

Functional and Phylogenetic Diversity

Luciano Brocchieri*

Department of Molecular Genetics and Microbiology, University of Florida, Gainesville, FL 32606, USA

Functional Diversity (FD) is a concept developed and widely explored in theoretical ecology (for critical reviews [1,2]). Measures of FD have been developed with the general idea of characterizing the functional aspects of biodiversity that inform the functioning of ecosystems in relation to environmental constraints. The simplest approach to describing FD is possibly the catalog of species presented in a community. The assumption is that "species richness" (the number of species in the community) correlates positively with the number of functionalities expressed by the community. Further developments can include accounting for the abundance of each species, and accounting for how similar species are in terms of their functionalities, i.e., describing the distribution of functional units rather than of species within the community. When a community is described in terms of functional units, it can be represented by a distribution of points in a functional-trait space, with as many dimensions as the number of traits considered. So, functional space can be represented by the values of one specific functionality of interest (a one-dimensional functional space), or as a multi-dimensional space of many continuous or discrete variables. Indices of FD are constructed to represent how the functional space is occupied by the community. Among many proposed indexes, some are based on pairwise diversities between all species [3,4] or between functional units [5], and are independent on the abundance of each species/functional unit. Rao's quadratic entropy, $Q = \sum \sum d_{ij} p_i p_j$ takes instead into account the frequencies of species in the community and their pairwise functional distance d_{ij} [6]. Functional distances can be derived based on the position of points in the functional-trait space [7], or can operate on representation of functional differences summarized by phenograms [8,9], in which the length of branches represents functional distance. FD indexes have been characterized as measuring three different qualities of diversity [2,7,10]: Functional richness, i.e., how much of the niche space is occupied by species or functional units; functional evenness, measuring how evenly distributed are traits within the occupied functional space; and functional divergence, measuring the differentiation of functionalities within the trait's space.

In contrast to indexes of FD, which operate on the observed distribution of functional features, indexes of Phylogenetic Diversity (PD) infer diversities among the species present in the community based on their phylogenetic relations represented by a taxonomic or phylogenetic tree. Genetic relatedness is expected to result in different species sharing functional traits inherited by descent from their common ancestors. It has been proposed that measures of PD can be used as a proxy for measures of FD (e.g., ref. [3]). I argue that, in fact, PD indexes differ from indexes of FD only in that instead of operating on the observed distribution of functional features, they estimate this distribution based on the tree, and that any index of PD can be used as an index of FD by letting it operate on observed frequencies.

Commonly used indices of PD include Faith's $PD = \sum_{i \in B_T} L_i$ where L_i is the length of the i -th branch of the tree, corresponding to the length of the phylogenetic tree [11]. PD is interpreted as the total amount of evolutionary history among the species considered in the tree. Implicit in this measure is the idea that the 'total amount of evolutionary history' is directly proportional to the total number of functional features. In contrast to Faith's PD, other measures of Faith's PD take into consideration the relative abundance of each species in the community. Among these is Rao's quadratic entropy $Q = \sum \sum d_{ij} p_i p_j$ [6]. As in its use as a measure of FD, p_i and p_j are the relative abundances of species i and

j , but d_{ij} is a measure of 'phylogenetic distance' defined using a rooted phylogenetic tree as the difference between the depth of the tree and the sum of the branch lengths common to both species. Phylogenetic entropy [12] $H_p = - \sum_i L_i a_i \log a_i$, considers the aggregate abundance a_i of all species descended from branch i . Pavoine et al. [13] parametric index of PD $I_q = \sum_{k=1}^N (t_k - t_{k-1}) H_{q,K}$ is based on a rooted ultrametric tree and species abundances. In constructing this index the tree is partitioned into N time intervals $(t_k - t_{k-1})$ separating consecutive bifurcation events. If P_{ik} is the aggregate frequency of the i -th character among n characters descended from the branches represented in the time interval $(t_k - t_{k-1})$ with $q \geq 0$ and $q \neq 1$:

$$H_{q,K} = \left[1 - \sum_{i=1}^n p_{i,K}^q \right] / (q-1)$$

and when $q=1$:

$$H_{1,K} = \lim_{q \rightarrow 1} H_{q,K} = - \sum_{i=1}^n p_{i,K} \ln p_{i,K}$$

I_q is related to the previous measures of PD by the relations $I_0 = PD - T$; $I_1 = H_p$; $I_2 = Q$. Chao and collaborators [14] derived a general class of measures called Mean PD of order q (${}^q\bar{D}$) taking into consideration species abundances and phylogenetic relations, related to Jost's diversity measures [15,16] based on Hill's numbers [17]. When $q \geq 0$ and $q \neq 1$, mean PD is defined as:

$${}^q\bar{D}(\bar{T}) = \frac{1}{\bar{T}} \left\{ \sum_{i \in B_T} L_i \left(\frac{a_i}{\bar{T}} \right)^q \right\}^{1/(1-q)}$$

and when $q=1$:

$${}^1\bar{D}(\bar{T}) = \lim_{q \rightarrow 1} {}^q\bar{D}(\bar{T}) = \exp \left[- \sum_{i \in B_T} \frac{L_i}{\bar{T}} a_i \log a_i \right]$$

In these formulations $\bar{T} = \sum_{i \in B_T} L_i a_i$ is the average depth of a rooted tree. With the appropriate choice of order q , mean phylogenetic diversities ${}^q\bar{D}$ have simple relations with other diversity indexes [14]. With Faith's PD when $q=0$: $PD = \bar{T}$; ${}^0\bar{D}$; with phylogenetic entropy H_p when $q=1$: ${}^1\bar{D}(\bar{T}) = \exp(H_p/\bar{T})$; and with Rao's quadratic entropy Q when $q=2$: ${}^2\bar{D}(\bar{T}) = \bar{T}/(\bar{T} - Q)$.

In the expression of PD indexes is implicit an estimate of the frequency of phenotypic (or functional) features in a community. This is accomplished assuming as a model for the evolution of traits Camin-Sokal parsimony [18], by which traits evolve only once in evolutionary history and persist through all descending lineages. With this assumption, each OTU is on average represented by a number of features proportional to the average depth \bar{T} of the tree, the length L_i

***Corresponding author:** Luciano Brocchieri, Department of Molecular Genetics and Microbiology, University of Florida, Cancer and Genetics Research Complex, 2033 Mowry Rd, Gainesville, FL 32606, USA, Tel: 352 284 5414; E-mail: lucianob@ufl.edu

Received October 18, 2016; **Accepted** October 20, 2016; **Published** October 26, 2016

Citation: Brocchieri L (2016) Functional and Phylogenetic Diversity. J Phylogenetics Evol Biol 4: e122. doi: [10.4172/2329-9002.1000e122](https://doi.org/10.4172/2329-9002.1000e122)

Copyright: © 2016 Brocchieri L. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

of a branch i are proportional to the number of features common to all OTUs descended from branch i , and $L_i a_i / \bar{T}$ is the relative aggregate abundance of those features in the community. These estimated frequencies are then used by the index to evaluate diversity. In the case of mean PD of order $q \geq 0$ and $q \neq 1$, this can be more clearly seen by rewriting it as:

$${}^q D(\bar{T}) = \left\{ \sum_{i \in B_T} \frac{L_i a_i}{\bar{T}} a_i^{q-1} \right\}^{-1/(q-1)}$$

Where $L_i a_i / \bar{T}$ is the estimated fraction of all features in the community that are shared by a fraction a_i of all individuals. If $E[f(a)]$ is the mean value of a transformation $f(a)$ of the frequency of features, the above relation can be expressed in terms of this mean. For the example above: ${}^q D(\bar{T}) = \{E(a^{q-1})\}^{-1/(q-1)}$. Similarly, when $q=1$ then

Since other PD indexes can be expressed as functions of mean phylogenetic diversities, they also operate on the corresponding means. To each index of PD can be then associated a corresponding index of FD. While the index of FD operates on estimated frequencies of features, the corresponding index of FD can be applied to observed frequencies of functional features.

Unique to PD indexes is how they infer frequencies of features from the phylogenetic tree. To obtain estimates for the frequency of features, many alternative evolutionary models could be utilized, and the choice of the Camin-Sokal parsimony model implicit in the current formulation of PD indexes can be questioned. With the availability of deep-sequencing technologies, molecular sequences (most often conserved proteins or ribosomal-RNA genes) are currently used to identify species within a community (as, for example, for human or environmental microbiomes), and to obtain evolutionary trees then used to evaluate the PD of the community. In these cases, the evolutionary history of protein or DNA sequences is used as a marker of the evolutionary history of the corresponding species. The possibility of using PD indexes on observed frequency of features, suggest that the multiple alignment of protein or DNA sequences used to obtain the phylogenetic tree could be used directly as a marker of phenotypic diversity within the community, bypassing the necessity to first infer a tree, and then to use the tree to infer distribution of features. If the alignment is not available, its diversity could be inferred using the same probabilistic evolutionary models used to produce the phylogenetic tree [19]. It may, however, be questioned if the evolution and diversification of protein or DNA sequences is a useful representative of diversification of functionalities within the community. For example, proteins of obligate parasites such as *Mycoplasma* or *Rickettsia* often evolve at higher rates than in other species, reflecting in longer terminal branches in the evolutionary tree. Under a Camin-Sokal parsimony model, these long branches would be interpreted as accumulation of functional features. However, these organisms have at the same time experienced genome-size reduction and loss of functionalities, and the higher rate of evolution of some of their proteins may in fact more likely represent loss rather than accumulation of functionalities. Phylogenetic and FD indexes have been scrutinized for how useful they are in describing the functional properties of a community (e.g., refs. [3,20] Flynn et al. 2011; Srivastava et al. 2012). I am not aware of any study to date investigating the usefulness of different markers (e.g., protein or DNA sequences) in predicting the differentiation of functional features across organisms, and how the choice of alternative evolutionary models can improve accuracy in estimating FD. I believe these will be interesting topics of future investigation.

References

- Magurran AE, McGill BJ (2011) Biological Diversity: Frontiers in Measurement and Assessment. Oxford University Press, Oxford, UK.
- Mouchet MA, Villéger S, Mason NWH, Moullot D (2010) Functional diversity measures: an overview of their redundancy and their ability to discriminate community assembly rules. *Funct Ecol* 24: 867-876.
- Flynn DFB, Mirotchnick N, Jain M, Palmer MI, Naeem S (2011) Functional and phylogenetic diversity as predictors of biodiversity-ecosystem-function relationships. *Ecology* 92: 1573-1581.
- Walker B, Kinzig A, Langridge J (1999) Plant attribute diversity, resilience and ecosystem function: the nature and significance of dominant and minor species. *Ecosystems* 2: 95-113.
- Schmera D, Erös T, Podani J (2009) A measure for assessing functional diversity in ecological communities. *J Aquat Ecol* 43: 157-167.
- Rao CR (1982) Diversity and dissimilarity coefficients: a unified approach. *Theor Popul Biol* 21: 24-43.
- Villéger S, Mason NWH, Moullot D (2008) New multidimensional functional diversity indices for a multifaceted framework in functional ecology. *Ecology* 89: 2290-2301.
- Petchey OL, Gaston KJ (2002) Functional diversity (FD), species richness and community composition. *Ecol Lett* 5: 402-411.
- Mouchet M, Guilhaumon F, Villéger S, Mason NWH, Tomasini JA, et al. (2008) Towards a consensus for calculating dendrogram-based functional diversity indices. *Oikos* 117: 794-800.
- Mason NWH, Moullot D, Lee WG, Wilson JB (2005) Functional richness, functional evenness and functional divergence: the primary components of functional diversity. *Oikos* 111: 112-118.
- Faith DP (1992) Conservation evaluation and phylogenetic diversity. *Biol Conserv* 61: 1-10.
- Allen B, Kon M, Bar-Yam Y (2009) A new phylogenetic diversity measure generalizing Shannon index and its application to phyllostomid bats. *Am Nat* 174: 236-243.
- Pavoine S, Love MS, Bonsall MB (2009) Hierarchical partitioning of evolutionary and ecological patterns in the organization of phylogenetically-structured species assemblages: application to rockfish (genus: *Sebastes*) in the Southern California Bight. *Ecol Lett* 12: 898-908.
- Chao A, Chiu CH, Jost L (2010) Phylogenetic diversity measures based on Hill Numbers. *Phil Trans R Soc B* 365: 3599-3609.
- Jost L (2006) Entropy and diversity. *OIKOS* 113: 363-375.
- Jost L (2007) Partitioning diversity into independent alpha and beta components. *Ecology* 88: 2427-2439.
- Hill MO (1973) Diversity and evenness: a unifying notation and its consequences. *Ecology* 54: 427-432.
- Camin JH, Sokal RR (1965) A method for deducing branching sequences in phylogeny. *Evolution* 19: 311-326.
- Brocchieri L (2015) Phylogenetic Diversity and the Evolution of Molecular Sequences. *J Phylogen Evolution Biol* 3: 1.
- Srivastava DS, Cadotte MW, MacDonald AA, Marushia RG, Mirotchnick N (2012) Phylogenetic diversity and the functioning of ecosystems. *Ecol Lett* 15: 637-648.