Ⓔ

A New Phylogenetic Diversity Measure Generalizing the Shannon Index and Its Application to Phyllostomid Bats

Benjamin Allen,[1,2,*] Mark Kon,[1,2] and Yaneer Bar-Yam[2]

1. Department of Mathematics and Statistics, Boston University, Boston, Massachusetts 02115;   2. New England Complex Systems Institute, Cambridge, Massachusetts 02138

ABSTRACT: Protecting biodiversity involves preserving the maximum number and abundance of species while giving special attention to species with unique genetic or morphological characteristics. In balancing different priorities, conservation policymakers may consider quantitative measures that compare diversity across ecological communities. To serve this purpose, a measure should increase or decrease with changes in community composition in a way that reflects what is valued, including species richness, evenness, and distinctness. However, counterintuitively, studies have shown that established indices, including those that emphasize average interspecies phylogenetic distance, may increase with the elimination of species. We introduce a new diversity index, the phylogenetic entropy, which generalizes in a natural way the Shannon index to incorporate species relatedness. Phylogenetic entropy favors communities in which highly distinct species are more abundant, but it does not advocate decreasing any species proportion below a community structure–dependent threshold. We contrast the behavior of multiple indices on a community of phyllostomid bats in the Selva Lacandona. The optimal genus distribution for phylogenetic entropy populates all genera in a linear relationship to their total phylogenetic distance to other genera. Two other indices favor eliminating 12 out of the 23 genera.

Keywords: diversity index, biodiversity, phylogenetic entropy, quadratic diversity, Shannon index.

Introduction

Conserving biodiversity is a central aim of environmental policy. In order to plan conservation strategies that effectively allocate limited resources, it is important to characterize the diversity present in a given community. However, there is no universally accepted biodiversity measure. Traditional indices, such as species richness, the Shannon

* Corresponding author; e-mail: benjcallen@gmail.com.

index (Shannon 1948), and the Simpson index (Simpson 1949), characterize diversity based on the number of species present (species richness) and the distribution of the number of organisms per species (species evenness). However, many authors (Ehrlich 1988; May 1990; Humphries et al. 1995; Crozier 1997) have argued that different species make unequal contributions to diversity and that priority in conservation decisions should be given to species with unique genetic or morphological characteristics. Newer diversity indices have been introduced to reflect these priorities, including indices that characterize only the relatedness or distinctness of species (Vane-Wright et al. 1991; Faith 1992; Nixon and Wheeler 1992; Solow et al. 1993), as well as measures that combine abundance and relatedness data. The most established of these is quadratic diversity, introduced by Rao (1982) and independently rediscovered under the name "taxonomic diversity" by Warwick and Clarke (1995), which measures the average taxonomic or phylogenetic distance between individual organisms. Other measures combining both types of data have been considered in recent research (Izsák and Papp 2000; Webb et al. 2002; Ricotta and Avena 2003; Ricotta 2004; Ricotta and Szeidl 2006; Weikard et al. 2006; Helmus et al. 2007).

A diversity measure can be considered as a valuation of the conservation priority of each species in a community. By considering how changes in species abundance affect a diversity measure, we determine the value the measure places on community members. A diversity index may favor reductions in the relative proportion of some species, for example, when a decrease in a dominant species yields a more even distribution by increasing the proportion of rare species, or a decrease in a less distinct species increases the relative abundances of more distinct species. However, it is reasonable to expect a diversity measure to favor retaining the members of a rare species, so as not to eliminate the species entirely. As is shown by Shimatani (2001), Wei-

Table 1: Diversity index formulas

Index	Symbol	Formula[a]	Source
Species richness	n	n	
Simpson index	D	$-\sum_i p_i^2$	Simpson 1949
Shannon index	H	$-\sum_i p_i \ln p_i$	Shannon 1948
Phylogenetic diversity	PD	$\sum_{\text{branches } b} \ell(b)$	Faith 1992
Quadratic diversity	Q	$\sum_{i<j} d_{ij} p_i p_j$	Rao 1982
Taxonomic entropy[b]	$H(P, K)$	$-\sum_i p_i \ln k_i$	Ricotta and Avena 2003
Unnamed[c]	H_d	$-\sum_i p_i \ln\left(1 - \sum_{j \neq i} d_{ij} p_j\right)$	Ricotta and Szeidl 2006

[a] n = number of species, p_i = proportion of individuals in species s_i, and d_{ij} = taxonomic or phylogenetic distance between species s_i and s_j; d_{ij} can be equated with the branch length of the shortest path between the corresponding leaves of a phylogenetic tree.

[b] $k_1, \ldots, k_n$ are species distinctness weights, normalized so that $\sum k_i = 1$.

[c] For this measure, the distances are normalized so that $d_{ij} \leq 1$.

kard et al. (2006), and our work below, this is not the case for some diversity measures, specifically for Q, the quadratic diversity index, and H_d, the index of Ricotta and Szeidl (2006; see table 1). There are circumstances when these indices favor eliminating species that are similar to abundant ones.

In this article, we introduce a new diversity index, the phylogenetic entropy H_P. Phylogenetic entropy places a high value on distinctive species but has the property that when members of a species become rare in proportion to other species, it is never desirable to eliminate them.

A direct formal relationship exists between H_P and the Shannon index, which is widely used, due in part to its deep mathematical roots and connections to information theory and physics. The H_P index generalizes the Shannon index in the same way that two other established indices, quadratic diversity and phylogenetic diversity, generalize Simpson's index and species richness, respectively (fig. 1). We prove these results and compare phylogenetic entropy to other extensions of Shannon's index (Casquilho et al. 1997; Ricotta and Avena 2003; Ricotta and Szeidl 2006).

To understand the differences between our index and other indices combining abundance and relatedness data, we analyzed the behavior of phylogenetic entropy, quadratic diversity, and H_d on bat communities in the Selva Lacandona in Chiapas, Mexico. We found that all three indices correlate with established measures of richness and evenness and prefer distributions with greater proportions of more distinct species. However, quadratic diversity and H_d favored elimination of a less distinct genus (i.e., they increased monotonically as the abundance of this genus was reduced to 0) and were optimized by distributions containing less than half of the available genera. Such situations cannot occur for the phylogenetic entropy. The optimized distribution should be understood as the logical consequence of policy decisions using a particular diversity measure. As evidence for the mathematical naturalness of our measure, an axiomatic characterization of phyloge-

netic entropy, generalizing a similar characterization of the Shannon index, is given in the appendix in the online edition of the *American Naturalist*.
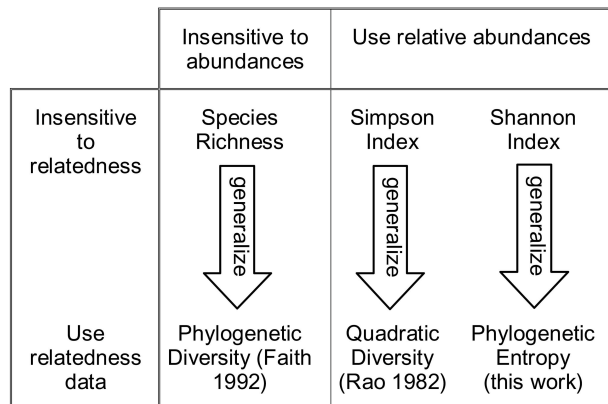
Definition and Analysis

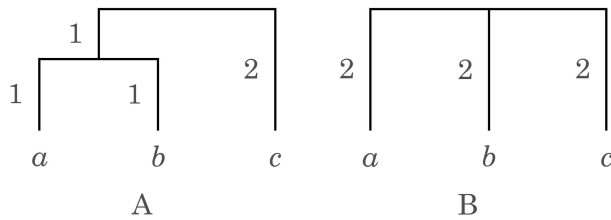We define the phylogenetic entropy index of a biological community as

$$H_P = -\sum_{\text{branches } b \text{ of } T} \ell(b) p(b) \ln p(b), \tag{1}$$

where T is a rooted phylogenetic tree for the community, $\ell(b)$ is the length of a branch b of T, and $p(b)$ is the proportion of individuals in the (present-day) community who are represented by leaves descending from b. The Shannon index,

$$H = -\sum_{i=1}^{n} p_i \ln p_i, \tag{2}$$

	Insensitive to abundances	Use relative abundances	
Insensitive to relatedness	Species Richness	Simpson Index	Shannon Index
	⬇ *generalize*	⬇ *generalize*	⬇ *generalize*
Use relatedness data	Phylogenetic Diversity (Faith 1992)	Quadratic Diversity (Rao 1982)	Phylogenetic Entropy (this work)

Figure 1: Relationships between diversity measures.

Figure 2: Phylogenetic trees for sample ecological communities. In *A*, species *c* is more distinct than *a* or *b*, whereas in *B*, the species are equally distinct.

arises as a special case of the phylogenetic entropy if all species are equally distinct or, equivalently, if *T* has uniform branch lengths.

Our index is superficially similar to the weighted Shannon index,

$$H_{\mathrm{w}} = -\sum w_i p_i \ln p_i,$$

used to measure the diversity of habitats in a landscape (Casquilho et al. 1997) and proposed as a measure of species diversity by Ricotta (2002) and Guiasu and Guiasu (2003). However, our index is evaluated in terms of the branch lengths $\ell(b)$ and proportions $p(b)$ reflecting the structure of the phylogenetic tree, whereas the weighted Shannon index uses a single weight for the diversity value of each species. We note that the branch proportions $p(b)$ used in phylogenetic entropy are not mutually exclusive (multiple species may descend from a single branch) and hence do not sum to 1 as the species proportions p_i do in weighted entropy. For trees with no internal branching (e.g., fig. 2*B*), the two indices can be made equal by setting w_i to be the distance from the root to species *i*. However, in general, there is no weighting scheme that would make the two indices equivalent.

Phylogenetic entropy increases with greater distinctness of species, due to its linear dependence on the branch lengths. Given a fixed set of species, it favors distributions that are biased toward, but not dominated by, the species that are more distinct. For example, given the phylogenetic tree shown in figure 2*A*, in which species *c* is more distinct than species *a* and *b*, H_{P} attains its maximum value for the distribution $p_a = p_b \approx 0.29$, $p_c \approx .41$. In the case of figure 2*B*, where all species are equally distinct, H_{P} reduces to twice the Shannon index, and its maximum value is attained by a uniform distribution.

The relationship between phylogenetic entropy and the Shannon index parallels relationships between other diversity indices. Phylogenetic entropy and five well-known indices (defined in table 1) can be organized into three

analogously related pairs: species richness and phylogenetic diversity, Simpson index and quadratic diversity, and Shannon index and phylogenetic entropy. As indicated in figure 1, the second index of each pair generalizes the first index to incorporate relatedness data, and moreover, these generalizations are mathematically parallel in the following sense: the first index of each pair can be written in the form

$$\sum_i F(p_i), \tag{3}$$

where $F(p)$ is, respectively, equal to $\chi_{(0,1]}$, $p(1 - p)$, and $-p \ln p$ for species richness, Simpson's index, and Shannon's index. The term $\chi_{(0,1]}$ is an indicator function that has value 1 for all $p > 0$ but has value 0 for $p = 0$, i.e., when there is no member of a species. Analogously, the second index can be written in the form

$$\sum_{\text{branches } b} \ell(b) F(p(b)), \tag{4}$$

using the corresponding F for the first index. As a consequence, the second index of each pair reduces to the first in the case where *T* has uniform branches of length 1. That these indices can be written in forms (3) and (4) is manifest except in the cases of the Simpson index, which was shown by Patil and Taillie (1982) to have form (3), and the quadratic diversity index, which is shown in the appendix.

The properties of a diversity index of the form (3) or (4) can be inferred from the shape of $F(p)$ (fig. 3). In particular, the sensitivity of an index to the introduction
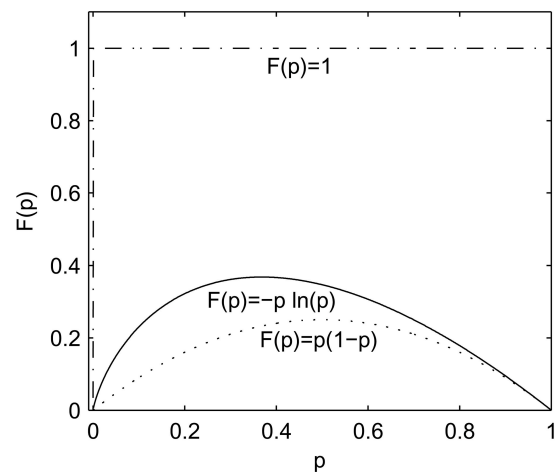
Figure 3: Choices for the function $F(p)$ describing diversity measure dependence on species proportion: $F(p) = \chi_{(0,1]}$ increases the most sharply at $p = 0$, followed by $F(p) = -p \ln p$ and $F(p) = p(1 - p)$.
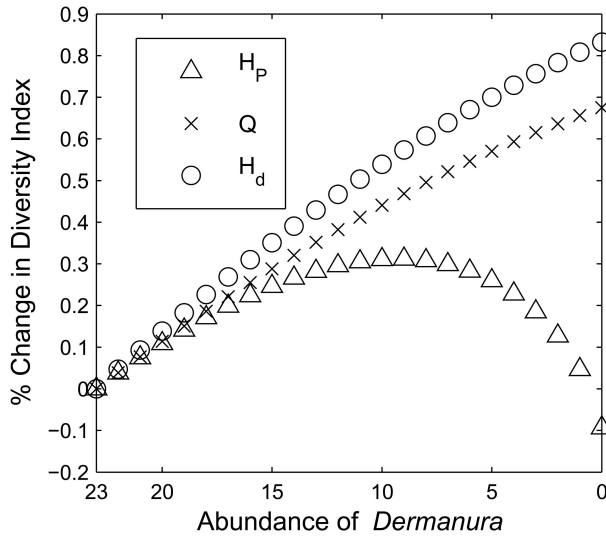
Figure 4: Change in diversity indices as the abundance of *Dermanura* in the forest habitat is reduced to 0.

or elimination of a species is dictated by the steepness by which F increases when p is increased from 0. Indices based on presence or absence of species use $F(p) = \chi_{(0,1]}$, which jumps discontinuously from 0 to 1 at $p = 0$. These indices are more sensitive than indices incorporating abundance, for which the corresponding func-

tions $F(p)$ are continuous. Among the continuous choices for $F(p)$, $F(p) = -p \ln p$ has an unbounded derivative for small p, which leads to greater sensitivity than indices using $F(p) = p(1 - p)$, the derivative of which is bounded.

Indices used to support conservation policy should embody the diversity characteristics we value, including the number of species, their taxonomic or genetic distinctness, and the evenness of their distribution. In the interest of promoting evenness and distinctness, diversity indices may favor reducing the relative proportions of more abundant or less distinct species. However, indices may be expected to protect individuals of rare species. This property, termed "weak species monotonicity" by Weikard et al. (2006), ensures that the goals of promoting diversity and maintaining species richness do not conflict.

It is shown in our case study (fig. 4) and elsewhere (Shimatani 2001; Izsák and Szeidl 2002; Weikard et al. 2006) that Q, H_d, and other indices combining abundance and relatedness data may increase monotonically as a less distinct species is reduced to 0 members, violating weak species monotonicity. We prove in the appendix that phylogenetic entropy always satisfies weak species monotonicity, due to the divergence of the derivative of $F(p) = -p \ln p$ as p is decreased to 0.

The taxonomic entropy of Ricotta and Avena (2003) also satisfies this property. However, due to the choice of

Table 2: Phyllostomid genus abundances in Selva Lacandona habitats

| Genus	Forest	Cacao	Oldfield	Cornfield
Artibeus (Neotropical fruit bat)	149	178	86	96
Carollia (short-tailed fruit bat)	121	259	342	85
Centurio (wrinkle-faced bat)	1	2		2
Chiroderma (big-eyed bat)			1	5
Chorotopterus (big-eared woolly bat)	5			
Dermanura (small fruit-eating bat)	23	17	17	18
Desmodus (vampire bat)	4	3		1
Diphylla (hairy-legged vampire bat)	1		1	
Enchisthenes (velvety fruit-eating bat)	1			
Glossophaga (long-tongued bat)	35	73	103	85
Hylonycteris (Underwood's long-tongued bat)				2
Lampronycteris (orange-throated bat)	1			
Lichonycteris (dark long-tongued bat)				1
Micronycteris (big-eared bat)		2	1	
Mimon (golden bat)	2	1	4	
Phyllostomus (spear-nosed bat)	5		5	
Platyrrhinus (broad-nosed bat)	6	10	17	18
Sturnia (yellow-shouldered bat)	56	137	102	234
Tonatia (round-eared bat)	1	4		1
Trachops (fringe-lipped bat)	2			
Uroderma (tent-making bat)	7	4	2	4
Vamyressa (yellow-eared bat)	1	3		
Vampyrodes (great stripe-faced bat)	1			
Total	422	693	681	552

Source: Adapted from Medellin et al. (2000).
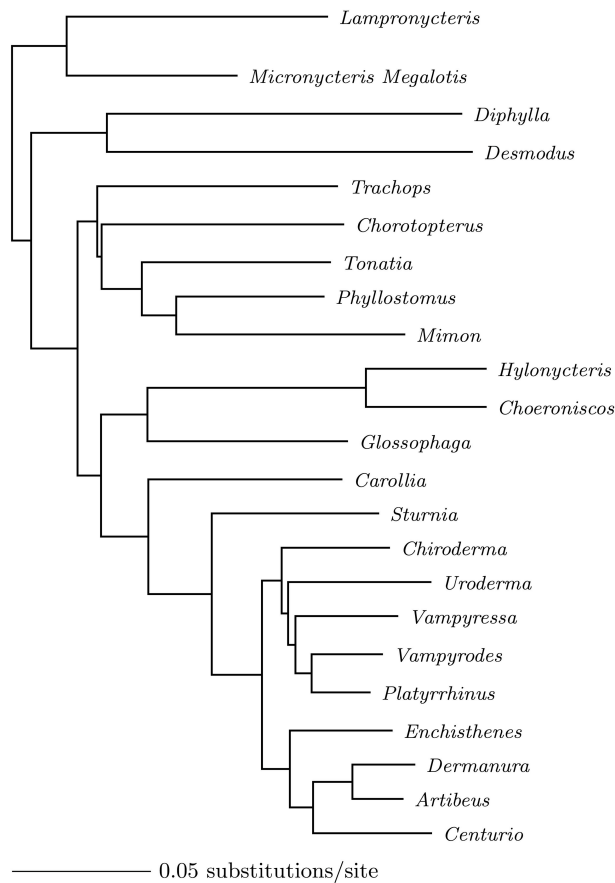
Figure 5: Molecular phylogeny of phyllostomid bat genera in Selva Lacandona habitats, adapted from Baker et al. (2003).

weights used in this index (table 1), it favors increasing the abundances of less distinct species over those of more distinct ones, counter to a conventional understanding of diversity.

Case Study

To further compare phylogenetic entropy with other indices, we investigated the behavior of H_p, Q, and H_d on communities of bats in the Selva Lacandona in Chiapas, Mexico. We used data provided by Medellin et al. (2000), who surveyed bat communities (primarily of the family Phyllostomidae, or leaf-nosed bats) in four habitats of the Selva Lacandona: rainforest, cacao plantations, inactive agricultural plantations with secondary vegetation, and corn plantations. (These habitats are hereafter referred to as forest, cacao, oldfield, and cornfield, respectively.) They found that the diversity of bats (measured using species richness, the Shannon index, number of rare species, and relative abundance of the most common species) decreases with increasing levels of habitat disturbance, indicating that bat diversity may be a useful indicator of habitat integrity.

We combined this abundance data with a phylogenetic study of Phyllostomidae conducted by Baker et al. (2003). Using evidence from mitochondrial DNA and Recombination-Activating Gene-2 sequences, this study organized 48 of the 53 phyllostomid genera into a phylogenetic tree with explicit branch lengths. Though the phylogeny of Phyllostomidae is still the subject of debate (Wetterer et al. 2000; Jones et al. 2002), the analysis of Baker et al. (2003) provides an ample starting point for the study of phylogenetic diversity measures on this family.

In order to apply the phylogeny of Baker et al. (2003) to the abundance data of Medellin et al. (2000), we considered only the phyllostomid bats that were surveyed and grouped them by genus. All genera found in the four habitats are represented in the phylogenetic study except *Lichonycteris* (dark long-tongued bats), of which one individual was found. In order to include this individual in our study, we considered it to belong instead to genus *Choeroniscus* (long-tailed bats), which has been shown in other studies (e.g., Carstens et al. 2002) to occupy a similar phylogenetic position relative to the other genera sampled. Because only a single individual of this genus was found, the effects of this substitution are minimal. The abundance data grouped by genus are shown in table 2, and the phylogenetic tree showing only the relevant genera is given in figure 5.

We used three numerical tests to analyze the indices. First, we evaluated their values in the four habitats and compared the results with other established indices. Second, we evaluated the change in the indices as the abundance of a close relative to a dominant genus was reduced to 0. Finally, we found the distributions that maximize the three indices, which indicate the logical consequence of policy decisions based on them. The indices are concave, so global maxima can be obtained using conventional algorithms (we used optimization commands in Maple). Since only genus-level information is available, the measures were evaluated for diversity at the genus rather than species level.

Table 3: Diversity index values for bat habitats

Index	Forest	Cacao	Oldfield	Cornfield
Genus richness	17	13	12	14
Simpson index	.7647	.7432	.6852	.7466
Shannon index	1.788	1.576	1.474	1.626
Phylogenetic diversity[a]	1.432	.9806	.9212	.972
Quadratic diversity[a]	.0614	.05951	.05658	.059
Phylogenetic entropy[a]	.1333	.1176	.1135	.1178
H_d	.4974	.4683	.4444	.4644

[a] Branch length units are in substitutions per site.

Table 4: Genus abundances in the maximizing assemblages for H_P and Q

Genus	H_P maximizer	Q and H_d maximizer[a]
Artibeus	.01	
Carollia	.035	
Centurio	.033	.08
Chiroderma	.016	
Chorotopterus	.054	.015
Dermanura	.019	.01
Desmodus	.118	.185
Diphylla	.114	.176
Enchisthenes	.015	
Glossophaga	.041	
Hylonycteris	.064	.106
Lampronycteris	.087	.124
Lichonycteris	.064	.106
Micronycteris	.041	
Mimon	.07	.112
Phyllostomus	.028	
Platyrrhinus	.004	
Sturnia	.031	
Tonatia	.039	
Trachops	.053	.005
Uroderma	.036	.081
Vamyressa	.017	
Vampyrodes	.01	

[a] The maximizing assemblages for Q and H_d agree to within 10^{-6} in the genus proportions.

Results

The diversity index values in the four habitats are given in table 3. All indices ranked the forest and oldfield habitats as the most and least diverse, respectively. Of the remaining two habitats, the cacao plantation was rated more diverse by four indices (genus richness, the Shannon index, the Simpson index, and phylogenetic entropy). The remaining three indices ranked the cornfield as more diverse.

In the forest habitat, the minor genus *Dermanura* is a close relative to the dominant genus *Artibeus* and may therefore be considered a lower conservation priority. Some diversity indices may favor reductions in *Dermanura* abundance in order to increase the relative abundances of more distinct species. We considered abundance values between 0 and 23 individuals for *Dermanura*, leaving other abundances fixed, and calculated the change in the indices (fig. 4). Both Q and H_d favor reducing the abundance to 0, whereas H_P is maximized for 9 individuals.

The genus assemblages that maximize each index are given in table 4. Interestingly, the distributions that maximize H_d and Q coincide to within 10^{-6} in the genus proportions. These distributions contain 11 genera, spread more or less evenly through the phyllostomid subfamilies. The distribution that maximizes H_P contains all 23 genera

and is ranked more diverse by every diversity index aside from Q and H_d (table 5).

The abundance of a genus in the maximizing distributions correlates with its total phylogenetic distance to all other genera (fig. 6), though the relationship is stronger for phylogenetic entropy ($R^2 = 0.946$) than for Q and H_d ($R^2 = 0.703$).

Discussion

Our findings suggest that the diversity indices we compared value richness, evenness, and distinctness. This can be seen from the index values in the four habitats, for which the indices correlated with genus richness, evenness indices (Shannon's index and Simpson's index), and phylogenetic diversity, a measure of distinctness. However, Q and H_d both favor complete elimination of a less distinct taxon, while H_P favors the preservation of all taxa.

The observation that the maximizing distributions for Q and H_d are similar suggests a close relationship between these indices. In the expression

$$H_d = -\sum_i p_i \ln\left(1 - \sum_{j \neq i} d_{ij} p_j\right),$$

the distances d_{ij} are normalized by dividing by the maximum possible distance. If the distances between species are small compared with the maximum distance, the values of $\sum_{j \neq i} d_{ij} p_j$ will be small. Using $\ln(1 - x) \approx -x$ for small x yields

$$H_d \approx -\sum_i p_i \left(-\sum_{j \neq i} d_{ij} p_j\right) = 2Q.$$

This suggests that H_d and Q are similar for many communities.

Shimatani (2001) was the first to recognize that maximizing quadratic diversity can decrease richness. As noted by Pavoine et al. (2005), this occurs only when the leaves of the phylogenetic tree are different distances from the

Table 5: Diversity index values on the index-maximizing assemblages

Index	H_P maximizer	Q and H_d maximizer
Genus richness	23	11
Simpson index	.935	.871
Shannon index	2.899	2.138
Phylogenetic diversity	1.6982	1.1253
Quadratic diversity	.1136	.1197
Phylogenetic entropy	.2963	.2628
H_d	1.227	1.347
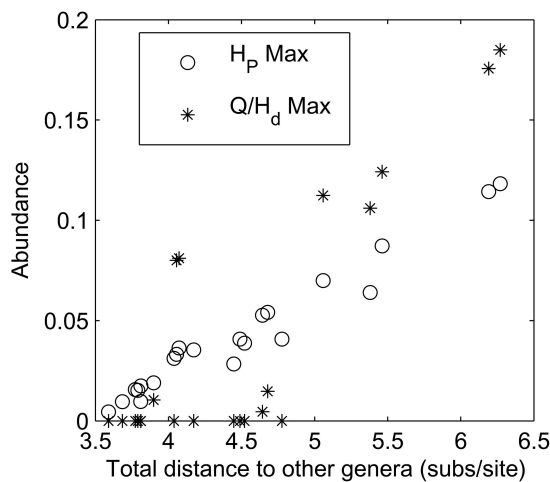
Figure 6: Genus abundances in the maximizing genus distributions for the three indices plotted against the total phylogenetic distance from a genus to all other genera.

root. In such situations, Q and H_d can favor eliminating the species that are less distinct (relative to the other species present), thereby maximizing interspecific distances. Phylogenetic entropy does not favor eliminating species for any tree and appears to favor distributions in which the abundance of a species is linearly related to its aggregate distance to all other species.

Conclusion

Phylogenetic entropy generalizes the Shannon index to reflect the priority of preserving highly distinct species. Unlike when using quadratic diversity and some other measures combining abundance and relatedness data, basing conservation priority on phylogenetic entropy also protects species richness. These qualities make our index attractive for diversity studies in which both abundance and phylogenetic data are relevant and available.

The general forms (3) and (4) can be used to create a variety of indices, some of which may have significant special properties in the context of biodiversity measurement. A more thorough investigation of the link between the properties of the function F and the behavior of the corresponding diversity index may yield useful indices as well as insight into mathematical properties of diversity.

Literature Cited

Baker, R., S. Hoofer, C. Porter, and R. Van Den Bussche. 2003. Diversification among New World leaf-nosed bats: an evolutionary hypothesis and classification inferred from digenomic congruence of DNA sequence. Occasional Papers of the Museum of Texas Tech University 230:1–32.

Carstens, B., B. Lundrigan, and P. Myers. 2002. A phylogeny of the Neotropical nectar-feeding bats (Chiroptera: Phyllostomidae) based on morphological and molecular data. Journal of Mammalian Evolution 9:23–53.

Casquilho, J., M. Neves, and F. Rego. 1997. Extensions of the Shannon function and the equilibria of proportions—an application to the land mosaic. Anais do Instituto Superior de Agronomia (Portugal) 46:77–99.

Crozier, R. 1997. Preserving the information content of species: genetic diversity, phylogeny, and conservation worth. Annual Review of Ecology and Systematics 28:243–268.

Ehrlich, P. 1988. The loss of diversity: causes and consequences. Pages 21–27 *in* E. Wilson, ed. Biodiversity. National Academy Press, Washington, DC.

Faith, D. P. 1992. Conservation evaluation and phylogenetic diversity. Biological Conservation 61:1–10.

Guiasu, R., and S. Guiasu. 2003. Conditional and weighted measures of ecological diversity. International Journal of Uncertainty, Fuzziness, and Knowledge-Based Systems 11:283–300.

Helmus, M., T. Bland, C. Williams, and A. Ives. 2007. Phylogenetic measures of biodiversity. American Naturalist 169:E68–E83.

Humphries, C., P. Williams, and R. Vane-Wright. 1995. Measuring biodiversity value for conservation. Annual Review of Ecology and Systematics 26:93–111.

Izsák, J., and L. Papp. 2000. A link between ecological diversity indices and measures of biodiversity. Ecological Modelling 130:151–156.

Izsák, J., and L. Szeidl. 2002. Quadratic diversity: its maximization can reduce the richness of species. Environmental and Ecological Statistics 9:423–430.

Jones, K., A. Purvis, A. MacLarnon, O. Bininda-Emonds, and N. Simmons. 2002. A phylogenetic supertree of the bats (Mammalia: Chiroptera). Biological Reviews 77:223–259.

May, R. 1990. Taxonomy as destiny. Nature 347:129–130.

Medellin, R., M. Equihua, and M. Amin. 2000. Bat diversity and abundance as indicators of disturbance in Neotropical rainforests. Conservation Biology 14:1666–1675.

Nixon, K., and Q. Wheeler. 1992. Measures of phylogenetic diversity. Pages 216–234 *in* M. Novacek and Q. Wheeler, eds. Extinction and phylogeny. Columbia University Press, New York.

Patil, G. P., and C. Taillie. 1982. Diversity as a concept and its measurement. Journal of the American Statistical Association 77:548–567.

Pavoine, S., S. Ollier, and D. Pontier. 2005. Measuring diversity from dissimilarities with Rao's quadratic entropy: are any dissimilarities suitable? Theoretical Population Biology 67:231–239.

Rao, C. R. 1982. Diversity and dissimilarity coefficients: a unified approach. Theoretical Population Biology 21:24–43.

Ricotta, C. 2002. Bridging the gap between ecological diversity indices and measures of biodiversity with Shannon's entropy: comment to Izsák and Papp. Ecological Modelling 152:1–3.

———. 2004. A parametric diversity measure combining the relative abundances and taxonomic distinctiveness of species. Diversity and Distributions 10:143–146.

Ricotta, C., and G. C. Avena. 2003. An information-theoretical measure of taxonomic diversity. Acta Biotheoretica 51:35–41.

Ricotta, C., and L. Szeidl. 2006. Towards a unifying approach to diversity measures: bridging the gap between the Shannon entropy and Rao's quadratic index. Theoretical Population Biology 70:237–243.

Shannon, C. 1948. A mathematical theory of communication. Bell Systems Technological Journal 27:379–423.

Shimatani, K. 2001. On the measurement of species diversity incorporating species differences. Oikos 93:135–147.

Simpson, E. H. 1949. Measurement of diversity. Nature 163:688.

Solow, A., S. Polasky, and J. Broadus. 1993. On the measurement of biological diversity. Journal of Environmental Economics and Management 24:60–68.

Vane-Wright, R., C. Humphries, and P. Williams. 1991. What to protect: systematics and the agony of choice. Biological Conservation 55:235–254.

Warwick, R. M., and K. R. Clarke. 1995. New "biodiversity" measures reveal a decrease in taxonomic distinctness with increasing stress. Marine Ecology Progress Series 129:301–305.

Webb, C., D. Ackerly, M. McPeek, and M. Donoghue. 2002. Phylogenies and community ecology. Annual Review of Ecology and Systematics 33:475–505.

Weikard, H., M. Punt, and J. Wesseler. 2006. Diversity measurement combining relative abundances and taxonomic distinctiveness of species. Diversity and Distributions 12:215–217.

Wetterer, A., M. Rockman, and N. Simmons. 2000. Phylogeny of phyllostomid bats (Mammalia: Chiroptera): data from diverse morphological systems, sex chromosomes, and restriction sites. Bulletin of the American Museum of Natural History 248:1–200.

Associate Editor: Ross H. Crozier
Editor: Monica A. Geber
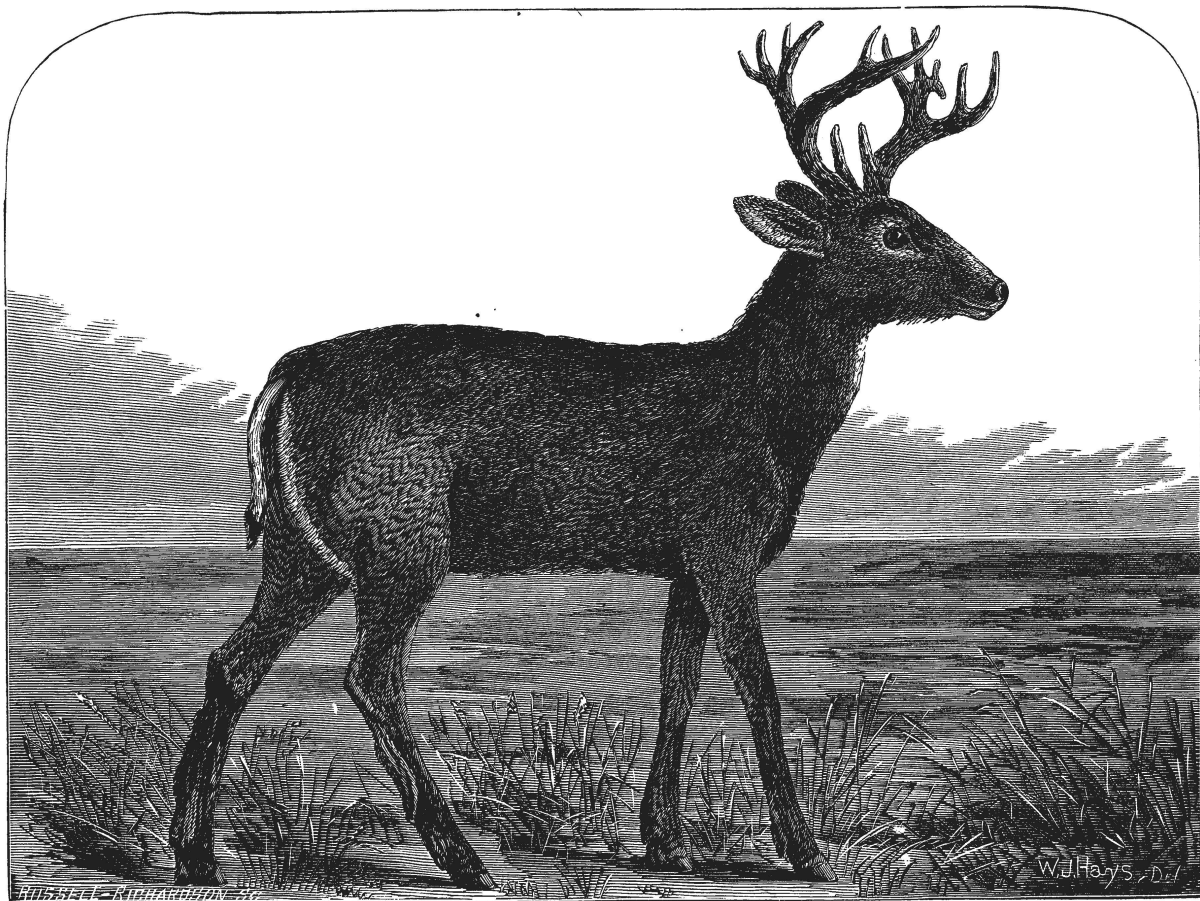
Mule deer (*Cervus macrotis*) from "The Mule Deer" by W. J. Hays (*American Naturalist*, 1869, 3:180–181)

Appendix from B. Allen et al., "A New Phylogenetic Diversity Measure Generalizing the Shannon Index and Its Application to Phyllostomid Bats"

(Am. Nat., vol. 174, no. 2, p. 000)

Proofs of Mathematical Results

Axiomatic Characterization of Phylogenetic Entropy

One argument for the naturalness of the Shannon index is that it is uniquely characterized by the following axioms, up to a constant multiplicative factor representing a choice of units (Shannon 1948): (1) H is a continuous function of the species proportions vector $\mathbf{p} = (p_1, \ldots, p_n)$. (2) If the species are evenly distributed so that each $p_i = 1/n$, H increases monotonically with the number of species n. (3) If the individuals in a single species s_i are replaced by an equal number of individuals in two new species s_j and s_k, H increases by the proportion $p_j + p_k$ of individuals in either of these species, times the diversity $H(p_j/(p_j + p_k), p_k/(p_j + p_k))$ of the subcommunity composed of these species.

The phylogenetic entropy has a similar axiomatic characterization, using axioms that generalize those above to incorporate relatedness data in the form of a phylogenetic tree: (1′) H_P is a function of the topology and branch lengths of T and the species proportions $\mathbf{p}$ and varies continuously with the branch lengths and species proportions. (2′) If the species are evenly distributed and equally distinct, H_P increases monotonically with the number of species. (3′) If the individuals in a single species are replaced by an equal number of individuals in (a) two new species, in such a way that a portion of a terminal branch is replaced by two new branches of the same length as the replaced portion, then H_P increases by the proportion of individuals in either of the new species times the diversity H_P of the subcommunity composed of these species; (b) one new species, in such a way that only the length of their terminal branch is changed, the change in H_P depends only on the change in branch length and the number of individuals in the species. When restricted to trees with branches of a fixed uniform length, each of the axioms above reduces to the corresponding axiom of Shannon.

The phylogenetic entropy is, up to a choice of units, the only mathematical index satisfying these axioms. The following proof demonstrates this result by considering a hypothetical function $F(T, \mathbf{p})$ that satisfies the axioms. We first show that $F(T, \mathbf{p}) = cH_P(T, \mathbf{p})$, for some constant c, whenever T has branches of uniform length or, equivalently, the pairwise distances between species are equal. We then show that this equality is preserved under the two operations of axiom 3′: splitting a portion of a terminal branch into two branches of equal length and changing the length of a terminal branch. Since any pair $(T, \mathbf{p})$ can be obtained by iterating these operations on an appropriately chosen initial tree with uniform branch lengths, this will show $F = cH_P$ for all trees and distributions.

If the branches of T are of uniform length ℓ, the unique characterization theorem for Shannon's index implies that $F(T, \mathbf{p}) = f(\ell)H(\mathbf{p})$, where $H(\mathbf{p})$ is the Shannon index of the community and $f(\ell)$ is a continuous function (independent of $\mathbf{p}$).

We now determine the function f. If ℓ is 0, the community merges into a single species, and the value of $H(\mathbf{p})$ is 0. Continuity implies that $F(T, \mathbf{p}) = f(\ell)H(\mathbf{p})$ approaches 0 as ℓ approaches 0. Since $H(\mathbf{p})$ is independent of ℓ for $\ell > 0$, $f(\ell)$ must approach 0 as ℓ approaches 0, and $f(0) = 0$. Furthermore, if ℓ is changed by any amount $\Delta\ell$, axiom 3′b (applied to each branch) implies that the resulting change in f should depend only on the amount of change $\Delta\ell$ and not on the initial length ℓ. In particular,

$$f(\ell + \Delta\ell) - f(\ell) = f(\Delta\ell) - f(0) = f(\Delta\ell),$$

or, rearranging terms, $f(\ell + \Delta\ell) = f(\ell) + f(\Delta\ell)$; that is, f is linear. So $f(\ell) = c\ell$ for some constant c, and it follows that for trees with uniform branch lengths,

1

$$F(T, \mathbf{p}) = c\ell H(\mathbf{p})$$

$$= -c\ell \sum_i p_i \ln p_i \tag{A1}$$

$$= -c \sum_{\text{branches } b} \ell p(b) \ln p(b).$$

Thus, the proposition is true whenever T has uniform branch lengths.

We now show that the equality $F(T, \mathbf{p}) = cH_{\text{P}}(T, \mathbf{p})$ is preserved under the two operations of axiom 3′, starting with the branch-splitting operation. Suppose that from a terminal branch b_0 in a given tree, a portion of length ℓ is split into two new branches b_1 and b_2, so that the length of the new branches is also ℓ and $p(b_0) = p(b_1) + p(b_2)$, in accordance with axiom 3′a. From equation (A1), the value of F in the subcommunity of these two species is

$$-c\ell \left(\frac{p(b_1)}{p(b_0)} \ln \frac{p(b_1)}{p(b_0)} + \frac{p(b_2)}{p(b_0)} \ln \frac{p(b_2)}{p(b_0)} \right)$$

$$= -c \frac{\ell}{p(b_0)} (p(b_1) \ln p(b_1) + p(b_2) \ln p(b_2) - (p(b_1) + p(b_2)) \ln p(b_0))$$

$$= -c \frac{\ell}{p(b_0)} (p(b_1) \ln p(b_1) + p(b_2) \ln p(b_2) - p(b_0) \ln p(b_0)).$$

Axiom 3′a tells us that the resulting increase in F is $p(b_0)$ times the above quantity, that is,

$$-c\ell (p(b_1) \ln p(b_1) + p(b_2) \ln p(b_2) - p(b_0) \ln p(b_0)).$$

This is equal to the increase in cH_{P} under the replacement of the portion of length ℓ of branch b_0 by b_1 and b_2. Thus $F = cH_{\text{P}}$ is preserved under branch splitting.

Now suppose the length of a terminal branch b_0 in a given tree is changed by an amount $\Delta\ell$. We show that the change in F is equal to the change in H_{P}, which is $-c\Delta\ell p(b_0) \ln p(b_0)$, demonstrating that $F = cH_{\text{P}}$ is preserved under the branch resizing operation as well.

By axiom 3′b, the change in F depends only on $\Delta\ell$ and $p(b_0)$ and not on the tree to which branch b_0 is attached. We write

$$\Delta F = -c\Delta\ell p(b_0) \ln p(b_0) + g(\Delta\ell, p(b_0))$$

for some unknown continuous function g, and we prove that $g(\Delta\ell, p) = 0$ for any values of $\Delta\ell$ and p.

Suppose that each branch of a tree with uniform branch lengths and proportions $p_1, \dots, p_n$ is lengthened by $\Delta\ell$. From equation (A1), the value of F changes by

$$-c\Delta\ell \sum_{i=1}^{n} p_i \ln p_i$$

under this lengthening. This implies that

$$\sum_{i=1}^{n} g(\Delta\ell, p_i) = 0,$$

for any value of $\Delta\ell$, whenever $p_1 + \dots + p_n = 1$. Substituting $p_i = 1/n$ yields $g(\Delta\ell, 1/n) = 0$ for arbitrary n. Then for any natural numbers $k < m$, setting one species proportion to k/m and all others to $1/m$ yields $g(\Delta\ell, k/m) = 0$. Since we have shown $g(\Delta\ell, p)$ is 0 for all rational values of p and all values of $\Delta\ell$, continuity guarantees that it is 0.

Alternate Form of Quadratic Diversity

In order to complete the proof of the correspondence of the six diversity indices shown in figure 1, we show that the quadratic diversity may be written as

$$Q = \sum_{\text{branches } b} \ell(b)p(b)(1 - p(b)). \tag{A2}$$

A special case of this result using taxonomically defined distances was proven by Shimatani (2001).

Consider a specific branch b of T. Let $i_1, \ldots, i_k$ index the species descended from b and $i_{k+1}, \ldots, i_n$ index the species not descended from b. Then the term $\ell(b)p(b)(1 - p(b))$ can be written as

$$\ell(b)(p_{i_1} + \ldots + p_{i_k})(p_{i_{k+1}} + \ldots + p_{i_n}).$$

This expression contains the term $p_i p_j$ only if species s_i is descended from branch b and species s_j is not, or vice versa. Therefore, in equation (A2), the overall coefficient of $p_i p_j$ will be the total length of all branches representing ancestors of species s_i but not species s_j or vice versa. This length is precisely the distance d_{ij} between species s_i and s_j on the tree. Thus,

$$\sum_{\text{branches } b} \ell(b)p(b)(1 - p(b)) = \sum_{i<j} d_{ij} p_i p_j,$$

which is the definition of quadratic diversity.

Weak Species Monotonicity

Phylogenetic entropy decreases when a rare species is removed from a community. This property, weak species monotonicity, is not shared by quadratic diversity or Ricotta and Szeidl's H_d index (fig. 4). The proof that H_P satisfies this property rests on the fact that the derivative of $F(p) = -p \ln p$, $F'(p) = -(1 + \ln p)$, diverges to positive infinity as p is decreased to 0.

Suppose, in a community with initial distribution $(p_1(0), p_2(0), \ldots, p_n(0))$, that the first species is reduced to zero abundance along the trajectory

$$(p_1(t), p_2(t), \ldots, p_n(t)) = \left(p_1(0) - t, p_2(0)\left(\frac{1 - p_1(0) + t}{1 - p_1(0)}\right), \ldots, p_n(0)\left(\frac{1 - p_1(0) + t}{1 - p_1(0)}\right)\right),$$

for t varying from 0 to $p_1(0)$. This corresponds to a linear decrease in the first species while the others remain in fixed proportion.

Let b_1 be the terminal branch above the first species, $b_2, \ldots, b_k$ the nonterminal branches between the first species and the root, and $b_{k+1}, \ldots, b_m$ the branches not lying above the first species. Then the branch proportions at time t are given by

$$p(b_i, t) = \begin{cases} p_1(0) - t & i = 1 \\ (p(b_i, 0) - p_1(0))\left(\dfrac{1 - p_1(0) + t}{1 - p_1(0)}\right) + p_1(0) - t & 2 \leq i \leq k \\ p(b_i, 0)\left(\dfrac{1 - p_1(0) + t}{1 - p_1(0)}\right) & k + 1 \leq i \leq m \end{cases}.$$

We compute the rate of change in H_P with t:

$$\frac{d}{dt} H_{\mathrm{P}} = \ell(b_1)[1 + \ln(p_1(0) - t)]$$

$$- \sum_{i=2}^{k} \ell(b_i) \left\{ 1 + \ln \left[(p(b_i, 0) - p_1(0)) \left(\frac{1 - p_1(0) + t}{1 - p_1(0)} \right) + p_1(0) - t \right] \right\}$$

$$\times \left(\frac{p(b_i, 0) - p_1(0)}{1 - p_1(0)} - 1 \right)$$

$$- \sum_{i=k+1}^{m} \ell(b_i) \left\{ 1 + \ln \left[p(b_i, 0) \left(\frac{1 - p_1(0) + t}{1 - p_1(0)} \right) \right] \right\} \frac{p(b_i, 0)}{1 - p_1(0)}.$$

As t approaches $p_1(0)$ (i.e., the abundance of the first species approaches 0), the above expression becomes dominated by $\ell(b_1) \ln(p_1(0) - t)$ and approaches negative ∞. Thus, there is a threshold beyond which any further decline in the first species will cause H_{P} to decrease.