



How should a dendrogram-based measure of functional diversity function? A rejoinder to Petchey and Gaston

János Podani and Dénes Schmera

J. Podani (podani@ludens.elte.hu), Dept of Plant Taxonomy and Ecology, Biological Inst., Eötvös Loránd Univ., Pázmány P. s. 1/c, HU-1117 Budapest, Hungary. – D. Schmera, Dept of Zoology, Plant Protection Inst., Hungarian Academy of Sciences, POB 102, HU-1525 Budapest, Hungary.

This paper is a response to Petchey and Gaston's criticism of our previous paper on the measurement of dendrogram-based functional diversity. In contrast to their suggestions, we insist that Euclidean distance is unsuitable to the analysis of variables with mixed scale types and maintain our earlier view that both the distance function and the clustering method can influence the dendrogram-based measure of functional diversity. We propose an extension of Gower's formula to accommodate nominal traits with non-exclusive categories and emphasize the necessity of the methodological standardization of functional diversity measures.

Petchey and Gaston (this issue) comment on dendrograms and derived functional diversity (FD) measures, as a response to our criticism (Podani and Schmera 2006) of their earlier work summarized in Petchey and Gaston (2002). The purpose of these authors is "... to refute, clarify, and provide more appropriate solutions to a number of issues ..." we raised. We feel, however, that while some of the new proposals contain useful elements and are welcome, refutation of most of our points fails and the way they clarify the problems increases, rather than diminishes the methodological confusion. Therefore, some comments on their paper are in order.

Comparison of the incomparable

Petchey and Gaston are fair enough to acknowledge at least implicitly the inappropriateness of Euclidean distance in the analysis of functional variables with mixed scale types. Nevertheless, with the obvious purpose of defending their old and bad strategy, they redraw the dendrogram for the arctic vegetation data (Chapin et al. 1996) which is created in a manner still unknown to us. Then, they try to show that correct analysis by the Gower measure adds relatively "minor" information and that general conclusions are insensitive to the distance

measure used. However, their attitude towards this issue is potentially dangerous and carries a bad message for the scientific community, for several reasons. 1) The fundamental problem is that results produced by an admittedly unsuitable method do not make sense, cannot justify any scientific argumentation and should not be compared to anything else to validate a bad approach. Ecology is often considered by many as "soft science", and the criticism is right when such inadmissible evaluations appear in the ecological literature. 2) Technical details of the computations are presented neither in Petchey and Gaston (2002) nor in their recent commentary, so that the results cannot be reproduced and tested by anyone else which would otherwise be an essential criterion in scientific research. We do not know how they handled nominal and ordinal variables and missing values in the data when calculating Euclidean distances – which would be interesting to see even if such calculations can only be incorrect. We therefore repeat our warnings that nominal and ordinal variables cannot be standardized in the classical way and cannot be subjected to metric computations. Details of calculating the Gower formula are also missing, so the reader cannot appreciate how Petchey and Gaston handled the single presence–absence variable (double zeros considered or disregarded?) and how they incorporated the six ordinal variables present in the arctic

vegetation dataset. The program package they used for clustering is not mentioned. Not surprisingly, we were therefore unable to reproduce exactly the results in Petchey and Gaston's Fig. 1a and 1b, although we tried several combinations of input options. 3) If we forget about the above reservations and insist on comparing the results, we find that the conclusions by Petchey and Gaston are not entirely right. The dendrograms of Fig. 1a and 1b, no matter how obtained, appear radically different in shape at first sight, although there are some agreements in minor details. The trends, displayed by the diagrams on the right of Fig. 1a and 1b, i.e. that FD, presumably standardized by the maximum, increases along with species richness, would be probably similarly monotonous for even more differing dendrograms. But the picture is in fact not as simple as suggested, because the Gower measure produced apparently smaller variances than Euclidean distance, an observation that escaped the attention of the authors (this figure was then corrected before printing). So, we do not understand the reply by Petchey and Gaston that the choice of the method "... may have little effect on quantitative patterns in FD." In their example, both the dendrograms and the scattergrams exhibit striking differences so the only pattern they may refer to is the monotonous trend, which is less surprising anyway. 4) We cannot interpret the scattergrams of Fig. 1c simply because no captions to vertical axes are given, so the reader can only guess what is going on there. The top diagram appears to be a graphical comparison between the Gower formula and the Euclidean distance, this time without standardization by the maxima (around 15 and 10, respectively). If this is the case, then the differences between the two coefficients are not negligible at all. The diagram on the bottom applies standardization again, making its interpretation difficult not only because the caption on the vertical axis is missing. In general, comparisons are hard to follow when the scale oscillates between the raw and standardized ones! 5) The statement that results are insensitive "... in fact also to the clustering algorithm ..." is not substantiated at all, the authors did not demonstrate this, leaving the readers totally unconvinced. In fact, the choice of the clustering method may or may not have substantial effects, which is never known a priori unless a mathematical proof is given. Table 2 in our paper (Podani and Schmera 2006) clearly shows that FD decreases when we shift from complete linkage (CL) to UPGMA and then to single linkage (SL).

The suggestion that the choice of distance coefficient (when measurement scales are uniform) and the cluster-

ing method is immaterial because it is the cophenetic correlation coefficient that needs to be maximized also deserves attention. Yes, it is quite possible that the highest cophenetic correlation is not the one produced by UPGMA for a given problem. This is not simply because there are some noted alternatives (centroid, median, SL, CL) but because there is an infinite series of distance optimizing clustering procedures available through the general strategy suggested by Lance and Williams (1966). Therefore, maximization would always require a complete screening of results derived by this general algorithm, of which UPGMA, SL and CL are only special cases, although absolute optima could not be found easily because the controlling parameter of this general procedure is continuous. That is, maximization of cophenetic correlation would overcomplicate the calculations unnecessarily. The same holds true for the choice of the distance coefficient, of which we can also imagine an infinite number of possibilities, such as those offered by the general formula of Faith (1984). That is, the theoretically well-sound requirement of maximizing the cophenetic correlation coefficient is not feasible in practice and standard use of the same clustering method (such as UPGMA) with Euclidean distance (when appropriate) or Gower formula (when appropriate) is preferable. In their abstract, Petchey and Gaston reject our suggestion to use the above combination of distances and clustering procedure, whereas in the next sentence they argue that the methods "... may have little effect on quantitative patterns in FD". But then, why worry so much about maximizing the cophenetic correlation?

Multiplicity of character states

We agree that categories of a nominal variable may often be non-exclusive, that is, a species may be characterized by the simultaneous presence of two or more character states. Petchey and Gaston propose that in such cases information is more appropriately coded by as many independent binary variables as the number of categories. The problem with this suggestion is that such functional variables would be overweighted in the analysis relative to those not having character state multiplicity. The value of a distance coefficient, however, should not be influenced by differences in the number of categories between the functional variables. A more elegant and very simple solution is therefore as follows. For a pair of species *j* and *k*, the disagreement in the given trait *i* with non-exclusive or exclusive states is given by

$$s_{ijk} = 1 - \frac{\text{number of categories characterizing both species } j \text{ and } k}{\text{number of categories characterizing at least one of the two species}}$$

This formula replaces Eq. 2–3 in Podani and Schmera (2006) so that character state multiplicity is now accommodated by the Gower formula without giving excess weight to any trait. s_{ijk} is 1.0 if the two species do not share any character state, whereas it becomes 0.75, for example, if the first species employs foraging strategies A and B, while the second one employs B, C and D. Alternative solutions to this problem, when character state multiplicity is expressed by fuzzy coding, rather than by presences and absences, are found in Bady et al. (2005), which can also be made compatible with the Gower formula easily.

The number of traits

The question whether the number of functional traits influences FD is considered irrelevant by Petchey and Gaston. This may be generally true that two different assemblages described in terms of different numbers of traits deserve no comparison, but our examples showed how FD changes when the same assemblage is analyzed with one trait removed at a time. We maintain that this is a meaningful ecological question, because in this way we can identify traits that are most influential in determining FD of a given community. Of course, our purpose with the comparison of the FD of Patagonian forbs with the FD of insectivorous birds was not to ask “. . . whether a kilometre is longer than an hour”, as Petchey and Gaston interpret the issue. In fact, we demonstrated via this example the lack of positive relationship of Gower-distance-based FD and the number of species, rather than the number of traits, so that Petchey and Gaston entirely misunderstood this part of our paper. (Actually, the Patagonian forbs data comprise 6 traits, the bird data include 24, and FD was higher for the bird data in which, however, the number of species is smaller than in the Patagonian forbs dataset.)

Species number

We are glad to see that the improved version of dendrogram-based FD is freed from the problems the old concept suffered from. It is especially welcome that zero diversity is obtained for the single-species case and we do not have to consider communities “without species”. The set monotonicity criterion is met by this new measure, but only if examined in “downwards” direction! There is a starting dendrogram derived for a given set of species, and species removals can only decrease – or leave unchanged – the total of remaining branch lengths (Ricotta 2005). This is an obvious advantage over our procedure which does not satisfy this requirement. However, the set monotonicity criterion

assumes in general that diversity should not decrease by the addition of a new species to a given species set. As we understand, the new FD does not apply to species additions, and therefore cannot be adapted to situations in which new species appear in a community. Thus, the set monotonicity condition only partially holds. The price that must be paid to guarantee set monotonicity is high because fundamental species pools, which can only be reduced in size according to the new definition of FD, are not always readily available in ecological studies. Also, there is good reason to assume that the functionality of a community may change considerably when, for example, a new invasive species occurs. Recomputing dendrograms may therefore still be a useful strategy in more realistic situations without a predefined species pool such that both species extinction and immigration are allowed. Petchey and Gaston’s argument that “. . . it is not the use of total branch length, but the recalculation of the functional dendrogram for each assemblage, that causes the violation of set monotonicity” by our method is merely a play with words, because total branch length is the statistic and the recalculation of the dendrogram is the process that causes change in this statistic, and these two are not separable in our case.

Conclusions

We maintain that measurement of functional diversity via dendrograms is a complex matter; data treatment, distance function and clustering method used are all potentially influential or even decisive, and that methodological standardization is still possible and necessary. It is difficult to find a procedure such that all mathematical requirements are met at the same time, especially if one thinks not only in terms of species extinction as the only factor affecting species richness. Surely enough, set monotonicity is not a condition that is superior to compatibility of distances with measurement scales. Correct answers to particular ecological questions require careful methodological thinking which does not mean that statistics is more important than ecology. If the ecologist gives more thought to statistical aspects, however, the statistical methods may only better serve the purposes of our field.

Acknowledgements – The financial support to J.P. in form of a Hungarian Research Grant, OTKA T43732 is gratefully acknowledged.

References

- Chapin, F. S. I. et al. 1996. Plant functional types as predictors of transient responses of arctic vegetation to global change. – *J. Veg. Sci.* 7: 347–358.

- Bady, P. et al. 2005. Use of invertebrate traits for the biomonitoring of European large rivers: the effects of sampling effort on genus richness and functional diversity. – *Freshwater Biol.* 50: 159–173.
- Faith, D. P. 1984. Patterns of sensitivity of association measures in numerical taxonomy. – *Math. Biosci.* 69: 199–207.
- Lance, G. N. and Williams, W. T. 1966. A generalized sorting strategy for computer classifications. – *Nature* 212: 218.
- Petchey, O. L. and Gaston, K. J. 2002. Functional diversity (FD), species richness and community composition. – *Ecol. Lett.* 5: 402–411.
- Podani, J. and Schmera, D. 2006. On dendrogram-based measures of functional diversity. – *Oikos* 115: 179–185.
- Ricotta, C. 2005. A note on functional diversity measures. – *Basic Appl. Ecol.* 6: 479–486.