# Measuring Biodiversity by Hill Numbers

## 1  Diversity of species, phylogeny and function

For an assemblage of individuals, its diversity can be measured in ways different in incorporation of species difference. In species diversity, all species are assumed to be equally distinct. In phylogenetic diversity, evolutionary or phylogenetic differences, *e.g.* taxonomic classification and well-supported phylogenetic tree, is taken into consideration. All else being equal, an assemblage of closely related species is less diverse in terms of phylogeny than an assemblage composed of highly divergent species. In functional diversity, species are described by a set of traits and species difference can be measured by dissimilarities of trait profiles. There are three main approach to functional diversity measures: distance based, dendrogram based and trait-value based.

## 2  Classic measures of species diversity

Species richness is a simple count of number of species present in an assemblage. It is an intuitive and frequently used diversity index and is a key metric in conservation biology. However, it does not incorporate information about species abundances and is difficult to estimate accurately from small samples.

Shannon index is another measure of diversity, defined as

$$H_{Sh} = -\sum_{i=1}^{S} p_i \ln p_i \tag{1}$$

where $S$ is species richness and $p_i$ is the relative abundance of $i$th species.

Gini-Simpson index is another popular measure of diversity, defined as

$$H_{GS} = 1 - \sum_{i=1}^{S} p_i^2 \tag{2}$$

It gives the probability that two randomly drawn individuals belong to different species.

Species richness, Shannon index and Gini-Simpson index can be united into a family of generalized

indexes, parametrized by a variable $q \geq 0$. The generalized index is defined as

$$^{q}H = \frac{1 - \sum\limits_{i=1}^{S} p_i^q}{q - 1} \tag{3}$$

Thus,

$$^{0}H = S - 1 \tag{4}$$

$$^{1}H = \lim_{q \to 1} {}^{q}H = H_{Sh} \tag{5}$$

$$^{2}H = H_{GS} \tag{6}$$

## 3 Classic measures of phylogenetic diversity

Faith's phylogenetic diversity (PD) (Faith 1992) is a widely used measure of phylogenetic diversity. It is defined as the sum of all branch lengths of a phylogenetic tree connecting all species present in an assemblage. Like species richness, Faith's PD does not incorporate information on species abundance.

Rao's quadratic entropy (Rao 1982) incorporates both phylogeny and species abundances, defined as

$$Q = \sum_{i=1}^{S} \sum_{i=1}^{S} d_{ij} p_i p_j \tag{7}$$

where $d_{ij}$ is the phylogenetic distance between $i$th and $j$th species.

Allen *et al.* (2009) proposed another measure of phylogenetic measure:

$$H_p = -\sum_i L_i a_i \ln a_i \tag{8}$$

where $L_i$ is the length of branch $i$, $a_i$ is the summed relative abundance of all species descended from branch $i$.

For an ultrametric tree in which all branch tips are the same distance (denote by depth $T$) to the root, Faith's PD, Rao's $Q$ and $H_p$ of Allen *et al.* can be generalized as (Pavoine *et al.* 2009)

$$^{q}I = \frac{T - \sum\limits_i L_i a_i^q}{q - 1} \tag{9}$$

2

Thus,

$$^0I = \sum_i L_i - T \tag{10}$$

$$^1I = \lim_{q \to 1} {}^qI = -\sum_i L_i a_i \ln a_i \tag{11}$$

$$^2I = T - \sum_i L_i a_i^2 \tag{12}$$

## 4 Classic measures of functional diversity (distance-based)

Functional attribute diversity (FAD) (Walker *et al.* 1999) is the sum of the pairwise distances between species:

$$FAD = \sum_{i=1}^{S}\sum_{j=1}^{S} d_{ij} \tag{13}$$

where $d_{ij}$ is the functional distance of species $i$ and $j$.

Rao's $Q$ (Equation 7) can be used as a measure of functional diversity, replacing $d_{ij}$ by functional distance. Guiasu (2011, 2012) proposed a weighted Gini-Simpson index for pairs of species as follows:

$$GS_D = \sum_{i=1}^{S}\sum_{j=1}^{S} d_{ij} p_i p_j = Q - \sum_{i=1}^{S}\sum_{j=1}^{S} d_{ij}(p_i p_j)^2 \tag{14}$$

Ricota and Szeidl (2009) and de Bello *et al.* (2010) transformed Rao's $Q$ for functional diversity to the effective number of species with a maximum species pairwise distance $d_{max}$. It is the theoretical species richness of a perfectly even assemblage with the same Rao's $Q$ as the original assemblage:

$$Q_e = \frac{1}{1 - Q/d_{max}} \tag{15}$$

## 5 Measure diversity by Hill numbers

Measures of species diversity, phylogenetic diversity and functional diversity can be incorporated into a unified framework based on Hill numbers (Chao and Chiu *et al.* 2014) Denote by $C$ an assemblage of entities (species or phylogenetic tree branch segments or species pairs), and by $u$ elements in $C$. For each $u$, denote by $v_u$ its attribute value and by $a_u$ its weight (*e.g.* biomass, cover area, abundance). The weighted mean attribute value is

$$\bar{V} = \sum_{u \in C} a_u v_u \tag{16}$$

3

The Hill number of order $q$ is defined as

$$^qAD(\bar{V}) = [\sum_{u \in C} v_u \times (\frac{a_u}{\bar{V}})^q]^{\frac{1}{1-q}} = [\sum_{u \in C} v_u \times (\frac{a_u}{\sum\limits_{u \in C} a_u v_u})^q]^{\frac{1}{1-q}} \tag{17}$$

AD is short for attribute diversity. Further,

$$^0AD(\bar{V}) = \sum_{u \in C} v_u \tag{18}$$

$$^1AD(\bar{V}) = \lim_{q \to 1} {}^1AD(\bar{V}) = exp\{-\sum_{u \in C} v_u \frac{a_u}{\bar{V}} \ln \frac{a_u}{\bar{V}}\} \tag{19}$$

$$^2AD(\bar{V}) = \frac{1}{\sum\limits_{u \in C} v_u (\frac{a_u}{\bar{V}})^2} \tag{20}$$

For species diversity, $C$ is an assemblage of species indexed by $i = 1, 2, 3, \ldots, S$. Denote by $p_i$ the relative abundance of $i$th species and let it be its weight. Since all species are equally distinct, they have same attribute value 1. Thus, Equation 17 reduces to ordinary Hill numbers (Hill 1973):

$$^qD = (\sum_{i=1}^{S} p_i^q)^{\frac{1}{1-q}} \tag{21}$$

For phylogenetic diversity, $C$ is the assemblage of all branch segments in a phylogenetic tree. Index each branch by $i = 1, 2, 3, \ldots, B$, and denote by $L_i$ the length of branch $i$. Let $a_i$ be the branch abundance, *i.e.* the summed relative abundance of all species descended from the branch $i$. The weighted mean attribute value is

$$\bar{T} = \sum_{i=1}^{B} L_i a_i \tag{22}$$

The Hill number for phylogenetic diversity is (Chao *et al.* 2010)

$$^q\bar{D}(\bar{T}) = [\sum_{i=1}^{B} L_i \times (\frac{a_i}{\sum\limits_{i=1}^{B} L_i a_i})^q]^{(\frac{1}{1-q})} \tag{23}$$

For functional diversity, $C$ is the assemblage of pairs of species. Denote by $S$ species richness and by $p_i$ relative abundance of $i$th species. For species pair $u = (i, j)$, its attribute value $v_u$ equals functional

distance of species $i$ and $j$, $i.e.$ $v_u = d_{ij}$. Its abundance $a_u = p_i p_j$. The weighted mean attribute value is

$$Q = \sum_{i,j=1}^{S} d_{ij} p_i p_j \tag{24}$$

The Hill number for functional diversity is (Chiu and Chao 2014)

$$^q FD(Q) = [\sum_{i,j=1}^{S} d_{ij} \times (\frac{p_i p_j}{\sum\limits_{i,j=1}^{S} d_{ij} p_i p_j})^q]^{\frac{1}{1-q}} \tag{25}$$

## 6 Replication principle

Assume there is a total of $N$ assemblages of entities (species or phylogenetic tree branch segments or species pairs), denoted by $C_1, C_2, \ldots, C_N$. These assemblages are completely distinct, $i.e.$ $\forall i, j = 1, 2, \ldots, N, C_i C_j = \emptyset$. For assemblage $C_i$, denote by by $v_{ui}$ the attribute value of element $u \in C_i$ and by $a_{ui}$ its weight. Assume all assemblages are identical in weighted mean attribute value and Hill number of same order $q$, $i.e.$

$$\bar{V} = \sum_{u \in C_i} a_{ui} v_{ui}, i = 1, 2, 3, \ldots, N \tag{26}$$

and

$$^q AD(\bar{V}) = [\sum_{u \in C_i} v_{ui} \times (\frac{a_{ui}}{\bar{V}})^q]^{\frac{1}{1-q}}, i = 1, 2, 3, \ldots, N \tag{27}$$

For pooled assemblage $\sum\limits_{i=1}^{N} C_i$, its weighted mean attribute value

$$\bar{V}_p = \sum_{i=1}^{N} \sum_{u \in C_i} a_{ui} v_{ui} = \sum_{i=1}^{N} \bar{V} = N \times \bar{V} \tag{28}$$

The Hill number of pooled assemblage is

$$\begin{aligned}
^q AD(\bar{V}_p) &= [\sum_{i=1}^{N} \sum_{u \in C_i} v_{ui} \times (\frac{a_{ui}}{\bar{V}_p})^q]^{\frac{1}{1-q}} \\
&= [\sum_{i=1}^{N} \sum_{u \in C_i} v_{ui} \times (\frac{a_{ui}}{\bar{V}})^q \times \frac{1}{N^q}]^{\frac{1}{1-q}} \\
&= \{\sum_{i=1}^{N} \frac{1}{N_q} \times [^q AD(\bar{V})]^{1-q}\}^{\frac{1}{1-q}} \\
&= N \times^q AD(\bar{V})
\end{aligned} \tag{29}$$

5

# 7  Decomposition of Hill numbers

Consider a pooled assemblage composed of $N$ assemblages. Denote by $z_{ik}$ the weight of $i$th entities (species or phylogenetic tree branch segments or species pairs) in $k$th assemblage, where $i = 1, 2, 3, \ldots, S$ and $k = 1, 2, 3, \ldots, N$. The total weight in $k$th assemblage is $z_{+k} = \sum_{i=1}^{S} z_{ik}$, and the total weight of $i$th entity in the pooled assemblage is $z_{i+} = \sum_{k=1}^{N} z_{ik}$. The total weight of all entities in the pooled assemblage is $z_{++} = \sum_{k=1}^{N} \sum_{i=1}^{S} z_{ik}$. The attribute value of $i$th entity is $v_i$.

The gamma diversity of the pooled assemblage given by Equation 17 is

$$^{q}AD_{\gamma}(\bar{V}) = \left( \sum_{i=1}^{S} v_i \times \left( \frac{z_{i+}}{\sum_{i=1}^{S} z_{i+} v_i} \right)^q \right)^{\frac{1}{1-q}} \tag{30}$$

The alpha diversity is (Chiu *et al.*)

$$^{q}AD_{\alpha}(\bar{V}) = \left( \sum_{i=1}^{S} v_i \times \sum_{k=1}^{N} \left( \frac{z_{ik}}{\sum_{i=1}^{S} z_{i+} v_i} \right)^q \right)^{\frac{1}{1-q}} \tag{31}$$

The beta diversity is

$$^{q}AD_{\beta}(\bar{V}) = \frac{1}{N} \frac{^{q}AD_{\gamma}(\bar{V})}{^{q}AD_{\alpha}(\bar{V})} \tag{32}$$

When the $N$ assemblages are completely identical, *i.e.* $z_{ik} = z$ for all $i = 1, 2, 3, \ldots, S$ and $k = 1, 2, 3, \ldots, N$, $^{q}AD_{\beta}(\bar{V})$ equals 1. When the $N$ assemblages are completely distinct, *i.e.* any two assemblages do not have shared entity, $^{q}AD_{\beta}(\bar{V})$ equals $N$.

## References

[1] Chao, A. , Chiu, C. H. ,  Jost, L. . (2014). Unifying species diversity, phylogenetic diversity, functional diversity, and related similarity and differentiation measures through hill numbers. Annual Review of Ecology Evolution  Systematics, 45(1), 297-324.