

*New Idea***To dendrogram or not? Consensus methods show that is the question needed to move functional diversity metrics forward****Mark S. Poesch**

Mark S. Poesch (poesch@ualberta.ca), University of Alberta, Department of Renewable Resources, 751 General Services Road, Edmonton, Alberta, T6G 2H1

Abstract

Functional diversity indices have become important tools for measuring variation in species characteristics that are relevant for ecosystem services. A frequently used dendrogram-based method for measuring functional diversity, ‘FD’, was shown to be sensitive to methodological choices in its calculation, and consensus methods have been suggested as an improvement. The objective of this study was to determine whether consensus methods can be used to reduce sensitivity when measuring FD. To calculate FD, a distance measure and a clustering method must be chosen. Using data from three natural communities, this study demonstrates that consensus methods were unable to resolve even simple choices of distance measure (Euclidean and cosine) and clustering method (UPGMA, complete and single linkage). Overall, there was low consensus, ranging from 41–45%, across choices inherent in functional diversity. Further, regardless of how FD was measured, or how many species were removed from the community, FD closely mirrored species richness. Future research on the impact of methodological choices, including choices inherent in producing a dendrogram and the statistical complications they produce, are needed to move functional diversity metrics forward.

Keywords: functional diversity; multivariate analysis; dendrogram; distance measures; ecosystem science; community ecology.

Introduction

Functional diversity is the amount of variation of functional units (e.g. traits) in multi-dimensional space (Villéger et al 2008). There is a growing consensus that functional diversity, and not species diversity, is likely to be the component of biodiversity most relevant to ecosystem function (Tilman 1997, Cadotte et al. 2011). Studies on functional diversity have concluded that ecosystem function tends to correlate more strongly with functional diversity indices than with species diversity indices (Loreau et al. 2001, Petchey and Gaston 2006). These results have led to a growing need to develop robust methodologies for quantifying functional diversity (Walker et al. 2008, Poos et al. 2009).

A frequently used index, known as dendrogram-based functional diversity, or more commonly as ‘FD’, measures functional diversity as the total branch length of a functional dendrogram (Petchey and Gaston 2002). To produce a dendrogram several decisions need to be made (Poos et al. 2009). First, the number and type of traits important to ecosystem function need to be identified. Second, a distance measure needs to be chosen that characterizes the relative differences among species based on their traits. Finally, a clustering algorithm is needed to produce a dendrogram that hierarchically segregates species into functional groups based on their relative distances (Petchey and Gaston 2002). FD has been criticized for having the additional subjective step of clustering traits onto a dendrogram due to its sensitivity in producing replicable results (Poos et al. 2009). Although standard methods may provide one

way to reduce subjectivity, it is unlikely that a single distance measure or clustering algorithm can be used in all circumstances. Therefore, methods that reduce the sensitivity of functional diversity are continually being sought (Mouchet et al. 2008, Cadotte et al. 2011).

Despite the common use of functional diversity metrics, claims related to whether such metrics can be used to derive ecologically robust conclusions have been quantitatively evaluated recently (Podani and Schmera 2006, Poos et al. 2009), as have concerns regarding its effectiveness for determining ecosystem properties (Cadotte et al. 2011). Consensus methods, which synthesize dendrograms into a single classification based on similar topologies, have been suggested as a standardized approach for dealing with methodological uncertainties with FD (Mouchet et al. 2008). The objective of this study is to determine whether consensus methods can be used to reduce the sensitivity in the measure of FD. For this purpose, sensitivity is defined as the persistence of dendrogram topologies across methodological choices. This definition requires that the produced patterns of species groupings are maintained. If functional diversity is not robust in this sense, it would suggest that the conclusions regarding functional diversity may be limited by the methodological choices inherent in its calculation. Furthermore, explicit recognition of the effects of using a dendrogram, and the decisions needed to get there (e.g. choosing a distance measure and clustering algorithm) need to be better understood, so appropriate guidelines for making these decisions can be formulated.

Materials and Methods

In this study, three data sets are used as in previous studies of FD (Petchey and Gaston 2002, Podani and Schmera 2006, Poos et al. 2009). These datasets represent a variation in the number and type of species (from 11 to 22), and the number and type of functional traits (from 6 to 27; Holmes et al. 1979, Jaksic´ and Medel 1990, Munoz and Ojeda 1997). Unlike these previous studies (Podani and Schmera 2006; Petchey and Gaston 2002), data sets with mixed data types were removed as they are inappropriate for use with multiple distance measures (Podani and Schmera 2006).

The metric FD is based on the total branch length of a dendrogram of functional traits. To obtain this dendrogram, species traits must be assigned a distance (or resemblance) measure and clustering algorithm. Distance measures quantify the association between two entities based on their characteristics (e.g. species based on their functional traits). There are a large number of distance measures from which to choose depending on the data. Two distance measures previously used to assess the sensitivity of FD were used: Euclidean distance and cosine distance (Poos et al. 2009). Cosine

distance down-weights the potential over-fit created by co-varying traits (Legendre and Legendre 2012), a problem often encountered when analyzing functional traits of species (Petchey and Gaston 2006), whereas, Euclidean distance emphasizes larger values, in particular where positive covariance exists between traits (Poos et al. 2009). All trait matrices were standardized so that all traits had a mean = 0 and variance = 1 (Petchey and Gaston 2002, Petchey and Gaston 2006).

Variability in ecological data is often associated with just a few entities of which clustering into key groupings can provide insight. Three clustering algorithms were used in this analysis, unpaired pair group method with arithmetic mean (UPGMA), single linkage (i.e. nearest neighbor) and complete linkage (i.e. maximum or farthest neighbor). These algorithms represent natural endpoints across a methodological continuum of hierarchical clustering algorithms where single linkage lies on one end, complete linkage on the other and UPGMA lies somewhere in the middle (Podani and Schmera 2006, Poos et al. 2009).

Using Consensus Methods to Reduce Uncertainty when Measuring FD

To determine whether distance measure or clustering algorithm influenced dendrogram topologies of FD, a routine was developed (in MatLAB version 7.1) to randomize the removal of n species from the dataset and recalculate FD for each species combination, clustering algorithm and distance measure. Each level of n species was replicated 1000 times. FD was calculated at each species richness interval as the total distance of branches in the dendrogram. As FD measures the total branch lengths of a functional dendrogram, which relies on clustering method and distance measure, all dendrograms were rescaled to value between 0-1 using the full species model. The range in FD for the full species level at each different clustering method and distance measure was summarized (Petchey and Gaston 2002).

Each variant distance measure/clustering algorithm dendrogram was compared using consensus trees (Rohlf 1982, Mouchet et al. 2008). Dendrograms were compared using the consensus index $CI(C)$ (Rohlf 1982). Unlike cophenetic correlation (Mouchet et al. 2008), which compares a dendrogram to the distance measure, the consensus index compares the similarity of dendrograms based on their cluster membership (Rohlf 1982). The 50% majority rule consensus index was used where a value of one indicates all subgroups share at least 50% membership (i.e. the consensus tree is completely bifurcated indicating similar topology between the original trees) and a value of zero indicates no subgroups are shared (Rohlf 1982). Although a more strict measure of consensus can be used, the use of a 50% majority rule leads to a more liberal assessment of

the similarity between trees than a strict measure would provide.

Results

The Relationship between FD, Distance Measure & Clustering Algorithm

There was large variation in the measured amount of FD across methodological choices (Figure 1). Overall, conclusions regarding the qualitative relationship of FD among communities were not robust to methodological choices (i.e. FD was not consistent across methodological choices). FD ranged 34.2% on average across clustering algorithms (Range: 21–61%) at maximum species richness; while distance measure ranged 20.5% (Range 12–41%; Figure 1). There was a strong relationship between FD and species richness, regardless of the combination of distance measure/clustering algorithm used (Figure 1). When species were randomly removed from the assemblage, in all cases, there was a negative linear decrease in functional diversity.

Identifying Sensitivity in FD Using Consensus Methods

Dendrogram topologies showed little resemblance across distances measures (Table 1). The overall low value of the consensus $CI(C)$ indicated that the decision of choosing a distance measure influences the overall dendrogram to such an extent that there was little resemblance between the dendrogram based on Euclidean distance and the dendrogram based on cosine distance (Table 1; Figure 1 inset). When mean consensus values are compared, approximately 37–52% of the dendrogram groupings were concordant, depending on the distance measure or clustering algorithm used (Table 1). The clustering algorithm did not improve the similarity between functional topologies. For example, single linkage, UPGMA and complete linkage all showed similar rates of consensus tree resemblance, regardless of the size of the tree or the dataset used (Table 1). Overall, dendrogram topologies were not robust to the choice of distance measure or clustering algorithm, and there was little consensus across methods (Table 1).

Discussion

Functional diversity has become an important, but controversial focus of research at the boundary between community and ecosystem ecology (Tilman 1997, Loreau et al. 2001). To calculate most functional diversity indices, a method is required for quantifying interspecific differences in functional traits. In cases where there is only one trait of interest, simple approaches may be appropriate, such as the coefficient

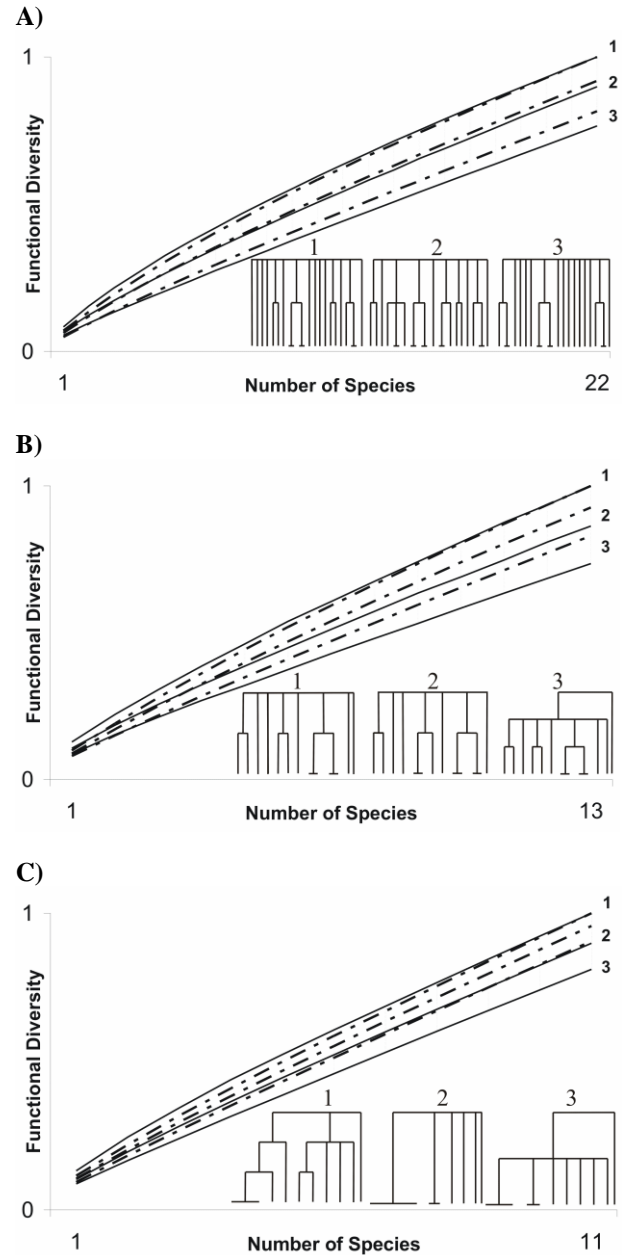


Figure 1. The relationship between species richness and functional diversity (FD), using differing clustering algorithms (1=complete linkage, 2=unweighted pair group method with arithmetic mean, 3=single linkage) and distance measures (solid lines = Euclidean distance, dashed lines=cosine distance), when species are individually removed. Three datasets are shown: **A)** Insectivorous birds (Holmes et al. 1979), **B)** Intertidal fish (Munoz and Ojeda 1997), and **C)** Predatory vertebrates (Jaksić and Medel 1990). Inset are 50% majority rule consensus trees demonstrating lack of consensus of species when calculating functional diversity using different distance measures, but the same clustering approach (numbered as above).

Table 1. Dendrogram group fidelity across distance measures (Euclidean and cosine) for each clustering algorithm: single linkage, unweighted pair group method with arithmetic means (UPGMA), and complete linkage. Group fidelity was determined by majority rules consensus trees using $CI(C)$ consensus index.

Data set	No. species	Complete Linkage	UPGMA	Single Linkage	Mean
Insectivorous birds (Holmes et al. 1979)	22	0.55	0.35	0.35	0.40
Intertidal fish (Munoz and Ojeda 1997)	13	0.55	0.45	0.55	0.52
Predatory vertebrates (Jaksić and Medel 1990)	11	0.22	0.55	0.33	0.37
Mean	—	0.44	0.45	0.41	0.43

of variation (Cadotte et al. 2011). However, the flexibility to use more than one trait often is required to understand even simple natural systems, and in such cases, the inclusion of trait matrices, distance measures and sometimes dendrograms, is required (Petchey and Gaston 2002). The use of these multivariate statistical procedures introduces complications that require researchers to make several key decisions for data analysis (Maire et al. 2015). Ultimately, these decisions should have minimal effect on patterns of species characteristics as they relate to ecosystem function. Unfortunately, this does not appear to be the case.

In this study, different methods of calculating FD lead to different dendrograms, and consequently different measures of functional diversity. For example, FD varied by a maximum of 61% based on different clustering algorithms, and 41% based on different distance measures (Figure 1). As no universal method is likely to be appropriate for all uses (Poos et al. 2009), consensus methods have been promoted to reduce sensitivity (Mouchet et al. 2008). Contrary to such assertions, this study demonstrates that developing consensus across even simple choices (i.e. two distance measures and three clustering algorithms) did not improve the calculation of FD (Table 1). Here, on average there was low consensus across dendrogram topologies, with only 37–52% agreement.

There is considerable debate regarding the most appropriate measure of functional diversity and the qualities that metric should possess (Loreau et al. 2001, Podani and Schmera 2006, Cadotte et al. 2011). Ideally methods of functional diversity should be independent of measures of species richness (Dalerum et al. 2012) and robust to decisions inherent in its calculation. Unfortunately, neither appears to be the case here. FD was closely related to species richness, regardless of: the community type, how species were removed from the community; or how FD was measured (Figure 1). Although there are several ways to measure functional diversity (Cadotte et al. 2011), quantitative comparisons of how functional diversity indices differ are rare (Petchey and Gaston 2006), and evaluations of other functional diversity indices are needed.

Future research regarding how function diversity should be measured—such as questions into how many dimensions are needed to explain functional diversity (Marie et al. 2015)? which species? what kind of diversity? and which ecosystem function (Cadotte et al. 2011)?—need to be expanded to include methodological considerations such as what distance measure? what clustering algorithm? and what constitutes a consensus? Perhaps most importantly, decisions are needed on whether the issues related to producing a dendrogram, and the statistical complications they produce (Petchey and Gaston 2006, Mouchet et al. 2008), improve the measure of FD. Although methodological decisions in FD have been highlighted previously (see Poos et al. 2009; Podani and Schmera 2006), the continual use of dendrogram-based metrics (Petchey and Gaston 2002) and its surrogates (Mouchet et al. 2008) in the functional diversity literature necessitate a reexamination as to whether producing a dendrogram, even with a consensus method, is merited for measuring functional diversity. As the clustering technique will produce dendrograms whether or not true groups exist (Legendre and Legendre 2012), the cumulative effect of these decisions to the relevance of the identified groups may be unknowingly large. Explicit recognition and justification of all methodological decisions is needed for improved functional diversity metrics, however particular attention is needed for producing a dendrogram due to the sensitivity shown in this study. Perhaps the idea that functional traits can be expressed within two-dimensional multivariate space, as done using a dendrogram, has reached its utility and true multi-dimensional techniques are needed (Maire et al. 2015). In Shakespearean parlance ‘to dendrogram or not: consensus methods show that is the question to move functional diversity metrics forward’.

Acknowledgements

Funding was provided by NSERC & OGS. Earlier drafts were improved by discussions with S. Walker and D.A. Jackson. I thank Karl Lamothe (ADPR) and one

anonymous reviewer for their helpful comments on earlier versions.

Referees

Caleb Hasler – cthasler@gmail.com
University of Illinois-Urbana

Karl Lamothe – karl.lamothe@mail.utoronto.ca
University of Toronto

References

Cadotte, M.W., Carscadden, K., and N. Mirotchnick. 2011. Beyond species: functional diversity and the maintenance of ecological processes and services. *Journal of Applied Ecology* 48:1079–1087. [CrossRef](#)

Dalerum, F., Cameron, E.Z., Kunkel, K. and M.J. Somers. 2012. Interactive effects of species richness and species traits on functional diversity and redundancy. *Theoretical Ecology* 5: 129–139. [CrossRef](#)

Holmes, R.T., Bonney, R.E.J., and S.W. Pacala. 1979. Guild structure of the Hubbard Brook bird community: a multivariate approach. *Ecology* 60: 512–520. [CrossRef](#)

Jaksić, F.M., and R.G. Medel. 1990. Objective recognition of guilds: testing for statistically significant species clusters. *Oecologia* 82: 87–92. [CrossRef](#)

Legendre, P. and L. Legendre. 2012. *Numerical Ecology*, 3rd edition. Elsevier, Oxford, UK.

Loreau, M., Naeem, S., Inchausti, P., Bengtsson, J., Grime, J.P., Hector, A., et al. 2001. Biodiversity and ecosystem functioning: Current knowledge and future challenges *Science* 294: 804–808. [CrossRef](#)

Maire, E., Grenouillet, G., Brosse, S., and S. Villeger. How many dimensions are needed to accurately assess functional diversity? A pragmatic approach for assessing the quality of functional spaces. *Global Ecology and Biogeography* 24: 728–740. [CrossRef](#)

Mouchet, M., Guilhaumon, F., Villéger, S., Mason, N.W.H., Tomasini, J.A. and D. Mouillot. 2008. Towards a consensus for calculating dendrogram-based functional diversity indices *Oikos* 117: 794–800. [CrossRef](#)

Munoz, A.A., and F.P. Ojeda. 1997. Feeding guild structure of a rocky intertidal fish assemblage in central Chile. *Environmental biology of fishes* 49: 471–479. [CrossRef](#)

Petchey, O.L., and K.J. Gaston. 2002. Functional diversity (FD), species richness and community composition. *Ecology Letters* 5: 402–411. [CrossRef](#)

Petchey, O.L., and K.J. Gaston. 2006. Functional diversity: back to basics and looking forward. *Ecology Letters* 9: 741–758. [CrossRef](#)

Podani, J., and D. Schmera. 2006. On dendrogram-based measures of functional diversity. *Oikos* 115: 179–185. [CrossRef](#)

Poos, M.S., Walker, S.W. and D.A. Jackson. 2009. Functional-diversity indices can be driven by methodological choices and species richness. *Ecology* 90: 341–347. [CrossRef](#)

Rohlf, F.J. 1982. Consensus indices for comparing classifications. *Mathematical Biosciences* 59: 131–144. [CrossRef](#)

Tilman, D. 1997. Distinguishing between the effects of species diversity and species composition. *Oikos* 80: 185. [CrossRef](#)

Villéger, S., Mason, N.W.H. and D. Mouillot. 2008. New multidimensional functional diversity indices for a multifaceted framework in functional ecology. *Ecology* 89: 2290–2301. [CrossRef](#)

Walker, S.W., Poos, M.S. and D.A. Jackson. 2008. Functional rarefaction: estimating functional diversity from field data. *Oikos* 117: 286–296. [CrossRef](#)