Decomposing phylogenetic entropy into $\alpha$, $\beta$ and $\gamma$ components

Maud A. Mouchet* and David Mouillot

UMR CNRS-UM2-IFREMER-IRD 5119 Ecosystèmes Lagunaires, Université, Montpellier 2 cc 093, 34 095 Montpellier Cedex 5, France
*Author for correspondence (maud.mouchet@umontp2.fr).

Measuring the phylogenetic diversity of communities has become a key issue for biogeography and conservation. However, most diversity indices that rely on interspecies phylogenetic distances may increase with species loss and thus violate the principle of weak monotonicity. Moreover, most published phylogenetic diversity indices ignore the abundance distribution along phylogenetic trees, even though lineage abundances are crucial components of biodiversity. The recently introduced concept of phylogenetic entropy overcomes these limitations, but has not been decomposed across scales, i.e. into $\alpha$, $\beta$ and $\gamma$ components. A full understanding of mechanisms sustaining biological diversity within and between communities needs such decomposition. Here, we propose an additive decomposition framework for estimating $\alpha$, $\beta$ and $\gamma$ components of phylogenetic entropy. Based on simulated trees, we demonstrate its robustness to phylogenetic tree shape and species richness. Our decomposition fulfils the requirements of both independence between components and weak monotonicity. Finally, our decomposition can also be adapted to the partitioning of functional diversity across different scales with the same desirable properties.

Keywords: branch; phyllobetadiversity; phylogenetic diversity; scale; tree shape

1. INTRODUCTION

Within the context of the global extinction crisis, it has been argued that protecting the Tree of Life is a key component of conservation prioritization [1]. However, we cannot begin to map the distribution of phylogenetic diversity or to measure how fast it is disappearing unless we can appropriately measure it.

Many estimators are available for approximating the phylogenetic diversity of a species assemblage [2], with Faith’s PD [3] being the most popular. However, some of these indices rely only on the phylogenetic links between species, neglecting relative abundances. Indices including abundances only consider the species level, whereas the abundance of lineages and the distribution of abundance along the entire phylogenetic tree deserve to be included. For instance, the potential loss of phylogenetic diversity is linked to lineage extinction probabilities [4], which depend on abundance distributions among tree branches. Other phylogenetic diversity indices, such as Rao’s entropy, do not satisfy the principle of weak monotonicity [5], i.e. the addition of a species may decrease the diversity value if other species abundances change marginally. This principle is critical for conservation issues, as it ensures that promoting the highest phylogenetic distinctiveness among species and maintaining species richness do not conflict [6]. In a recent paper, Allen et al. [6] presented a new phylogenetic diversity index overcoming these limitations; this method and biogeographically weighted evolutionary distinctness [7] are the only indices based on both tree branches as biological units and abundances. Moreover, it satisfies the principle of weak monotonicity.

There is a growing consensus that biodiversity across different scales should be considered, e.g. [8]. Traditionally, ecologists have defined three levels of biodiversity: (i) local biodiversity, termed $\alpha$-diversity, (ii) regional biodiversity, or $\gamma$-diversity, and (iii) the differentiation among local biodiversities, called $\beta$-diversity (or biological turnover). The turnover between local diversities has been recognized as a key determinant of biodiversity establishment and maintenance at regional [9] and global [10] scales.

To overcome the limitations of previous indices, we propose decomposing the phylogenetic entropy introduced by Allen et al. [6] into $\alpha$, $\beta$ and $\gamma$ components. Using simulations, we show that this decomposition provides a measure of phyllobetadiversity [8], robust to phylogenetic tree shape and species richness. We also indicate, through simulations, that the decomposition into $\alpha$-, $\beta$- and $\gamma$-phylogenetic diversities yields independent components.

2. MATERIAL AND METHODS

Allen et al. [6] proved that the phylogenetic entropy generalizes the Shannon index in the same way that two other widely used phylogenetic diversity indices—quadratic diversity [11] and phylogenetic diversity [3]—generalize Simpson’s index and species richness, respectively (table 1). Thus, the properties previously demonstrated for the Shannon index are conserved [12,13].

Jost [12] demonstrated that the decomposition of diversity indices into $\alpha$, $\beta$ and $\gamma$ components has often produced $\beta$ estimates with a hidden dependence on $\alpha$, preventing comparisons and eventually leading to spurious results. As an alternative, Jost [12] proposed using ‘number equivalents’ permitting the decomposition of any diversity index into independent components. The obtained decomposition is not systematically additive or multiplicative, but must be analytically derived for each diversity index. The author showed that the decomposition of the Shannon entropy is additive, and is the only standard diversity measure partitionable into meaningful, independent $\alpha$ and $\beta$ components when community weights are unequal. Furthermore, this concave function always satisfies the condition that $\alpha$-diversity does not exceed $\gamma$-diversity [12,14].

We thus used this additive decomposition to propose $\alpha$, $\beta$ and $\gamma$ components adapted from the phylogenetic entropy of Allen et al. [6].

Let $A_{k,b}$ be the abundance (number of individuals or biomass) of species $i$ within site $k$, the regional pool having a total of $S$ species and $K$ sites. The weight $f$ of each site $k$ is proportional to its relative abundance over sites:

$$f_k = \frac{\sum_{i=1}^{S} A_{ik}}{\sum_{k=1}^{K} \sum_{i=1}^{S} A_{ik}}. \tag{2.1}$$

Local phylogenetic entropy (within site $k$), following Allen et al. [6], is given by:

$$H_k = -\sum_{i=1}^{T} \frac{I(b_i) - p_s(b_i) \times I(a) - p_s(a)}{p_s(b_i)} | p_s(b_i)|, \tag{2.2}$$

where $T$ is a phylogenetic tree for the species regional pool, $I(b_i)$ is the probability that species $b_i$ is included in the phylogenetic tree $T$, and $p_s(b_i)$ is the probability that species $b_i$ is sampled.
Figure 1. A hypothetical case study illustrating the calculation of $\alpha$, $\beta$- and $\gamma$-phylogenetic diversities. (a) Phylogenetic relationships among three species in a regional pool. Phylogenetic distances (branch lengths) among pairs of species are indicated on the dendrogram. (b) Local abundances of species A, B, and C in two hypothetical regions, each divided into two communities (C1 and C2). The total abundance per species and per site is indicated. (c) $H_{\alpha\text{C1}}$: local phylogenetic diversity; $H_{\beta}$: regional phylogenetic diversity; $H_{\alpha}$: mean local phylogenetic diversity estimated on local communities; $H_{\beta\text{C2}}$: phylogenetic turnover.

Table 1. Phylogenetic diversity estimators available to measure $\alpha$, $\gamma$ and $\beta$ components.

<table>
<thead>
<tr>
<th>Presence/absence</th>
<th>$\alpha$-diversity</th>
<th>$\gamma$-diversity</th>
<th>$\beta$-diversity</th>
<th>Simpson index</th>
<th>Shannon index</th>
</tr>
</thead>
</table>

length of a branch $b_i$ and $p_s(b_i)$ is the local proportion of abundance:

$$p_s(b_i) = \frac{\sum_{k=1}^{K} A_{bk}}{\sum_{k=1}^{K} A_k},$$

(2.3)

where $S_i$ is the number of species or leaves descending from $b_i$. The regional phylogenetic entropy ($\gamma$) was defined as:

$$H_\gamma = -\sum_{i} l(b_i) \times p_s(b_i) \times \ln[p_s(b_i)],$$

(2.4)

where the regional proportion of abundance belonging to branch $b_i$ is:

$$p_s(b_i) = \frac{\sum_{k=1}^{K} l(b_i) \times p_s(b_i) \times \ln[p_s(b_i)]}{\sum_{k=1}^{K} \sum_{b=1}^{K} A_{bk}}.$$  

(2.5)

Combining formulae (2.1), (2.3) and (2.5), we can express $p_s(b_i)$ as:

$$p_s(b_i) = \frac{\sum_{k=1}^{K} p_s(b_i) \times \sum_{b=1}^{K} A_{bk}}{\sum_{k=1}^{K} \sum_{b=1}^{K} A_{bk}} = \sum_{k=1}^{K} p_s(b_i) \times f_k).$$

(2.6)

The regional proportion of abundance belonging to branch $b_i$ is thus the weighted sum of local proportions of abundance of this branch $b_i$. The average within-site ($\alpha$) phylogenetic entropy, $H_\alpha$, is presented as the expectation of $H_k$ over all sites according to the weight $w_k$:

$$H_\alpha = -\sum_{k=1}^{K} w_k \sum_{i=1}^{T} l(b_i) \times p_s(b_i) \times \ln[p_s(b_i)].$$

(2.7)

The weight is $w_k = 1/K$ if each site $k$ is given the same weight regardless of its regional contribution, or $w_k = f_k$, if each site $k$ weight corresponds to its regional contribution. In any case:

$$\sum_{k=1}^{K} w_k = 1.$$

The aim of our study is not to discuss these alternative approaches to site weighting. Both methods have been proposed, and the final choice depends on the question investigated and the data available [15,16].

Following the additive partitioning of Shannon diversity, $\beta$ phylogenetic entropy is estimated by:

$$H_\beta = H_\gamma - H_\alpha.$$  

Accordingly, $H_\beta$ can be expressed as:

$$H_\beta = \sum_{i=1}^{T} l(b_i) \times p_s(b_i) \times \ln[p_s(b_i)]$$

$$+ \sum_{k=1}^{K} w_k \sum_{i=1}^{T} l(b_i) \times p_s(b_i) \times \ln[p_s(b_i)].$$

(2.8)

Thus, we obtain:

$$H_\beta = \sum_{i=1}^{T} l(b_i)$$

$$\times p_s(b_i) \times \ln[p_s(b_i)] - \sum_{k=1}^{K} w_k p_s(b_i) \times \ln[p_s(b_i)].$$

(2.9)

$H_\beta = 0$, when all local proportions of abundance belonging to branch $b_i$ are equivalent to the corresponding regional proportions, i.e. for each site $k$, $p_s(b_i) = p_s(b_i)$. In this case, there is no phylogenetic turnover among sites, because all branches have similar abundances. By contrast, when species compositions or local relative abundances belonging to branch $b_i$ vary among sites, then $H_\beta$ increases. Pavoine et al. [17] demonstrated that partitioning any

conceivable index into $\alpha$, $\beta$ and $\gamma$ components, using an ultrametric phylogenetic tree and abundances at the $\gamma$ scale including those of local communities ($\alpha$ scale), prevents negative $H_\beta$ values. Since this approach corresponds to the weighted method in equation (2.7), we have:

$$p_\gamma(b_i) \times \ln[p_\gamma(b_i)] - \sum_{k=1}^{K} w_k p_k(b_i) \times \ln[p_k(b_i)] \leq 0$$

(2.10)

and $H_\beta \geq 0$.

As an illustration, we propose two simple decompositions in figure 1 (see electronic supplementary material, S1 for details).

To investigate the potential influence of tree shape and community size on the phylogenetic entropy decomposition, we simulated artificial trees with different species richness values and tree structure. We produced 50-species and 100-species trees using two contrasted models, either the Yule model or the proportional-to-distinguishable arrangements (PDA) model (electronic supplementary material, S2). To explore the influence of species richness on our decomposition, we built communities by randomly sampling species along a gradient of richness (from 1 to 50 or 1 to 100 species), which were then compared with a given 20-species community. $\beta$- and $\gamma$-diversity values were calculated on pairs of communities composed of one 20-species community and one $n$-community ($n$ ranging from 1 to 50 or 1 to 100 species depending on the regional tree size). $\beta$-diversity is expressed as the proportion of biological turnover (i.e. $H_{\beta\alpha} = 1 - (H_\beta/H_\gamma)$).

**Figure 2.** $\alpha$-, $\beta$- and $\gamma$-diversity values estimated on communities sampled among (a) 50-species or (b) 100-species trees generated under the Yule (closed circles) or PDA (open circles) model. $\alpha$- (based on average local diversity values (a)), $\beta$- and $\gamma$-diversity values were calculated on pairs of communities composed of one 20-species community and one $n$-community ($n$ ranging from 1 to 50 or 1 to 100 species depending on the regional tree size). $\beta$-diversity is expressed as the proportion of biological turnover (i.e. $H_{\beta\alpha} = 1 - (H_\beta/H_\gamma)$).
levels (excepting low levels), total tree sizes and diversification models, allowing comparisons. Such a decomposition contributes to a better understanding of the phylogenetic diversity distribution across space and time, and thus to define appropriate conservation efforts towards the preservation of the Tree of Life. Our decomposition could be adapted to the partitioning of functional diversity when species are clustered along functional trees.

16 Villégé, S. & Mouillot, D. 2008 Additive partitioning of diversity including species differences: a comment


