

LETTER

Phylogenetic diversity metrics for ecological communities: integrating species richness, abundance and evolutionary history

Marc W. Cadotte,^{1,2*} T. Jonathan Davies,¹ James Regetz,¹ Steven W. Kembel,^{3,4} Elsa Cleland^{1,5} and Todd H. Oakley⁶

¹National Center for Ecological Analysis and Synthesis, University of California, 735 State St., Santa Barbara, CA 93101, USA

²Department of Biological Sciences, University of Toronto – Scarborough, 1265 Military Trail, Scarborough, ON M1C 1A4, Canada

³Integrative Biology, University of California, Berkeley, CA 94720, USA

⁴Center for Ecology and Evolutionary Biology, University of Oregon, Eugene, OR 97403, USA

⁵Ecology, Behavior and Evolution, University of California – San Diego, La Jolla, CA 92093, USA

⁶Ecology, Evolution and Marine Biology, University of California, Santa Barbara, CA 93107, USA

*Correspondence: E-mail: mcadotte@utsc.utoronto.ca

Abstract

Phylogenetic information is increasingly being used to understand the assembly of biological communities and ecological processes. However, commonly used metrics of phylogenetic diversity (PD) do not incorporate information on the relative abundances of individuals within a community. In this study, we develop three indices of PD that explicitly consider species abundances. First, we present a metric of phylogenetic-abundance evenness that evaluates the relationship between the abundance and the distribution of terminal branch lengths. Second, we calculate an index of hierarchical imbalance of abundances at the clade level encapsulating the distribution of individuals across the nodes in the phylogeny. Third, we develop an index of abundance-weighted evolutionary distinctiveness and generate an entropic index of phylogenetic diversity that captures both information on evolutionary distances and phylogenetic tree topology, and also serves as a basis to evaluate species conservation value. These metrics offer measures of phylogenetic diversity incorporating different community attributes. We compare these new metrics to existing ones, and use them to explore diversity patterns in a typical California annual grassland plant community at the Jasper Ridge biological preserve.

Keywords

Abundance distribution, biodiversity, ecophylogenetic diversity, evenness, evolutionary history, Jasper Ridge biological preserve, null models, Shannon Index.

Ecology Letters (2010) 13: 96–105

INTRODUCTION

Increasingly, ecologists have come to recognize that species are not independent entities, but rather their functional and ecological similarities are shaped by patterns of common ancestry (Felsenstein 1985; Harvey & Pagel 1991). Over the past several years the literature has witnessed a new coordination between evolutionary history and community ecology (Webb *et al.* 2002, 2006; Cavender-Bares & Wilczek 2003; Cavender-Bares *et al.* 2009). Phylogenetic trees and the tools to estimate them have become widely available,

and ecologists are increasingly using this information to understand ecological patterns. Phylogenies are being used in ecological studies in two complementary ways. First, measurements of the phylogenetic distinctiveness are being used to assign conservation value to individual species (Vanewright *et al.* 1991; Pavoine *et al.* 2005; Redding & Mooers 2006; Isaac *et al.* 2007) or to assemblages (Faith 1992; Crozier 1997; Crozier *et al.* 2005; Davies *et al.* 2008). Secondly, the evolutionary relationships among species are used to inform our understanding of the mechanisms driving patterns of co-occurrence (Webb *et al.* 2002, 2006;

Cavender-Bares & Wilczek 2003; Silvertown *et al.* 2006). The key questions are whether environmental sorting or competitive exclusion produce community patterns and at what spatial and phylogenetic scales sorting and exclusion are strongest (Cavender-Bares & Wilczek 2003). More recently, phylogenetic information has been used to determine if the evolutionary relationships among members of a community or trophic level affect other ecological processes (Cadotte *et al.* 2008). For example, the underlying phylogenetic structure of a plant community might be informative for predicting patterns of abundance and community composition of insect herbivores specialized to certain groups of plant species.

Evolutionary relationships may affect ecological processes and dynamics because phylogenies reflect the integrated phenotypic differences among taxa (Felsenstein 1985; Harvey & Pagel 1991; Faith 1992), and may be a more encapsulating measure than singular, discretely measured traits (Vanewright *et al.* 1991).

Taxonomic and phylogenetic information therefore reveals the presence or absence of evolutionary lineages and the genes or traits they represent. A rich history of research on community structure has shown that there is marked variation in the abundances of species within communities, typically following a hollow curve distribution with many rare species and few abundant species (Preston 1948; MacArthur 1957; Hubbell 2001). As a consequence of variation in abundance, the proportional representation of traits within a community will also differ.

A number of diversity metrics that combine species richness and abundance distributions have been proposed (Magurran 2003); two well-known examples include the Simpson and Shannon biodiversity indices. More recently, diversity measures incorporating functional traits (Petchev & Gaston 2002; Vileger *et al.* 2008) or the evolutionary history of species have been developed (e.g. Faith 1992, 2006; Webb *et al.* 2002; Hardy & Senterre 2007; Helmus *et al.* 2007). In particular, Faith's (1992) phylogenetic diversity (PD), which measures the shared phylogenetic history among taxa occurring in a sample, has been perhaps the most widely used measure of PD. However, Faith's PD does not incorporate information on the relative abundances of taxa within communities, but calculates PD based on the presence or absence of species. The ecological interactions within communities are a product of both the abundances and the evolutionary histories of constituent species. What matters then is the PD among individuals, as opposed to among species, i.e. how well represented are the different phylogenetic lineages among-individual members of a community? If we consider a plant community, we may develop differing predictions of herbivore diversity if 99% of the community is dominated by a single plant species and its representative traits, in contrast to a community with

more even abundance distributions (Hillebrand *et al.* 2008). In addition, predictions may also differ depending on whether abundances are phylogenetically clustered, so that one or a few clades dominate, or are evenly dispersed across the phylogenetic tree connecting the species.

Previous attempts to incorporate abundance information into phylogenetic measures have focused on pairwise taxon distances (e.g. Pavoine *et al.* 2005; Hardy & Senterre 2007; but see Helmus *et al.* 2007; Allen *et al.* 2009) as opposed to total community PD which may be more directly comparable with the commonly used traditional community diversity metrics. In this study, we derive a series of phylogenetic community diversity metrics that capture information on both phylogenetic tree topology and branch lengths simultaneously across the set of species within a community. First, we calculate a metric of phylogenetic-abundance evenness (PAE) that measures the relationship between species evolutionary distinctiveness (ED) and abundance. Secondly, we measure the imbalance of abundances at higher clades (IAC). Finally, we determine the evolutionary uniqueness of individuals and use that to generate an entropic index encapsulating species ED (abundance-weighted evolutionary distinctiveness; AED). These three metrics offer alternative measures of ecophylogenetic diversity based on different assumptions. All three may be useful for exploring ecological processes and patterns, depending upon the question of interest. In addition, our metrics may provide a guide for practical conservation, and in particular conservation efforts designed to maintain ecosystem services. All metrics and example data were coded and examined using R 2.9.1 (R Development Core Team 2009) and will be part of a forthcoming package called *ecoPD* (J. Regetz, M.W. Cadotte and T.J. Davies, unpublished data).

ABUNDANCE-INFORMED ECOPHYLOGENETIC METRICS

Community PD is defined here as the sum of the phylogenetic branch lengths connecting all taxa present in a community without regard for any larger regional species pool (Crozier *et al.* 2005; Cadotte *et al.* 2008), which is subtly different than another metric, Faith's PD, which measures total phylogenetic branch lengths within habitats while retaining the root from a larger regional phylogeny (Faith 1992, 2006; Crozier 1997; Crozier *et al.* 2005; Pardi & Goldman 2007). Our measure of community PD is the logical measure for assessing coexistence or functional differences among species because community-level processes operate on extant members; whereas Faith's PD is the desired measure for evaluating habitat conservation priorities as it measures the amount or proportion of total evolutionary history contained within habitats (Faith 1992,

2006; Crozier 1997; Crozier *et al.* 2005; Pardi & Goldman 2007).

Local PD values are only affected by presence/absence patterns. There are several ways to add abundance information to PD measures. The first step is to estimate intraspecific PD. If researchers had information on total genetic diversity among all individuals in a community, then one could perform phylogenetic analyses at the individual level or calculate diversity at the allele level without regard for species (e.g. Mace *et al.* 2003; Sherwin *et al.* 2006). For example, in a time-calibrated tree, PD could be calculated to the species-level plus the sum of the intraspecific branch lengths connecting individuals. However, such information does not exist for whole communities, or even whole assemblages within a single trophic level. Lacking individual-level genetic information, one alternative would be to derive the expected evolutionary divergence times assuming some general coalescent or splitting model (see Appendix 1). However, such approaches require problematic assumptions and likely over-estimate PD (Appendix 1). The second method, and the one we advocate, is to describe how PD is parsed among members of a community.

Phylogenetic-abundance evenness

A simple way to calculate the amount of evolutionarily unique information represented by an individual of species i , is the unshared branch length or the length of the phylogenetic branch connecting it to its most recent divergence in the community, λ_i . We can scale λ_i by the number of individuals or abundance, n_i , of species i (Fig. 1b):

$$EI_i = \lambda_i \cdot n_i \quad (1)$$

The absolute value of EI is not very meaningful, but it can be used to evaluate whether individuals are evenly

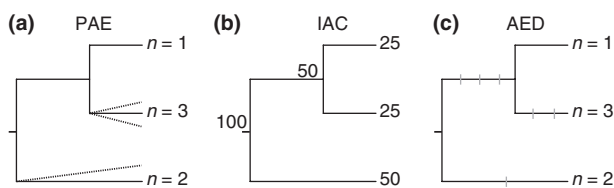


Figure 1 Graphical representation of the three abundance-weighted phylogenetic diversity metrics. The first metric, phylogenetic-abundance evenness, scales terminal branches in proportion to species abundance (a). Secondly, imbalance of abundances at the clade, creates an expected abundance based on splitting abundances among higher clades (b). Finally, abundance-weighted evolutionary distinctiveness partitions branch lengths among descendent individuals within a community (c).

distributed among lineages, in comparison with a null model, for example where individuals are evenly distributed across species. We derive this measure of evenness (PAE; see Table 1 for definitions of equation terms for all metrics) as:

$$PAE = \frac{PD + \sum_{i=1}^S \lambda_i \cdot (n_i - 1)}{PD + (\bar{n} - 1) \sum_{i=1}^S \lambda_i} \quad (2)$$

where PD is the sum of all branch lengths represented by the species in the community of S species and \bar{n} is the mean number of individuals per species. PAE falls between 0 and 1 when individuals are clustered into relatively short terminal branches, whereas $PAE > 1$ when individuals are clustered into long terminal branches. When individuals are evenly distributed across terminal branch lengths, $PAE = 1$. Figure 2b gives PAE values for the example tree.

A randomization test can be used to evaluate whether PAE differs significantly from 1, while retaining the abundance distribution of species. Observed abundances are repeatedly shuffled among the species in the community to construct a null distribution of PAE values. This test assumes the distribution of species relative abundances is fixed and thus multiple communities are compared where only their phylogenetic structures affect the deviation from 1. If randomizations encompass both 1 and the observed PAE , no significant trend is assumed. An alternative might be to draw null abundances from *a priori* distributions, such as the log-normal (Preston 1948), Fisher's log series (Fisher *et al.* 1943), the broken-stick (MacArthur 1957) or Hubbell's (2001) zerosum multinomial, which would allow researchers to infer the affect of the abundance distribution on PAE values.

Imbalance of abundances at the clade level

While PAE measures the phylogenetic-abundance distributions among terminal (i.e. unshared) branches, alternative metrics that quantify imbalance higher in the phylogeny or that do not rely on branch length estimates can be developed. A number of metrics have been used to estimate imbalance of higher-level distributions of clade richness (e.g. Agapow & Purvis 2002). We extend this approach to determine the IAC by quantifying the relative deviation in the abundance distribution from a null case where individuals are evenly partitioned between clade splits. As we move from the root down the phylogeny, the summed abundance of all descendent species is parsed by the number of splits at each node. For example, if total community abundance, N , is 100, at the first split the sum of abundances within each sister clade has expectation $n = 50$ (Fig. 1c). Thus the

Table 1 List of symbols and acronyms used in the equations

Symbol	Description
PD _c	Sum of all phylogenetic branches including within-species estimates of divergences
PAE	Phylogenetic evenness of the abundance distribution scaled by branch lengths
IAC	Relative deviation from null expectation of phylogenetically balanced abundances
AED	Evolutionary distinctiveness of individuals with a community
PD	Sum of all phylogenetic branches connecting species together within a community
Faith's PD	Sum of all phylogenetic branches connecting species together within a community plus the root from a larger regional phylogeny
PD _{mi}	PD within-species <i>i</i> , estimated using a polytomy or Yule model
H'	Shannon diversity index, $H' = -\sum_{i=1}^S p_i \cdot \ln p_i$, an entropy measure
H _{ED}	Entropic measure of diversity of evolutionary distinctiveness among species
E _{ED}	Equitability of H _{ED}
H _{AED}	Entropic measure of diversity of evolutionary distinctiveness among individuals
E _{AED}	Equitability of H _{AED}
δ _{<i>i</i>}	The age or phylogenetic distance within-species <i>i</i>
<i>n_i</i>	The abundance of species <i>i</i>
<i>n_k</i>	The total abundance of all species originating from node <i>k</i>
<i>N</i>	Total community abundance
EL _{<i>i</i>}	Unshared PD represented by species <i>i</i> scaled by <i>n_i</i>
\hat{n}_i	The expected abundance of species <i>i</i>
η _{<i>k</i>}	The number of lineages originating from node <i>k</i>
<i>v</i>	The number of nodes
λ _{<i>e</i>}	Branch length of edge <i>e</i> and where λ _{<i>i</i>} is the terminal branch length for species <i>i</i>

expected number of individuals at node *k*, \hat{n}_k under this null is:

$$\hat{n}_k = \frac{N}{\prod_{k \in s(T, k, r)} \eta_k} \quad (3)$$

where, η is the number of lineages originating at node *k* out of *v* nodes in the set $s(T, k, r)$, which is the number of nodes between node *k* and the root *r* in tree *T*. When *k* is a terminal node, \hat{n}_k becomes \hat{n}_i , the expected abundance of species *i*. For a fully resolved phylogeny without any hard polytomies, all η = 2, and so the denominator becomes 2^{*v*}. The index, IAC is the relative deviation from this null expectation, standardized by the number of splits (nodes), allowing one to compare values from multiple phylogenies with differing topologies and number of species.

$$IAC = \frac{\sum_{i=1}^S |n_i - \hat{n}_i|}{v} \quad (4)$$

This metric is the relative per-node imbalance in the distribution of individuals, and is always ≥ 0. If IAC = 0 then the distribution of individuals is balanced among higher clades. If strength of competition was proportional to phylogenetic relatedness, we might predict that IAC would tend towards 0, because this would minimize the summed 'competitive load' experienced by individuals within the community; that is, closely related species are more likely to parse resources. Conversely, if environmental filtering was key to community structure (and assuming phylogenetic conservation of ecological traits), we might predict IAC >> 0, because conditions favouring high abundance within a particular species, would also favour higher abundance among its close relatives (Cadotte *et al.* 2009). A shortcoming of this measure is that it does not take into account branch length – which we explore in the following metric, AED. Figure 2c shows the distribution of IAC values across the three hypothetical communities. Significant trends can be detected by randomizing abundances, as for PAE.

Both the IAC and PAE create indices – relaying information on the distribution of individuals with respect to the phylogenetic distinctiveness of the species within the community being analysed. In a related metric incorporating species' abundances, PSE, Helmus *et al.* (2007) quantify phylogenetic signal by comparison with a hypothetical complete polytomy. However, PSE, depends on both the topology and abundance distribution, whereas our metric, AIC, assumes the tree to be fixed.

Abundance-weighted evolutionary distinctiveness

Our final measure assigns an index of ED to species or individuals based on abundances and phylogenetic distances. We derive this metric from the ED method of Isaac *et al.* (2007), which is similar to the evolutionary splits (ES) method of Redding and colleagues (Redding & Mooers 2006; Redding *et al.* 2008), except that ED partitions deep branch lengths by the number of descendant species as opposed to the number of proximate subtending clades, as in ES. We develop two complementary metrics at differing phylogenetic levels, the species and the individual levels. We start at the species level using the ED calculation following Isaac *et al.* (2007), but there notable alternatives (Vanewright *et al.* 1991; Pavoine *et al.* 2005; Redding & Mooers 2006). For species *i* in tree *T*, ED is calculated as:

$$ED(T, i) = \sum_{e \in s(T, i, r)} \left(\lambda_e \cdot \frac{1}{S_e} \right) \quad (5)$$

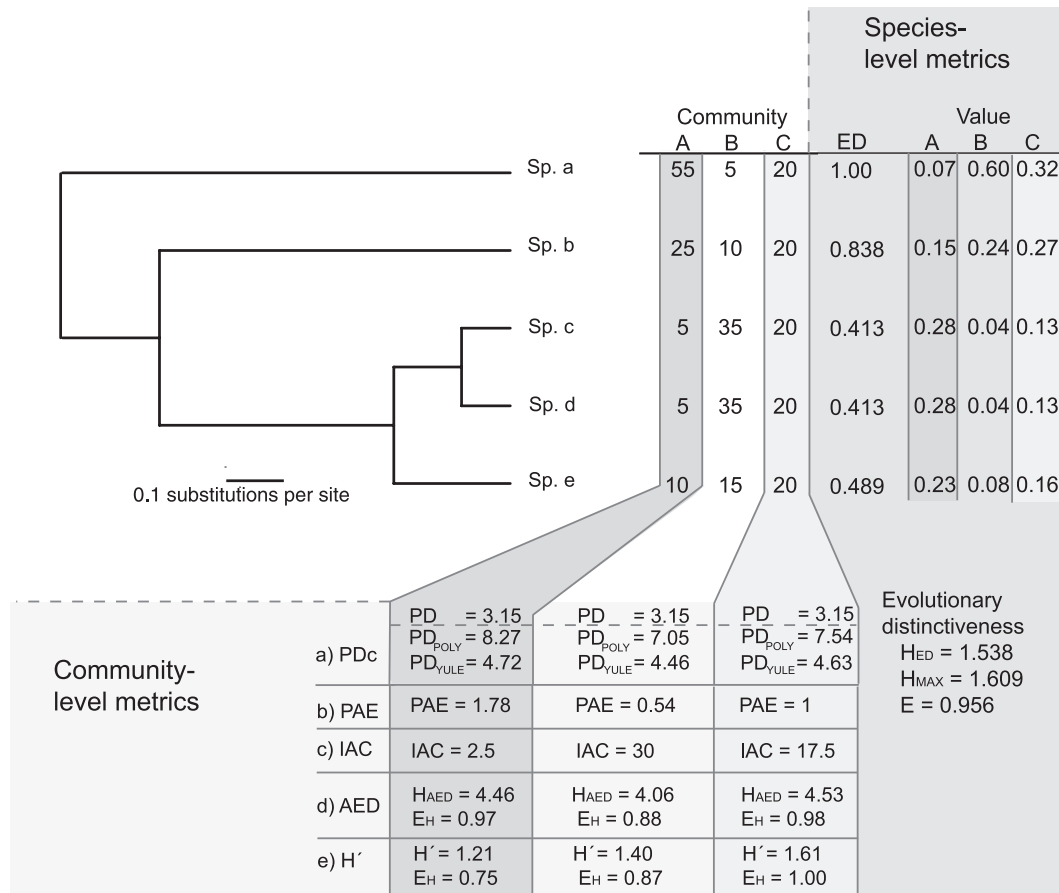


Figure 2 Using hypothetical communities and their phylogeny, we show how the various metrics estimate diversity. We use five community-level diversity metrics. (a) PD and PD_c (defined in Appendix 1), which includes estimates of among-individual divergences, reveal that community A potentially contains the most evolutionary information. Community A was also the furthest from even (phylogenetic-abundance evenness) because abundance is skewed towards long branch lengths (b) and, as a consequence, also the most balanced (IAC) (c). Further, this community was highly diverse according to abundance-weighted evolutionary distinctiveness (d) in contrast to having low diversity using the standard Shannon Index (e). For the species-level metrics, species A was the most distinct (ED) but its conservation value ($AED_i / \sum_{i=1}^S AED_i$) depended on its community abundance.

where e is an edge of length λ in the set $s(T, i, r)$ connecting species i to the root, r and S_e is the number of species that descend from edge e . ED is the proportional amount of evolutionary history encapsulated by species i and thus total PD is simply:

$$PD = \sum_{i=1}^S ED_i \tag{6}$$

From this an entropy measure, H'_{ED} , can be calculated, which represents the evolutionary information content of a randomly sampled species from the community:

$$H'_{ED} = - \sum_{i=1}^S \frac{ED_i}{PD} \cdot \ln \frac{ED_i}{PD} \tag{7}$$

Here equitability is:

$$E_{ED} = \frac{H'_{ED}}{\ln S} \tag{8}$$

We can incorporate abundance information at the individual level by splitting evolutionary information among individuals (Fig. 1d). If we assume that the J individuals in species i have phylogenetic branches of length zero (we provide equations when individuals have branch lengths > 0 in Appendix 2), but still contain evolutionary information, then the ED of a species is proportional not only to phylogenetic distances but also the distribution of individuals beneath each edge. Here AED for each individual within-species i , AED_i is:

$$AED(T, i) = \sum_{e \in s(T, i, r)} \frac{\lambda_e}{n_e} \tag{9}$$

where n_e is the total abundance of species that descend from edge e . For the AED values, total PD is again the sum:

$$PD = \sum_{i=1}^S AED_i \cdot n_i \quad (10)$$

The AED metric may have direct conservation applications. The relative average value of an individual ($AED_i / \sum_{i=1}^S AED_i$) can be used by managers to identify individuals, and by extension species, whose loss corresponds to the greatest loss of evolutionary information. If, as has been proposed, evolutionary history captures functional diversity necessary for ecosystem processes and services (e.g. see Cadotte *et al.* 2008), minimizing this loss of evolutionary diversity might maximize the preservation of ecosystem function.

Further, we can also derive a community entropy measure (H_{AED}) at the individual level rather than at the traditional species level.

$$H_{AED} = - \sum_{i=1}^S \frac{n_i \cdot AED_i}{PD} \cdot \ln \frac{n_i \cdot AED_i}{PD} \quad (11)$$

where equitability is:

$$E_{AED} = \frac{H_{AED}}{\ln N} \quad (12)$$

E_{AED} is maximized when AED values are equal, meaning that species abundance is scaled to ED. Thus low H_{AED} communities should also tend to have high IAC values, because the distribution of abundances is phylogenetically imbalanced, and more derived (younger) species tend to have higher abundances. In Fig. 2d, community A, which has the lowest Shannon diversity value (H'), is shown to be highly diverse with high equitability when accounting for the distribution of evolutionary information among individuals (Fig. 2). Community B, which is moderately diverse when using the Shannon Index, has the lowest AED diversity because the most abundant species is sharing disproportionately little evolutionarily unique information, and is also the most imbalanced (IAC = 30), and has a low PAE value.

In a recent paper, Allen *et al.* (2009) also proposed a phylogenetically informed entropy measure H_p , similar to our H_{AED} . However, H_p differs from H_{AED} in that it depends on the summed product of two independent values, branch lengths and proportional abundances. Since H_p is influenced both by changes in branch lengths and abundances, interpreting differences between communities is complicated, and will be sensitive to the units by which branch lengths are scaled. As a result, and as we show in our simulations (below), H_{AED} and H_p are sensitive to different aspects of community phylogenetic structure.

COMPARING METRICS WITH SIMULATED DATASETS

To provide an overview of how the various metrics relate to each other and tree shape we simulated a set of 100 random tree topologies with branch lengths drawn from a random uniform distribution for each of four richness levels; 4, 8, 16 and 32 species. Trees were then made ultrametric using nonparametric rate smoothing (Sanderson 1997) and rescaled so that tip to root distance = 1. Tips were assigned an abundance from a log-normal distribution with mean of 2 and SD of 1 on a log scale, and rescaled so that total abundance = 100, and we created 100 such random abundance distributions for each tree, generating 40 000 simulated data sets. We then ranked communities by each of our metrics as well as a set of broadly comparable diversity metrics drawn from the literature, and used the Euclidean distances between rankings to hierarchically cluster the alternative metrics using the *complete linkage* method as implemented in R.

We present the results of our cluster analysis as a dendrogram (Fig. 3). The precise relationship among the various metrics is sensitive to the clustering algorithm employed; nonetheless, several broad groupings can be discerned. Tree imbalance (IC), which excludes information on abundance and divergence times, is perhaps unsurprisingly distant from all other measures. Interestingly, IAC,

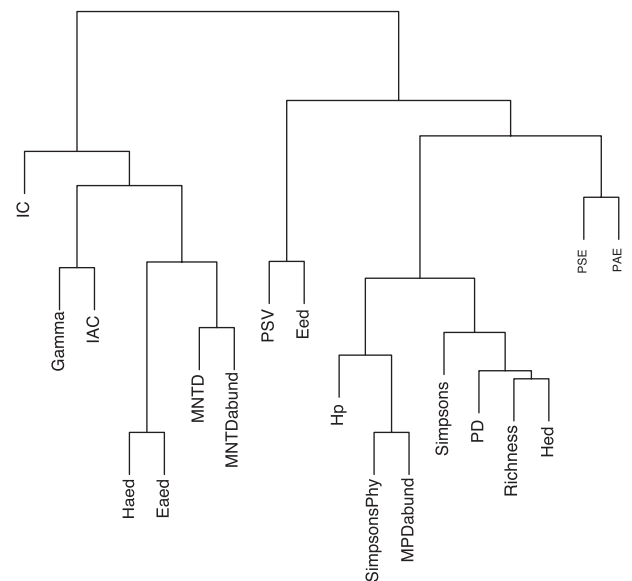


Figure 3 Hierarchically clustered Euclidean distances of the community ranking for 18 diversity metrics. Besides the metrics introduced in this paper, we contrast abundance-weighted MPD and MNTD, and non-abundance-weighted MNTD (Webb *et al.* 2008); phylogenetic species variability (PSV) and phylogenetic species evenness (PSE) (Helmus *et al.* 2007); H_p (Allen *et al.* 2009); and SimpsonsPhy (Hardy & Senterre 2007).

which measures phylogenetic imbalance in abundances, is the most similar metric to gamma. Gamma (Pybus & Harvey 2000) is a measure of the deviation of the phylogenetic topology from a null expectation of a constant species birth rate, and captures the distribution of branch lengths from the tips to the root of the tree. Species richness, PD, and H_{ED} provide similar rankings, with the sensitivity of all of these methods to species richness likely driving the relationship. It is of interest that our preferred metrics of community PD (H_{AED} and E_{AED}) also provide distinct rankings from the more traditional diversity metrics, while several computationally equivalent metrics (Simpsons-Phy; Hardy & Senterre 2007; PSE; Helmus *et al.* 2007; MPDabund; Webb *et al.* 2008; Allen *et al.* 2009) perform very similarly.

EXPLORING DIVERSITY IN THE JASPER RIDGE BIOLOGICAL PRESERVE PLANT COMMUNITY

Data for this case study were collected in May, 2007 in an area of sandstone-derived soils located near the main entrance of Stanford University's Jasper Ridge biological preserve. This site is located in central coastal California (37.4°N, 122.2°W), at an elevation of 120 m, with a long-term average precipitation of 655 mm year⁻¹ (Dukes *et al.* 2005). The plant community is dominated by naturalized European annual grasses, but also contains forbs, legumes and perennial grasses. The naturally assembled plant community was surveyed for species composition and abundance in 30.1 × 1 m² plots, using a visual cover estimation technique. The plots were arranged in three blocks of 10 plots each, spaced at 5-m intervals. All taxa encountered were identified to the species level if possible, but the genus of senesced or non-flowering taxa was used if species-level identification was not possible. We constructed a phylogeny for the species recorded in the Jasper Ridge plots, which included a total of 34 species (see Appendix 3 for details).

Plots at Jasper Ridge varied from 9 to 20 species recorded, with an average of 14.03 (SD = 2.65). Abundance, measured as per cent cover, for the plots ranged from 95 to 216, with an average of 145.03 (SD = 32.5). We compared the performance of the various metrics by scaling them to a mean of 0 and SD of 1 to allow the comparison of different metrics. We used Pearson correlation coefficients to assess similarities in metric values for plots. Three PD measures, which do not combine phylogenetic and abundance information (PD, H_{ED} and E_{ED}), all ranked communities similarly to richness (Fig. 4), and were all positively correlated with species richness ($r = 0.81, 0.97$ and 0.38 respectively; $P < 0.05$, see Appendix 4). However, our alternative abundance-weighted PD metrics ranked communities differently than richness (Fig. 4). IAC was

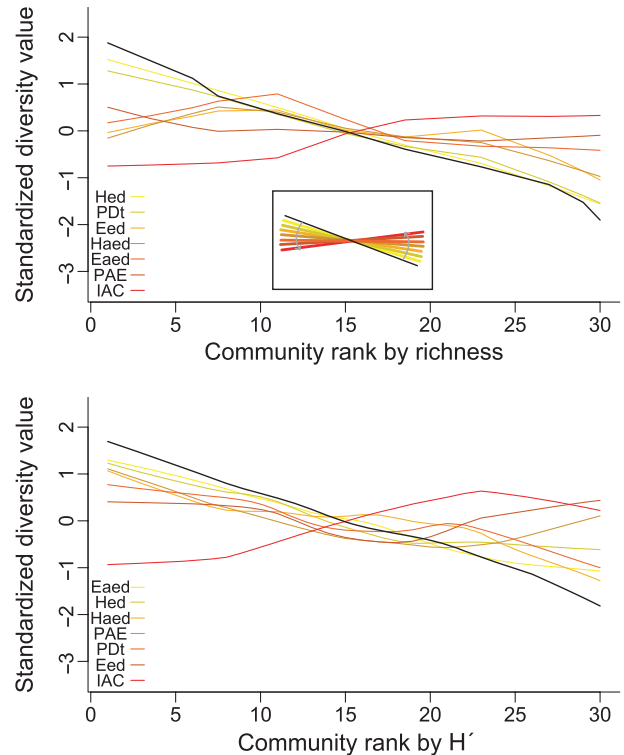


Figure 4 Deviations in community ranks by the abundance-weighted phylogenetic metrics relative to richness (top panel – black line) and H' (bottom panel – black line). Metric were LOWESS smoothed, and colours are ranked according to the magnitude of the absolute differences between the two rankings with red indicating the greatest differences. The inset in the top panel shows idealized differences where red-spectrum colours tend towards a lack of correlation in community rankings. Thus, red metrics rank communities according to differing criteria.

negatively correlated to richness ($r = -0.46$, $P = 0.011$; Appendix 4), meaning that more species-rich communities tended to be more imbalanced, whereas H_{AED} was positively related to community richness ($r = 0.42$, $P = 0.022$; Appendix 4). Neither PAE nor E_{AED} were significantly correlated with richness (Appendix 4), PAE and E_{AED} therefore provide information on community structure and diversity not captured by richness alone.

Our abundance-weighted phylogenetic metrics are correlated with the Shannon Index (H' , Fig. 4) to some degree (PD, H_{ED} , H_{AED} and E_{AED} , positively so and IAC negatively –Appendix 4), but ranked communities substantially differently. The diversity measures that do not integrate abundance and phylogenetic information (PD, H' , H_{ED} and E_{ED}) were uncorrelated with total community abundance (Appendix 4). However, of the four abundance-weighted metrics, two (PAE and H_{AED}) were also uncorrelated with abundance, while IAC increased ($r = 0.68$, $P < 0.001$) and E_{AED} decreased

($r = -0.41$, $P = 0.023$) with total community abundance (Appendix 4). There has been much recent interest in the use of phylogenetic measures of community structure, the Jasper Ridge data illustrates how our metrics weight differently information for calculating community diversity. Differences among metrics could open the door to further hypotheses testing.

DISCUSSION

We developed these metrics because the incorporation of species abundances into measures of PD has been largely lacking, but of recent interest (e.g. see: Pavoine *et al.* 2005; Sherwin *et al.* 2006; Hardy & Senterre 2007). Abundance-informed measures of PD are potentially important for several reasons. First, ecological processes and dynamics can be sensitive to community abundance distributions (Hillebrand *et al.* 2008). For example, rates of soil nutrient depletion or insect herbivore diversity may be significantly affected by the distribution of traits within a plant community, and the distribution of traits is a product of both the evolution of those traits and the relative abundances of the plant species that possess them.

Secondly, phylogenetic effects on community coexistence have traditionally focused on the presence/absence of species (e.g. Webb 2000; Cavender-Bares *et al.* 2006; Cadotte *et al.* 2008), but knowing abundances can potentially inform our understanding of mechanisms affecting coexistence. Species that co-occur despite identical or highly overlapping niche requirements should each have lower abundances than when not together (Chesson 2000). If niche differences/similarities have a phylogenetic signal, then our metrics could reveal evidence for competition not apparent from species presence–absence data. In addition, it would also be possible to explore deviations from expected abundance distributions. These metrics open the door to new sets of hypothesis tests.

Finally, conservation decision-making is increasingly considering phylogenetic information (e.g. Faith 1992; Crozier 1997; Redding & Mooers 2006; Isaac *et al.* 2007). Our metrics allow conservation biologists to quantify both community phylogenetic distinctiveness and rarity simultaneously. Most germane is our relative AED metric that scores the relative importance (in terms of information value) of each species in a community. For example, the highest importance score in our hypothetical communities (Fig. 2) was given to *Sp. a* in community B because it is both the most phylogenetically distinct and rare. Using such a metric across a number of sites could give priority to species that may be overlooked by traditional rarity indices.

Using the Jasper Ridge survey data, we show that our metrics of PD rank communities differently compared to traditional diversity metrics. No single diversity metric

encapsulates all aspects of community diversity, thus the choice of metric should be conditional on the question of interest. Our simulations illustrate that PD metrics fall into a few broad categories that capture different aspects of phylogenetic tree shape and community structure.

LIMITATIONS AND QUALIFICATIONS

The metrics proposed here are potentially sensitive to phylogenetic construction methods and assumptions. The use of backbone phylogenetic trees (e.g. Davies *et al.* 2004) to construct community phylogenies using tools such as Phylomatic (Webb & Donoghue 2005) inevitably produces polytomies below the family level. Assuming equal branch lengths among closely related taxa will result in reduced variation in our metrics that use branch length information (i.e. PAE and AED). However, the influence of subfamilial and subgeneric branch length estimates matters only when important trait or niche differences have evolved at those levels compared with differences represented deeper in the phylogeny. Some recent work has found that using a fully resolved molecular phylogeny vs. one obtained from a supertree (with polytomies) has little effect on the ability to detect correlations between PD and ecological patterns across numerous species (Cadotte *et al.* 2008, 2009). However, it is likely that complex ecosystem processes represent an aggregate effect of multiple traits constrained at different phylogenetic tree depths. Nonetheless, the generality of patterns of niche evolution remains unresolved (Losos 2008), as do phylogenetic issues such as the effect of variation in, for example, taxon sampling or branch length estimation methods (i.e. using nucleotide substitution rates vs. a time-scaled ultrametric tree) on our ability to detect the ecological effects of niche evolution.

What our metrics provide beyond basic measures of PD is the potential for additional explanatory power in the interpretation of community assembly and ecosystem functionality. This additional information suggests that abundance is important, and that ecological and evolutionary processes operating below the species-level need to be considered when studying community-level interactions and diversity (Johnson & Stinchcombe 2007; Hughes *et al.* 2008). Each of our metrics provides different information, potentially useful with different phylogenies or for testing differing hypotheses; however, we think that our final measure, AED, may be the most powerful, as it incorporates both branch length and topological information.

In exploring these metrics, we have neither explicitly commented on the abundance measures used nor their interpretation. All these metrics are conceptually intended to work with count data, where n_i is the number of individuals of species I in a given community. These metrics work with other measures of abundance such as biomass or per cent

cover, but one needs to be cognizant of the resulting interpretation of the metrics. When abundance is measured as the number of individuals, then each metric can be interpreted as partitioning phylogenetic information among individuals. However, with other abundance measures, the partitioning of phylogenetic information is among these units (e.g. g m^{-2} or % cover). Abundance in the Jasper Ridge community plots was per cent cover, hence, phylogenetic information was partitioned by per cent cover.

CONCLUSION

Incorporating abundance information into PD provides a link to existing, non-phylogenetic measures of diversity that are often based on equitability of relative abundances rather than presence within a community. Just as abundance-weighted, community-level measures of functional diversity are better predictors of ecosystem processes than those based on unweighed metrics (Vile *et al.* 2006; Cingolani *et al.* 2007), PD measures that incorporate abundance information may improve our ability to understand and predict the impact of evolutionary history on ecological interactions.

ACKNOWLEDGEMENTS

MWC wishes to thank Jon Lovett-Doust for inspiring him to think about evolutionary history and geographical abundances. We wish to thank Stephano Allesina for providing insights into developing an entropic measure and to Arne Mooers and an anonymous reviewer for comments on an earlier version of the manuscript. For MWC and TJD, part of this work was conducted as Postdoctoral Associates at the National Center for Ecological Analysis and Synthesis, a Center funded by NSF (Grant #DEB-0553768), the University of California, Santa Barbara, and the State of California. SWK was supported by a NSERC postdoctoral fellowship.

REFERENCES

- Agapow, P.M. & Purvis, A. (2002). Power of eight tree shape statistics to detect nonrandom diversification: a comparison by simulation of two models of cladogenesis. *Syst. Biol.*, 51, 866–872.
- Allen, B., Kon, M. & Bar-Yam, Y. (2009). A new phylogenetic diversity measure generalizing the Shannon index and its application to Phyllostomid bats. *Am. Nat.*, 174, 236–243.
- Cadotte, M.W., Cardinale, B.J. & Oakley, T.H. (2008). Evolutionary history and the effect of biodiversity on plant productivity. *Proc. Natl Acad. Sci. USA*, 105, 17012–17017.
- Cadotte, M.W., Hamilton, M.A. & Murray, B.R. (2009). Phylogenetic relatedness and plant invader success across two spatial scales. *Divers. Distrib.*, 15, 481–488.
- Cavender-Bares, J. & Wilczek, A. (2003). Integrating micro- and macroevolutionary processes in community ecology. *Ecology*, 84, 592–597.
- Cavender-Bares, J., Keen, A. & Miles, B. (2006). Phylogenetic structure of floridian plant communities depends on taxonomic and spatial scale. *Ecology*, 87, S109–S122.
- Cavender-Bares, J., Kozak, K., Fine, P.V.A. & Kembel, S.W. (2009). The merging of community ecology and phylogenetic biology. *Ecol. Lett.*, 12, 693–715.
- Chesson, P. (2000). Mechanism of maintenance of species diversity. *Annu. Rev. Ecol. Syst.*, 31, 343–366.
- Cingolani, A.M., Cabido, M., Gurvich, D.E., Renison, D. & Diaz, S. (2007). Filtering processes in the assembly of plant communities: are species presence and abundance driven by the same traits? *J. Veg. Sci.*, 18, 911–920.
- Crozier, R.H. (1997). Preserving the information content of species: genetic diversity, phylogeny, and conservation worth. *Annu. Rev. Ecol. Syst.*, 28, 243–268.
- Crozier, R.H., Dunnnett, L.J. & Agapow, P.M. (2005). Phylogenetic biodiversity assessment based on systematic nomenclature. *Evol. Bioinform.*, 1, 11–36.
- Davies, T.J., Barraclough, T.G., Chase, M.W., Soltis, P.S., Soltis, D.E. & Savolainen, V. (2004). Darwin's abominable mystery: insights from a supertree of the angiosperms. *Proc. Natl Acad. Sci. USA*, 101, 1904–1909.
- Davies, T.J., Fritz, S.A., Grenyer, R., Orme, C.D.L., Bielby, J., Bininda-Emonds, O.R.P. *et al.* (2008). Phylogenetic trees and the future of mammalian biodiversity. *Proc. Natl Acad. Sci. USA*, 105, 11556–11563.
- Dukes, J.S., Chiariello, N.R., Cleland, E.E., Moore, L.A., Shaw, M.R., Thayer, S. *et al.* (2005). Responses of grassland production to single and multiple global environmental changes. *PLoS Biol.*, 3, 1829–1837.
- Faith, D.P. (1992). Conservation evaluation and phylogenetic diversity. *Biol. Conserv.*, 61, 1–10.
- Faith, D.P. (2006). The role of the phylogenetic diversity measure, PD, in bioinformatics: getting the definition right. *Evol. Bioinform.*, 2, 277–283.
- Felsenstein, J. (1985). Phylogenies and the comparative method. *Am. Nat.*, 125, 1–15.
- Fisher, R.H., Corbet, A.S. & Williams, C.B. (1943). The relation between the number of species and the number of individuals in a random sample of animal population. *J. Anim. Ecol.*, 12, 42–58.
- Hardy, O.J. & Senterre, B. (2007). Characterizing the phylogenetic structure of communities by an additive partitioning of phylogenetic diversity. *J. Ecol.*, 95, 493–506.
- Harvey, P.H. & Pagel, M. (1991). *The Comparative Method in Evolutionary Biology*. Oxford University Press, Oxford.
- Helmus, M.R., Bland, T.J., Williams, C.K. & Ives, A.R. (2007). Phylogenetic measures of biodiversity. *Am. Nat.*, 169, E68–E83.
- Hillebrand, H., Bennett, D. & Cadotte, M.W. (2008). The consequences of dominance: a review of the effects of evenness on local and regional ecosystem processes. *Ecology*, 89, 1510–1520.
- Hubbell, S.P. (2001). *The Unified Neutral Theory of Biodiversity and Biogeography*. Princeton University Press, Princeton, NJ.
- Hughes, A.R., Inouye, B.D., Johnson, M.T.J., Underwood, N. & Vellend, M. (2008). Ecological consequences of genetic diversity. *Ecol. Lett.*, 11, 609–623.

- Isaac, N.J.B., Turvey, S.T., Collen, B., Waterman, C. & Baillie, J.E.M. (2007). Mammals on the EDGE: conservation priorities based on threat and phylogeny. *PLoS ONE*, 2, e296.
- Johnson, M.T.J. & Stinchcombe, J.R. (2007). An emerging synthesis between community ecology and evolutionary biology. *Trends Ecol. Evol.*, 22, 250–257.
- Losos, J.B. (2008). Phylogenetic niche conservatism, phylogenetic signal and the relationship between phylogenetic relatedness and ecological similarity among species. *Ecol. Lett.*, 11, 995–1007.
- MacArthur, R.H. (1957). On the relative abundance of bird species. *Proc. Natl Acad. Sci. USA*, 43, 293–295.
- Mace, G.M., Gittleman, J.L. & Purvis, A. (2003). Preserving the tree of life. *Science*, 300, 1707–1709.
- Magurran, A.E. (2003). *Measuring Biological Diversity*. Blackwell Science, Oxford.
- Pardi, F. & Goldman, N. (2007). Resource-aware taxon selection for maximizing phylogenetic diversity. *Syst. Biol.*, 56, 431–444.
- Pavoine, S., Ollier, S. & Dufour, A.-B. (2005). Is the originality of a species measurable? *Ecol. Lett.*, 8, 579–586.
- Petchey, O.L. & Gaston, K.J. (2002). Functional diversity (FD), species richness and community composition. *Ecol. Lett.*, 5, 402–411.
- Preston, F.W. (1948). The commonness and rarity of species. *Ecology*, 29, 254–283.
- Pybus, O. & Harvey, P.H. (2000). Testing macro-evolutionary models using incomplete molecular phylogenies. *Proc. R. Soc. B Biol. Sci.*, 267, 2267–2272.
- R Development Core Team (2009). *R: A Language and Environment for Statistical Computing*. In: R Foundation for Statistical Computing Vienna, Austria.
- Redding, D.W. & Mooers, A.O. (2006). Incorporating evolutionary measures into conservation prioritization. *Conserv. Biol.*, 20, 1670–1678.
- Redding, D.W., Hartmann, K., Mimoto, A., Bokal, D., DeVos, M. & Mooers, A.O. (2008). Evolutionarily distinctive species often capture more phylogenetic diversity than expected. *J. Theor. Biol.*, 251, 606–615.
- Sanderson, M.J. (1997). A nonparametric approach to estimating divergence times in the absence of rate constancy. *Mol. Biol. Evol.*, 14, 1218–1231.
- Sherwin, W.B., Jabot, F., Rush, R. & Rossetto, M. (2006). Measurement of biological information with applications from genes to landscapes. *Mol. Ecol.*, 15, 2857–2869.
- Silvertown, J., Dodd, M., Gowing, D., Lawson, C. & McConway, K. (2006). Phylogeny and the hierarchical organization of plant diversity. *Ecology*, 87, S39–S49.
- Vanewright, R.I., Humphries, C.J. & Williams, P.H. (1991). What to protect – systematics and the agony of choice. *Biol. Conserv.*, 55, 235–254.
- Vile, D., Shipley, B. & Garnier, E. (2006). Ecosystem productivity can be predicted from potential relative growth rate and species abundance. *Ecol. Lett.*, 9, 1061–1067.
- Villegger, S., Mason, N.W.H. & Mouillot, D. (2008). New multidimensional functional diversity indices for a multifaceted framework in functional ecology. *Ecology*, 89, 2290–2301.
- Webb, C.O. (2000). Exploring the phylogenetic structure of ecological communities: an example for rain forest trees. *Am. Nat.*, 156, 145–155.
- Webb, C.O. & Donoghue, M.J. (2005). Phylomatic: tree assembly for applied phylogenetics. *Mol. Ecol. Notes*, 5, 181–183.
- Webb, C.O., Ackerly, D.D., McPeck, M.A. & Donoghue, M.J. (2002). Phylogenies and community ecology. *Annu. Rev. Ecol. Syst.*, 33, 475–505.
- Webb, C.O., Losos, J.B. & Agrawal, A.A. (2006). Integrating phylogenies into community ecology. *Ecology*, 87, S1–S2.
- Webb, C.O., Ackerly, D.D. & Kembel, S.W. (2008). Phylocom: software for the analysis of phylogenetic community structure and trait evolution. *Bioinformatics*, 24, 2098–2100.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Appendix S1 Estimating total phylogenetic diversity including among-individual divergences.

Appendix S2 Calculating AED when intraspecific divergences are estimated.

Appendix S3 Construction of example and Jasper Ridge phylogenies.

Appendix S4 Correlations among metrics from the Jasper Ridge data.

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.

Editor, Ross Crozier

Manuscript received 2 September 2009

First decision made 16 September 2009

Manuscript accepted 1 October 2009