Formal Links between Feature Diversity and Phylogenetic Diversity

Kristina Wicke[1], Arne Mooers[2], and Mike Steel[3,*]

[1]*Institute of Mathematics and Computer Science, University of Greifswald, Greifswald, 17489, Germany*
[2]*Department of Biological Sciences and the Crawford Lab for Evolutionary Studies, Simon Fraser University, Burnaby, British Columbia, V5A1S6, Canada*
[3]*Biomathematics Research Centre University of Canterbury, Christchurch, 8140, New Zealand*
Corresponding author: mike.steel@canterbury.ac.nz

Abstract.—The extent to which phylogenetic diversity (PD) captures feature diversity (FD) is a topical and controversial question in biodiversity conservation. In this short paper, we formalize this question and establish a precise mathematical condition for FD (based on discrete characters) to coincide with PD. In this way, we make explicit the two main reasons why the two diversity measures might disagree for given data; namely, the presence of certain patterns of feature evolution and loss, and using temporal branch lengths for PD in settings that may not be appropriate (e.g., due to rapid evolution of certain features over short periods of time). Our article also explores the relationship between the "Fair Proportion" index of PD and a simple index of FD (both of which correspond to Shapley values in cooperative game theory). In a second mathematical result, we show that the two indices can take identical values for any phylogenetic tree, provided the branch lengths in the tree are chosen appropriately. [Evolutionary distinctiveness; feature diversity; phylogenetic diversity; shapley value.]

1 Introduction

Almost 30 years ago, Dan Faith published a seminal paper that laid out how phylogenies might aid in identifying sets of species with maximal "feature diversity" (Faith 1992). Faith's stated goal was to support practical biodiversity conservation in the face of limited resources, coupled with the assumption that maximising feature diversity (the total number of unique character states represented by a set of taxa) was a desirable conservation target.

Drawing on the call of Vane-Wright et al. (1991) to consider taxonomic distinctiveness when prioritizing species, Faith introduced the phylogenetic diversity (PD) metric, simply the sum of the edge lengths of the minimal subtree linking a subset of species to the root of the encompassing phylogeny (also called the "minimum spanning path" by Faith (1992)). Importantly, these edge lengths were given in units of reconstructed character changes under maximum parsimony on the cladogram representing a character state matrix with no homoplasy. Faith showed, with an example, that the sum of these reconstructed edge lengths would lead to the same total feature diversity as that calculated from the character matrix itself. Importantly, if these cladistic edge lengths are representative of all features, then maximizing PD (e.g., over a given subset size) would maximise feature diversity, even in the face of some homoplasy. The bulk of Faith's (1992) paper was devoted to introducing the machinery to maximize PD.

Efficient algorithms for finding maximum PD sets are available (Bordewich et al. 2008), the metric has been extended to networks (Minh et al. 2009), and there are countless case studies that both measure and optimize PD for conservation (see, e.g., Pollock et al. 2017); Faith's original paper has been cited in excess of 2000 times. A recent review (Tucker et al. 2019) considered the literature concerning both the empirical correlations between PD and feature diversity, and the expected relationship between PD and various conservation values.

Surprisingly, though, the necessary conditions under which PD will capture feature diversity have never been formalized. Here, by using discrete characters, a model with no homoplasy, and appropriate edge lengths, we prove that the PD of a subtree does indeed measure feature diversity as defined by (Faith 1992). This proof allows us to state more formally when PD does not necessarily capture feature diversity, thereby allowing for further modeling and statistical evaluation of the expected relationship under more realistic models. Given the close connection between PD and taxonomic distinctiveness, we also consider the conditions under which its phylogenetic measure (specifically, the Shapley value of evolutionary isolation) can capture its feature-based analog.

2 Preliminaries

2.1 Feature Diversity

Consider a set X of taxa with $|X| = n$, and suppose that each taxon $x \in X$ has an associated finite set F_x of "features." To allow extra generality, we will assume that each element $f \in F_x$ has a corresponding positive score $\mu(f) \in \mathbb{R}^{>0}$, which might be viewed as a measure of the complexity, novelty, or richness of f (the default option is to set $\mu(f) = 1$ for all f). Let $\mathcal{F}$ denote the set of all features present amongst the taxa in the collection X, and let $\mathbb{F} = (F_x : x \in X)$ be the ordered n-tuple containing the feature sets of the taxa in X. We will sometimes call $\mathbb{F}$ a *feature assignment* as it summarizes how a set of features is assigned to each taxon in X.

Note that $\mathbb{F}$ provides the same information as a table showing the presence and absence of features across

1

TABLE 1.   A standard character state matrix (0=absence, 1=presence) representing the assignment of four features (f_1,f_2,f_3,f_4) across three taxa (a,b,c).

Taxon	f_1	f_2	f_3	f_4
a	1	1	0	1
b	1	0	1	0
c	0	1	1	0

taxa. So if $X=\{a,b,c\}$, then the feature assignment $\mathbb{F}=(F_a,F_b,F_c)$, where $F_a=\{f_1,f_2,f_4\}$, $F_b=\{f_1,f_3\}$, and $F_c=\{f_2,f_3\}$, corresponds to a standard character state matrix where there are two states per feature: presence (1) or absence (0) (see Table 1).

Given a subset Y of X, let

$$\nu_{\mathbb{F}}(Y):=\sum_{f\in\cup_{x\in Y}F_x}\mu(f).$$

Thus, $\nu_{\mathbb{F}}(Y)$ is the sum of the values of the features that are present in at least one taxon in Y. We refer to $\nu_{\mathbb{F}}(Y)$ as the *feature diversity* (FD) of Y, Note that in this sum, each feature is counted only once if present, and that our FD is closer to "trait diversity" as used by (Tucker et al. 2019), and distinct from "functional diversity" as used more generally in ecology.

The function $\nu_{\mathbb{F}}$ (which assigns each subset Y of X a non-negative real value $\nu_{\mathbb{F}}(Y)$) clearly satisfies the following two properties: $\nu_{\mathbb{F}}(\emptyset)=0$ and $\nu_{\mathbb{F}}$ is monotone (i.e., $Y\subseteq Y'\Rightarrow\nu_{\mathbb{F}}(Y)\leq\nu_{\mathbb{F}}(Y')$). Moreover, $\nu_{\mathbb{F}}$ also satisfies the submodularity inequality:

$$\nu_{\mathbb{F}}(Y\cup Y')+\nu_{\mathbb{F}}(Y\cap Y')\leq\nu_{\mathbb{F}}(Y)+\nu_{\mathbb{F}}(Y'),\qquad(2.1)$$

and a proof is provided in the Appendix.

2.2 Phylogenetic Diversity

Now consider a rooted phylogenetic X-tree $T=(V,E)$ with root ρ, leaf set X, and edge length assignment $\ell:E\to\mathbb{R}^{\geq0}$. For technical reasons (by allowing greater generality in the statement of our results, in particular, when later on considering features that are present in *all* taxa under consideration), we assume that T has an additional "stem edge" (ρ',ρ), where ρ' is a degree-1 vertex and ρ has in-degree 1 and out-degree at least 2. Note that we do not consider ρ' as a leaf of T, that is, as a present-day species, but as an ancestral species that lived sometime in the past. The reason for including this stem edge is that features may have evolved ancestrally to some clade, and this stem edge provides a way to formally accommodate this possibility. The *phylogenetic diversity* (PD) of a subset Y of X is usually defined as the sum of the lengths of the edges in the minimal subtree of T that contains the leaves in Y and the root ρ of T. Here, we extend this definition by also including the length of the stem edge (ρ',ρ) in the calculation of PD for any subset $Y\subseteq X$ with $|Y|\geq1$. This adds a constant, namely $\ell((\rho',\rho))$, to all subsets $Y\subseteq X\setminus\emptyset$ but does not affect properties of PD, such as its monotonicity or submodularity.

3 LINKING FEATURE DIVERSITY TO PHYLOGENETIC DIVERSITY

The extinction of species at the present leads to both the loss of feature diversity and phylogenetic diversity. This is illustrated in Figure 1, for a set $\mathbb{F}=\{\alpha,\beta,\gamma,\delta\}$ of four features.

Now consider a model, based on a rooted phylogenetic X-tree T in which (i) each feature in $\mathcal{F}$ arises on exactly one edge of T and (ii) each feature that arises is never lost and is passed down to all descendant vertices (including the leaves). This is just a model where every feature is a perfect synapomorphy.

We can describe this more precisely by specifying a map $h:\mathcal{F}\to E$, which indicates which edge each given feature arises on (note that several features may arise on the same edge). Thus, $h^{-1}(e)$ denotes the set of features that arise on edge e. Here, we assume that $h^{-1}(e)\neq\emptyset$ for all interior edges of T (i.e., each interior edge of T gives rise to at least one feature). Notice that this is equivalent to allowing interior edges with "no event" (i.e., without a feature arising on them) and then contracting all interior "no event" edges.

Note, however, that there may be pendant edges incident to leaves of T on which no features arise. Similarly, it may be the case that no feature arises on the stem edge; in particular, no features arise on this stem edge precisely when there is no feature that is present in every taxon.

Under this model, F_x is then equal to the union of the sets $h^{-1}(e)$ over all the edges e on the (unique) path from ρ' to leaf x.

When a feature assignment $\mathbb{F}$ can be realized in this way, we will denote this by writing $\mathbb{F}=F[T,h]$. Not every feature assignment $\mathbb{F}$ can be realized in this way (on any tree). As an example, consider the feature assignment described by the character matrix in Table 1. In this case, there is no rooted phylogenetic X–tree $T=(V,E)$ and map $h:\{f_1,f_2,f_3,f_4\}\to E$ for which $(F_a,F_b,F_c)=F[T,h]$.

Fortunately, it is easy to characterize precisely when a feature assignment $\mathbb{F}$ can be realized as $F[T,h]$, and where T is either stipulated or not. The required condition corresponds to the well-known structure of characters necessary (and sufficient) to perfectly fit a common phylogenetic tree, namely that character states are arranged among taxa as a set of nested apomorphies.

To describe this, we first introduce some additional notation. Let $X_f:=\{x\in X:f\in F_x\}$ denote the subset of taxa in X that have feature f. Moreover, let $\mathcal{C}_{\mathcal{F}}:=\{X_f:f\in\mathcal{F}\}$ be the collection of the sets X_f. The following result easily follows from other well-known results in phylogenetics (in particular, Proposition 2.1 in (Steel 2016)).

Proposition 1.

(i) $\mathbb{F}=F[T,h]$ *for some* $h:\mathcal{F}\to E$ *if and only if* X_f *corresponds to a cluster of T for each feature $f\in\mathcal{F}$. Moreover, when $\mathbb{F}=F[T,h]$, the map h is uniquely determined: for each $f\in\mathcal{F}$, $h(f)$ is the edge directly above the most recent common ancestor of the taxa in X_f.*
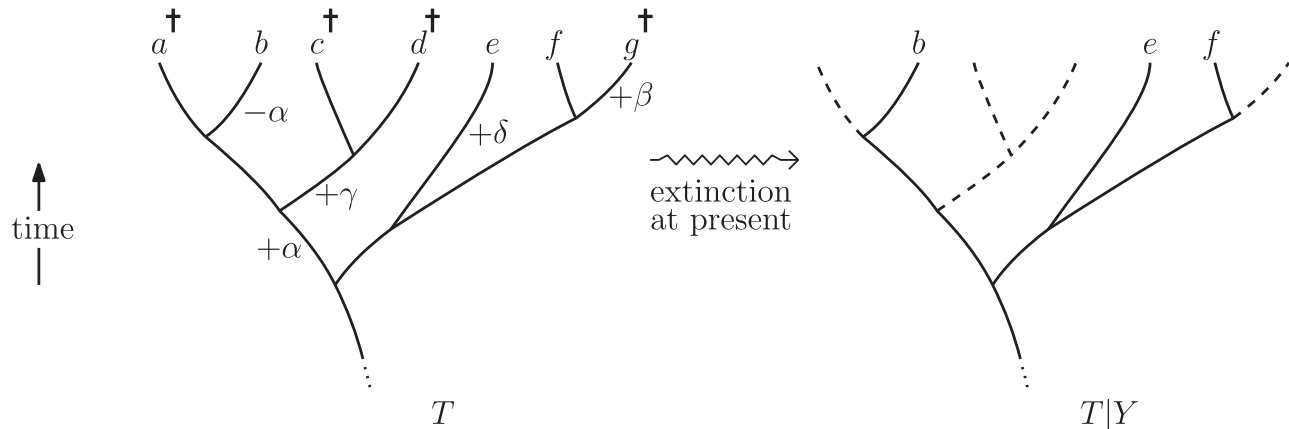
FIGURE 1.   The extinction of species a, c, d, and g at the tips (leaves) of the phylogenetic tree T leads to a PD loss equal to the sum of the lengths of dashed edges absent from the "pruned" tree $T|Y$ of the surviving species $Y = \{b, e, f\}$. In addition, the four features (α, β, γ, and δ) that have evolved in T under some birth–death processes (where $+ =$ birth, $- =$ death) are all present among the pre-extinction set of tip species, but only δ is still present following extinction, so feature diversity (FD) has also declined. In general, an edge may have several birth and/or death events.

(ii) *There exists a tree T and map h such that $\mathbb{F} = F[T, h]$ if and only if $\mathcal{C}_\mathcal{F}$ is a hierarchy on X. In other words, for all pairs $X_f, X_{f'} \in \mathcal{C}_\mathcal{F}$, we have $X_f \cap X_{f'} \in \{\emptyset, X_f, X_{f'}\}$ (i.e., X_f and $X_{f'}$ are either disjoint or nested).*

4   FIRST MAIN RESULT

We can now describe the relationship between FD and PD in a precise way. The proof is provided in the Appendix.

Theorem 1. *Let T be a rooted phylogenetic X-tree and let $\mathbb{F}$ be an assignment of features across the taxa in X.*

(i) *$\mathbb{F} = F[T, h]$ for some function $h : \mathcal{F} \to E$ if and only if $\nu_\mathbb{F}$ is exactly equal to the PD function for some edge length assignment ℓ of T that assigns strictly positive lengths to all interior edges of T and non-negative lengths to all pendant edges and the stem edge (ρ', ρ) (i.e., $\nu_\mathbb{F}(Y) = \mathrm{PD}_{(T,\ell)}(Y)$ for all subsets Y of X).*

(ii) *When (i) holds, h and ℓ are both uniquely determined. In particular, $\ell = \ell_h$, where, for each edge e of T, $\ell_h(e) := \sum_{f:h(f)=e} \mu(f)$ (and $\ell_h(e) = 0$ for each pendant edge e of T with $h^{-1}(e) = \emptyset$).*

As a simple illustration of Theorem 1 (and Proposition 1) consider $\nu_\mathbb{F}$, where $\mathbb{F}$ consists of the four characters in Table 1, together with the rooted phylogenetic tree $T = ((a, b), c)$. By Proposition 1, it is easily checked that $\mathbb{F}$ cannot be realized as $F[T, h]$ for any map h. It follows from Theorem 1 that $\nu_\mathbb{F}$ cannot be made to correspond exactly to phylogenetic diversity on the tree T using non-negative edge lengths that are strictly positive on the interior edges (moreover, this holds for any other tree on leaf set $\{a, b, c\}$). In general, if $\mathbb{F}$ does

not equal $F[T, h]$ for some map h then $\nu_\mathbb{F}$ cannot be made to correspond exactly to phylogenetic diversity on the tree T.

5   DIVERSITY INDICES

A *diversity index* for FD (or PD) is a non-negative score assigned to each taxon $x \in X$ that sums to the total FD (or PD, respectively) of X. Diversity indices can be viewed as a way to apportion FD (or PD) fairly among the extant taxa. Although there are various ways to do this, we focus on one that is characterized by simple axioms, namely, the Shapley value (from cooperative game theory), which coincides, in the PD setting, with the well-known Fair Proportion index (described below).

5.1   Feature Diversity Index

Given $\mathbb{F}$, let

$$\varphi_\mathbb{F} : X \to \mathbb{R}^{\geq 0}$$

be the function defined by:

$$\varphi_\mathbb{F}(x) := \sum_{f \in F_x} \frac{\mu(f)}{n(f)},$$

where $n(f)$ is the number of taxa that have feature f (i.e., $n(f) = |X_f|$). In words, $\varphi_\mathbb{F}(x)$ assigns to each taxon x a sum of scores—one score for each of its features—where the score for feature f is $\mu(f)$ if x is the only taxon having this feature; otherwise, the score equals $\mu(f)$ divided by the total number of taxa having feature f.

The following result provides a formal justification for regarding $\varphi_\mathbb{F}$ as a natural index of FD. Note that this index does not depend on any underlying phylogeny, or on assumptions concerning feature evolution. It is also easily computed.

The result is phrased within the general framework of cooperative game theory (a topic more well-known in economics than biology, though it has recently been applied to PD, as we discuss below). In this general framework, one has a finite set X and a function s that assigns to each subset Y of X a corresponding score $s(Y)$ with $s(\emptyset)=0$ (in our current setting $s(Y)=\nu_{\mathbb{F}}(Y)$). Given the pair (X,s) (also called a "cooperative game"), one seeks to apportion the score of the full set X among each of its elements according to an index (i.e., a value for each element of X) in a way that reflects the contribution each element makes to the total score. In this general framework, there is a particular index, called the *Shapley value*, that is uniquely determined by well-motivated axioms, and which is given by an explicit (if somewhat complex) combinatorial expression (Shapley 1953). To be precise, given the pair (X,s) and an element $x \in X$, the Shapley value for x is defined as:

$$SV_s(x) = \frac{1}{n!} \sum_{S \subseteq X : x \in S} (|S|-1)!(n-|S|)! \Delta(S,x),$$

where $\Delta(S,x)=s(S)-s(S\setminus\{x\})$, and where $n=|X|$. In words, the Shapley value for an element (here, a leaf) is its expected contribution to the score (here, feature diversity) of all possible future subsets of elements from the set (here, subsets of taxa), where subset sizes are equiprobable (Steel 2016). The Shapley value satisfies four desirable axioms (in fact, it is uniquely characterized by them), namely efficiency, symmetry, a dummy axiom, and additivity. Efficiency states that the sum of Shapley values equals the score of X, that is, $\sum_{x \in X} SV_s(x)=s(X)$. Symmetry implies that two elements $x,y \in X$ with $x \neq y$ that contribute the same amount of worth (in our case, of feature diversity) to any subset they join receive the same Shapley value (formally, for all $x,y \in X$ with $x \neq y$ and for all $S \subseteq X \setminus \{x,y\}$, if $s(S \cup \{x\})=s(S \cup \{y\})$, then $SV_s(x)=SV_s(y)$). The dummy axiom, on the other hand, states that an element $x \in X$ that does not contribute any worth to any subset obtains a Shapley value of zero (formally, if for all $S \subseteq X \setminus \{x\}$, $s(X \cup \{x\})=s(X)$, then $SV_s(x)=0$). Finally, additivity constitutes the technical property that given a set X and two different score functions, say s_1 and s_2, we have for all $x \in X$: $SV_{s_1+s_2}(x)=SV_{s_1}(x)+SV_{s_2}(x)$.

Our next result states that the simple FD index $\varphi_{\mathbb{F}}$ introduced above coincides with the Shapley value for $s=\nu_{\mathbb{F}}$ (the proof is provided in the Appendix).

Proposition 2. *The FD index $\varphi_{\mathbb{F}}$ is precisely the Shapley value for the pair $(X,\nu_{\mathbb{F}})$. In particular, $\sum_{x \in X} \varphi_{\mathbb{F}}(x)=\nu_{\mathbb{F}}(X)$.*

5.2 Phylogenetic Diversity Index

Given the pair (T,ℓ), the *Fair Proportion index* (FP) (from Redding (2003) and Redding et al. (2007), see also (Isaac et al. 2007)) for taxon x is given by:

$$FP_{(T,\ell)}(x) = \sum_{e \in P(T;\rho',x)} \frac{1}{n(e)} \cdot \ell(e),$$

where $P(T;\rho',x)$ denotes the unique path from ρ' to x and where $n(e)$ is the number of leaves descending from the endpoint of edge e closest to the leaves.

It turns out that the FP index coincides exactly with the Shapley value based on PD (i.e., when PD is used as the characteristic function in the underlying cooperative game), a result first shown by Fuchs and Jin (2015). As $\varphi_{\mathbb{F}}$ is (by Proposition 2) equivalent to the Shapley value based on FD, Theorem 1 thus has the interesting implication that if a feature assignment $\mathbb{F}$ can be realized on a tree (i.e., if $\mathbb{F}=F[T,h]$), then the Shapley values based on PD and FD coincide.

Proposition 3. *If $\mathbb{F}=F[T,h]$, then $\varphi_{\mathbb{F}}(x)$ is equal to the Fair Proportion index for taxon x on tree T for the edge length assignment ℓ_h.*

The proof of this statement is given in the Appendix. We now establish a further result. We show that $\varphi_{\mathbb{F}}(x)$ can always be interpreted as $FP_{(T,\ell)}(x)$ for any tree T (even if $\mathbb{F} \neq F[T',h]$ for any tree T').

Theorem 2. *Let $\mathbb{F}$ be a feature assignment such that $F_x \neq \emptyset$ for all $x \in X$, and let T be any rooted phylogenetic X-tree (with additional stem edge). Then, there exists an edge length assignment $\ell : E \to \mathbb{R}^{>0}$ that assigns strictly positive lengths to all edges of T, such that $\varphi_{\mathbb{F}}(x)=FP_{(T,\ell)}(x)$ for all $x \in X$.*

A simple example to illustrate Theorem 2 is provided in Figure 2 for the assignment of features given in Table 1. Taking $\mu(f_i)=1$ for $i=1,2,3,4$, we have $\varphi_{\mathbb{F}}(a)=\frac{1}{2}+\frac{1}{2}+1=2$ and $\varphi_{\mathbb{F}}(b)=\varphi_{\mathbb{F}}(c)=\frac{1}{2}+\frac{1}{2}=1$. Then, $\varphi_{\mathbb{F}}$ can be realized as a FP index on each of the four possible phylogenetic trees on leaf set $\{a,b,c\}$ by strictly positive edge lengths as shown. Other choices of strictly positive edge lengths for the trees in (ii)–(iv) are possible, and for the star tree we require $0 < \epsilon < 1 = \min_{x \in X} \varphi_{\mathbb{F}}(x)$.

The proof of Theorem 2 is provided in the Appendix; however, we provide an outline of the argument here. First observe that if T is a star tree then Theorem 2 clearly holds, since we can simply assign edge length $\varphi_{\mathbb{F}}(x)-\epsilon$ to the edge incident with leaf x and length $|X|\epsilon$ to the stem edge (for $0 < \epsilon < \min_{x \in X} \varphi_{\mathbb{F}}(x)$) and obtain $\varphi_{\mathbb{F}}(x)=FP_{(T,\ell)}(x)$ for all $x \in X$. If T is not a star tree, then we could assign edge length 0 to all the interior edges and the stem edge and length $\varphi_{\mathbb{F}}(x)$ to the pendant edge incident with leaf x in order to obtain the required identity, but this does not satisfy the additional condition promised in the theorem. The non-trivial part of the proof of Theorem 2 is to show that one can "lift" some fraction of the lengths of the pendant edges so as that (i) all edges of T have strictly positive length, and in such a way that (ii) the required identity between the FD and PD diversity indices holds for each taxon x. The formal proof of Theorem 2 actually establishes a slightly stronger result to that stated, and one that is equally valid for rooted phylogenetic X-trees *without* a stem edge.
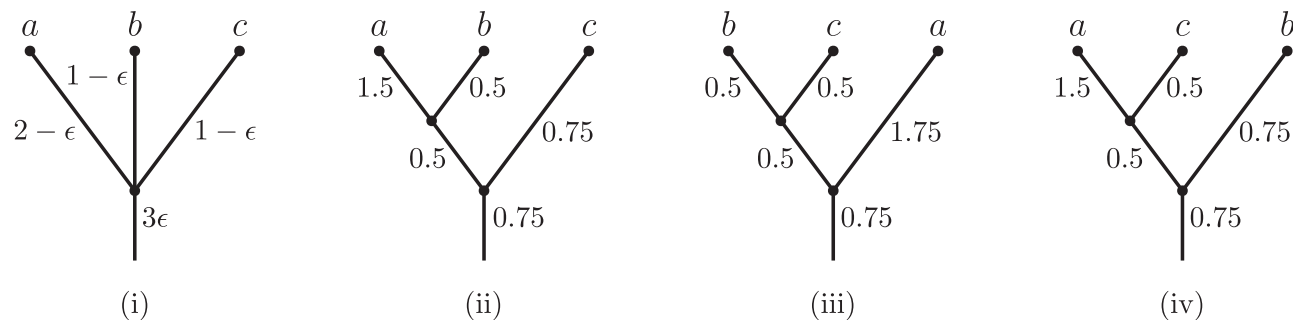
FIGURE 2.   The feature diversity index $\varphi_{\mathbb{F}}$ for the features in Table 1 (each having unit μ value) can be represented as a FP index on any tree topology with positive edge lengths (see text for further details).

6   DISCUSSION

It would be a mistake to interpret Theorem 1 above as stating that feature diversity coincides with phylogenetic diversity (on a given tree with suitably chosen branch lengths) only under evolutionary scenarios in which features arise once in the tree and are never lost. Instead, Theorem 1 states that these two measures coincide precisely when the distribution of features across taxa can be described by such a single-gain-and-no-loss model, even if the underlying reality might be different. For instance, a feature can arise along a stem edge, be lost in one of the two descendant edges, but arise again in its descendants such that the entire crown clade expresses the feature. The feature's true history is obscured but its distribution is still perfectly congruent with the underlying tree and thus meets the conditions of Proposition 1.

While it is possible for all features to appear to be perfect synapomorphies even though there have been undetectable losses and regains, given a rate of evolution high enough and a state space small enough it is unlikely that all homoplasy will remain hidden, and, in such cases, the required conditions (and conclusion) of Theorem 1 will not hold.

A second (related) reason why Theorem 1 allows PD and FD to diverge in applications is that even when $\mathbb{F} = F[T,h]$, the edge lengths must be suitably chosen. For instance, we need a stochastic process in which features arise independently at a constant (and very small) rate r, such that, conditional on a feature arising (at least) once in the tree, as $r \to 0$, the expected number of features that arise on an edge will be proportional to the temporal length of that edge (and each trait will arise exactly once in the tree).

The divergence between PD and FD also has computational consequences. Given a set of taxa X, a phylogenetic X–tree T with branch lengths, and any number k, finding a set of k species whose survival of an extinction event would conserve the largest possible PD is easy (the greedy algorithm provably works; Pardi and Goldman 2005; Steel 2005). On the other hand, if branch lengths do not precisely capture feature evolution, identifying k species whose survival would maximize feature diversity is, in general, NP-hard (by a reduction from the NP-complete problem SET COVER).

6.1   Implications for Empirical Studies

The constraints on when PD and FD will and will not coincide should be considered in light of expected empirical patterns. For instance, time-calibrated ultrametric phylogenetic trees are generally used when comparing PD scores (see, e.g., Pollock et al. 2017; Gumbs et al. 2020) and FP scores (see, e.g., Tonini et al. 2016; Forest et al. 2018; Gumbs et al. 2018; Stein et al. 2018). (Faith 1992) astutely pointed out that convergent features (i.e., homoplasies) "are not predictive of similarities of other features," such that "greater phylogenetic diversity will, on average, imply greater feature diversity as defined by any particular collection of features." In other words, there is an implicit appeal to the stochastic model outlined above being a good approximation for the majority of relevant features. However, the evolution of some important subset of features may not be captured with this model at all (Mazel et al. 2017), or, more prosaically, may simply evolve at such a high rate that the time-calibrated ultrametric tree edge lengths are not predictive of the number and placement of features (e.g., due to saturation). Indeed, measures of PD and measured functional diversity (a subset of feature diversity) need not coincide (Devictor et al. 2010; Brum et al. 2017; see also the discussion in (Winter et al., 2013)). One critical empirical question flowing from this work is whether there are subsets of features that are simultaneously (i) more valuable to conservation than the average feature and (ii) convergent or likely to be convergent, perhaps due to parallel adaptation or high evolutionary rates (Mazel et al. 2018, 2019; Owen et al. 2019). To the extent that there are, the force of Faith's all-important "average" PD = FD statement weakens. However, answering the question is non-trivial because it requires that we know about the mode of evolution of conservation-relevant features in a focal clade. The only attempt to test this we know of is by Forest et al. (2007) for Southern African

plants. Here, the authors catalogued known uses (e.g., for medicine or food) for genera of plants and asked whether sequential maximization of PD on the Southern African plant phylogeny consistently captured more useful genera (and so, more currently conservation-relevant features) than if species were chosen at random. PD did capture more useful genera, supporting Faith's average argument. We echo the call from (Tucker et al. 2019) for more tests of the power of PD to capture conservation-relevant features, for example, linking bespoke phylogenies with known species-specific uses (see, e.g., Ernst et al. 2015, 2016).

We make several comments about the measure of feature diversity $\varphi_{\mathbb{F}}(x)$. The first is that its compact expression does not require any particular model of feature evolution on a tree: different features and different subtrees can be governed by different processes. This might expand its usefulness, because we need not even map the features onto any particular tree (as in Figure 1). Moreover, given a feature assignment $\mathbb{F}$, we can realize $\varphi_{\mathbb{F}}(x)$ as the FP index on *every* phylogenetic X-tree (with suitably chosen edge lengths) even if $\mathbb{F}$ itself cannot be realized on *any* tree (*cf.* Theorem 2 and Fig. 2). Importantly, we still require (Faith 1992)'s "average" argument, namely, that the distribution of measured features mirrors the features of conservation concern more generally. Here again, both theoretical and empirical tests using features of known conservation value are needed.

A second comment is a reminder that the Shapley value is quite specific in what it measures, that is, the expected contribution (here, of features) to all possible future subsets of taxa, where subset sizes are equiprobable (Steel 2016). Although the measure has a long pedigree and is tractable, this is not a natural distribution of subsets of taxa under, for example, a field-of-bullets model of future extinction (see also Faith 2008). More work is needed on what reasonable subset distributions might be, for example, based on phylogenies of regional floras and faunas and, perhaps, reasonable projections of species loss from the landscape.

Finally in this area, the exploration of how well PD captures FD needs to be expanded, both to continuous characters (see, e.g., Tucker et al. 2018), and, for both discrete and continuous characters, to more sophisticated and realistic models of evolution (see, e.g., table 3 in (Tucker et al. 2019)).

In conclusion, our article provides a precise mathematical framework to help address some fundamental questions and possible future approaches concerning the link between feature and phylogenetic diversity, a critical connection for phylogeny-oriented conservation triage.

A.1 APPENDIX: MATHEMATICAL PROOFS

Proof of Theorem 1. The proof of Theorem 1 relies on the following three lemmas.

Lemma 1. *Given a rooted phylogenetic X-tree T, suppose that $\mathbb{F} = F[T,h]$ for some map $h: \mathcal{F} \to E$. Let $\ell_h: E \to \mathbb{R}^{\geq 0}$ be defined by setting*

$$\ell_h(e) := \sum_{f:h(f)=e} \mu(f),$$

for each edge e of T (where $\ell_h(e) := 0$ if $h^{-1}(e) = \emptyset$). Then, for all subsets Y of X we have:

$$\nu_{\mathbb{F}}(Y) = PD_{(T,\ell_h)}(Y).$$

Proof. Suppose that $\mathbb{F} = F[T,h]$. For each $f \in \mathcal{F}$, let X_f be the set of taxa that have feature f. Then,

$$\nu_{\mathbb{F}}(Y) = \sum_{\substack{f \in \mathcal{F}: \\ X_f \cap Y \neq \emptyset}} \mu(f) = \sum_{\substack{e \in E(T): \\ \exists f \in \mathcal{F}: h(f)=e \text{ and } X_f \cap Y \neq \emptyset}} \ell_h(e), \quad \text{(A1)}$$

where the last equality follows from the fact that $\mathbb{F} = F[T,h]$.

On the other hand,

$$PD_{(T,\ell_h)}(Y) = \sum_{\substack{e \in E(T): \\ c_T(e) \cap Y \neq \emptyset}} \ell_h(e), \quad \text{(A2)}$$

where $c_T(e)$ denotes the set of leaves of T that are separated from the root of T by e. Now, as $\mathbb{F} = F[T,h]$, when $e = h(f)$, $c_T(e)$ corresponds to the set X_f. Thus, for $e \in E(T)$, we can conclude that $c_T(e) \cap Y \neq \emptyset$ precisely if

- $\exists f \in \mathcal{F}: h(f) = e$ and $X_f \cap Y \neq \emptyset$, or

- $\nexists f \in \mathcal{F}: h(f) = e$ and e is a pendant edge incident to a leaf $y \in Y$ (in which case $\ell_h(e) = 0$).

Thus, we can re-write Eqn. (A2) as

$$PD_{(T,\ell_h)}(Y) = \sum_{\substack{e \in E(T): \\ \exists f \in \mathcal{F}: h(f)=e \text{ and } X_f \cap Y \neq \emptyset}} \ell_h(e) = \nu_{\mathbb{F}}(Y),$$

where the last equality follows from Eqn. (A1). This completes the proof. □

Lemma 2. *Given a rooted phylogenetic X-tree T, suppose that the identity*

$$\nu_{\mathbb{F}}(Y) = PD_{(T,\ell)}(Y)$$

holds for all subsets $Y \subseteq X$, where $\ell: E \to \mathbb{R}^{\geq 0}$ is such that the interior edges of T are assigned strictly positive lengths and pendant edges (including the stem edge) are assigned non-negative lengths. Then, there exists a map $h: \mathcal{F} \to E$ such that $\mathbb{F} = F[T,h]$.

Proof. We will prove this statement by contradiction. Thus, assume that $\nu_{\mathbb{F}}(Y) = PD_{(T,\ell)}(Y)$ for all $Y \subseteq X$ but there is no map $h: \mathcal{F} \to E$ such that $\mathbb{F} = F[T,h]$. We now distinguish two cases: (i) $\mathbb{F}$ cannot be explained by T, but by some other tree T', that is, $\mathbb{F} = F[T',h']$, or (ii) the collection of sets $\mathcal{C}_{\mathcal{F}} = \{X_f : f \in \mathcal{F}\}$ does not form a hierarchy and cannot be explained by any tree (cf. Proposition 1, Part (ii)).

(i) First, suppose that $\mathbb{F} \neq F[T,h]$ but $\mathbb{F} = F[T',h']$, and $\nu_{\mathbb{F}}(Y) = PD_{(T,\ell)}(Y)$ for all $Y \subseteq X$. Now, as $T \neq T'$, there must be some $i,j,k \in X$ such that restricting T and T' to $\{i,j,k\}$ yields distinct trees. More precisely, there exist $i,j,k \in X$ such that

    (a) $T|_{\{i,j,k\}}$ is either the caterpillar tree on three leaves with cherry $[i,j]$ or the rooted star tree on $\{i,j,k\}$,

    (b) $T'|_{\{i,j,k\}}$ is either the caterpillar tree on three leaves with cherry $[i,k]$ or the rooted star tree on $\{i,j,k\}$,

    (c) $T|_{\{i,j,k\}} \neq T'|_{\{i,j,k\}}$ (in particular, $T|_{\{i,j,k\}}$ and $T'|_{\{i,j,k\}}$ are not both star trees).

Let $\Delta_{\mathbb{F}}(x,x') := \nu_{\mathbb{F}}(\{x\}) + \nu_{\mathbb{F}}(\{x'\}) - \nu_{\mathbb{F}}(\{x,x'\})$, for each distinct pair $x, x' \in X$. Then as $\nu_{\mathbb{F}}(Y) = PD_{(T,\ell)}$ for all $Y \subseteq X$, we have from (a) that:

$$\Delta_{\mathbb{F}}(i,j) \begin{cases} = \Delta_{\mathbb{F}}(i,k), & \text{if } T|_{\{i,j,k\}} \text{ is a star tree;} \\ > \Delta_{\mathbb{F}}(i,k), & \text{otherwise.} \end{cases}$$
(A3)

On the other hand, as $\mathbb{F} = F[T',h']$, we have by Lemma 1, that $\nu_{\mathbb{F}}(Y) = PD_{(T',\ell_{h'})}(Y)$ for all $Y \subseteq X$ (where $\ell_{h'}(e) = \sum_{f:h'(f)=e} \mu(f)$; in particular, $\ell_{h'}(e) > 0$ for each interior edge e of T'). This implies that:

$$\Delta_{\mathbb{F}}(i,k) \begin{cases} = \Delta_{\mathbb{F}}(i,j), & \text{if } T'|_{\{i,j,k\}} \text{ is a star tree;} \\ > \Delta_{\mathbb{F}}(i,j), & \text{otherwise.} \end{cases}$$
(A4)

Comparing Eqns (A3) and (A4), and using the fact that $T|_{\{i,j,k\}}$ and $T'|_{\{i,j,k\}}$ cannot both be star trees, this yields a contradiction. As (i,j,k) was an arbitrary triple of leaves for which $T|_{\{i,j,k\}} \neq T'|_{\{i,j,k\}}$, this contradiction implies that the initial assumption was wrong. In particular, $\mathbb{F} = F[T,h]$.

(ii) Now, assume that $\nu_{\mathbb{F}}(Y) = PD_{(T,\ell)}(Y)$ for all $Y \subseteq X$, but $\mathcal{C}_{\mathcal{F}} = \{X_f : f \in \mathcal{F}\}$ does not form a hierarchy. This implies that there exists $f_1, f_2 \in \mathcal{F}$ such that

    (a) There exists a taxon $x_1 \in X$ such that $x_1 \in X_{f_1} \cap X_{f_2}$.

    (b) There exists a taxon $x_2 \in X$ such that $x_2 \in X_{f_1} \setminus X_{f_2}$.

    (c) There exists a taxon $x_3 \in X$ such that $x_3 \in X_{f_2} \setminus X_{f_1}$.

We now partition the feature set $\mathcal{F}$ into eight pairwise disjoint subsets $A, \dots, G$, where

$$A := \{f \in \mathcal{F} : f \in F_{x_1} \setminus (F_{x_2} \cup F_{x_3})\}$$
$$B := \{f \in \mathcal{F} : f \in F_{x_2} \setminus (F_{x_1} \cup F_{x_3})\}$$
$$C := \{f \in \mathcal{F} : f \in F_{x_3} \setminus (F_{x_1} \cup F_{x_2})\}$$
$$D := \{f \in \mathcal{F} : f \in (F_{x_1} \cap F_{x_2}) \setminus F_{x_3}\}$$
$$E := \{f \in \mathcal{F} : f \in (F_{x_1} \cap F_{x_3}) \setminus F_{x_2}\}$$
$$F := \{f \in \mathcal{F} : f \in (F_{x_2} \cap F_{x_3}) \setminus F_{x_1}\}$$
$$G := \{f \in \mathcal{F} : \bigcap_{i=1}^{3} F_{x_i}\}$$
$$H := \{f \in \mathcal{F} : f \notin \bigcup_{i=1}^{3} F_{x_i}\}.$$

Note that $D \neq \emptyset$ (because by (a)–(c), $f_1 \in (F_{x_1} \cap F_{x_2}) \setminus F_{x_3}$). Analogously, $E \neq \emptyset$ (because $f_2 \in (F_{x_1} \cap F_{x_3}) \setminus F_{x_2}$).

Given a set of features S, let $\mu(S) := \sum_{f \in S} \mu(f)$ denote the sum of scores of features present in S. As $\mu(f) > 0$ for all $f \in \mathcal{F}$, by the preceding argument, in particular $\mu(D), \mu(E) > 0$.

We now compute $\nu_{\mathbb{F}}(Y)$ for all $Y \subseteq \{x_1, x_2, x_3\}$ with $|Y| \geq 1$, and compare it to $PD_{(T,\ell)}(Y)$. Recall that $PD_{(T,\ell)}(Y)$ for $Y \subseteq X$ is computed by considering the sum of edge lengths in the minimum subtree of T connecting the taxa in Y and ρ'. Without loss of generality, we can assume that the subtree induced by $\{x_1, x_2, x_2\}$ has the structure depicted in Fig. A.1 (otherwise, we exchange leaf labels).

Now, by assumption $\nu_{\mathbb{F}}(Y) = PD_{(T,\ell)}(Y)$ for all $Y \subseteq X$. For $Y \subseteq \{x_1, x_2, x_3\}$ with $|Y| \geq 1$, this gives rise to a system of 7 linear equations (where $\ell(p)$ denotes the length of path p):

$$\nu_{\mathbb{F}}(\{x_1\}) = PD_{(T,\ell)}(\{x_1\})$$
$$\Leftrightarrow \mu(A) + \mu(D) + \mu(E) + \mu(G) = \ell(p_1) + \ell(p_5) + \ell(e_\rho)$$
$$\nu_{\mathbb{F}}(\{x_2\}) = PD_{(T,\ell)}(\{x_2\})$$
$$\Leftrightarrow \mu(B) + \mu(D) + \mu(F) + \mu(G) = \ell(p_2) + \ell(p_4) + \ell(p_5) + \ell(e_\rho)$$
$$\nu_{\mathbb{F}}(\{x_3\}) = PD_{(T,\ell)}(\{x_3\})$$
$$\Leftrightarrow \mu(C) + \mu(E) + \mu(F) + \mu(G) = \ell(p_3) + \ell(p_4) + \ell(p_5) + \ell(e_\rho)$$
$$\nu_{\mathbb{F}}(\{x_1, x_2\}) = PD_{(T,\ell)}(\{x_1, x_2\})$$
$$\Leftrightarrow \mu(A) + \mu(B) + \mu(D) + \mu(E) + \mu(F) + \mu(G) = \ell(p_1) + \ell(p_2) + \ell(p_4) + \ell(p_5) + \ell(e_\rho)$$
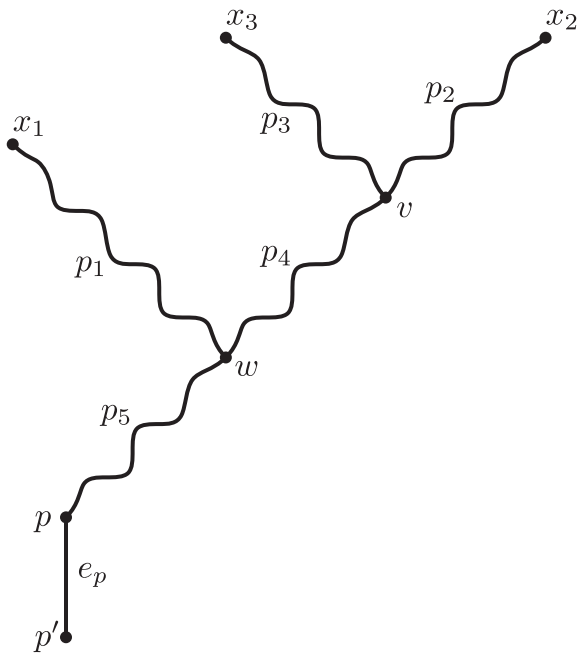
FIGURE A.1. Subtree induced by taxa x_1, x_2, and x_3 in the proof of Lemma 2. v denotes the most recent common ancestor of x_2 and x_3. Analogously, w denotes the most recent common ancestor of x_1, x_2, and x_3. Furthermore, p_1 denotes the unique path from w to x_1, p_2 denotes the unique path from v to x_2 and so forth.

$$\nu_{\mathbb{F}}(\{x_1, x_3\}) = PD_{(T,\ell)}(\{x_1, x_3\})$$

$$\Leftrightarrow \mu(A) + \mu(C) + \mu(D) + \mu(E) + \mu(F) + \mu(G)$$

$$= \ell(p_1) + \ell(p_3) + \ell(p_4) + \ell(p_5) + \ell(e_\rho)$$

$$\nu_{\mathbb{F}}(\{x_2, x_3\}) = PD_{(T,\ell)}(\{x_2, x_3\})$$

$$\Leftrightarrow \mu(B) + \mu(C) + \mu(D) + \mu(E) + \mu(F) + \mu(G)$$

$$= \ell(p_2) + \ell(p_3) + \ell(p_4) + \ell(p_5) + \ell(e_\rho)$$

$$\nu_{\mathbb{F}}(\{x_1, x_2, x_3\}) = PD_{(T,\ell)}(\{x_1, x_2, x_3\})$$

$$\Leftrightarrow \mu(A) + \mu(B) + \mu(C) + \mu(D) + \mu(E) + \mu(F) + \mu(G)$$

$$= \ell(p_1) + \ell(p_2) + \ell(p_3) + \ell(p_4) + \ell(p_5) + \ell(e_\rho).$$

Solving this system of linear equations for $\mu(A), \ldots, \mu(G)$ yields $\mu(A) = \ell(p_1)$, $\mu(B) = \ell(p_2)$, $\mu(C) = \ell(p_3)$, $\mu(D) = \mu(E) = 0$, $\mu(F) = \ell(p_4)$, and $\mu(G) = \ell(p_5) + \ell(e_\rho)$.

However, as our assumption implies that $\mu(D), \mu(E) > 0$, this is a contradiction. Thus, the initial assumption was false. In particular, $\{X_f : f \in \mathcal{F}\}$ forms a hierarchy. Thus, by Proposition 1, Part (ii), there exist T' and h' such that $\mathbb{F} = F[T', h']$. Now, by case (i) of this proof, this implies $\mathbb{F} = F[T, h]$. This completes the proof.

□

Lemma 3. *Let T be a rooted phylogenetic X-tree (with additional stem edge). Then, the edge lengths of T are uniquely*
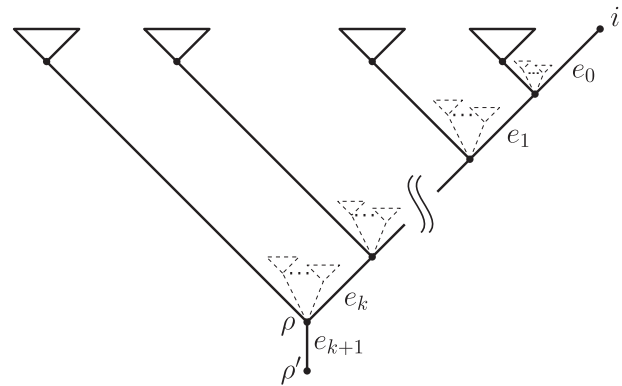

FIGURE A.2. Representing a phylogenetic X-tree T relative to a reference leaf i. Note that T is not assumed to be binary.

determined by the induced PD scores of all subsets $Y \subseteq X$ with $|Y| \leq 2$.

Proof. Let T be a rooted phylogenetic X-tree (with additional stem edge), and assume that $PD_T(Y)$ is given for all $Y \subseteq X$ with $|Y| \leq 2$. We now show that we can uniquely infer the edge lengths of T from these scores. Let $i \in X$ be a leaf of T. Then, there is a unique path $e_{k+1}, e_k, \ldots, e_1, e_0$ from ρ' to i in T (see Fig. A.2), and we can infer the lengths of these edges in a 'top-down' approach (i.e., starting with edge e_{k+1} and moving down the tree towards edge e_0).

For $\ell(e_{k+1})$, let j be a leaf that is not a descendant of edge e_k (in other words, j is not in the same maximal pending subtree as i). Then, clearly,

$$PD_T(\{i, j\}) = PD_T(\{i\}) + PD_T(\{j\}) - \ell(e_{k+1}),$$

(because $\ell(e_{k+1})$ contributes twice to the sum $PD_T(\{i\}) + PD_T(\{j\})$, but only once to $PD_T(\{i, j\})$). In other words, $\ell(e_{k+1}) = PD_T(\{i\}) + PD_T(\{j\}) - PD_T(\{i, j\})$.

Now, let $e_i = (u, v)$ be an interior edge in the path from ρ' to i, for which the lengths of its preceding edges are already determined, that is, $\ell(e_{k+1}), \ldots, \ell(e_{i+1})$ are known. Moreover, let j be a leaf that is a descendant from e_i, but not from e_{i-1}.

Then, with a similar argument as in the previous case, we have

$$PD_T(\{i, j\}) = PD_T(\{i\}) + PD_T(\{j\}) - \ell(P(T; \rho', v)),$$

where $\ell(P(T; \rho', v))$ denotes the length of the unique path from ρ' to v in T (which contributes twice to the sum $PD_T(\{i\}) + PD_T(\{j\})$, but only once to $PD_T(\{i, j\})$). In other words, $\ell(P(T; \rho', v)) = PD_T(\{i\}) + PD_T(\{j\}) - PD_T(\{i, j\})$. On the other hand, $\ell(P(T; \rho', v)) = \ell(e_{k+1}) + \ell(e_k) + \cdots + \ell(e_{i+1}) + \ell(e_i)$, and as $\ell(e_{k+1}), \ldots, \ell(e_{i+1})$ are known, we can uniquely infer $\ell(e_i)$.

Finally, after inferring the lengths of the edges $e_{k+1}, e_k, \ldots, e_1$ as described above, we can also uniquely infer the length of the pendant edge e_0 incident to i as $\ell(e_0) = PD_T(\{i\}) - \sum_{j=1}^{k+1} \ell(e_j)$.

In summary, we can uniquely infer all edge lengths of edges in the path from ρ' to i from the PD scores of

subsets of X of size at most 2. As i was an arbitrary leaf of T, this completes the proof. $\qquad\square$

We now show that Theorem 1 follows from these lemmas. Part (i) of Theorem 1, namely

$$\mathbb{F} = F[T, h] \Leftrightarrow \nu_{\mathbb{F}}(Y) = PD_{(T, \ell)}(Y) \forall Y \subseteq X$$

follows from Lemmas 1 and 2 (the "only if" implication is implied by Lemma 1 and the "if" implication is implied by Lemma 2).

For Part (ii), the uniqueness of ℓ (i.e., $\ell = \ell_h$), follows by combining Lemmas 1 and 3. More precisely, Lemma 1 states that assigning edge lengths according to ℓ_h induces the equality of $\nu_{\mathbb{F}}(Y)$ and $PD_{(T, \ell_h)}(Y)$ for all $Y \subseteq X$, whereas, by Lemma 3, the edge lengths of a given tree T are uniquely determined by the induced PD scores of all $Y \subseteq X$ (indeed, even those with size at most 2 suffice). Moreover, the uniqueness of h is implied by Proposition 1, Part (i). This completes the proof. $\qquad\square$

Proof of Theorem 2. In order to prove Theorem 2, we prove a slightly stronger statement. To this end, let T be a rooted phylogenetic X-tree with $X = \{x_1, \dots, x_n\}$ and let $(y_1, \dots, y_n)$ be an ordered n-tuple of strictly positive real numbers. We claim that there exists an edge length assignment $\ell : E \to \mathbb{R}^{>0}$ that assigns strictly positive lengths to all edges of T such that $y_i = FP_{(T, \ell)}(x_i)$ for all $x_i \in X$. By choosing $y_i = \varphi_{\mathbb{F}}(x_i)$ for each $x_i \in X$, the statement of the theorem immediately follows (due to the assumed condition $F_{x_i} \neq \emptyset$ for all $x_i \in X$, along with the fact that μ takes strictly positive values, we have that $\varphi_{\mathbb{F}}(x_i)$ is a strictly positive real number for each $x_i \in X$ and thus meets the conditions above.)

First, note that we can always achieve $y_i = FP_{(T, \ell')}(x_i)$ for all $x_i \in X$ when we consider an edge length assignment ℓ' that allows edges to be assigned length zero because, in this case, if e_{x_i} denotes the pendant edge incident to x_i, we can set $\ell'(e_{x_i}) = y_i$ for each $x_i \in X$, and $\ell'(e) = 0$ for all interior edges and the stem edge, which clearly results in $y_i = FP_{(T, \ell')}(x_i)$ for all $x_i \in X$.

We now show that we can obtain an edge length assignment ℓ assigning strictly positive lengths to all edges of T from ℓ' by redistributing lengths in a "bottom-up" approach (i.e., moving from pendant edges towards the stem edge).

First, for each pendant edge e_{x_i}, set $\ell(e_{x_i}) = \ell'(e_{x_i}) = y_i$, which is strictly positive as $y_i > 0$ by assumption. Now, let e be an edge of T such that all edges descending from e already have strictly positive lengths, whereas all edges above e (if they exist) still have length zero. Let $e_1, \dots, e_k$ denote the descending edges incident to e, and let $t_1, \dots, t_k$ denote the subtrees pending from e (where tree t_i has stem edge e_i for $i = 1, \dots, k$). Moreover, for $i = 1, \dots, k$, let $\delta_{e_i} := \frac{\ell(e_i)}{n(e_i)}$ denote the ratio between the length of e_i and the number of leaves descending from it. Without loss of generality, we may assume that edge e_1 minimizes this ratio (else we exchange edge labels). Furthermore, let $0 < c < 1$. We now re-assign edge lengths to $e_1, \dots, e_k$

and e as follows (where $\ell_{\text{old}}(e_i)$ refers to the edge length e_i is currently assigned):

1. $\ell_{\text{new}}(e) := (1 - c) \cdot \ell_{\text{old}}(e_1) \cdot \frac{n(e)}{n(e_1)}$

2. $\ell_{\text{new}}(e_1) := c \cdot \ell_{\text{old}}(e_1)$,

3. $\ell_{\text{new}}(e_i) := \ell_{\text{old}}(e_i) - \frac{n(e_i)}{n(e)} \cdot \ell_{\text{new}}(e)$ for $i = 2, \dots, k$.

Now, in order to show that this is a valid re-distribution of edge lengths, we need to show that

(i) $\ell_{new}(e) > 0$ and $\ell_{\text{new}}(e_i) > 0$ for $i = 1, \dots, k$.

(ii) $FP_{(T, \ell_{\text{old}})}(x) = FP_{(T, \ell_{\text{new}})}(x)$ for all $x \in X$, that is, the FP indices are not affected by the re-assignment of edge lengths.

First, consider (i). As $\ell_{\text{old}}(e_1) > 0$ by assumption, and $0 < c < 1$, we clearly have $\ell_{\text{new}}(e) > 0$, and $\ell_{\text{new}}(e_1) > 0$. Now, consider e_i for $i \in \{2, \dots, k\}$. Here, we have

$$\begin{aligned} \ell_{\text{new}}(e_i) &= \ell_{\text{old}}(e_i) - \frac{n(e_i)}{n(e)} \cdot \ell_{\text{new}}(e) \\ &= \ell_{\text{old}}(e_i) - \frac{n(e_i)}{n(e)} \cdot (1 - c) \cdot \ell_{\text{old}}(e_1) \cdot \frac{n(e)}{n(e_1)} \\ &= \ell_{\text{old}}(e_i) - \frac{\ell_{\text{old}}(e_1)}{n(e_1)} \cdot n(e_i) \cdot (1 - c) \\ &\geq \ell_{\text{old}}(e_i) - \frac{\ell_{\text{old}}(e_i)}{n(e_i)} \cdot n(e_i) \cdot (1 - c) \\ &\quad (\text{because } e_1 \text{ minimizes } \delta_{e_i} = \ell_{\text{old}}(e_i)/n(e_i)) \\ &= \ell_{\text{old}}(e_i) - \ell_{\text{old}}(e_i) \cdot (1 - c) = c \cdot \ell_{\text{old}}(e_i) > 0, \end{aligned}$$

where the last inequality again follows from the fact that (by assumption) $\ell_{\text{old}}(e_i) > 0$ and $0 < c < 1$. This completes the proof of (i).

For (ii) note that the FP indices of taxa not descending from e are not affected by the re-assignment of edge lengths, so it suffices to consider all $x \in c_T(e)$. In the following, let $t_i \setminus e_i$ be the rooted phylogenetic tree obtained from t_i by deleting its stem edge. Then, we clearly have for all $x \in c_T(e)$:

$$FP_{(T, \ell_{\text{old}})}(x) = FP_{(t \setminus e_i, \ell_{\text{old}})}(x) + \frac{\ell_{\text{old}}(e_i)}{n(e_i)},$$

(because by assumption all edges above e_i have length zero before the re-assignment of edge lengths according to steps 1–3). On the other hand, we have for all $x \in c_T(e)$:

$$FP_{(T, \ell_{\text{new}})}(x) = FP_{(t \setminus e_i, \ell_{\text{new}})}(x) + \frac{\ell_{\text{new}}(e_i)}{n(e_i)} + \frac{\ell_{\text{new}}(e)}{n(e)}.$$

Note that $FP_{(t \setminus e_i, \ell_{\text{old}})}(x) = FP_{(t \setminus e_i, \ell_{\text{new}})}(x)$ for all $x \in c_T(e)$ (because the lengths of edges in $t_i \setminus e_i$ are not changed). We now show that $FP_{(T, \ell_{\text{old}})}(x) = FP_{(T, \ell_{\text{new}})}(x)$ for all $x \in X$.

First, let $x \in t_1$. Then, we have

$$FP_{(T,\ell_{\text{new}})}(x) = FP_{(t\backslash e_1, \ell_{\text{new}})}(x) + \frac{\ell_{\text{new}}(e_1)}{n(e_1)} + \frac{\ell_{\text{new}}(e)}{n(e)}$$

$$= FP_{(t\backslash e_1, \ell_{\text{old}})}(x) + \frac{c \cdot \ell_{\text{old}}(e_1)}{n(e_1)} + \frac{(1-c) \cdot \ell_{\text{old}}(e_1) \cdot \frac{n(e)}{n(e_1)}}{n(e)}$$

$$= FP_{(t\backslash e_1, \ell_{\text{old}})}(x) + c \cdot \frac{\ell_{\text{old}}(e_1)}{n(e_1)} + (1-c) \cdot \frac{\ell_{\text{old}}(e_1)}{n(e_1)}$$

$$= FP_{(t\backslash e_1, \ell_{\text{old}})}(x) + \frac{\ell_{\text{old}}(e_1)}{n(e_1)} = FP_{(T,\ell_{\text{old}})}(x).$$

Now, let $x \in t_i$ for $i \in \{2, \dots, k\}$. Then, we have

$$FP_{(T,\ell_{\text{new}})}(x) = FP_{(t\backslash e_i, \ell_{\text{new}})}(x) + \frac{\ell_{\text{new}}(e_i)}{n(e_i)} + \frac{\ell_{\text{new}}(e)}{n(e)}$$

$$= FP_{(t\backslash e_i, \ell_{\text{old}})}(x) + \frac{\ell_{\text{old}}(e_i) - \frac{n(e_i)}{n(e)} \cdot \ell_{\text{new}}(e)}{n(e_i)} + \frac{\ell_{\text{new}}(e)}{n(e)}$$

$$= FP_{(t\backslash e_i, \ell_{\text{old}})}(x) + \frac{\ell_{\text{old}}(e_i)}{n(e_i)} - \frac{\ell_{\text{new}}(e)}{n(e)} + \frac{\ell_{\text{new}}(e)}{n(e)}$$

$$= FP_{(T,\ell_{\text{old}})}(x).$$

In summary, re-assigning edge lengths according to the conditions 1–3 (listed above) is valid (because conditions (i) and (ii) hold). Thus, for each edge e whose length was changed, we now simply set $\ell(e) = \ell_{\text{new}}(e)$ and repeat the procedure. In this way, we can construct an edge length assignment ℓ that assigns strictly positive lengths to *all* edges of T (including pendant edges and the stem edge), such that $y_i = FP_{(T,\ell)}(x_i)$ for all $x_i \in X$. This completes the proof. $\square$

Proof of Proposition 2. Notice that both $\varphi_{\mathbb{F}}(x)$ and $\nu_{\mathbb{F}}(Y)$ (with $Y \subseteq X$) are linear functions in $\mu(f)$. More precisely,

$$\varphi_{\mathbb{F}}(x) = \sum_{f \in \mathcal{F}} \gamma(x,f) \cdot \mu(f),$$

where $\quad \gamma(x,f) = \begin{cases} \frac{1}{n(f)}, & \text{if } f \in F_x; \\ 0, & \text{otherwise.} \end{cases}$

Analogously,

$$\nu_{\mathbb{F}}(Y) = \sum_{f \in \mathcal{F}} \gamma'(Y,f) \cdot \mu(f),$$

where $\quad \gamma'(Y,f) = \begin{cases} 1, & \text{if } f \in \cup_{x \in Y} F_x; \\ 0, & \text{otherwise.} \end{cases}$

Thus, by linearity (see also Lemma 6.14 in Steel (2016)), it suffices to show the statements for the case that one element of $\mathcal{F}$, say f_i, has score $\mu(f_i) = 1$, whereas $\mu(f_j) = 0$ for all $f_j \in \mathcal{F} \backslash \{f_i\}$. Note that μ was earlier defined to be strictly positive, but we are relaxing this here as it allows us to prove a slightly stronger claim (i.e. the result holds even when μ takes zero values), using a more direct argument.

For the first part of the proof, recall that given the pair $(X, \nu_{\mathbb{F}})$, the Shapley value of $x \in X$ is given by

$$SV_{\nu_{\mathbb{F}}}(x) = \frac{1}{n!} \sum_{S \subseteq X : x \in S} (|S|-1)!(n-|S|)! \Delta(S,x),$$

where $\Delta(S,x) = \nu_{\mathbb{F}}(S) - \nu_{\mathbb{F}}(S \backslash \{x\})$. We now show that $SV_{\nu_{\mathbb{F}}}(x) = \varphi_{\mathbb{F}}(x)$ (where we assume that $\mu(f_i) = 1$ and $\mu(f_j) = 0$ for all $f_j \in \mathcal{F} \backslash \{f_i\}$).

We can distinguish two cases:

- If $f_i \notin F_x$, then $\Delta(S,x) = 0$ for all S, and thus, $SV_{\nu_{\mathbb{F}}}(x) = 0$. On the other hand, we clearly also have $\varphi_{\mathbb{F}}(x) = 0$.

- If $f_i \in F_x$, then $\Delta(S,x) = 1$ if (i) $x \in S$ and (ii) there is no $y \in S$ with $f_i \in F_y$; otherwise $\Delta(S,x) = 0$. Let $C \subseteq X$ be the set of taxa that have feature f_i, i.e., $C = \{y \in X : f_i \in F_y\}$, and so $n(f_i) = |C|$. Then, $SV_{\nu_{\mathbb{F}}}(x)$ can be written as

$$\frac{1}{n!} \sum_{\substack{S : x \in S \\ S \backslash \{x\} \subseteq X \backslash C}} (|S|-1)!(n-|S|)! \cdot 1$$

$$= \frac{1}{n!} \sum_{k=1}^{n-n(f_i)+1} \binom{n-n(f_i)}{k-1}(k-1)!(n-k)! = \frac{1}{n(f_i)},$$

where the last equality follows from the fact that $\frac{1}{n!}\sum_{j=0}^{n-r}\binom{n-r}{j}j!(n-1-j)! = \frac{1}{r}$ for $1 \le r \le n$ (Lemma 6.15 in Steel (2016)) (here: $j = k-1$ and $r = n(f_i)$). On the other hand, $\varphi_{\mathbb{F}}(x) = \frac{\mu(f_i)}{n(f_i)} = \frac{1}{n(f_i)}$, which completes the proof.

The second part of Proposition 2, follows directly from properties of the Shapley value, namely from the efficiency axiom; however we give a direct proof. Again, it suffices to consider the case where $\mu(f_i) = 1$ and $\mu(f_j) = 0$, for all $j \ne i$, in which case we obtain the required equality:

$$\sum_{x \in X} \varphi_{\mathbb{F}}(x) = \sum_{\substack{x \in X: \\ f_i \in F_x}} \varphi_{\mathbb{F}}(x) + \sum_{\substack{x \in X: \\ f_i \notin F_x}} \varphi_{\mathbb{F}}(x) = \sum_{\substack{x \in X: \\ f_i \in F_x}} \frac{\mu(f_i)}{n(f_i)} + 0$$

$$= \mu(f_i) = 1 = \nu_{\mathbb{F}}(X).$$

$\square$

Proof of Proposition 3. Let $\mathbb{F} = F[T,h]$ and let $x \in X$. As noted above, we have:

$$FP_{(T,\ell_h)}(x) = \sum_{e \in P(T; \rho', x)} \frac{1}{n(e)} \cdot \ell_h(e).$$

Importantly, we can also write $\varphi_{\mathbb{F}}(x)$ as follows:

$$\varphi_{\mathbb{F}}(x) = \sum_{f \in F_x} \frac{\mu(f)}{n(f)} = \sum_{\substack{e \in E(T): \\ \exists f \in F_x \text{ with } h(f)=e}} \frac{\ell_h(e)}{n(f)}.$$

Now, since $\mathbb{F} = F[T,h]$, all the edges on which features present in F_x arise must lie on the unique path from ρ' to x. Moreover, a feature f' not contained in F_x, cannot have arisen on this path. More precisely, if a feature f arises on edge e, then a taxon $x \in X$ has this feature if and only if it is a descendant of e. In particular, $n(f) = n(e)$. In summary, this implies that $\varphi_\mathbb{F}(x) = FP_{(T, \ell_h)}(x)$. $\qquad \square$

Proof of Inequality (2.1). Let W be the function that assigns to each subset Y of X the union of the sets of features present amongst the taxa in Y. Thus, $W(Y) := \bigcup_{x \in Y} F_x$. From the proof of Proposition 2 we have:

$$\nu_\mathbb{F}(Y) = \sum_{f \in \mathcal{F}} \gamma'(Y, f) \cdot \mu(f),$$

where $\quad \gamma'(Y, f) = \begin{cases} 1, & \text{if } f \in W(Y); \\ 0, & \text{otherwise.} \end{cases}$

Now,

$$W(Y \cup Y') = W(Y) \cup W(Y'),$$

and

$$W(Y \cap Y') \subseteq W(Y) \cap W(Y'),$$

(and the containment can be strict). It follows that for all $f \in \mathcal{F}$ and all $Y, Y' \subseteq X$:

$$\gamma'(Y \cup Y', f) + \gamma'(Y \cap Y', f) - \gamma'(Y, f) - \gamma'(Y', f) \leq 0.$$

Since $\nu_\mathbb{F}(Y \cup Y) + \nu_\mathbb{F}(Y \cap Y') - \nu_\mathbb{F}(Y) - \nu_\mathbb{F}(Y')$ is a positive weighted sum of the corresponding γ' quantities above, Inequality (2.1) now follows. $\qquad \square$

REFERENCES

Bordewich, M., A. Rodrigo, and C. Semple. 2008. Selecting taxa to save or sequence: Desirable criteria and a greedy solution. Syst. Biol. 57:825–834.

Brum, F. T., C. H. Graham, G. C. Costa, S. B. Hedges, C. Penone, V. C. Radeloff, C. Rondinini, R. Loyola, and A. D. Davidson. 2017. Global priorities for conservation across multiple dimensions of mammalian diversity. Proc. Natl. Acad. Sci. USA 114: 7641–7646.

Devictor, V., D. Mouillot, C. Meynard, F. Jiguet, W. Thuiller, and N. Mouquet. 2010. Spatial mismatch and congruence between taxonomic, phylogenetic and functional diversity: the need for integrative conservation strategies in a changing world. Ecol. Lett. 13:1030–1040.

Ernst, M., O. M. Grace, C. H. Saslis-Lagoudakis, N. Nilsson, H. T. Simonsen, and N. Rønsted. 2015. Global medicinal uses of *Euphorbia* L. (Euphorbiaceae). J. Ethnopharmacol. 176:90–101.

Ernst, M., C. H. Saslis-Lagoudakis, O. M. Grace, N. Nilsson, H. T. Simonsen, J. W. Horn, and N. Rønsted. 2016. Evolutionary prediction of medicinal properties in the genus *Euphorbia* L. Sci. Rep. 6.

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

Faith, D. P. 2008. Threatened species and the potential loss of phylogenetic diversity: conservation scenarios based on estimated extinction probabilities and phylogenetic risk analysis. Conserv. Biol. 22:1461–1470.

Forest, F., R. Grenyer, M. Rouget, and et. al. 2007. Preserving the evolutionary potential of floras in biodiversity hotspots. Nature 445:757–760.

Forest, F., J. Moat, E. Baloch, and *et al*. 2018. Gymnosperms on the EDGE. Sci. Rep. 8:6053.

Fuchs, M. and E. Y. Jin. 2015. Equality of Shapley value and fair proportion index in phylogenetic trees. J. Math. Biol. 71:1133–1147.

Gumbs, R., C. L. Gray, M. Böhm, M. Hoffmann, R. Grenyer, W. Jetz, S. Meiri, U. Roll, N. R. Owen, and J. Rosindell. 2020. Global priorities for conservation of reptilian phylogenetic diversity in the face of human impacts. Nat. Commun. 11:2616.

Gumbs, R., C. L. Gray, O. R. Wearn, and N. R. Owen. 2018. Tetrapods on the EDGE: Overcoming data limitations to identify phylogenetic conservation priorities. PLoS One 13:e0194680.

Isaac, N. J. B., S. T. Turvey, B. Collen, C. Waterman, and J. E. M. Baillie. 2007. Mammals on the edge: conservation priorities based on threat and phylogeny. PLoS One 2:e296.

Mazel, F., A. O. Mooers, G. V. D. Riva, and M. W. Pennell. 2017. Conserving phylogenetic diversity can be a poor strategy for conserving functional diversity. Syst. Biol. 66:1019–1027.

Mazel, F., M. W. Pennell, M. W. Cadotte, S. Diaz, G. V. D. Riva, R. Grenyer, F. Leprieur, A. O. Mooers, D. Mouillot, C. M. Tucker, and W. D. Pearse. 2018. Prioritizing phylogenetic diversity captures functional diversity unreliably. Nat. Commun. 9:2888.

Mazel, F., M. W. Pennell, M. W. Cadotte, S. Diaz, G. V. D. Riva, R. Grenyer, F. Leprieur, A. O. Mooers, D. Mouillot, C. M. Tucker, and W. D. Pearse. 2019. Reply to: "Global conservation of phylogenetic diversity captures more than just functional diversity". Nat. Commun. 10:858.

Minh, B. Q., S. Klaere, and A. von Haeseler. 2009. Taxon selection under split diversity. Syst. Biol. 58:586–594.

Owen, N. R., R. Gumbs, C. L. Gray, and D. P. Faith. 2019. Global conservation of phylogenetic diversity captures more than just functional diversity. Nat. Commun. 10:859.

Pardi, F. and N. Goldman. 2005. Species choice for comparative genomics: being greedy works. PLoS Genetics 1:e71.

Pollock, L. J., W. Thuiller, and W. Jetz. 2017. Large conservation gains possible for global biodiversity facets. Nature 546:141–144.

Redding, D., K. Hartmann, A. Mimoto, D. Bokal, M. Devos, and A. O. Mooers. 2007. Evolutionarily distinct species capture more phylogenetic diversity than expected. J. Theor. Biol. 25: 606–615.

Redding, D. W. 2003. Incorporating genetic distinctness and reserve occupancy into a conservation priorisation approach. Master's thesis, University of East Anglia, Norwich, UK.

Shapley, L. S. 1953. A value for *n*-person games. Pages 307–317 *in* Contributions to the Theory of Games (AM-28), Volume II. Princeton University Press, Princeton, New Jersey, United States.

Steel, M. 2005. Selecting taxa to save or sequence: desirable criteria and a greedy solution. Syst. Biol. 54:527–529.

Steel, M. 2016. Phylogeny: discrete and random processes in evolution. Society for Industrial and Applied Mathematics, Philadelphia PA.

Stein, R. W., C. G. Mull, T. S. Kuhn, and *et al*. 2018. Global priorities for conserving the evolutionary history of sharks, rays and chimaeras. Nat. Ecol. Evol. 2:288–â£"298.

Tonini, J. F. R., K. H. Beard, R. B. Ferreira, W. Jetz, and R. A. Pyron. 2016. Fully-sampled phylogenies of squamates reveal evolutionary patterns in threat status. Biol. Conserv. 204:23–31.

Tucker, C., T. Davies, M. Cadotte, and W. Pearse. 2018. On the relationship between phylogenetic diversity and trait diversity. Ecology 99:1473–1479.

Tucker, C. M., T. Aze, M. W. Cadotte, J. L. Cantalapiedra, C. Chisholm, S. Díaz, R. Grenyer, D. Huang, F. Mazel, W. D. Pearse, M. W. Pennell, M. Winter, and A. O. Mooers. 2019. Assessing the utility of conserving evolutionary history. Biol. Rev. 94:1740–1760.

Vane-Wright, R., C. Humphries, and P. Williams. 1991. What to protect?—Systematics and the agony of choice. Biol. Conserv. 55:235–254.

Winter, M., V. Devictor, and O. Schweiger. 2013. Phylogenetic diversity and nature conservation: where are we? Trends Ecol. Evol. 28:199–204.