**Basic premise (near as I can tell):**

"A system of priorities that reflects the value of taxonomic diversity can be achieved by setting priorities such that the subset of taxa that is protected has **maximum underlying feature diversity**. Such feature diversity of taxon subsets is difficult to estimate directly, but can be predicted by the cladistic/phylogenetic relationships among the taxa."

---

**Basic Metric**

Phylogenetic Diversity (PD) is just sum of weights* of edges of the subtree defined by subset S of X, |X|=n.

\[ PD_S = \sum_{e \in S} \lambda_e \]

Here, \( S = \{A, B, E\}, PD = 10 \)

* so these weights must reflect "underlying feature diversity"
Broad Recipe

1. Dimensions of biodiversity: $\alpha, \beta, \sum x, \text{var}(x)$
2. Sum of edge weights on trees
3. Isolation as a surrogate for maximizing expected PD
4. Other dimensions of the tree and conservation
5. What does an edge weight represent?

---

**Dimensions of Phylogenetic Biodiversity**

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Scale</th>
<th>Richness: How much</th>
<th>Divergence: How different</th>
<th>Regularity: How regular</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diversity of sample(s) [alpha]</td>
<td>edge weights (PD)</td>
<td>mean(patr. dist.)</td>
<td>mean(evol. isol.)</td>
<td>var(patr. dist.); tree shape, $l_c$</td>
</tr>
<tr>
<td>Differences among samples [beta]</td>
<td>sum of unshared edgeweights 'Unifrac'</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Phylogenetic Diversity (PD)** is just sum of weights of edges of the subtree defined by subset $S$ of $X$, $|X|=n$

$$PD_S = \sum_{e \in S} \lambda_e$$

Original goal was one of optimization, generally to find maxPD given constraints, e.g. when choosing $k$ of $n$ tips: for $k=3$, max(PD) is reached for $S = \{A,D,E\}$ (by inspection)

following Pavoine & Bonsall, 2011, sPhy
Maximizing PD ("max-sum") for $k$ of $n$ problem is possible with a greedy algorithm (Steel 2005; Pardi & Goldman 2005)

But, common PD setting is to choose among a defined set of $S$ (e.g. representing conservation areas); this, and related problems are hard (Faller, Pardi & Steel, 2007)*

Given $S = \{A,D\} \{A,B,C\} \{C,E\}$
maxPD for 1 set is $\{A,B,C\}$
maxPD for 2 sets is $\{C,E\} \{A,D\}$

*but see Chernomor et al. poster 4

Recast $k$ of $n$ problem using "Expected PD"

$$E(PD) = \sum_{e \in T} \lambda_e \left( 1 - \prod_{j=1}^{Ce} p(\text{ext})_j \right)$$

$Ce$: # leaves that ultimately subtend edge $e$
$p(\text{ext})_j$: probability of extinction of leaf $j$

Witting & Loeschcke, 1995
Hartmann & Steel, 2006

For $k$ of $n$, we assign $p(\text{ext}) = 0$ to leaves in $S$ and $p(\text{ext}) = 1.0$ to leaves not in $S$.

Expected PD

$p(\text{ext}) = 0.9 \ 0.7 \ 0.5 \ 0.5 \ 0.6$

$$E(PD) = \sum_{e \in T} \lambda_e \left( 1 - \prod_{j=1}^{Ce} p(\text{ext})_j \right)$$

$Ce$: # leaves that ultimately subtend edge $e$
$p(\text{ext})_j$: probability of extinction of leaf $j$

$$E(PD) = .2 + .3 + .5 + .75 + 1.6 + .74 + .75 + .8425 = 5.6825$$

If we could halve the $p(\text{extinction})$ of a single species, which should it be?
Expected PD

\[ p(\text{ext}) = 0.9 \ 0.7 \ 0.5 \ 0.5 \ 0.6 \]

\[ E(PD) = \sum_{e \in E} \lambda_e (1 - \prod_{j=1}^{C_e} p(\text{ext})_j) \]

\[ C_e: \# \text{ leaves that ultimately subtend edge } e \]

\[ p(\text{ext})_j: \text{ probability of extinction of leaf } j \]

\[ E(PD) = 0.2 + 0.3 + 0.5 + 0.75 + 1.6 + 0.74 + 0.75 + 0.8425 = 5.6825 \]

If we could halve the p(extinction) of a single species, which should it be? (answer: A)

Can also produce the mirror of E(PD) tree (the E(Loss tree) and, e.g., map that on the landscape

Maximizing E(PD) when changing p(\text{ext}) of leaves at specified costs and budget: "Noah’s Ark Problem”*

*Weitzman, 1998: cited >150 times
Again, generally a hard problem
(Hartmann & Steel 2006, Fardi & Goldman 2007, Billionnet 2013)

\[
E(PD|S) = \sum_{e \in T} \lambda_e (1 - \prod_{j \notin S} p(\text{ext})_j \prod_{k \in S} p(\text{ext})'_k)
\]

Find \(S\) that maximizes \(E(PD)\), given all \(k\) members of \(S\) now have \(p(\text{ext})' < p(\text{ext})\) (are conserved).
Each \(\text{delta}(p(\text{ext}))\) has a cost \(c_k\), and \(\sum c_k \leq B\), the overall budget.

What about networks?

\[
SD_S = \sum_e \lambda_e (X|Y)
\]

\(X|Y\) is a set of splits; \(X\) is in \(S\), and \(Y\) is in \(S\)

Consider edges on a network as defining "splits" and sum of splits is total SD. We can visualise with a circular network

Here, for \(S = \{A,D\}\),
\(SD = 7\)
\(\max(SD)\) for \(k=2\)

Spillner et al. 2008
Minh et al. 2009

\[PD = \text{sum of edge weights}\]
\[
\text{Expected}(PD) = \text{sum or weighted edge weights}\]
\[
PD \text{ complementarity} = \text{a leaf's contribution to a tree (ie to the sum of edge weights)}\]

\[
PD_{\text{comp}_x} = PD_T - PD_{T-x}
\]

\(T\) is tree
\(T-x\) is subtree without leaf \(x\)

This is just the pendant edge weight or PE (for a rooted tree, need to include a zero-length outgroup leaf to get E)
PD complementarity and evolutionary isolation

One can define other sets that a leaf can complement, e.g., its contribution to the fictitious $E(PD)$ of the tree

$E(PD)_{\text{comp}_x} = E(PD)_{p(\text{ext})_x=0} - E(PD)_{p(\text{ext})_x=1}$

Christened "heightened evolutionary distinctness" by Steel et al. 2007. This formulation by Faith 2008.

The original motivation for this measure comes from:
1. Ad-hoc measures of "evolutionary isolation"
2. A specific measure of complementarity (Shapley):

$$\psi^s_{x} = \frac{1}{n!} \sum_{S \subseteq X, x \subseteq S} (|S|-1)!(|X|-|S|)!(PD_S - PD_{S-x})$$

Averaged over all equally-weighted subsets! PD complementarity of $x$ to subset $S$

The original motivation for considering \( E(PD) \text{comp}_x \):

1. Ad-hoc measures of "evolutionary isolation"

2. A specific measure of complementarity:

\[
FP_x = \sum_{e \in s(T, x, r)} \frac{\lambda_e}{C_e}
\]

Fair proportion divides the tree up among its leaves

5 of 12 total possible \( S' \), all set sizes equally likely

\( \lambda_e \): set of edges from \( x \) to root \( r \)
\( C_e \): # leaves descending from \( e \)

Redding 2003, Redding et al. 2008
Shapley and FP are the same in expectation for large $n$, and also in practice.

FP assigns all of the pendant edge to a leaf, and then some diminishing proportion of "deeper" edges.

$$FP_x = \sum_{e \in s(T, x, r)} \frac{\lambda_e}{C_e}$$

$s$: set of edges from $x$ to root $r$

$C_e$: # leaves descending from $e$

$FP$ is weighted towards local tree shape nearest the leaves, and so is strongly correlated with PE.

Can prioritize populations using splits on networks.

Can use Shapley if no $p(\text{ext})$ or $E(\text{SD}) \text{comp.}$ if we have $p(\text{ext})$.

Hartmann 2008, 2012

Volkmann, (Spillner, Moulton) 2014
\[ FP_x = \sum_{e \in s(T,x,r)} \frac{\lambda_e}{C_e} \]

\( s \): set of edges from \( x \) to root \( r \)
\( C_e \): # leaves descending from \( e \)

\( FP \) is weighted towards local tree shape nearest the leaves, and so is strongly correlated with \( PE \).

Quickly asymptotes as we move deeper in the tree and \( C_e \) increases. This makes absolute values broadly comparable across large groups (e.g. mammals vs. birds vs. plants).

Motivation for identifying isolated species is that they are (i) uniquely identifiable for conservation (ii) sets of isolated species spans much of the tree.

But this assumes \( p(\text{ext})=1 \) for all other species!

Motivation for identifying isolated species is that they are (i) uniquely identifiable for conservation (ii) sets of isolated species spans much of the tree.

But this assumes \( p(\text{ext})=1 \) for all other species!

Sets of isolated species may have high \( PD \), but not high \( PD_{\text{comp}} \) (or high \( \text{exp}(PD)_{\text{comp}} \)).

This pair offers highest complementarity to the rest of the tree.

Redding et al. submitted

Redding et al. 2008

Kakapo

Shane McInnis
For this to be generally the case, high FP species are rarely close relatives.
Under what diversification models is this not rare (Morlon, Etienne)*?

*because it does happen, e.g. tailed frogs

Ok, why are we concerned with a sum of edge weights? (anyway)
Ok, why are we concerned with a sum of edge weights? (anyway)

1. Wilson’s exhortation to preserve "information"
2. Practical extension to ecosystem services

A practical reason to conserve PD (sum of edge weights)

1. Across 29 global experiments, PD > 20 AIC units better predictor of productivity than species richness or functional group representation.

2. Across a set of famous experimental plots testing biodiversity and ecosystem functioning, PD > 10 AIC units better predictor of productivity than 13 other measures, including species richness.

3. Experiment designed to test whether PD predicts productivity confirm these other analyses.

Cadotte et al. 2008, 2009; others; Cadotte 2013

David Tilman's Cedar Creek plots
"The ecological consequences of biodiversity loss can be predicted from evolutionary history."

Emerging trend 4 of 4, Cardinale et al. 2012 Nature

Most controversial conservation paper of 2013?
Vellend et al. 2013 PNAS

No net loss of local biodiversity over time!

Controversial conservation paper from 2014
Domelas et al. 2014 Science

Low beta diversity

High beta diversity
**Dimensions of Phylogenetic Biodiversity**

<table>
<thead>
<tr>
<th>Dimension Scale</th>
<th>Richness: How much? Sum of...</th>
<th>Divergence: How different? Mean of...</th>
<th>Regularity: How regular? Variance of...</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diversity of sample(s) [alpha]</td>
<td>edge weights (PD)</td>
<td>mean(patr. dist.)</td>
<td>var(patr. dist.); tree shape, (I_c)</td>
</tr>
<tr>
<td>Differences among samples [beta]</td>
<td>sum of unshared edgeweights 'Unifrac'</td>
<td>?</td>
<td>?</td>
</tr>
</tbody>
</table>

This suggests conservation biologists might consider effects of "beta diversity" (e.g. between samples A & B)*

*various speakers, CSEE meeting 2014

\[
Unifrac = \frac{2 \cdot PD_{A\cup B} - PD_A - PD_B}{PD_{A\cup B}}
\]

Is this the best measure? does not, e.g. differentiate how shared edges are organized

Luzupone & Knight 2005
also Ferrier et al. 2007
Bryant (Morlon) et al. 2008

**Dimensions of Phylogenetic Biodiversity**

<table>
<thead>
<tr>
<th>Dimension Scale</th>
<th>Richness: How much? Sum of...</th>
<th>Divergence: How different? Mean of...</th>
<th>Regularity: How regular? Variance of...</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diversity of sample(s) [alpha]</td>
<td>edge weights (PD)</td>
<td>mean(patr. dist.)</td>
<td>var(patr. dist.); tree shape, (I_c)</td>
</tr>
<tr>
<td>Differences among samples [beta]</td>
<td>sum of unshared edgeweights 'Unifrac'</td>
<td>?</td>
<td>?</td>
</tr>
</tbody>
</table>

following Pavoine & Bonsall, 2011, sPhy

**Broad Recipe**

1. Dimensions of biodiversity: \(\alpha, \beta, \sum, \bar{x}, \text{var}(x)\)
2. Sum of edge weights on trees
3. Isolation as a surrogate for maximizing expected PD
4. Other dimensions of the tree and conservation
5. What does an edge weight represent?
All the dimensions and scales use edge weights

"...underlying feature diversity of taxon subsets...can be predicted by...phylogenetic relationships among the taxa." (Faith 1992)

for discussion, see, e.g. Diniz-Filho et al. 2013

Observation: there are 13 published measures of evolutionary isolation (some redundant like FP&Shapley)

Acronyms don't matter; weight leaf-ward and root-ward edge weights differently (on birth-death trees*)

Redding et al. submitted

Obviously, the score you use matters...top EDGE spp

Evolution happens over time

Redding et al. submitted
This tree will capture "feature diversity" under one very common model: if many features evolve under independent Brownian motion processes.

Rootward edges contain more information on feature diversity.

Tipward edges contain more information about feature diversity.

raise all entries in var-covar matrix* (by unknowable factor)

*Pagel, 1999, Blomberg 2003, Harmon et al., 2010
All PD calculations flow directly, and several problems go from hard to less hard.

This is the general case (additive tree) considered by Faith and others since. Because the edge lengths are idiosyncratic, there is no "transformation" possible. Usually, single or few traits used to produce additive tree.

Under particular models of feature evolution on additive trees, PD will not capture maximum feature diversity. Features arise linearly with time and are lost exponentially with rate lambda. On this tree, if characters are lost a rate = 2000x slower than they arise up to 10x faster.

$M(S) = \min\{PD_{x,y} : x, y \in S\}$

maximizing $M(S)$ for $|S| = k$ as alternate function.

Bordewich et al. 2008 (see also Moulton et al. 2007)
r^2 = 0.82 for gut specialists of similar genome size (yellow line)

This suggests that edge weights on additive trees for single markers predicts "total information"

Across a broader array of communities, 16S genetic distance predicts (with suitable preparation of data) functional components of genomes
Is there *any* way to measure the shared and unique "total information content" of a genome* (or feature diversity of an organism)?

With this, we could see how well additive (or, better-still, ultrametric) trees (or splits on networks) predicted the difference in information/features between leaves.

*using, e.g. compression algorithms? (I. Martyn, pers. comm.)