by Caccianiga et al. (2006) and have been used in many previous papers for exploring community assembly rules along ecological gradients (e.g., Caccianiga et al., 2006; Ricotta et al., 2016; Ricotta et al., 2018). The data set can be found in Appendices S2 and S5 and contains the abundances of 45 alpine plant species collected in 59 vegetation plots of approximately 25 m². All species abundances were measured with a five-point ordinal scale transformed to ranks. The plots were divided into three successional stages based on the age of the moraine deposits: 17 early-successional plots, 32 mid-successional plots and 10 late-successional plots. For additional details, see Caccianiga et al. (2006).



Based on the species abundances in each plot, we calculated the Euclidean distance between all pairs of plots. Next, using Q_W and Q_{WS} as test statistics, we tested for overall and species-wise compositional differences among the three successional stages with dissimilarity-based MANOVA. All calculations were done with a new R function available in Appendices S3 and S6. Significance is computed by randomly permuting the 59 plots (data vectors) among the three successional stages with 9,999 replicates. For all species that showed significant compositional difference among the three groups of plots, we also tested for pairwise differences among all pairs of groups (9,999 permutations) using the same approach as described in Pillar and Orlóci (1996) for multivariate contrasts.

4 | RESULTS

The overall compositional differences among the three groups of plots and between all pairs of groups proved highly significant with the minimum possible p -value for 9,999 randomizations ($p = 0.0001$) showing that the three successional stages are compositionally well distinguishable. The results of the single-species MANOVA are shown in Figure 2, together with the species abundances in each group and the contributions of single species to the compositional differences among the groups (SES_j).

From a total of 45 species, we identified 32 with a clear preference for certain successional stages ($p < 0.05$). It therefore appears that more than two thirds of the species sampled have rather narrow ecological requirements allowing them to colonize only the most ecologically suitable habitats along the primary succession. Caccianiga et al. (2006) showed that the establishment of the first pioneer species is associated with random dispersal mechanisms that drive the colonization of the glacial deposits by early-successional ruderal forbs, such as *Cerastium uniflorum*, *Oxyria digyna*, or *Tussilago farfara*, whereas the late-successional stages are preferentially colonized by stress-tolerant graminoids (sensu Grime, 1974), such as *Carex curvula*, or *Carex sempervirens*. Mid-successional plots include species with intermediate (ruderal/stress-tolerant) strategies, such as *Achillea moschata* or *Trifolium pallescens*, which are among the species that contribute most to the compositional differences among the successional stages (Figure 2).

5 | DISCUSSION

In this paper, we showed that when we test for compositional differences among two or more groups of plots with one-factor db-MANOVA, the within-group sum of squares Q_W can be additively

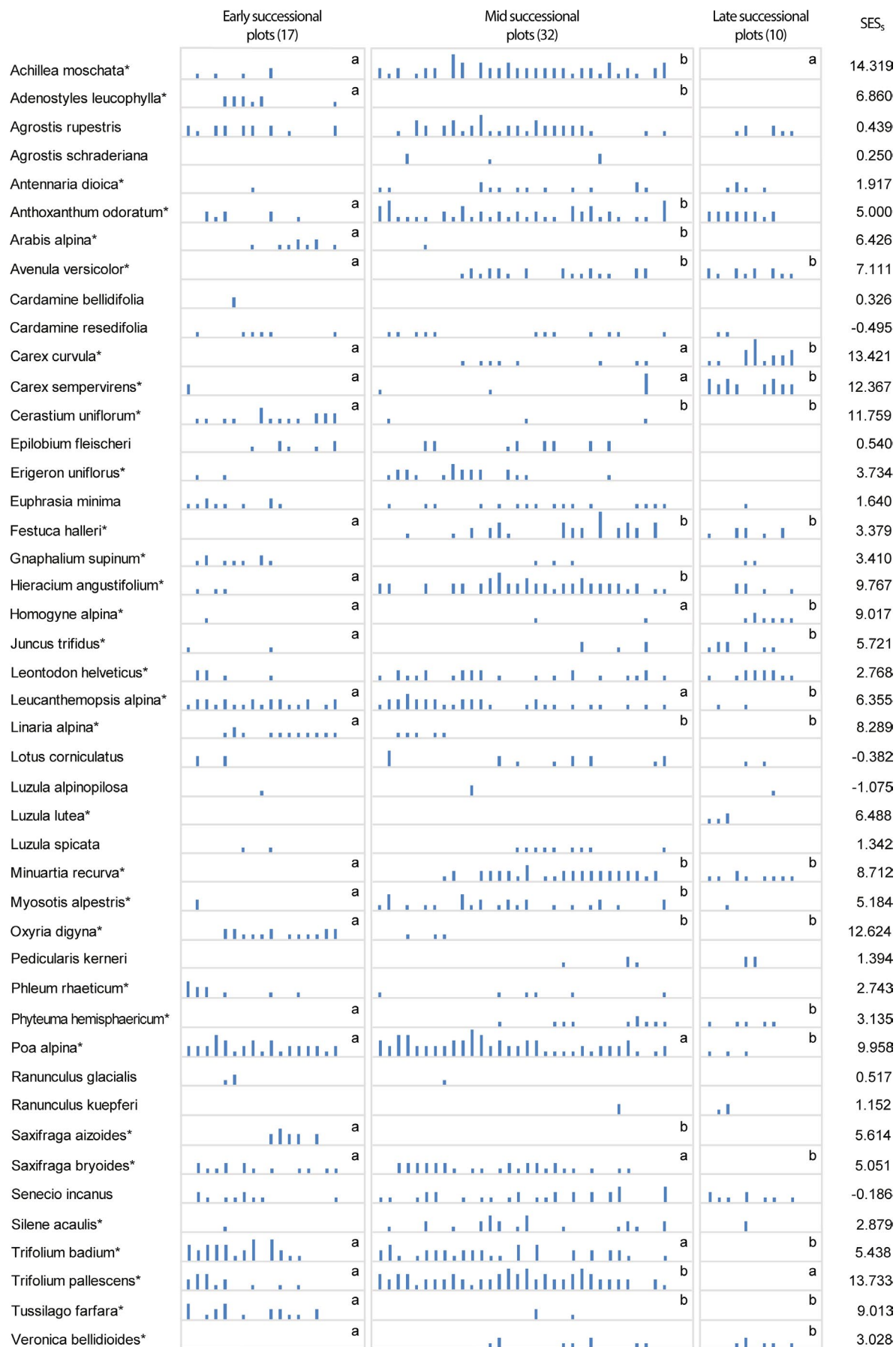
decomposed into species-level values, provided that the squared dissimilarities d_{ij}^2 can also be additively decomposed into their species-level contributions d_{ijs}^2 such that $d_{ij}^2 = \sum_{s=1}^S d_{ijs}^2$. In the remainder we will call a dissimilarity coefficient d_{ij} that conforms to this property 'squared decomposable' (S-decomposable).

The species that contribute most to the compositional differences among the groups can be appropriately called indicator species as they are preferentially concentrated in particular groups of plots. Note however that, while classical indicator species analysis usually identifies indicators of single groups of plots, with the proposed method some species may be related to more than one group. By taking into account combinations of groups of plots and adopting a multiple-contrast approach (Pillar & Orlóci, 1996), this method provides an extra flexibility to model the habitat preferences of species with varying niche breadths (De Cáceres et al., 2010). Readers interested in comparing the results of the proposed method with the results of classical indicator species analysis using the same data of our worked example are addressed to Ricotta et al. (2020).

A desirable aspect of this method for identifying indicator species is that it is based on the same dissimilarity measure used for multivariate analysis of variance (see Pavoine et al., 2013). This puts indicator species analysis and multivariate analysis of variance under the same mathematical umbrella. Regarding the most appropriate dissimilarity coefficient for db-MANOVA, being S-decomposable, the Euclidean distance appears to be a natural choice. Likewise, all Euclidean distances computed on transformed species abundances, such as the Chord or the Hellinger distance, may be equally adequate (see Legendre & Gallagher, 2001). Note that, according to Pillar (2013), a comprehensive evaluation of the power and accuracy of one-factor db-MANOVA showed that the Euclidean distance gave better results compared to other dissimilarity coefficients in spite of its well-known limitations for summarizing compositional dissimilarity among plots (see Orlóci, 1978; Legendre & Gallagher, 2001).

However, the same species-level decomposition can also be extended to virtually any dissimilarity measure of the form $\sqrt{\delta_{ij}}$ provided that δ_{ij} is additively decomposable (A-decomposable) such that $\delta_{ij} = \sum_{s=1}^S \delta_{ijs}$. A-decomposable measures include among others the Manhattan distance $\sum_{s=1}^S |x_{is} - x_{js}|$, the Canberra distance $\sum_{s=1}^S \frac{|x_{is} - x_{js}|}{(x_{is} + x_{js})}$, or the Bray-Curtis dissimilarity $\sum_{s=1}^S \frac{|x_{is} - x_{js}|}{\sum_{s=1}^S (x_{is} + x_{js})}$. Note however that since the Bray-Curtis dissimilarity summarizes the difference in species abundances between plots i and j compared to the total species abundance in both plots (see Ricotta & Podani, 2017), the contribution of each species depends on the values of other species. This is conceptually distinct from classical indicator species analysis which is usually species-specific.

FIGURE 2 Results of the distance-based MANOVA for the 45 species of the Rutor data set. The species abundances in each group of plots on a five-point ordinal scale and the contribution of each species to the compositional differences among the groups (SES_j) are also shown. The species showing significant compositional differences among the three groups of plots ($p < 0.05$; 9,999 permutations) are marked with an asterisk. For those species, we also tested for significant pairwise differences among all pairs of groups. For each species, letters indicate individually significant tests ($p < 0.05$; 9,999 permutations) after correction for multiple tests according to Benjamini and Yekutieli (2001)





If the square root of an A-decomposable dissimilarity coefficient $\sqrt{\delta_{ij}}$ is used in one-factor db-MANOVA, we obtain:

$$Q_W = \sum_{k=1}^K Q_{Wk} = \sum_{k=1}^K \frac{1}{N_k} \sum_{i>j}^N \delta_{ij} \epsilon_{ijk} = \sum_{s=1}^S \sum_{k=1}^K \frac{1}{N_k} \sum_{i>j}^N \delta_{ijs} \epsilon_{ijk} \quad (9)$$

Therefore, dealing with A-decomposable dissimilarities, Equation 9 tells us that the within-group sum of squares of any dissimilarity measure of the form $\sqrt{\delta_{ij}}$ can be additively decomposed into species-level values. At the same time, the term Q_W in Equation 9 can also be interpreted as the within-group sum of non-squared A-decomposable dissimilarities δ_{ij} . As such, Q_W is a special case of the “average dissimilarity within groups” used by Mielke and Berry (2001) in multi-response permutation procedures (MRPP), which is a class of multivariate permutation methods for testing for compositional differences among a-priori-defined groups of plots. This makes the differences between randomization tests based on squared and non-squared dissimilarities much fuzzier than one might think. A short overview of MRPP and its relationship with db-MANOVA is shown in Appendix S4.

Although randomization tests based on A-decomposable non-squared dissimilarities have been rarely used in ecology (partly because they are not included in common statistical packages), they can be equally adopted in one-factor designs for testing for compositional differences among a-priori-defined groups of plots and for identifying the indicator species that contribute most to the compositional differences among the groups. This allows to expand the number of suitable dissimilarities to the whole arsenal of A-decomposable coefficients already available in the ecological literature (Legendre & De Cáceres, 2013). Also, the partitioning of other test statistics could be evaluated, such as the between-groups sum of squares (Q_B) defined in Equation 1. In this case, the proposed approach could be further extended to multi-factor group comparisons (see Pillar & Orlóci, 1996), when it may be relevant to identify indicator species of each group considering the factors separately and their interactions.

Which of these dissimilarities will show the best performance in statistical and biological terms (i.e., accuracy and power vs biological interpretability)? Can the proposed approach be further extended to functional and phylogenetic dissimilarities, such as the large class of A-decomposable coefficients introduced by Pavoine and Ricotta (2019)? These are relevant questions, and their answers can help shed light on the effects of ecological processes on community composition.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

AUTHOR CONTRIBUTIONS

CR conceived the idea. CR, SP and VP developed the methodology. CR, SP and BC analyzed the data. SP wrote the R script. CR took the lead in writing the main text; all authors revised the manuscript critically and approved the final version.

DATA AVAILABILITY STATEMENT

Data on the alpine vegetation used in the worked example are available in Appendices S2 and S5; the R scripts for the dissimilarity-based MANOVA decomposition are available in Appendices S3 and S6.

ORCID

Carlo Ricotta  <https://orcid.org/0000-0003-0818-3959>

Sandrine Pavoine  <https://orcid.org/0000-0003-2767-6484>

Bruno E. L. Cerabolini  <https://orcid.org/0000-0002-3793-0733>

Valério D. Pillar  <https://orcid.org/0000-0001-6408-2891>

REFERENCES

- Anderson, M.J. (2001) A new method for non-parametric multivariate analysis of variance. *Austral Ecology*, 26, 32–46.
- Benjamini, Y. & Yekutieli, D. (2001) The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics*, 29, 1165–1188.
- Botta-Dukát, Z. (2018) Cautionary note on calculating standardized effect size (SES) in randomization test. *Community Ecology*, 19, 77–83. <https://doi.org/10.1556/168.2018.19.1.8>.
- Caccianiga, M., Luzzaro, A., Pierce, S., Ceriani, R.M. & Cerabolini, B. (2006) The functional basis of a primary succession resolved by CSR classification. *Oikos*, 112, 10–20.
- Cai, L. (2006) Multi-response permutation procedure as an alternative to the analysis of variance: An SPSS implementation. *Behavior Research Methods*, 38, 51–59.
- Chytrý, M., Tichý, L., Holt, J. & Botta-Dukát, Z. (2002) Determination of diagnostic species with statistical fidelity measures. *Journal of Vegetation Science*, 13, 79–90.
- De Cáceres, M. & Legendre, P. (2009) Associations between species and groups of sites: indices and statistical inference. *Ecology*, 90, 3566–3574.
- De Cáceres, M., Legendre, P. & Moretti, M. (2010) Improving indicator species analysis by combining groups of sites. *Oikos*, 119, 1674–1684. <https://doi.org/10.1111/j.1600-0706.2010.18334.x>.
- Dufrêne, M. & Legendre, P. (1997) Species assemblages and indicator species: the need for a flexible asymmetrical approach. *Ecological Monographs*, 67, 345–366.
- Edgington, E.S. (1987) *Randomization tests*. New York, NY: Marcel Dekker.
- Gotelli, N.J. & McCabe, D.J. (2002) Species co-occurrence: a meta-analysis of J.M. Diamond's assembly rules model. *Ecology*, 83, 2091–2096.
- Gower, J.C. & Legendre, P. (1986) Metric and Euclidean properties of dissimilarity coefficients. *Journal of Classification*, 3, 5–48.
- Grime, J.P. (1974) Vegetation classification by reference to strategies. *Nature*, 250, 26–31.
- Legendre, P. & De Cáceres, M. (2013) Beta diversity as the variance of community data: Dissimilarity coefficients and partitioning. *Ecology Letters*, 16, 951–963. <https://doi.org/10.1111/ele.12141>.
- Legendre, P. & Gallagher, E.D. (2001) Ecologically meaningful transformations for ordination of species data. *Oecologia*, 129, 271–280.
- Manly, B.F.J. (1991) *Randomization and Monte Carlo methods in biology*. London: Chapman and Hall.
- McArdle, B.H. & Anderson, M.J. (2001) Fitting multivariate models to community data: a comment on distance-based redundancy analysis. *Ecology*, 82, 290–297.
- Mielke, P.W. & Berry, J.A. (2001) *Permutation methods: A distance approach*. New York: Springer-Verlag.
- Orlóci, L. (1978) *Multivariate analysis in vegetation research*. The Hague, NL: Junk.



- Pavoine, S. & Ricotta, C. (2019) Measuring functional dissimilarity among plots: Adapting old methods to new questions. *Ecological Indicators*, 97, 67–72. <https://doi.org/10.1016/j.ecolind.2018.09.048>.
- Pavoine, S., Gasc, A., Bonsall, M.B. & Mason, N.W.H. (2013) Correlations between phylogenetic and functional diversity: Mathematical artefacts or true ecological and evolutionary processes? *Journal of Vegetation Science*, 24, 781–793. <https://doi.org/10.1111/jvs.12051>.
- Pillar, V.D. (2013) How accurate and powerful are randomization tests in multivariate analysis of variance? *Community Ecology*, 14, 153–163. <https://doi.org/10.1556/ComEc.14.2013.2.5>.
- Pillar, V.D. & Orlóci, L. (1996) On randomization testing in vegetation science: Multifactor comparisons of relevé groups. *Journal of Vegetation Science*, 7, 585–592. <https://doi.org/10.2307/3236308>.
- Podani, J. & Csányi, B. (2010) Detecting indicator species: some extensions of the IndVal measure. *Ecological Indicators*, 10, 1119–1124. <https://doi.org/10.1016/j.ecolind.2010.03.010>.
- Ricotta, C., Acosta, A.T.R., Caccianiga, M., Cerabolini, B.E.L., Godefroid, S. & Carboni, M. (2020) From abundance-based to functional-based indicator species. *Ecological Indicators*, 118, 106761. <https://doi.org/10.1016/j.ecolind.2020.106761>.
- Ricotta, C., Bacaro, G., Caccianiga, M., Cerabolini, B. & Pavoine, S. (2018) A new method for quantifying the phylogenetic redundancy of biological communities. *Oecologia*, 186, 339–346. <https://doi.org/10.1007/s00442-017-4026-x>.
- Ricotta, C., Carboni, M. & Acosta, A.T.R. (2015) Let the concept of indicator species be functional!. *Journal of Vegetation Science*, 26, 839–847. <https://doi.org/10.1111/jvs.12291>.
- Ricotta, C., de Bello, F., Moretti, M., Caccianiga, M., Cerabolini, B. & Pavoine, S. (2016) Measuring the functional redundancy of biological communities: A quantitative guide. *Methods in Ecology and Evolution*, 7, 1386–1395. <https://doi.org/10.1111/2041-210X.12604>.

- Ricotta, C. & Podani, J. (2017) On some properties of the Bray-Curtis dissimilarity and their ecological meaning. *Ecological Complexity*, 31, 201–205. <https://doi.org/10.1016/j.ecocom.2017.07.003>.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

Appendix S1. On the relationship between $SES(Q_{Ws})$ and $SES(Q_{Bs})$.

Appendix S2. Community composition matrix of the Alpine vegetation used in the worked example.

Appendix S3. R scripts.

Appendix S4. Overview of MRPP and its relationship with db-MANOVA.

Appendix S5. Community composition matrix of the Alpine vegetation in text format.

Appendix S6. R scripts in text format.

How to cite this article: Ricotta C, Pavoine S, Cerabolini BEL, Pillar VD. A new method for indicator species analysis in the framework of multivariate analysis of variance. *J Veg Sci*. 2021;32:e13013. <https://doi.org/10.1111/jvs.13013>

Supporting information to the paper

Ricotta, C. et al. A new method for indicator species analysis in the framework of multivariate analysis of variance. *Journal of Vegetation Science*.

Appendix S1. On the relationship between $SES(Q_{Ws})$ and $SES(Q_{Bs})$

The standardized effect size (SES) of a variable I is defined as (Gotelli and McCabe 2002):

$$SES = \frac{Obs(I) - Mean(I)}{SD(I)} \quad (A1)$$

where $Obs(I)$ is the observed value of I , $Mean(I)$ is the mean of the null distribution of the variable in random assemblages, and $SD(I)$ is the standard deviation of the null distribution. However, since the actual values of the species-level within-group sum of squared dissimilarities Q_{Ws} are usually lower than corresponding null values, we typically have: $Mean(Q_{Ws}) > Obs(Q_{Ws})$. Therefore, in this paper we calculated the contribution of a given species s to the within-group sum of squares Q_W as the standardized effect size of Q_{Ws} changed in sign (see Eq. 8 of the main text):

$$SES_s = -SES(Q_{Ws}) = \frac{Mean(Q_{Ws}) - Obs(Q_{Ws})}{SD(Q_{Ws})} \quad (A2)$$

Likewise, since the between-group sum of squares Q_B is the difference between the total sum of squares and the within-group sum of squares: $Q_B = Q_T - Q_W$, using the Euclidean distance, the decomposition of Q_B into additive species-level values Q_{Bs} gives:

$$\begin{aligned} Q_T &= \sum_{s=1}^S \frac{1}{N} \sum_{i>j}^N (x_{si} - x_{sj})^2 \\ Q_B &= Q_T - Q_W = \sum_{s=1}^S (Q_{Ts} - Q_{Ws}) = \sum_{s=1}^S \left(\frac{1}{N} \sum_{i>j}^N (x_{si} - x_{sj})^2 - \sum_{k=1}^K \frac{1}{N_k} \sum_{i>j}^N (x_{si} - x_{sj})^2 \varepsilon_{ijk} \right) \\ Q_{Bs} &= Q_{Ts} - Q_{Ws} = \frac{1}{N} \sum_{i>j}^N (x_{si} - x_{sj})^2 - \sum_{k=1}^K \frac{1}{N_k} \sum_{i>j}^N (x_{si} - x_{sj})^2 \varepsilon_{ijk} \end{aligned}$$

where x_{si} and x_{sj} are the abundances of species s ($s = 1, 2, \dots, S$) in plots i and j ($i, j = 1, 2, \dots, N$), respectively, N_k is the number of plots in group k ($k = 1, 2, \dots, K$) such that $\sum_{k=1}^K N_k = N$, Q_{Ts} is the species-level contribution to Q_T and ε_{ijk} is an indicator variable that takes the value one if plots i and j are both in group k . Otherwise it is zero.

Since in one-factor db-MANOVA Q_{Ts} is a constant by the permutation scheme chosen (randomly permuting plots across groups), we have:

$$\begin{aligned}
SES(Q_{Bs}) &= \frac{Obs(Q_{Ts} - Q_{Ws}) - Mean(Q_{Ts} - Q_{Ws})}{SD(Q_{Ts} - Q_{Ws})} \\
&= \frac{Q_{Ts} - Obs(Q_{Ws}) - Q_{Ts} + Mean(Q_{Ws})}{SD(Q_{Ws})} \\
&= \frac{Mean(Q_{Ws}) - Obs(Q_{Ws})}{SD(Q_{Ws})}
\end{aligned}$$

and hence

$$SES_s = -SES(Q_{Ws}) = SES(Q_{Bs}) \quad (A3)$$

References

Gotelli, N.J. & McCabe, D.J. 2002. Species co-occurrence: a meta-analysis of J.M. Diamond's assembly rules model. *Ecology* 83: 2091–2096.

Supporting information to the paper
Ricotta, C. et al. A new method for indicator species analysis in the framework of multivariate analysis of variance. *Journal of Vegetation Science*.

Appendix S2 Community composition matrix of the Alpine vegetation used in the worked example
Species abundances (% cover) were summarized with a five-point ordinal scale transformed to ranks

Species	Early-successional Plots																	Mid-successional Plots																															Late-successional plots																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																						
Achillea moschata	0	1	0	1	0	0	1	0	0	2	0	0	0	0	0	0	0	2	1	2	0	1	2	2	0	4	3	0	3	2	2	3	2	2	2	2	2	2	1	2	2	1	3	1	2	1	0	2	3	0	0	0	0	0	0	0	0	0																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																							
Adenostyles leucophylla	0	0	0	0	2	2	2	1	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																							
Agrostis rupestris	2	1	0	2	2	0	2	2	0	2	0	1	0	0	0	0	2	0	0	1	0	3	2	0	2	3	1	2	4	1	1	2	2	1	3	2	2	2	2	2	1	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																					
Agrostis schraderiana	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																				
Antennaria dioica	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	2	1	1	0	1	1	0	1	0	0	1	0	1	0	0	0	0	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																				
Anthoxanthum odoratum	0	0	2	1	2	0	0	0	0	2	0	0	0	1	0	0	0	3	4	1	1	1	1	0	2	1	3	1	2	1	2	1	2	1	2	1	1	0	3	2	3	1	2	1	0	1	1	0	4	2	2	2	2	2	1	2	0	0	0	0	0																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																				
Arabis alpina	0	0	0	0	0	0	0	1	0	0	1	1	2	1	2	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	

Supporting information to the paper

Ricotta, C. et al. A new method for indicator species analysis in the framework of multivariate analysis of variance. *Journal of Vegetation Science*.

Appendix S3. R scripts

Consider N plots distributed among K groups; plots are composed of species whose abundances within each plot are known. The Q_T , Q_B and Q_W statistics defined in the main text aim at evaluating the average difference, in terms of species identity and abundance, between any two plots (Q_T), between two plots within a group (Q_W), and the gap between Q_T and Q_W ($=Q_B$) due to compositional differences between groups of plots. R function "dbMANOVAspecies" calculates Q_T , Q_B and Q_W as in the main text and the species-centered components Q_{Ts} , Q_{Bs} and Q_{Ws} . It also calculates the SES values (equation 7) associated with Q_B and Q_{Bs} and allows tests for the significance of the SES values (H_0 = the SES value is similar as that expected by randomly permuting plots among groups of plots and H_1 = the SES value of is greater than that expected by randomly permuting plots among groups of plots). R function "dbMANOVAspecies_pairwise" complements function "dbMANOVAspecies" by performing post-hoc tests for all pairs of groups. It must be executed in the same R environment (workspace) as function "dbMANOVAspecies". A third function named `summary.dbMANOVAspecies_pairwise` provides a short summary of the results of function `dbMANOVAspecies_pairwise` (with SES and P values).

This program is free software: you can redistribute it and/or modify it under the terms of the GNU General Public License <http://www.gnu.org/licenses/>.

It will be integrated in version 2.1 of the `adiv` package of R (Pavoine, 2020): <https://cran.r-project.org/web/packages/adiv/index.html>. The functions were checked and applied with R-4.0.2.

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

Dependencies: `ade4` (Thioulouse et al. 2018).

Usage: `dbMANOVAspecies(comm, groups, nrep = 999, method = c("Euclidean", "Manhattan", "Canberra", "BrayCurtis"), global = TRUE, species = TRUE, padjust = "none", tol = 1e-8)`

`dbMANOVAspecies_pairwise(dbobj, signif = TRUE, salpha = 0.05, nrep = NULL)`

`summary.dbMANOVAspecies_pairwise(dbpair, DIGITS = 3)`

Arguments:

`comm`: a matrix of N plots $\times$ S species containing the relative or absolute abundance of all species. Columns are species and plots are rows.

`groups`: a vector of characters or a factor with the names of the groups associated with plots.

Names of groups must be listed in the same order as plots in object `comm`. For example, the first value of `groups` gives the name of the group for the first plot (first row in object `comm`).

nrep: a numeric that gives the number of permutations to be done. `nrep` can be set to `NULL` in function `dbMANOVAspecies_pairwise`, in which case the number of permutations used to build object `dbobj` (first argument of function `dbMANOVAspecies_pairwise`) is used by default.

method: a string, one of "Euclidean" (to perform the Euclidean distance), "Manhattan" (to select the Manhattan distance) , "Canberra" (to perform the Canberra distance), "BrayCurtis" (to opt for the BrayCurtis distance).

global: a logical. If `TRUE` the global test (using the SES of Q_B) is performed. The global test can be performed in addition to one test per species.

species: a logical. If `TRUE` one test per species is performed. If `species = FALSE`, then argument `global` is ignored and the global test is performed.

padjust: a method of correction for multiple tests one of `p.adjust.methods` (see `?p.adjust.methods` for possible choices); choose "none" if you don't need any correction for multiple tests. The argument `padjust` is ignored if `species = FALSE`.

tol: a numeric tolerance threshold. Any value between `-tol` and `tol` will be considered as equal to zero.

dbobj: an object of class "dbMANOVAspecies" obtained with function `dbMANOVAspecies`.

signif: a logical. If `TRUE`, the post-hoc tests are performed only for species that were associated with significant tests in object `dbobj`.

salpha: a numeric. The level of significance (nominal alpha error) for P values (must be between 0 and 1). Ignored if `signif = FALSE`.

dbspair: an object of class "dbMANOVAspecies_pairwise" obtained with function `dbMANOVAspecies_pairwise`.

DIGITS: integer indicating the number of decimal places to retain when displaying the results. If `NULL`, all decimals are retained.

Value:

`dbMANOVAspecies` returns a list with the following objects

\$observations: a vector or a data frame with the observed values of statistics (Q_T , Q_B and Q_W) (if `global = TRUE`) and/or the contribution each species has to these statistics (if `species = TRUE`)

\$test: an object of class `randtest` or `krandtest` with the results of the test for the global differences, in terms of species composition, between groups of plots (if `global = TRUE`) and/or for the contribution each species has in these differences (if `species = TRUE`).

`dbMANOVAspecies_pairwise` provides in a list the same objects as function `dbMANOVAspecies` but for each pair of groups (see examples below).

If only a global test was performed, `summary.dbMANOVAspecies_pairwise` provides a data frame with two rows named "SES" for the SES values (of the Q_B statistic) and "pvalue" for the P values and as many columns as there are pairs of groups of plots. Columns are named according to the two groups that are compared. For example, if there are two groups named Group1 and Group2, they are compared in column named "Group1:Group2". Else, it provides a data frame with species names as rows (in addition to a row named GLOBAL if a global test was also performed) and as columns the SES, P values, and if relevant adjusted P values, for each combination of two groups. Consider that Group1 and Group2 are the names of two groups. The column names that correspond to the comparison between Group1 and Group2 are written Group1:Group2.SES (for the SES values), Group1:Group2.pvalue (for the P values), and Group1:Group2.adj.pvalue (for the P values adjusted after correction for multiple tests, if a correction was done). NAs (= missing values) may be present in the data frame for species that are absent from two compared groups of plots. NAs may also be present in the data frame if `signif = TRUE` was used in function `dbMANOVAspecies_pairwise` for species that were associated to non-significant test a non-significant test when performing function `dbMANOVAspecies`.

Example:

Install and load packages `ade4` and `adiv`.

```
install.packages("ade4")
install.packages("adiv")
library(ade4)
library(adiv)
```

```
data(RutorGlacier)
```

The data set `RutorGlacier` is a list of objects. One of these objects named `Abund` contains the abundance of plant species over plots (plots as rows, species as columns). Another one, named `Fac`, indicates which successional stage each plot belongs to: "early" = early-successional stage, "mid" = mid-successional stage and "late" = late-successional stage.

We calculate below the Q_T , Q_B and Q_W statistics and the contribution each species has to these statistics. We also test for the significance of the differences in species compositions between groups of plots (the successional stages) and for the significance of the contribution each species has to these differences.

This first series of tests was done without any correction for multiple tests:

```
Qspecies <- dbMANOVAspecies(RutorGlacier$Abund, RutorGlacier$Fac,
nrep=9999, global=TRUE)
```

```
Qspecies
```

```
$observations
```

	Global	Achillea_moschata	Adenostyles_leucophylla	Agrostis_rupestris	
QT	1485.5932	72.84746	16.305085	62.576271	
QB	377.5973	37.84930	4.187438	3.066345	
QW	1107.9960	34.99816	12.117647	59.509926	
	Agrostis_schraderiana	Antennaria_dioica	Anthoxanthum_odoratum	Arabis_alpina	
QT	8.5762712	19.220339	66.13559	12.305085	
QB	0.3575212	1.904163	13.43155	3.101041	
QW	8.2187500	17.316176	52.70404	9.204044	
	Avenula_versicolor	Cardamine_bellidifolia	Cardamine_resedifolia		
QT	38.23729	3.9322034	13.220339		

QB	10.16854	0.1674975	0.237986
QW	28.06875	3.7647059	12.982353
	Carex_curvula	Carex sempervirens	Cerastium_uniflorum Epilobium_fleischeri
QT	50.54237	56.64407	25.22034 31.796610
QB	22.94237	25.10436	10.50159 1.710213
QW	27.60000	31.53971	14.71875 30.086397
	Erigeron_uniflorus	Euphrasia_minima	Festuca_halleri Gnaphalium_supinum
QT	32.033898	18.644068	74.88136 14.677966
QB	5.050442	1.626421	11.28136 2.123922
QW	26.983456	17.017647	63.60000 12.554044
	Hieracium_angustifolium	Homogyne_alpina	Juncus_trifidus Leontodon_helveticus
QT	66.98305	10.305085	21.661017 37.661017
QB	24.39371	3.388908	4.777561 4.709914
QW	42.58934	6.916176	16.883456 32.951103
	Leucanthemopsis_alpina	Linaria_alpina	Lotus_corniculatus Luzula_alpinopilosa
QT	38.983051	14.101695	31.2203390 5.72881356
QB	9.390404	4.353533	0.6865155 0.01263709
QW	29.592647	9.748162	30.5338235 5.71617647
	Luzula_lutea	Luzula_spicata	Minuartia_recurva Myosotis_alpestris
QT	5.728814	8.9491525	47.55932 38.745763
QB	1.328814	0.7156967	15.78432 7.862307
QW	4.400000	8.2334559	31.77500 30.883456
	Oxyria_digyna	Pedicularis_kernerii	Phleum_rhaeticum Phyteuma_hemisphaericum
QT	23.22034	12.915254	21.66102 13.186441
QB	10.50159	1.015254	2.66837 1.811441
QW	12.71875	11.900000	18.99265 11.375000
	Poa_alpina	Ranunculus_glacialis	Ranunculus_kuepferi Saxifraga_aizoides
QT	79.62712	5.7288136	8.5762712 22.94915
QB	29.11719	0.2894753	0.6012712 5.06680
QW	50.50993	5.4393382	7.9750000 17.88235
	Saxifraga_bryoides	Senecio_incanus	Silene_aucaulis Trifolium_badium
QT	36.677966	42.949153	38.406780 84.91525
QB	7.238628	1.212755	4.990603 18.51084
QW	29.439338	41.736397	33.416176 66.40441
	Trifolium_pallescent	Tussilago_farfara	Veronica_bellidioides
QT	96.64407	35.52542	17.186441
QB	48.17348	11.86550	2.317691
QW	48.47059	23.65993	14.868750

\$test

class: krandtest lightkrandtest

Monte-Carlo tests

Call: as.krandtest(sim = THEO, obs = OBS, alter = "greater",
p.adjust.method = padjust)

Number of tests: 46

Adjustment method for multiple comparisons: none

Permutation number: 9999

	Test	Obs	Std.Obs	Alter	Pvalue
1	GLOBAL	377.59726446	22.3933693	greater	0.0001
2	Achillea_moschata	37.84929586	14.3195540	greater	0.0001
3	Adenostyles_leucophylla	4.18743769	6.8605361	greater	0.0010
4	Agrostis_rupestris	3.06634472	0.4399014	greater	0.2444
5	Agrostis_schraderiana	0.35752119	0.2507189	greater	0.3304
6	Antennaria_dioica	1.90416251	1.9172993	greater	0.0438
7	Anthoxanthum_odorum	13.43154910	5.0003738	greater	0.0021
8	Arabis_alpina	3.10104063	6.4268457	greater	0.0014
9	Avenula_versicolor	10.16853814	7.1114719	greater	0.0002
10	Cardamine_bellidifolia	0.16749751	0.3267220	greater	0.4508
11	Cardamine_resedifolia	0.23798604	-0.4954167	greater	0.6596
12	Carex_curvula	22.94237288	13.4213793	greater	0.0001
13	Carex sempervirens	25.10436191	12.3673491	greater	0.0001
14	Cerastium_uniflorum	10.50158898	11.7599037	greater	0.0001
15	Epilobium_fleischeri	1.71021311	0.5404468	greater	0.2382
16	Erigeron_uniflorus	5.05044242	3.7340798	greater	0.0072
17	Euphrasia_minima	1.62642074	1.6405498	greater	0.0677
18	Festuca_halleri	11.28135593	3.3796018	greater	0.0109

19	Gnaphalium_supinum	2.12392198	3.4103469	greater	0.0120
20	Hieracium_angustifolium	24.39371261	9.7674020	greater	0.0001
21	Homogyne_alpina	3.38890828	9.0177216	greater	0.0003
22	Juncus_trifidus	4.77756107	5.7215847	greater	0.0018
23	Leontodon_helveticus	4.70991401	2.7683030	greater	0.0225
24	Leucanthemopsis_alpina	9.39040379	6.3551633	greater	0.0006
25	Linaria_alpina	4.35353315	8.2893262	greater	0.0003
26	Lotus_corniculatus	0.68651545	-0.3825797	greater	0.5787
27	Luzula_alpinopilosa	0.01263709	-1.0752639	greater	1.0000
28	Luzula_lutea	1.32881356	6.4883833	greater	0.0043
29	Luzula_spicata	0.71569666	1.3428603	greater	0.1016
30	Minuartia_recurva	15.78432203	8.7128217	greater	0.0002
31	Myosotis_alpestris	7.86230683	5.1848426	greater	0.0016
32	Oxyria_digyna	10.50158898	12.6248343	greater	0.0001
33	Pedicularis_kernerii	1.01525424	1.3945066	greater	0.0832
34	Phleum_rhaeticum	2.66836989	2.7432176	greater	0.0221
35	Phyteuma_hemisphaericum	1.81144068	3.1353861	greater	0.0180
36	Poa_alpina	29.11719217	9.9580604	greater	0.0001
37	Ranunculus_glacialis	0.28947532	0.5178989	greater	0.2059
38	Ranunculus_kuepferi	0.60127119	1.1529990	greater	0.1077
39	Saxifraga_aizoides	5.06679960	5.6141114	greater	0.0040
40	Saxifraga_bryoides	7.23862787	5.0510040	greater	0.0017
41	Senecio_incanus	1.21275548	-0.1865420	greater	0.4439
42	Silene_aucaulis	4.99060319	2.8797721	greater	0.0207
43	Trifolium_badium	18.51084247	5.4386693	greater	0.0017
44	Trifolium_pallidum	48.17347956	13.7331477	greater	0.0001
45	Tussilago_farfara	11.86549726	9.0139118	greater	0.0002
46	Veronica_bellidifolia	2.31769068	3.0280995	greater	0.0218

```
$call
dbMANOVAspecies(comm = RutorGlacier$Abund, groups = RutorGlacier$Fac,
  nrep = 9999, global = TRUE)
```

```
$method
[1] "Euclidean"
```

```
$p.adjust
[1] "none"
```

```
$tol
[1] 1e-08
```

```
attr(,"class")
[1] "dbMANOVAspecies"
```

Column "Std.Obs" above stands for the SES values.

The argument p.adjust of function dbMANOVAspecies allows us to add a correction for multiple tests, as shown below:

```
Qspecies_adj <- dbMANOVAspecies(RutorGlacier$Abund, RutorGlacier$Fac,
  nrep=9999, global=FALSE, padj = "BY")
```

```
Qspecies_adj$test
```

```
class: krandtest lightkrandtest
Monte-Carlo tests
Call: as.krandtest(sim = THEO, obs = OBS, alter = "greater",
  p.adjust.method = padjust)
```

```
Number of tests: 45
```

```
Adjustment method for multiple comparisons: BY
Permutation number: 9999
```

	Test	Obs	Std.Obs	Alter	Pvalue	Pvalue.adj
1	Achillea_moschata	37.84929586	14.5476173	greater	0.0001	0.001977727

2	Adenostyles_leucophylla	4.18743769	6.8676030	greater	0.0014	0.016287161
3	Agrostis_rupestris	3.06634472	0.4447499	greater	0.2438	1.000000000
4	Agrostis_schraderiana	0.35752119	0.2270552	greater	0.3308	1.000000000
5	Antennaria_dioica	1.90416251	1.9691571	greater	0.0416	0.257104465
6	Anthoxanthum_odoratum	13.43154910	5.0911948	greater	0.0018	0.018736358
7	Arabis_alpina	3.10104063	6.5351682	greater	0.0018	0.018736358
8	Avenula_versicolor	10.16853814	7.0132161	greater	0.0003	0.004563985
9	Cardamine_bellidifolia	0.16749751	0.3245258	greater	0.4523	1.000000000
10	Cardamine_resedifolia	0.23798604	-0.4994032	greater	0.6578	1.000000000
11	Carex_curvula	22.94237288	12.9765654	greater	0.0001	0.001977727
12	Carex sempervirens	25.10436191	12.3132479	greater	0.0001	0.001977727
13	Cerastium_uniflorum	10.50158898	11.8397618	greater	0.0001	0.001977727
14	Epilobium_fleischeri	1.71021311	0.5685664	greater	0.2323	1.000000000
15	Erigeron_uniflorus	5.05044242	3.7316699	greater	0.0081	0.066748275
16	Euphrasia_minima	1.62642074	1.5525536	greater	0.0742	0.444688841
17	Festuca_halleri	11.28135593	3.4584432	greater	0.0114	0.090184335
18	Gnaphalium_supinum	2.12392198	3.4100552	greater	0.0123	0.093561684
19	Hieracium_angustifolium	24.39371261	10.0349870	greater	0.0001	0.001977727
20	Homogyne_alpina	3.38890828	9.1413930	greater	0.0003	0.004563985
21	Juncus_trifidus	4.77756107	5.9383581	greater	0.0013	0.016069029
22	Leontodon_helveticus	4.70991401	2.7065112	greater	0.0258	0.164597896
23	Leucanthemopsis_alpina	9.39040379	6.1554185	greater	0.0005	0.007063309
24	Linaria_alpina	4.35353315	8.3619294	greater	0.0002	0.003595867
25	Lotus_corniculatus	0.68651545	-0.3721485	greater	0.5721	1.000000000
26	Luzula_alpinopilosa	0.01263709	-1.0815886	greater	1.0000	1.000000000
27	Luzula_lutea	1.32881356	6.6665182	greater	0.0030	0.026656316
28	Luzula_spicata	0.71569666	1.3595150	greater	0.0971	0.548677880
29	Minuartia_recurva	15.78432203	8.9913067	greater	0.0001	0.001977727
30	Myosotis_alpestris	7.86230683	5.1019674	greater	0.0023	0.021660816
31	Oxyria_digyna	10.50158898	12.5376496	greater	0.0001	0.001977727
32	Pedicularis_kernerii	1.01525424	1.3413867	greater	0.0864	0.502575243
33	Phleum_rhaeticum	2.66836989	2.6091377	greater	0.0253	0.164597896
34	Phyteuma_hemisphaericum	1.81144068	3.1256667	greater	0.0160	0.117198616
35	Poa_alpina	29.11719217	10.0485265	greater	0.0001	0.001977727
36	Ranunculus_glacialis	0.28947532	0.5423829	greater	0.2073	1.000000000
37	Ranunculus_kuepferi	0.60127119	1.2027124	greater	0.1038	0.570244518
38	Saxifraga_aizoides	5.06679960	5.8061748	greater	0.0031	0.026656316
39	Saxifraga_bryoides	7.23862787	4.9125044	greater	0.0022	0.021660816
40	Senecio_incanus	1.21275548	-0.1863817	greater	0.4462	1.000000000
41	Silene_aucaulis	4.99060319	2.9742563	greater	0.0187	0.132083887
42	Trifolium_badium	18.51084247	5.5267739	greater	0.0006	0.007910907
43	Trifolium_pallesces	48.17347956	13.7624679	greater	0.0001	0.001977727
44	Tussilago_farfara	11.86549726	9.1856793	greater	0.0001	0.001977727
45	Veronica_bellidioides	2.31769068	3.0687226	greater	0.0198	0.135030992

Column "Std.Obs" above contains the SES values.

Now for all species that showed significant compositional difference among the three groups of plots (with a nominal α error of 5%), we can test for pairwise differences among all pairs of groups thanks to function dbMANOVAspecies_pairwise as shown below:

without any correction for multiple tests

```
Qspeciespairwise <- dbMANOVAspecies_pairwise(Qspecies)
summary(Qspeciespairwise)
```

	early:late.SES	early:late.pvalue	early:mid.SES	early:mid.pvalue	late:mid.SES	late:mid.pvalue
GLOBAL	14.670	0.000	17.316	0.000	16.096	0.000
Achillea_moschata	1.150	0.234	13.908	0.000	12.819	0.000
Adenostyles_leucophylla	2.265	0.062	8.187	0.001	NA	NA
Agrostis_rupestris	NA	NA	NA	NA	NA	NA
Agrostis_schraderiana	NA	NA	NA	NA	NA	NA
Antennaria_dioica	3.174	0.036	3.120	0.024	-0.680	1.000
Anthoxanthum_odoratum	4.491	0.012	6.634	0.001	-0.738	1.000
Arabis_alpina	2.731	0.049	7.797	0.001	-0.560	1.000
Avenula_versicolor	11.096	0.000	6.269	0.002	1.063	0.139
Cardamine_bellidifolia	NA	NA	NA	NA	NA	NA
Cardamine_resedifolia	NA	NA	NA	NA	NA	NA
Carex_curvula	9.088	0.000	2.753	0.043	11.653	0.000
Carex sempervirens	10.394	0.000	-0.844	0.900	13.289	0.000
Cerastium_uniflorum	5.963	0.002	12.984	0.000	-0.019	0.576
Epilobium_fleischeri	NA	NA	NA	NA	NA	NA
Erigeron_uniflorus	0.218	0.512	2.977	0.034	2.687	0.044
Euphrasia_minima	NA	NA	NA	NA	NA	NA

Festuca_halleri	6.273	0.003	5.248	0.003	-0.584	0.671
Gnaphalium_supinum	0.543	0.214	5.034	0.007	-0.156	0.581
Hieracium_angustifolium	1.504	0.110	12.604	0.000	3.973	0.010
Homogyne_alpina	5.881	0.004	-0.799	1.000	9.563	0.001
Juncus_trifidus	5.197	0.006	-0.736	1.000	5.883	0.006
Leontodon_helveticus	3.954	0.016	1.659	0.087	0.731	0.191
Leucanthemopsis_alpina	9.509	0.000	0.902	0.174	5.254	0.003
Linaria_alpina	5.928	0.004	8.161	0.000	0.525	0.318
Lotus_corniculatus	NA	NA	NA	NA	NA	NA
Luzula_alpinopilosa	NA	NA	NA	NA	NA	NA
Luzula_lutea	3.489	0.043	NA	NA	6.004	0.011
Luzula_spicata	NA	NA	NA	NA	NA	NA
Minuartia_recurva	11.758	0.000	11.438	0.000	-0.149	0.422
Myosotis_alpestris	-1.074	1.000	5.000	0.004	3.349	0.018
Oxyria_digyna	6.377	0.002	13.886	0.000	-0.016	0.568
Pedicularis_kernerii	NA	NA	NA	NA	NA	NA
Phleum_rhaeticum	2.000	0.077	2.417	0.062	0.872	0.310
Phyteuma_hemisphaericum	6.861	0.003	3.105	0.040	-0.022	0.492
Poa_alpina	10.880	0.000	-0.533	0.671	11.802	0.000
Ranunculus_glacialis	NA	NA	NA	NA	NA	NA
Ranunculus_kuepferi	NA	NA	NA	NA	NA	NA
Saxifraga aizoides	1.821	0.113	6.964	0.003	NA	NA
Saxifraga_bryoides	5.884	0.002	-0.454	0.578	6.943	0.001
Senecio_incanus	NA	NA	NA	NA	NA	NA
Silene_acaulis	-0.323	0.760	3.850	0.016	0.851	0.170
Trifolium_badium	6.045	0.002	0.753	0.166	6.896	0.001
Trifolium_pallescens	2.978	0.025	10.574	0.000	16.363	0.000
Tussilago_farfara	4.564	0.005	10.042	0.000	-0.328	0.821
Veronica_Bellidioides	6.410	0.003	2.067	0.063	0.775	0.241

NAs are present in the data frame above for species that were associated to non-significant test in object Qspecies (where tests were done over all groups); and also, for species that are absent from the two compared groups of plots (e.g. species *Adenostyles leucophylla*, in mid- and late-successional stages).

with correction for multiple test

```
Qspeciespairwise_adj <- dbMANOVAspecies_pairwise(Qspecies_adj)
summary(Qspeciespairwise_adj)
```

	early:late.SES	early:late.pvalue	early:late.adj.pvalue
Achillea_moschata	1.123	0.240	1.000
Adenostyles_leucophylla	2.235	0.066	0.356
Agrostis_rupestris	NA	NA	NA
Agrostis_schraderiana	NA	NA	NA
Antennaria_dioica	3.041	0.039	0.241
Anthoxanthum_odorum	4.597	0.011	0.079
Arabis_alpina	2.708	0.046	0.263
Avenula_versicolor	11.445	0.000	0.003
Cardamine_bellidifolia	NA	NA	NA
Cardamine_resedifolia	NA	NA	NA
Carex_curvula	9.155	0.000	0.003
Carex sempervirens	10.055	0.000	0.004
Cerastium_uniflorum	6.107	0.001	0.020
Epilobium_fleischeri	NA	NA	NA
Erigeron_uniflorus	0.192	0.521	1.000
Euphrasia_minima	NA	NA	NA
Festuca_halleri	6.220	0.003	0.029
Gnaphalium_supinum	0.521	0.227	1.000
Hieracium_angustifolium	1.495	0.110	0.549
Homogyne_alpina	5.999	0.003	0.029
Juncus_trifidus	5.162	0.006	0.047
Leontodon_helveticus	4.142	0.013	0.088
Leucanthemopsis_alpina	9.824	0.000	0.003
Linaria_alpina	6.148	0.003	0.029
Lotus_corniculatus	NA	NA	NA
Luzula_alpinopilosa	NA	NA	NA
Luzula_lutea	3.462	0.044	0.260
Luzula_spicata	NA	NA	NA
Minuartia_recurva	12.017	0.000	0.003
Myosotis_alpestris	-1.056	1.000	1.000
Oxyria_digyna	6.528	0.001	0.013
Pedicularis_kernerii	NA	NA	NA
Phleum_rhaeticum	2.007	0.071	0.370
Phyteuma_hemisphaericum	6.831	0.003	0.029
Poa_alpina	10.979	0.000	0.003
Ranunculus_glacialis	NA	NA	NA
Ranunculus_kuepferi	NA	NA	NA
Saxifraga aizoides	1.758	0.121	0.581
Saxifraga_bryoides	5.780	0.002	0.023
Senecio_incanus	NA	NA	NA
Silene_acaulis	-0.331	0.760	1.000
Trifolium_badium	6.052	0.001	0.019

Trifolium_pallesces	2.959	0.025	0.165
Tussilago_farfara	4.617	0.006	0.046
Veronica_bellidioides	6.309	0.003	0.029
	early:mid.SES	early:mid.pvalue	early:mid.adj.pvalue
Achillea_moschata	14.390	0.000	0.002
Adenostyles_leucophylla	7.885	0.001	0.016
Agrostis_rupestris	NA	NA	NA
Agrostis_schraderiana	NA	NA	NA
Antennaria_dioica	3.250	0.022	0.153
Anthoxanthum_odoratum	6.777	0.001	0.015
Arabis_alpina	7.731	0.001	0.015
Avenula_versicolor	6.317	0.002	0.022
Cardamine_bellidifolia	NA	NA	NA
Cardamine_resedifolia	NA	NA	NA
Carex_curvula	3.006	0.037	0.222
Carex sempervirens	-0.838	0.892	1.000
Cerastium_uniflorum	12.625	0.000	0.002
Epilobium_fleischeri	NA	NA	NA
Erigeron_uniflorus	3.048	0.035	0.218
Euphrasia_minima	NA	NA	NA
Festuca_halleri	5.263	0.003	0.024
Gnaphalium_supinum	5.034	0.007	0.055
Hieracium_angustifolium	12.649	0.000	0.002
Homogyne_alpina	-0.804	1.000	1.000
Juncus_trifidus	-0.732	1.000	1.000
Leontodon_helveticus	1.711	0.083	0.431
Leucanthemopsis_alpina	0.863	0.183	0.879
Linaria_alpina	8.335	0.000	0.008
Lotus_corniculatus	NA	NA	NA
Luzula_alpinopilosa	NA	NA	NA
Luzula_lutea	NA	NA	NA
Luzula_spicata	NA	NA	NA
Minuartia_recurva	11.466	0.000	0.002
Myosotis_alpestris	5.153	0.003	0.027
Oxyria_digyna	14.007	0.000	0.002
Pedicularis_kernerii	NA	NA	NA
Phleum_rhaeticum	2.460	0.062	0.345
Phyteuma_hemisphaericum	3.253	0.034	0.218
Poa_alpina	-0.532	0.666	1.000
Ranunculus_glacialis	NA	NA	NA
Ranunculus_kuepferi	NA	NA	NA
Saxifraga aizoides	7.001	0.002	0.022
Saxifraga_bryoides	-0.445	0.570	1.000
Senecio_incanus	NA	NA	NA
Silene_aucaulis	3.897	0.017	0.122
Trifolium_badium	0.799	0.171	0.853
Trifolium_pallesces	10.203	0.000	0.004
Tussilago_farfara	10.052	0.000	0.002
Veronica_bellidioides	2.071	0.064	0.345
	late:mid.SES	late:mid.pvalue	late:mid.adj.pvalue
Achillea_moschata	12.897	0.000	0.003
Adenostyles_leucophylla	NA	NA	NA
Agrostis_rupestris	NA	NA	NA
Agrostis_schraderiana	NA	NA	NA
Antennaria_dioica	-0.680	1.000	1.000
Anthoxanthum_odoratum	-0.723	1.000	1.000
Arabis_alpina	-0.556	1.000	1.000
Avenula_versicolor	1.069	0.141	1.000
Cardamine_bellidifolia	NA	NA	NA
Cardamine_resedifolia	NA	NA	NA
Carex_curvula	11.245	0.000	0.007
Carex sempervirens	12.439	0.000	0.003
Cerastium_uniflorum	-0.009	0.566	1.000
Epilobium_fleischeri	NA	NA	NA
Erigeron_uniflorus	2.804	0.042	0.356
Euphrasia_minima	NA	NA	NA
Festuca_halleri	-0.590	0.677	1.000
Gnaphalium_supinum	-0.158	0.582	1.000
Hieracium_angustifolium	4.161	0.009	0.101
Homogyne_alpina	9.392	0.001	0.015
Juncus_trifidus	5.870	0.006	0.068
Leontodon_helveticus	0.774	0.178	1.000
Leucanthemopsis_alpina	5.229	0.004	0.048
Linaria_alpina	0.564	0.316	1.000
Lotus_corniculatus	NA	NA	NA
Luzula_alpinopilosa	NA	NA	NA
Luzula_lutea	5.951	0.011	0.106

Luzula_spicata	NA	NA	NA
Minuartia_recurva	-0.144	0.428	1.000
Myosotis_alpestris	3.375	0.016	0.148
Oxyria_digyna	0.003	0.565	1.000
Pedicularis_kernerii	NA	NA	NA
Phleum_rhaeticum	0.796	0.304	1.000
Phyteuma_hemisphaericum	-0.045	0.502	1.000
Poa_alpina	11.690	0.000	0.003
Ranunculus_glacialis	NA	NA	NA
Ranunculus_kuepferi	NA	NA	NA
Saxifraga aizoides	NA	NA	NA
Saxifraga_bryoides	6.754	0.001	0.015
Senecio_incanus	NA	NA	NA
Silene_aucaulis	0.832	0.177	1.000
Trifolium_badium	6.516	0.001	0.021
Trifolium_pallescens	16.158	0.000	0.003
Tussilago_farfara	-0.328	0.809	1.000
Veronica_bellidioides	0.780	0.239	1.000

Here again, NAs are present in the data frame above for species that were associated to non-significant test in object `Qspecies_adj` (where tests were done over all groups); and also, for species that are absent from the two compared groups of plots (e.g. species *Adenostyles leucophylla*, in mid- and late-successional stages).

References

- Pavoine, S. 2020. `adiv`: an R package to analyse biodiversity in ecology. *Methods in Ecology and Evolution* 11: 1106–1112.
- Thioulouse, J., Dray, S., Dufour, A.-B., Siberchicot, A., Jombart, J. & Pavoine, S. 2018. *Multivariate Analysis of Ecological Data with ade4*. Springer, New York.

Supporting information to the paper

Ricotta, C. et al. A new method for indicator species analysis in the framework of multivariate analysis of variance. *Journal of Vegetation Science*.

Appendix S4. Overview of MRPP and its relationship with db-MANOVA

Given a community composition matrix with the abundances of S species ($s = 1, 2, \dots, S$) in N plots ($j = 1, 2, \dots, N$), let the number of plots in group k ($k = 1, 2, \dots, K$) be N_k such that $\sum_{k=1}^K N_k = N$ and δ_{ij} be any coefficient of compositional dissimilarity (i.e. not necessarily A-decomposable) between plot i and plot j such that $\delta_{ij} = \delta_{ji}$ and $\delta_{ii} = 0$.

Mielke et al. (1976) proposed a class of multivariate permutation methods called multiresponse permutation procedures (MRPP) for testing for compositional differences among a-priori defined groups of plots. In MRPP, the test statistic is the weighted average dissimilarity within groups $\bar{\delta}$, given by:

$$\bar{\delta} = \sum_{k=1}^K C_k \xi_k \quad (\text{A1})$$

Geometrically, a smaller value of $\bar{\delta}$ denotes higher within-group concentration and larger between-group differences (Cai 2006). In Eq. A1, C_k is the weight for group k , such that $\sum_{k=1}^K C_k = 1$, and ξ_k is the average within-group distance for group k such that:

$$\xi_k = \frac{2}{N_k(N_k - 1)} \sum_{i>j}^N \delta_{ij} \varepsilon_{ijk} \quad (\text{A2})$$

The indicator variable ε_{ijk} takes the value one if plots i and j are both in group k . Otherwise it is zero.

Mielke and Berry (2001) provide a throughout discussion of the choice of group weights. If $C_k = N_k / N$ we have:

$$\bar{\delta} = \sum_{k=1}^K C_k \xi_k = \frac{2N}{(N_k - 1)} \sum_{i>j}^N \delta_{ij} \varepsilon_{ijk} \quad (\text{A3})$$

If $C_k = (N_k - 1) / (N - K)$ we have:

$$\bar{\delta} = \sum_{k=1}^K C_k \xi_k = \sum_{k=1}^K \frac{2}{N - K} \times \frac{1}{N_k} \sum_{i>j}^N \delta_{ij} \varepsilon_{ijk} \quad (\text{A4})$$

The term $2/(N-K)$ in Eq. A4 is not necessary in randomization tests, as it remains constant. Therefore, for $C_k = (N_k - 1)/(N - K)$, $\bar{\delta}$ in Eq. A4 is functionally equivalent to the within-group sum of non-squared dissimilarities Q_w in Eq. 9 of the main text.

According to Eq. A4, if the plot-to-plot dissimilarities δ_{ij} are A-decomposable, the test statistics $\bar{\delta}$ can be additively decomposed into the contributions of single species, thus enabling to identify the indicator species that contribute most to the compositional differences among the groups with the same dissimilarity measure used for MRPP.

References

- Cai, L. 2006. Multi-response permutation procedure as an alternative to the analysis of variance: An SPSS implementation. *Behavior Research Methods* 38: 51–59.
- Mielke, P.W. & Berry, J.A. 2001. *Permutation Methods: A Distance Approach*. Springer, New York.
- Mielke, P.W., Berry, K.J. & Johnson, E.S. 1976. Multi-response permutation procedures for a priori classifications. *Communications in Statistics: Theory & Methods* 5: 1409–1424.

Cardamine bellidifolia	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0																		
Cardamine resedifolia	0	1	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	1
0	1	1	0	1	1	1	0	0	0	0	0	0	0	0	0	0	1	1	1
0	0	0	1	0	1	1	0	0	0	0	1	0	1	1	0	0	0	0	0
0	0																		
Carex curvula	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	1	0	1	1	1	0	1	0	0	0	0
0	0	0	0	1	0	0	0	0	1	1	0	0	1	1	0	0	3	5	1
2	3																		2
Carex sempervirens	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	4	0	0	3	2	3	2	0	0	2	3
2	2																		
Cerastium uniflorum	0	1	1	0	1	1	0	0	3	1	1	1	1	1	1	0	2	2	2
0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
0	0																		
Epilobium fleischeri	0	0	0	0	0	0	0	1	0	0	2	1	0	0	0	0	1	0	2
0	0	0	0	0	2	2	0	0	0	0	0	0	0	1	2	0	0	2	2
0	0	0	2	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0																		
Erigeron uniflorus	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	1	2	2	1	0	0	1	3	2	2	2	0	0	2	1	1	0	0	0
0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0																		
Euphrasia minima	1	1	2	1	1	0	1	0	0	2	1	0	0	0	0	0	0	0	0
1	0	0	0	1	1	0	0	0	0	1	0	1	0	1	1	1	0	1	1
1	0	1	0	0	0	0	1	1	1	1	0	0	0	0	1	0	0	0	0
0																			
Festuca halleri	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	1	0	0	0	0	1	0	2	0	2	3	1	0	0	0	0	0
3	2	2	0	5	0	2	3	2	0	3	0	1	0	0	2	2	0	1	0
2	0																		
Gnaphalium supinum	0	1	2	0	1	1	1	0	2	1	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1

[illegible]

[illegible]

[illegible]

```
dbMANOVASpecies <- function(comm, groups, nrep = 999, method =
c("Euclidean", "Manhattan", "Canberra", "BrayCurtis"), global = TRUE,
species = TRUE, padjust = "none", tol = 1e-8){
```

```
  global <- global[1]
  species <- species[1]
  if(!inherits(global, "logical")) stop("Incorrect definition of
argument global")
  if(!inherits(species, "logical")) stop("Incorrect definition of
argument species")
  if(!species) global <- TRUE
  if(!inherits(comm, "data.frame") && !inherits(comm, "matrix"))
stop("comm must be a data frame or a matrix")
  if(is.null(colnames(comm))) colnames(comm) <- paste("species",
1:ncol(comm), sep="")
  if(is.null(rownames(comm))) rownames(comm) <- paste("community",
1:nrow(comm), sep="")
  if(any(comm < -tol)) stop("comm must have nonnegative values")
  if(any(colSums(comm) < tol)){
    comm <- comm[, colSums(comm) > tol, drop=FALSE]
    warning("species with zero abundance over all considered
communities have been removed")
  }
  if(nrow(comm) != length(groups)) stop("The length of argument groups
should be equal to the number of rows of argument comm")
  if(any(rowSums(comm) < tol)){
    comm2 <- comm[rowSums(comm) > tol, , drop=FALSE]
    groups <- groups[rowSums(comm) > tol]
    comm <- comm2
    warning("empty communities with zero species abundance have been
removed")
  }
  comm[comm < 0] <- 0
  if(!inherits(groups, "factor") && !inherits(groups, "character"))
stop("Incorrect definition for argument groups")
  if(!inherits(groups, "factor") | length(levels(groups)) !=
length(unique(groups)))
    groups <- factor(groups)
  method <- method[1]
  if(!method%in%c("Euclidean", "Manhattan", "Canberra", "BrayCurtis"))
stop("Incorrect definition of argument method")
  Nplots <- nrow(comm)
```

```
if(!species){
```

```
distabund <- function (abund, meth)
{
```

```
  d <- matrix(0, nrow(abund), nrow(abund))
  funEuclidean <- function(x) {
    sum((abund[x[1], ] - abund[x[2], ])^2)
  }
  funManhattan <- function(x) {
    sum(abs(abund[x[1], ] - abund[x[2], ]))
  }
  funCanberra <- function(x) {
    sum(abs(abund[x[1], ] - abund[x[2], col]))/(abund[x[1], ] +
abund[x[2], ]), na.rm=TRUE)
  }
}
```

```

funBrayCurtis <- function(x) {
  sum(abs(abund[x[1], ] - abund[x[2], ])/sum(abund[c(x[1],x[2]), ]))
}
index <- cbind(col(d)[col(d) < row(d)], row(d)[col(d) < row(d)])
if (meth == "Euclidean")
  d <- unlist(apply(index, 1, funEuclidean))
else if (meth == "Manhattan")
  d <- unlist(apply(index, 1, funManhattan))
else if (meth == "Canberra")
  d <- unlist(apply(index, 1, funCanberra))
else
  d <- unlist(apply(index, 1, funBrayCurtis))
attr(d, "Size") <- nrow(abund)
attr(d, "Diag") <- FALSE
attr(d, "Upper") <- FALSE
attr(d, "method") <- meth
class(d) <- "dist"
return(d)
}

distances <- distabund(comm, method)
funSTATs <- function(dis, gro){
  QT <- sum(as.vector(dis))/Nplots
  Distances <- as.matrix(dis)
  ldistancesW <- lapply(levels(gro), function(x) Distances[gro==x,
gro==x])
  ldistancesW <- lapply(ldistancesW, as.dist, diag = FALSE, upper =
FALSE)
  lQW <- lapply(ldistancesW, function(x)
sum(as.vector(x)/attributes(x)$Size))
  QW <- sum(as.vector(unlist(lQW)))
  QB <- QT - QW
  res <- c(QT, QB, QW)
  names(res) <- c("QT", "QB", "QW")
  return(res)
}
funperm <- function(i){
  egroups <- sample(groups)
  theo <- funSTATs(distances, egroups)[2]
  return(theo)
}
THEO <- sapply(1:nrep, funperm)
OBS <- funSTATs(distances, groups)[2]
simu <- as.randtest(sim = THEO, obs = OBS, alter = "greater")
tabobs <- funSTATs(distances, groups)
}

else{
distabund <- function (abund, col, meth)
{
  d <- matrix(0, nrow(abund), nrow(abund))
  funEuclidean <- function(x) {
    (abund[x[1], col] - abund[x[2], col])^2
  }
  funManhattan <- function(x) {
    abs(abund[x[1], col] - abund[x[2], col])
  }
}

```

```

funCanberra <- function(x) {
  if((abund[x[1], col] + abund[x[2], col]) < tol) return(0)
  else
    return(abs(abund[x[1], col] - abund[x[2], col])/(abund[x[1], col]
+ abund[x[2], col]))
}
funBrayCurtis <- function(x) {
  if(is.null(attributes(abund)$internal.p.rowSums))
    abs(abund[x[1], col] - abund[x[2], col])/sum(abund[c(x[1],x[2]),
])
  else
    abs(abund[x[1], col] - abund[x[2], col])/sum(attributes(abund)$rowSums[c(x[1],x[2])])
}
index <- cbind(col(d)[col(d) < row(d)], row(d)[col(d) < row(d)])
if (meth == "Euclidean")
  d <- unlist(apply(index, 1, funEuclidean))
else if (meth == "Manhattan")
  d <- unlist(apply(index, 1, funManhattan))
else if (meth == "Canberra")
  d <- unlist(apply(index, 1, funCanberra))
else
  d <- unlist(apply(index, 1, funBrayCurtis))
attr(d, "Size") <- nrow(abund)
attr(d, "Diag") <- FALSE
attr(d, "Upper") <- FALSE
attr(d, "method") <- meth
class(d) <- "dist"
return(d)
}

listDIS <- lapply(1:ncol(comm), function(i) as.matrix(distabund(comm,
i, method)))
if(global) {
  distances <- listDIS[[1]]
  for(i in 2:length(listDIS)) distances <- distances +
listDIS[[i]]
  distances <- as.dist(distances, diag = FALSE, upper = FALSE)
}
listDIS <- lapply(listDIS, as.dist, diag = FALSE, upper = FALSE)
funSTATs <- function(dis, gro){
  QT <- sum(as.vector(dis))/Nplots
  Distances <- as.matrix(dis)
  ldistancesW <- lapply(levels(gro), function(x) Distances[gro==x,
gro==x])
  ldistancesW <- lapply(ldistancesW, as.dist, diag = FALSE, upper =
FALSE)
  lQW <- lapply(ldistancesW, function(x)
sum(as.vector(x)/attributes(x)$Size))
  QW <- sum(as.vector(unlist(lQW)))
  QB <- QT - QW
  res <- c(QT, QB, QW)
  names(res) <- c("QT", "QB", "QW")
  return(res)
}
funperm <- function(i){
  egroups <- sample(groups)
  if(global) theoglobal <- funSTATs(distances, egroups)[2]
  theospecies <- as.vector(unlist(lapply(listDIS, function(dd)
funSTATs(dd, egroups)[2])))

```



```

        if(global) theo <- c(theoglobal, theospecies)
        else theo <- theospecies
        return(theo)
    }
    THEO <- t(cbind.data.frame(sapply(1:nrep, funperm)))
    if(global) colnames(THEO) <- c("GLOBAL", colnames(comm))
    else colnames(THEO) <- colnames(comm)
    obsspecies <- as.vector(unlist(lapply(listDIS, function(dd)
funSTATs(dd, groups)[2])))
    if(global){
        obsglobal <- funSTATs(distances, groups)[2]
        OBS <- c(obsglobal, obsspecies)
    }
    else OBS <- obsspecies
    simu <- as.krandtest(sim = THEO, obs = OBS, alter = "greater",
p.adjust.method = padjust)
    obsspecies <- cbind.data.frame(lapply(listDIS, function(dd)
funSTATs(dd, groups)))
    colnames(obsspecies) <- colnames(comm)
    if(global){
        obsglobal <- funSTATs(distances, groups)
        tabobs <- cbind.data.frame(Global = obsglobal, obsspecies)
    }
    else tabobs <- obsspecies
}
    RES <- list(observations = tabobs, test = simu, call = match.call(),
method = method, padjust = padjust, tol = tol)
    class(RES) <- "dbMANOVASpecies"
    return(RES)
}

```

```

dbMANOVASpecies_pairwise <- function(dboobj, signif = TRUE, salpha = 0.05,
nrep = NULL){

```

```

    tol <- dboobj$tol
    salpha <- salpha[1]
    signif <- signif[1]
    if(salpha > 1 | salpha < 0) stop("salpha must be a numeric value between
0 and 1")
    if(!inherits(signif, "logical")) stop("Incorrect definition of
argument signif")
    if(!inherits(dboobj, "dbMANOVASpecies")) stop("Argument dboobj must be
of class dbMANOVASpecies")
    padjust <- dboobj$test$adj.method
    if(is.null(nrep)) nrep <- max(dboobj$test$rep)
    LAAA <- as.list(dboobj$call)
    comm <- eval.parent(LAAA$comm)
    sauvcomm <- comm
    groups <- eval.parent(LAAA$groups)

    if(is.null(colnames(comm))) colnames(comm) <- paste("species",
1:ncol(comm), sep="")
    if(is.null(rownames(comm))) rownames(comm) <- paste("community",
1:nrow(comm), sep="")
    if(any(colSums(comm) < tol)){
        comm <- comm[, colSums(comm) > tol, drop=FALSE]
    }
    if(any(rowSums(comm) < tol)){
        comm2 <- comm[rowSums(comm) > tol, , drop=FALSE]
        groups <- groups[rowSums(comm) > tol]
    }

```

```

    comm <- comm2
  }
  comm[comm < 0] <- 0
  if(!inherits(groups, "factor") | length(levels(groups)) !=
length(unique(groups)))
    groups <- factor(groups)

  method <- dbobj$method

  if(length(dbobj$test$pvalue) ==1){
    species <- FALSE
    global <- TRUE
    signif <- FALSE
  }
  else{
    species <- TRUE
    if(length(dbobj$test$pvalue) > ncol(comm))
      global <- TRUE
    else
      global <- FALSE
  }
  if(method == "BrayCurtis" && signif){
    if(global) pval <- dbobj$test$pvalue[-1]
    else pval <- dbobj$test$pvalue
    if(all(pval > salpha)) stop("None of the species tests was
significant")
    commE <- comm[, pval <= salpha]
    attributes(commE)$internal.p.rowSums <- rowSums(comm)
    comm <- commE
  }
  if(method != "BrayCurtis" && signif){
    if(global) pval <- dbobj$test$pvalue[-1]
    else pval <- dbobj$test$pvalue
    if(all(pval > salpha)) stop("None of the species tests was
significant")
    comm <- comm[, pval <= salpha]
  }
  FF <- levels(groups)
  combFF <- combn(FF, 2)
  if(!species) {
    FUNcomb <- function(i){
      commC <- sauvcomm[groups%in%combFF[,i], ]
      groupsC <- factor(groups[groups%in%combFF[,i]])
      return(dbMANOVAspecies(commC, groupsC, nrep = nrep, method =
method, global = TRUE, species = FALSE, padjust = padjust, tol = tol))
    }
    simuglobal <- lapply(1:ncol(combFF), FUNcomb)
    names(simuglobal) <- apply(combFF, 2, function(x) paste(x,
collapse=":"))
    simu <- simuglobal
    class(simu) <- "dbMANOVAspecies_pairwise"
    return(simu)
  }
  if(!signif | !global) {
    FUNcomb <- function(i){
      commC <- comm[groups%in%combFF[,i], ]
      groupsC <- factor(groups[groups%in%combFF[,i]])
      return(dbMANOVAspecies(commC, groupsC, nrep = nrep, method =
method, global = global, species = TRUE, padjust = padjust, tol = tol))
    }
  }

```

```

simu <- lapply(1:ncol(combFF), FUNcomb)
names(simu) <- apply(combFF, 2, function(x) paste(x, collapse=":"))
}
else {
FUNcomb <- function(i){
  commC <- comm[groups%in%combFF[,i], ]
  groupsC <- factor(groups[groups%in%combFF[,i]])
  return(dbMANOVASpecies(commC, groupsC, nrep = nrep, method =
method, global = FALSE, species = TRUE, padjust = padjust, tol = tol))
}
simuspecies <- lapply(1:ncol(combFF), FUNcomb)
names(simuspecies) <- apply(combFF, 2, function(x) paste(x,
collapse=":"))
FUNcomb <- function(i){
  commC <- sauvcomm[groups%in%combFF[,i], ]
  groupsC <- factor(groups[groups%in%combFF[,i]])
  return(dbMANOVASpecies(commC, groupsC, nrep = nrep, method =
method, global = TRUE, species = FALSE, padjust = padjust, tol = tol))
}
simuglobal <- lapply(1:ncol(combFF), FUNcomb)
names(simuglobal) <- apply(combFF, 2, function(x) paste(x,
collapse=":"))
simu <- list(global_test = simuglobal, per_species_test = simuspecies)
}
attributes(simu)$species.names <- colnames(sauvcomm)
class(simu) <- "dbMANOVASpecies_pairwise"
return(simu)
}

```

```

summary.dbMANOVASpecies_pairwise <- function(dbspair, DIGITS = 3){

  if(!inherits(dbspair, "dbMANOVASpecies_pairwise")) stop("Argument
dbspair must be of class dbMANOVASpecies_pairwise")
  DIGITS <- DIGITS[1]

```

```

  funsum <- function(LList) {
    TT <- LList$test
    if(length(TT$pvalue)>1){
      if(!is.null(TT$adj.pvalue) && !TT$adj.method%in%"none")
      TAB <- as.data.frame(t(cbind.data.frame(SES = TT$expvar[, 1],
pvalue = TT$pvalue, adj.pvalue = TT$adj.pvalue)))
    }
    else
      TAB <- as.data.frame(t(cbind.data.frame(SES = TT$expvar[, 1],
pvalue = TT$pvalue)))
    colnames(TAB) <- TT$names
    return(TAB)
  }
  else{
    VEC <- c(TT$expvar[1], TT$pvalue)
    names(VEC) <- c("SES", "pvalue")
    return(VEC)
  }
}
funsumglobal <- function(LList) {
  TT <- LList$test
  VEC <- c(SES = TT$expvar[1], pvalue = TT$pvalue)
  return(VEC)
}

```

```

if(names(dbspair)[1] == "global_test"){

```

```

Npairs <- length(dbspair[[2]])
LLspecies <- dbspair[[2]]
RESSpecies <- lapply(LLspecies, funsum)
LLglobal <- dbspair[[1]]
RESglobal <- lapply(LLglobal, funsumglobal)
if(nrow(RESSpecies[[1]]) == 3)
  RESglobal <- lapply(RESglobal, function(x) c(x, "NA"))
RES <- list()
for(i in 1:length(RESglobal)){
  res <- cbind.data.frame(GLOBAL = RESglobal[[i]],
RESSpecies[[i]])
  rownames(res) <- rownames(RESSpecies[[i]])
  RES[[i]] <- res
}
names(RES) <- names(RESSpecies)
}
else {
  Npairs <- length(dbspair)
  RES <- lapply(dbspair, funsum)
}

if(inherits(RES[[1]], "data.frame")){
  if(colnames(RES[[1]])[1] == "GLOBAL")
    spenames <- c("GLOBAL", attributes(dbspair)$species.names)
  else
    spenames <- attributes(dbspair)$species.names
  FUNttab <- function(TTAB){
    ttab <- as.data.frame(matrix(NA, length(spenames),
nrow(RES[[1]])))
    rownames(ttab) <- spenames
    colnames(ttab) <- rownames(RES[[1]])
    ttab[colnames(TTAB), rownames(TTAB)] <- t(TTAB)
    return(ttab)
  }
  RES <- lapply(RES, FUNttab)
  nniv <- length(RES)
  ncol <- ncol(RES[[1]])
  nam1 <- names(RES)
  nam2 <- colnames(RES[[1]])
  nam1 <- rep(nam1, rep(ncol, nniv))
  nam2 <- rep(nam2, nniv)
  RES <- cbind.data.frame(RES)
  colnames(RES) <- paste(nam1, nam2, sep=".")
}
if(!inherits(RES[[1]], "data.frame")){
  RES <- cbind.data.frame(RES)
}
if(!is.null(DIGITS)){
  RES[!is.na(RES)] <- round(RES[!is.na(RES)], digits = DIGITS)
}
return(RES)
}

```