Hedging Our Bets: The Expected Contribution of Species to Future Phylogenetic Diversity

Mike Steel[1], Aki Mimoto[2] and Arne Ø. Mooers[2, 3]

[1]Biomathematics Research Centre, University of Canterbury, Christchurch, New Zealand. [2]IRMACS, Simon Fraser University, Burnaby, Canada. [3]Institute for Advanced Study, Berlin, Germany.

Abstract: If predictions for species extinctions hold, then the 'tree of life' today may be quite different to that in (say) 100 years. We describe a technique to quantify how much each species is likely to contribute to future biodiversity, as measured by its expected contribution to phylogenetic diversity. Our approach considers all possible scenarios for the set of species that will be extant at some future time, and weights them according to their likelihood under an independent (but not identical) distribution on species extinctions. Although the number of extinction scenarios can typically be very large, we show that there is a simple algorithm that will quickly compute this index. The method is implemented and applied to the prosimian primates as a test case, and the associated species ranking is compared to a related measure (the 'Shapley index'). We describe indices for rooted and unrooted trees, and a modification that also includes the focal taxon's probability of extinction and which links two complementary approaches to conserving phylogenetic diversity.

Keywords: phylogenetic diversity, extinction, biodiversity conservation, Shapley index

Introduction

Within a given taxonomic group, individual biological species are generally considered to be of equal or near-equal biodiversity value. So, for instance, areas with a greater number of species are more valuable than those with fewer (Myers et al. 2000). When wild species are ranked by value, this is usually based on their threat of extinction (see e.g. SARA, 2002). However, as pointed out by Cousins (1991), species are discovered and identified because they are different from other species, which suggests that they may differ in value. In the context of conservation, Avise (2005) has highlighted five different currencies for valuing species: rarity, distribution, ecology, charisma, and phylogeny. Here, we consider the value of a species based on its position in a phylogeny. A phylogeny is the directional, acyclic graph depicting relationships between leaves (species), which we define formally in the next section. A phylogeny generally has a root (which assigns direction) and edge weights that can represent unique feature diversity (e.g. as measured by evolutionary time or genetic distance). Species can be defined by the features they possess, and one measure of their worth is the expected contribution of their genetic, morphological or evolutionary distinctiveness. To this end, we can use a phylogeny to assign a measure of evolutionary value to a species, since branch lengths can be chosen to correspond to genetic, morphological or evolutionary distance. Because of the highly imbalanced shape of the Tree of Life, some species in a phylogeny will have far fewer close relatives than others in that phylogeny (Mooers and Heard, 1997), and these more distantly-related species will be expected to contribute more unique features (Faith, 1992).

Phylogenetic measures of conservation value have a long pedigree (see e.g. Altschul and Lipman, 1990; May, 1990) and have begun to be explored in some detail (Haake et al. 2005; Hartmann and Steel, 2007; Pavoine et al. 2005a, 2005b; Redding and Mooers, 2006). So, for example, Pavoine and colleagues presented one new phylogenetic measure of conservation value, a set of sampling weights such that the expected pairwise distance on the tree is maximized. Haake and colleagues extended the 'Shapley value' (Shapley, 1953) from co-operative game theory to the conservation setting to calculate the average distance of a focal species to all possible subsets of taxa. For both measures, species with high conservation scores are those expected to contribute more to the resulting sets. Yet another measure that uniquely apportions the tree to its tips (Redding, 2003; Isaac et al. 2007) and which is the focus of a new international conservation initiative (the EDGE initiative, Zoological Society of London) scales

Correspondence: Mike Steel, Biomathematics Research Centre, University of Canterbury, Christchurch, New Zealand. Tel: +64-33667001; Fax: +64-33642587; Email: m.steel@math.canterbury.ac.nz

almost perfectly with the Shapley value (unpublished results).

One question with these measures concerns the set of species that individual species are asked to complement. For instance, given known extinction probabilities for species, some future sets of species are much more likely than others and so some species will be more valuable because their close relatives are less likely to be included in future sets of species. Here we formalize this idea to extend the Shapley value of a species to include preassigned extinction probabilities, and show how this value can be efficiently computed. We then compare our measure with the original Shapley value using the prosimian primates as a test case.

In this paper we consider only extinction (neglecting possible speciation), because our focus is on the impact of current high rates of extinction over relatively short time frames (hundreds of years) for which little speciation may be expected. Also, although our indices rank species for conservation, we emphasise that such conservation includes safeguarding natural habitats.

Definition Let $\mathcal{T}$ be a rooted or unrooted phylogenetic tree with leaf set X, together with an assignment of positive lengths to the edges (branches) of $\mathcal{T}$. We let $l(e)$ denote the length of edge e, and let $E(\mathcal{T})$ denote the set of edges of $\mathcal{T}$. For a subset S of X, let $PD(S)$ denote the *phylogenetic diversity* of S defined as follows. If $\mathcal{T}$ is unrooted then $PD(S)$ is the sum of the lengths of the edges (branches) of $\mathcal{T}$ in the minimal subtree that connects S. If $\mathcal{T}$ is rooted, then $PD(S)$ is the sum of the lengths of the edges of $\mathcal{T}$ in the minimal subtree of $\mathcal{T}$ that connects S and the root of the tree. Figure 1 illustrates these concepts, and includes values at the tips that we will use in the next section. Note that although the branch lengths in this example are clock-like, this assumption is not required in any of the results we describe.

The HED Index

For a leaf $i \in X$, and a subset $S \subseteq X - \{i\}$ let

$$\Delta_{PD}(S, i) := PD(S \cup \{i\}) - PD(S).$$

The quantity $\Delta_{PD}(S, i)$ measures how much phylogenetic diversity i contributes to the tree that one obtains from $\mathcal{T}$ once species not in S have been pruned out (for example if they go extinct).
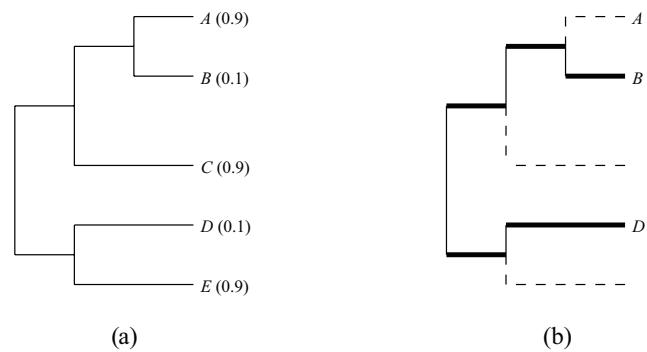
Figure 1. (**a**) A small rooted tree with edge lengths (of 2 units for the terminal edges incident with C, D, E, and 1 unit for the other five edges). Each tip j has an associated extinction probability $P(ext) = \epsilon_j$. (**b**) For a subset $S = \{B, D\}$ of taxa that are extant at some future time, the phylogenetic diversity score $PD(S)$ is the sum of the lengths of the edges indicated in bold. The dashed edges lead to extinct taxa.

Alternatively, $\Delta_{PD}(S, i)$ is the marginal increase in phylogenetic diversity of S if i is added.

Now, suppose that each species has an associated extinction probability $P(ext)$ (which may vary from species to species)—for example, this may be the probability that the species is extinct in (say) 100 years from now (either globally, or in some specified community). We will denote this $P(ext)$ value for species j by ϵ_j. In this paper we consider the simplest model which assumes that the extinction of each species in X comprise independent events. Given $i \in X$, let $\mathcal{S}_i$ denote the random subset of species in $X - \{i\}$ which survive (i.e. do not go extinct).

By the independence assumption we have:

$$\mathbb{P}[\mathcal{S}_i = S] = \prod_{j \in S}(1 - \epsilon_j) \times \prod_{j \in X - \{i\} - S} \epsilon_j.$$

For $i \in X$, let ψ_i denote the expected value of $\Delta_{PD}(\mathcal{S}_i, i)$. That is,

$$\psi_i = \mathbb{E}\left[\Delta_{PD}(\mathcal{S}_i, i)\right]$$
$$= \sum_{S \subseteq X - \{i\}} \mathbb{P}[\mathcal{S}_i = S] \Delta_{PD}(S, i). \qquad (1)$$

We call ψ_i the *heightened evolutionary distinctiveness* of species i, and the function $i \mapsto \psi_i$ the *heightened evolutionary distinctiveness* (HED) index for $\mathcal{T}$. Notice that if all the species in $X - \{i\}$ were guaranteed to survive, then ψ_i would be just the length of the pendant edge incident with

leaf i, however random extinctions mean that ψ_i will tend to be increased ('heightened') over this pendant edge length.

The quantity ψ_i measures the expected additional phylogenetic diversity species i would contribute at some future time if it is extant rather than extinct, given our uncertainty about which other species may have also be extinct at this future time. It is an example of a type of 'distinctiveness' measure described by Equation (2) of Weitzmann (1998). Notice that the extinction risk of species i does not influence ψ_i, as this quantity depends only on the extinction risks of other species. We will discuss this issue (and describe a related index that does incorporate the extinction probability of species i) later in the paper.

A related but different index, based on the Shapley value in co-operative game theory, has recently been described by Haake et al. (2005). This index, denoted here as ψ^{sh} can be defined (for unrooted trees) as follows: For $i \in X$,

$$\psi_i^{\text{sh}} = \frac{1}{|X|} \sum_{S \subseteq X - \{i\}} \binom{|X|-1}{|S|}^{-1} \Delta_{PD}(S, i). \quad (2)$$

This index has certain appealing properties. In particular, $\sum_{i \in X} \psi_i^{\text{sh}} = PD(X)$, and there is a simple formula for quickly computing ψ_i^{sh}. The index ψ^{sh} also has a stochastic interpretation, but this is not based on extinction or survival of species, rather on the expected contribution to PD of each species under all possible orderings of the total set of species (for details see Haake et al. 2005). The index ψ^{sh} allocates existing PD 'fairly' amongst the species, whereas ψ quantifies the expected contribution of each species to future PD.

Computing the HED Index

Computing the HED index directly via (1) could be problematic as it requires summation over all the subsets of $X - \{i\}$ and this grows exponentially with $|X|$. However we now show that the index can be readily and quickly computed, both for rooted and unrooted trees. This polynomial-time algorithm for computing ψ thus complements (but is quite different to) the polynomial-time algorithm described by Haake et al. (2005) for computing ψ^{sh}.

Rooted trees

For a rooted phylogenetic X–tree $\mathcal{T}$, and one of its edges e, let $C(e)$ denote the set of species in X that are descended from e (i.e. the clade that results from deleting e from $\mathcal{T}$). For $i \in X$, let $e_1, e_2, ..., e_k (k = k(i) \geq 1)$ denote the edges (branches) on the path from i to the root of $\mathcal{T}$, listed in the order they are visited by that path. Recall that $l(e)$ denotes the length of edge e. The proof of the following theorem is given in the Appendix.

Theorem 3.1

$$\psi_i = \sum_{r=1}^{k} l(e_r) \prod_{j \in C(e_r) - \{i\}} \epsilon_j$$

Note that in this (and the next) theorem we adopt the convention $\prod_{j \in \emptyset} \epsilon_j = 1$, which is relevant for the first term ($r = 1$) in the sum as $C(e_r) - \{i\}$ is empty. Thus the first term in the summation expression for ψ_i given by Theorem 3.1 is simply $l(e_1)$, the length of the pendant edge of $\mathcal{T}$ incident with species i.

Example

We can apply the HED index to the members of the rooted tree depicted in Figure 1. For example, to compute ψ_A by using Theorem 3.1 we have $\psi_A = 1 + 1 \cdot \epsilon_B + 1 \cdot \epsilon_B \epsilon_C = 1.19$. By inspection, we can see that the most valuable species will be D, since it shares an edge with only one other species above the root, and that this species (E) has a high $P(ext)$. At the other extreme, A shares its path to the root with two other species, and one of them (B) has a low $P(ext)$. It should therefore receive a low HED value. The computed values are $\psi_D = 2.9$, $\psi_B = 2.71$, $\psi_E = 2.1$, $\psi_C = 2.09$, and $\psi_A = 1.19$. Using the Shapley index (Haake et al. 2005), D and E are ranked first (with value $= 2.63$), followed by C (2.33) and then A and B (1.75). Pavoine's QE metric (Pavoine et al. 2005) returns the same ranking as does the Shapley. A portal for computing HED is available at http://www.disconti.nu/-phylo/emd.dpf

Unrooted trees, and properties of the index

We now provide a similar formula for efficiently computing the HED index for unrooted trees.

Given a leaf i of $\mathcal{T}$ and an edge e of $\mathcal{T}$, e induces a split of X into two disjoint subsets, and one of these subsets, which we denote as $C_i(e)$, contains i. The proof of the following theorem is given in the Appendix.

Theorem 3.2

$$\psi_i = \sum_{e \in E(T)} l(e) \cdot \left(\prod_{j \in C_i(e)-\{i\}} \epsilon_j \right) \cdot \left(1 - \prod_{j \in X - C_i(e)} \epsilon_j \right)$$

Notice that the rooted HED index is just a special case of the unrooted HED index (indeed Theorem 3.1 can be deduced from Theorem 3.2). To see this, given a rooted tree $\mathcal{T}$ attach a new leaf ρ to the root via a new edge to obtain an unrooted tree, and assign the new edge length 0. Let $\epsilon_\rho = 0$. Then it is easily seen that the HED index for $\mathcal{T}$ is just the HED index for the derived unrooted tree.

Using Theorem 3.1 it can be shown that if $\mathcal{T}$ is a rooted phylogenetic tree then the condition:

$$\sum_{i \in X} \psi_i = PD(X), \tag{3}$$

holds for all selections of positive branch lengths and ϵ's if and only if $\mathcal{T}$ is a 'star tree' (that is, every leaf is adjacent to the root). Moreover Theorem 3.2 shows that there is no unrooted phylogenetic tree $\mathcal{T}$ for which (3) holds for all positive branch lengths and ϵ values (of course (3) may hold on phylogenetic trees—either rooted or unrooted—if the branch lengths and ϵ values take certain values). This contrasts with the index ψ^{sh} which satisfies $\sum_{i \in X} \psi_i^{sh} = PD(X)$ on all unrooted phylogenetic trees and choices of branch lengths, a property that is referred to as the Pareto efficiency axiom by Haake et al. (2005). In the setting of this paper we should not be surprised that (3) holds for ψ only in very special cases since we are not trying to divide out existing PD amongst present taxa (one motivation behind ψ^{sh}) but rather quantify the expected contribution each species makes to future PD.

Application
We compared the HED index with the Shapley (Haake et al. 2005) values for the Prosimians (Mammalia: Primata), a group of approximately 50 species with a broad range of extinction probabilities. This group includes the Aye-Aye, the lemurs, the lorises and galagos. We made use of a recent dated Supertree of the order Primates (Vos and Mooers, 2004; Vos, 2006), see Figure 2, and Red List risk designations from the IUCN (www.iucnredlist.org, accessed February 2006). Following Isaac et al. (2007) and Redding and Mooers (2006), we first converted the five categories of risk (CR, EN, VU, NT, and LC) to probabilities of extinction. Under the IUCN criteria, the species in the VU category are given a $P(ext) = 0.1$ over the next 100 years. We gave the lowest and highest threat categories very conservative probabilities of extinction over the next 100 years of 0.001 and 0.9 respectively, leaving $P(ext) = 0.5$ for EN, and $P(ext) = 0.01$ for NT: this scale is very similar to that calculated from real population viability analyses for birds (Redding and Mooers, 2006). We are primarily interested in how the ranking of species changes using different approaches.

The bivariate correlation between the metrics is high (0.94). Both measures chose the Aye-Aye (*Daubentonia madagascarensis*) as the most important species, followed by *Perodicticus potto*. Interestingly, the three most highly ranked species under current conservation policy (the critically endangered lemurs *Propithecus tattersalli, Hapalemur simus, H. aureus*) are nested well up in the tree (Fig. 2) such that none of them were chosen in the top ten for either SV or HED. If we compare the rest of the rankings for these two metrics, the largest single difference is for the two *Arctocebus* species: they rank twelfth under SV (being relatively isolated on the tree), but only twenty-sixth under HED: because neither is severely threatened, the chances are good that their common path will persist.

On the prosimian tree, both measures are heavily influenced by the pendant edge (PE) length of the focal species (with correlations of PE vs. SV = 0.94, and of PE vs. HED = 0.98). PE is always part of the marginal increase to ψ, while interior edges are most likely represented with high probability, especially for larger and more balanced trees. Here, both polytomies and the use of a pure birth model for estimating unknown edge lengths bias pendant edges to be long. Simulated trees under more realistic models return significant but weaker correlations (e.g. for $d = 0.9b$, $N = 500$ species and a right-skewed distribution of extinction probabilities, correlations of PE vs. SV ~ 0.82, and PE vs. HED ~ 0.84). Even here,
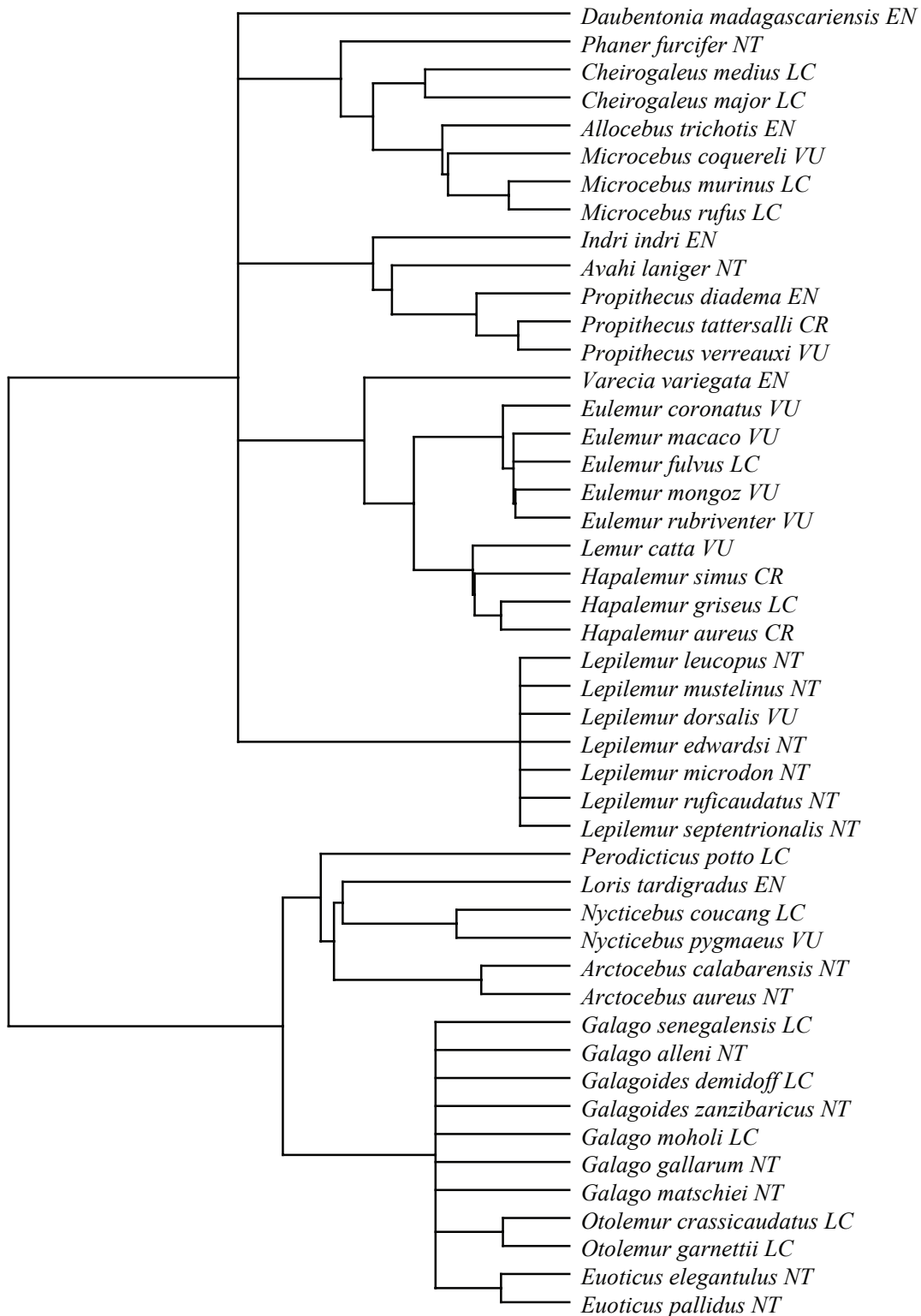
Figure 2. Prosimian species tree and associated IUCN threat categories. *CR*: critically endangered, *P*(*ext*) = 0.9; *EN*: endangered, *P*(*ext*) = 0.5; *VU*: vulnerable, *P*(*ext*) = 0.1; *NT*: near threatened, *P*(*ext*) = 0.01; *LC*: least concern, *P*(*ext*) = 0.001. Edge lengths are on an arbitrary scale that represents time since divergence.

PE is a poor predictor of HED for the genus *Propithecus*, for *H. griseus*, and for *Nycticebus couang* (Fig. 2). The first two groups contain the three most endangered species, increasing the value of close relatives. *Nycticebus* is an isolated genus, and *N. couang*'s sister species is listed as vulnerable (*P(ext)* = 0.1). Likewise for *P. verreauxi* —although it has close relatives and so a short PE, these relatives are at high risk of extinction, which increases its value; this is what we saw with species *D* in Figure 1.

Incorporating the focal taxon's extinction risk (HEDGE scores)

The effect of close relatives' risk status on one's own value is precisely the strength of the HED approach. However, the fact that the extinction risk of other species affects a focal species, but its own risk does not is somewhat counter-intuitive. We address this by showing how it is possible to write the HED index as the sum of two terms each of which takes into account the extinction risk of the focal species. To describe this further, let I be the random variable which takes the value i if the focal species i survives (at the future time under consideration) and which otherwise takes the value of the emptyset (i.e. $\emptyset$) if i goes extinct.
Let

$$\psi_i' = \mathbb{E}\left[PD\left(\mathcal{S}_i \cup \{i\}\right) - PD\left(\mathcal{S}_i \cup I\right) \right],$$

where, as before $\mathcal{S}_i$ is the random subset of species in $X - \{i\}$ that survive. In words, ψ_i' is the increase in the expected PD score if we condition on the event that species i survives. This quantity has also been investigated recently by Dan Faith (Faith, 2007) where it is referred to as (*expected*) *PD complementarity*.
Similarly, let

$$\psi_i'' = \mathbb{E}\left[PD(\mathcal{S}_i \cup I) - PD(\mathcal{S}_i) \right].$$

In words, ψ_i'' is the decrease in the expected PD score if we condition on the event that species i becomes extinct. The following result describes how to compute these two indices easily from the HED index, and verifies that they add together to give the HED index (its proof is given in the Appendix).

Theorem 4.1
  (i) $\psi_i' = \epsilon_i \cdot \psi_i$,
  (ii) $\psi_i'' = (1 - \epsilon_i) \cdot \psi_i$,
  (iii) $\psi_i' + \psi_i'' = \psi_i$.

The approach of assigning a value to a species which is a function of its phylogenetic distinctiveness and its extinction probability has been referred to as 'expected loss' by Redding and Mooers (2006) and, more evocatively, an 'EDGE' score (Evolutionarily Distinct and Globally Endangered) by Isaac et al. (2007).

In the same spirit we will call ψ' and ψ'' (which extend our HED index ψ) HEDGE (*heightened evolutionary distinctiveness and globally endangered*) scores. The HEDGE score ψ_i' is more relevant when evaluating actions that might save species, whereas the HEDGE score ψ_i'' is appropriate when evaluating actions that might cause the extinction of species (such as building a dam). Our measures link species-specific EDGE-type scores (Isaac et al. 2007) with the complementarity framework strongly advocated by Faith (2007).

One potential advantage of HED and HEDGE over previous scores is their flexibility in designing conservation scenarios. So for instance, we can choose IUCN-ranked species for which conservation is cheap and/or already partially successful, set their *P(ext)* to 0, and see how rankings of other species change. Alternatively, we might want to increase the *P(ext)* to 1.0 for certain species to see how others are affected.

Most generally, HED and HEDGE could be incorporated in an assessment of species value that included many factors besides risk and future contribution, e.g. the ecological, distributional and aesthetic values enumerated by Avise (2005), and the costs of recovery and probability of its success. We present these metrics in the hope that they will be used to promote the preservation of species and their natural habitats.

Acknowledgments

References

Altschul, S.F. and Lipman, D.J. 1990. Equal animals. *Nature*, 348:493–4.

Avise, J.C. 2005. Phylogenetic units and currencies above and below the species level. Pages 77–100 *in* Purvis A, Gittleman JL, and Brooks T, eds. Phylogeny and conservation. Cambridge University Press, Cambridge, UK.

Cousins, S.H. 1991. Species diversity measurement: choosing the right index. *Trends in Ecology and Evolution,* 6:190–2.

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

Faith, D.P. 2007. Phylogenetic diversity and conservation. *In* S.P. Carroll and C.Fox, eds. Conservation Biology: Evolution in action. Oxford University Press, New York, NY (in press).

Haake, C-J, Kashiwada, A. and Su. F.E. 2005. The Shapley value of phylogenetic trees. IMW Working Paper 363 (363). ArXiv (q-bio. QM/0506034)

Hartmann, K. and Steel, M. 2007. Phylogenetic diversity: from combinatorics to ecology. Pages 171–196. *In* O. Gascuel and M. Steel, eds. Reconstruction evolution: New mathematical and computational advances. Oxford University Press.

Isaac, N.J.B., Turvey, S.T., Collen, B., Waterman, C. and Baillie, J.E.M. 2007. Mammals on the EDGE: conservation priorities based on threat and phylogeny. *PLoS ONE* 2(3):e296.

May, R.M. 1990. Taxonomy as Destiny. *Nature,* 347:129–30.

Mooers, A.O. and Heard, S.B. 1997. Inferring process from phylogenetic tree shape. *Quarterly Review of Biology*, 72:31–54.

Myers, N., Mittermeier, R.A., Mittermeier, C.G., da Fonseca, G.A.B. and Kent. J. 2000. Biodiversity hotspots for conservation priorities. *Nature,* 403:853–8.

Pavoine, S., Ollier, S. and Pontier, D. 2005. Measuring diversity from dissimilarities with Rao's quadratic entropy: Are any dissimilarities suitable? *Theoretical Population Biology*, 67:231–9.

Pavoine, S., Ollier, S. and Dufour, A.B. 2005. Is the originality of a species measurable? *Ecology Letters*, 8:579–86.

Redding, D.W. 2003. Incorporating genetic distinctness and reserve occupancy into a conservation priorisation approach. Masters Thesis. University Of East Anglia, Norwich. UK.

Redding, D.W. and Mooers, A.O. 2006. Incorporating evolutionary measures into conservation prioritization. *Conservation Biology*, 20:1670–8.

SARA. 2002. Bill C-5, An Act respecting the protection of wildlife species at risk in Canada. Government of Canada, Ottawa.

Shapley, L.S. 1953. A value for n-person games. *Annals of Mathematical Studies*, 28:307–17.

Vos, R.A. and Mooers, A.O. 2004. Reconstructing divergence times for supertrees: a molecular approach. Pages 281–99 *in* O. R. P. Bininda-Emonds, ed. Phylogenetic supertrees: combining information to reveal the Tree of Life. Kluwer Academic Press, Dordrecht.

Vos, R. 2006. Inferring large phylogenies: the big tree problem. PhD. Thesis, Simon Fraser University.

Weitzman, M.L. 1998. The Noah's Ark problem. *Econometrica* 66(6):1279–98.

Appendix: Proofs of theorems

Proof of Theorem 3.1

First observe that the only edge lengths that contribute to $\Delta_{PD}(S, i)$ are those from the set $\{l(e_1), l(e_2), \ldots, l(e_k)\}$. Consequently, for the random set $\mathcal{S}_i$ of surviving species of $X - \{i\}$ we have

$$\Delta_{PD}(\mathcal{S}_i, i) = \sum_{r \geq 1} l(e_r) \cdot I_r(\mathcal{S}_i)$$

where $I_r(\mathcal{S}_i)$ is the 0,1 indicator random variable that takes the value 1 precisely if e_r is not an edge of the subtree of $\mathcal{T}$ connecting the taxa in $\mathcal{S}_i$ and the root of $\mathcal{T}$; since this is the only situation for which e_r lies in the subtree of $\mathcal{T}$ connecting $\mathcal{S}_i \cup \{i\}$ but not in the subtree of $\mathcal{T}$ connecting $\mathcal{S}_i$. Thus, by linearity of expectation,

$$\psi_i = \mathbb{E}\left[\Delta_{PD}(\mathcal{S}_i, i)\right] = \mathbb{E}\left[\sum_{r \geq 1} l(e_r) \cdot I_r(\mathcal{S}_i)\right]$$
$$= \sum_{r \geq 1} l(e_r) \cdot \mathbb{E}\left[I_r(\mathcal{S}_i)\right],$$

and since $I_r(\mathcal{S}_i)$ takes the values 0 and 1, $\mathbb{E}\left[I_r(\mathcal{S}_i)\right] = \mathbb{P}\left[I_r(\mathcal{S}_i) = 1\right]$.
Thus,

$$\psi_i = \sum_{r \geq 1} l(e_r) \cdot \mathbb{P}\left[I_r(\mathcal{S}_i) = 1\right].$$

Now, the event '$I_r(\mathcal{S}_i) = 1$' occurs precisely if none of the elements in $C(e_r) - \{i\}$ survive, and this latter event has probability $\prod_{j \in C(e_r) - \{i\}} \epsilon_j$. Substituting this into the previous equation establishes the theorem.

Proof of Theorem 3.2

For $i \in X$ and the random subset $\mathcal{S}_i \subseteq X - \{i\}$, we have

$$\Delta_{PD}(S_i, i) = \sum_{e \in E(\mathcal{T})} l(e) \cdot I_e(\mathcal{S}_i)$$

where $I_e(\mathcal{S}_i)$ is the 0,1 indicator random variable taking the value 1 precisely if $\mathcal{S}_i$ consists of no elements of $C_i(e) - \{i\}$ and at least one element of $X - C_i(e)$.

Thus,

$$\psi_i = \sum_{e \in E(\mathcal{T})} l(e) \cdot \mathbb{P}[I_e(\mathcal{S}_i) = 1]$$

and by the independence assumption

$$\mathbb{P}[I_e(\mathcal{S}_i) = 1] = \mathbb{P}[\mathcal{S}_i \cap (C_i(e) - \{i\}) = \emptyset] \\ \times \mathbb{P}[\mathcal{S}_i \cap (X - C_i(e)) \neq \emptyset].$$

and so

$$\mathbb{P}[I_e(\mathcal{S}_i) = 1] = \left(\prod_{j \in C_i(e) - \{i\}} \epsilon_j \right) \cdot \left(1 - \prod_{j \in X - C_i(e)} \epsilon_j \right),$$

as claimed.

Proof of Theorem 4.1
By definition,

$$\psi_i = \mathbb{E}[PD(\mathcal{S}_i \cup \{i\}) - PD(\mathcal{S}_i)],$$

and so

$$\mathbb{E}[PD(\mathcal{S}_i \cup \{i\})] = \psi_i + \mathbb{E}[PD(\mathcal{S}_i)].$$

Now we can write the unconditional expectation $\mathbb{E}[PD(\mathcal{S}_i \cup I)]$ as the weighted sum of conditional expectations $\mathbb{E}[PD(\mathcal{S}_i \cup I) | I = \{i\}] \mathbb{P}(I = \{i\}) + \mathbb{E}[PD(\mathcal{S}_i \cup I) | I = \emptyset] \mathbb{P}(I = \emptyset)$
and so

$$\mathbb{E}[PD(\mathcal{S}_i \cup I)] = (1 - \epsilon_i) \mathbb{E}[PD(\mathcal{S}_i \cup \{i\})] \\ + \epsilon_i \mathbb{E}[PD(\mathcal{S}_i)].$$

Parts (i) and (ii) now follow by applying these equations (and the linearity of expectation) to the definitions of ψ_i' and ψ_i''. Part (iii) follows directly from parts (i) and (ii).