FLORA FL, WELLHAM PAD.

# *PHYLOGENETIC AND ECOLOGICAL COEFFICIENTS FOR BIODIVERSITY ASSESSMENTS*

1ST MARCH 2021

fortifyscience.com/2021-03-01-flora

# PHYLOGENETIC AND ECOLOGICAL COEFFICIENTS FOR BIODIVERSITY ASSESSMENTS

F. L. Flora[1] & P. A. D. Wellham[2]

[1] BioFortify Research Institute, Nottingham, United Kingdom
[2] School of Pharmacy, University of Nottingham, Nottingham United Kingdom

The coefficients and numerical formulae described in this paper underlie an improved way from convention to assess biological diversity via phylogenetics and ecological understanding, and more broadly further a more contextual basis for working with biological information.

## *Branch Factor*

Branch factor describes the number of branching events in a given area of a cladogram, across given taxa and taxonomic ranks. The branch factor for a given taxon is the number of taxa following its parent taxon, i.e. the number of sister taxa plus itself. The branch factor between two given taxa is calculated by multiplying branch factors between the taxa as appropriate. For example, in a cladogram featuring species, genera, and families, the branch factor for a given species up to its family is defined by

$$B_{S\text{\textasciicircum}F} = B_S \;.\; B_G$$

Where $S$ is the given species, $B_S$ is the number of species in the parent genus, and $B_G$ is the branch factor for the parent genus (within the parent family in turn). Note that $B_F$ is not included.

Hence in a cladogram with multiple taxonomic ranks of which junior taxonomic rank $q$ and senior taxonomic rank $m$ are featured, branching factor $B$ for given taxon of rank $q$ is defined by

$$B_{q\text{\textasciicircum}m} = B_{m-1} \,.\, B_{m-2} \,.\, \dots \,.\, B_{m-r} \,.\, B_q$$

Where $r = q+1$.

Approximations for branch factor may be made using an intermediate terminal taxonomic rank $t$. The factor above can therefore be approximated to

$$B_{q\{t\}\text{\textasciicircum}m} = B_{m-1} \,.\, B_{m-2} \,.\, \dots \,.\, B_{m-u} \,.\, N_{q(t)}$$

Where number $N$ of $q$ taxa of given taxon of rank $t$ approximates $B_{q\text{\textasciicircum}t}$ , and $u = t+1$. This equates simply to

$$B_{q\{t\}\text{\textasciicircum}m} = B_{t\text{\textasciicircum}m} \,.\, N_{q(t)} \,.$$

For example, if junior taxonomic rank $q$ is species $S$ and senior taxonomic rank is phylum $P$, the relevant branch factor $B_{S^\wedge P}$ can be approximated to the following:

$$B_{S\{C\}^\wedge P} = B_{C^\wedge P} \,.\, N_{S(C)}$$

$$B_{S\{O\}^\wedge P} = B_{O^\wedge P} \,.\, N_{S(O)}$$

$$B_{S\{F\}^\wedge P} = B_{F^\wedge P} \,.\, N_{S(F)}$$

$$B_{S\{G\}^\wedge P} = B_{G^\wedge P} \,.\, N_{S(G)}$$

where each alternative is successively more accurate an approximation than the last. The natural conclusion of this sequence is not an approximation but the equation

$$B_{S\{S\}^\wedge P} = B_{S^\wedge P} \,.\, N_{S(S)} = B_{S^\wedge P} \,.$$

In a cladogram, terminal taxa can be designated at in appropriate places to allow approximations for branch factors if the overall number of taxa would otherwise be too exhaustive. The termini taxa may be set at a consistent taxonomic rank throughout the cladogram, or may be defined by rules based upon $B$ values. In an example of the latter, $t$ may be set as defined by the lowest taxonomic rank in each case where $B_{t^\wedge m}$ may not exceed a given number. It is often desired to work with species level when assessing biodiversity, which therefore is the junior taxonomic rank to be most relevant in formulae from now on. This can be swapped out for a variable $q$ if necessary.

### *Phylogenetic Coefficient*

The phylogenetic coefficient is defined by

$$PhyCo_{q\{t\}^\wedge m} = 1 - 0.99 \,.\, Log_{10}[B_{q\{t\}^\wedge m}] \,/\, Log_{10}[max\ (B_{q\{t\}^\wedge m})]$$

Where $max\ (B_{q\{t\}^\wedge m})$ is the greatest branch factor score of its type in the cladogram or in the dataset in question. Note here that the use of 0.99 as a factor against the ratio sets the lowest possible $PhyCo$ to 0.01 (1%). This $PhyCo$ corresponds to the terminal taxon $t$ which gives the highest $B_{q\{t\}^\wedge m}$ branch factor. The highest possible $PhyCo$ is 1 (100%).

## *Ecological Extension of the Phylogenetic Coefficient*

An ecological extension *EcoPhyCo* can be applied to the phylogenetic coefficient to take into consideration ecological interactions between species of different taxa. This can include predator-prey, parasite/parasitoid-host, herbivore-plant, commensal-host, and obligate and facultative mutualistic interactions. For these purposes, in each case a dependent (*d*) and host (*h*) are defined – in each case respectively for the types of interactions listed. In mutualistic associations, the major partner is defined as the host. If this is disputed or ambiguous, both partners may be defined as the either the dependent and/or host. *EcoPhyCo* is defined as

$$EcoPhyCo_h = ( PhyCo_h + PhyCo_h \, . \, PhyCo_d ) / 2 \, .$$

For taxa where no ecological interactions are known or recorded,

$$EcoPhyCo_h = PhyCo_h / 2$$

For ecological interactions to be used in this extension, both dependent and host species involved must be present in the cladogram or dataset. For example, if a cladogram was constructed from species observed in a given habitat, and the *EcoPhyCo* coefficient used for biodiversity calculations, both species of any interaction included would have to be observed. Therefore, unlike standard *PhyCo*, which is affected only by cladogram structure, *EcoPhyCo* scores are also affected by observations made on a case-by-case basis.

If a host species has more than one dependent partner species – say *n* species, the formula for $EcoPhyCo_h$ is adjusted as follows:

$$EcoPhyCo_h = ( PhyCo_h + \Sigma \, (PhyCo_h \, . \, PhyCo_d) ) / n \, .$$

## *Heritage Extension of the Phylogenetic Coefficient*

The heritage extension *HerPhyCo* applied to *PhyCo* accounts for both age and diversity of relatively larger taxa. It is defined by

$$HerPhyCo_m = Mean \, ( PhyCo_{q(t)\text{^}m} ) \, . \, ( A_m / A_{m+1} )$$

Where *A* is the age of a given taxon. The heritage score is thus defined by the choice of junior taxonomic rank *q* and the terminal *t* .

### *Biodiversity and Richness Scores*

Scores for richness and biodiversity are defined by formulae

$$Richness = \Sigma Co$$

$$Diversity = (\Sigma(Co \,.\, N)) \,/\, \Sigma Co$$

Where *Co* can represent coefficients *PhyCo*, *EcoPhyCo*, or *HerPhyCo* as appropriate.

### *Grouping vs. Branching*

In all and any of the above concepts, a cladogram is required to be the basis of the assessments made. The construction of a cladogram is down to discretion and can be done using a trade-off between the accuracy using the most detailed information available, and practicalities of building complex structures. For instance, a cladogram may in a simpler form group together a set of families under one order, all of which have the same branch factor $B_{F\text{^}O}$. We can call this a *grouping* approach. Conversely, using the most detailed understanding from phylogenetic data, one may construct a phylogeny in which every possible dichotomy is made between taxa. In this arrangement there would be no consistent value for $B_{F\text{^}O}$, and actually the concept of family and order would be redundant for the calculation of branch factor values. Instead it would be more practical to number divisions rather than use taxonomic ranks. Here the most derived families would have higher branch factor values up to their order, with more basal ones having lower branch factor values. We can call this a *branching* approach. It would be important to note that in this instance, for the case of the phylogenetic coefficient *PhyCo*, the setting of the formula to the lowest possible coefficient value to 0.01 prevents these values becoming exponential after serial dichotomies in a cladogram constructed with *branching* approach.

Using the above concepts may a biologist properly assess biodiversity – in terms of phylogenetics, ecology and evolution, rather than just at the observable species level. Combined with suitable data collection methods, ecological understanding and phylogenetic methods, this may inform relevant policy and diagnose effects on the living environment.