Review

Author for correspondence:
Mike Steel
e-mail: mathmomike@gmail.com

Neutral phylogenetic models and their role in tree-based biodiversity measures

Mike Steel

Biomathematics Research Centre, University of Canterbury, Christchurch, New Zealand

MS, 0000-0001-7015-4644

A wide variety of stochastic models of cladogenesis (based on speciation and extinction) lead to an identical distribution on phylogenetic tree shapes once the edge lengths are ignored. By contrast, the distribution of the tree's edge lengths is generally quite sensitive to the underlying model. In this paper, we review the impact of different model choices on tree shape and edge length distribution, and its impact for studying the properties of phylogenetic diversity (PD) as a measure of biodiversity, and the loss of PD as species become extinct at the present. We also compare PD with a stochastic model of feature diversity, and investigate some mathematical links and inequalities between these two measures plus their predictions concerning the loss of biodiversity under extinction at the present.

This article is part of the theme issue "'A mathematical theory of evolution": phylogenetic models dating back 100 years'.

1. Introduction

Stochastic tree models play a key role in phylogenetics at several levels. First, they describe the process of speciation and extinction represented by phylogenetic trees, a topic that dates back to Yule's 1925 paper [1]. Further stochastic processes operate within the branches of phylogenetic trees, for example to model the evolution of genetic sequences within species (under Markovian site substitution processes), or to describe the process of incomplete lineage sorting of gene trees within a species tree (under the multispecies coalescent) [2]. The investigation of these stochastic models has provided biologists with greatly enhanced techniques for inferring evolutionary trees from genomic data. Stochastic processes have also been applied to model lateral gene transfer (e.g. in prokaryotes) where genes move between the edges of a phylogenetic tree, with the goal of inferring a central tree-like signal from a large number of conflicting gene trees [3].

In this paper, we focus on some generic properties of simple stochastic models that generate trees and that operate within them. We first review some properties of neutral evolutionary models which induce a distribution on tree shapes that depends solely on the number of leaves. We describe some aspects of tree shape that depend on how speciation is viewed and compare the predictions of the neutral model concerning tree shape with the corresponding predictions of the uniform model on phylogenetic trees. We also describe some recent findings regarding the lengths of the edges (shortest, longest, average) in birth–death trees.

Next we focus on phylogenetic measures of biodiversity that are relevant to conservation biology theory (namely, phylogenetic diversity (PD) and feature diversity (FD)). The latter involves a birth–death type process operating within a (fixed or birth–death) tree. We explore the relationship between these two measures on fixed trees and on classical birth–death trees, with particular attention to predicting the relative loss of diversity owing to rapid extinction at the present. This includes some recent results and possible questions for further investigation.

(a) Types of trees

In this paper, we deal with both fixed phylogenetic trees and random birth–death trees. A phylogenetic tree T here refers to a rooted directed tree (having its edges directed away from the root) with a set of labelled leaves (vertices of out-degree 0) and with the remaining interior vertices unlabelled and of out-degree at least 2. If X is the set of leaves of T, we say that T is a phylogenetic X-tree. A star tree is a phylogenetic tree with a single interior vertex. Mostly we deal with binary trees in which all interior vertices have out-degree exactly 2. Although we can view a phylogenetic tree as a (discrete) graph without edge lengths, we will often also assign a (positive real-valued) length ℓ_e to each edge e of the tree (and if these edge lengths are proportional to time they are said to be ultrametric). If we remove the labels on the leaves of a binary phylogenetic tree (and remove any edge lengths) we obtain a tree shape denoted τ. For example, when $|X| = 4$, there are precisely two tree shapes: the balanced ('fork') shape and the unbalanced ('caterpillar') shape.

Notation: Given a rooted phylogenetic X-tree T, an edge of T, and a leaf x of T we let:
- $n = |X|$; the number of leaves of T;
- $C_T(e)$ denote the set of leaves of T that are descendants of edge e;
- $n_e = |C_T(e)|$; and
- $P_T(x)$ denote the set of edges on the path from the root of T to leaf x.

A birth–death tree $\mathcal{T}_t$ is a rooted tree generated by a stochastic process of binary splitting and death of lineages, starting from a single lineage at time 0 and acting for time t. The tree $\mathcal{T}_t$ is complete if it includes lineages that become extinct by time t, whereas a reduced birth–death tree is obtained from the complete tree by removing these extinct lineages (this reduced tree is also referred to in the literature as the reconstructed tree). Notice that a reduced birth–death tree with extant species set X induces a discrete phylogenetic X-tree (by ignoring the edge lengths and the 'stem' edge at the top of the tree), as indicated in figure 1 (lower left). For either class of trees, a pendant edge in a tree is an edge incident with a leaf; all other edges are interior.

By convention, a vertex of out-degree 2 in a phylogenetic tree represents an ancestral speciation event. This would typically involve a lineage splitting off from an existing lineage and forming a new species. However, there is usually no distinction made in a birth–death (or phylogenetic) tree as to which of the two edges corresponds to the formation of a 'new' species, rather than a continuation of the existing one. This distinction has been noted by others (e.g. [4,5]) and invoking the asymmetry between parent and offspring lineages (as Yule did in his original model of genera evolution) can change how one measures various quantities. To address this, we use the term evolutionary tree for a tree that makes explicit which of the two edges directed out of a degree-2 vertex corresponds to the newly forming species (the horizontal edges in figure 1 upper (right)).

Evolutionary trees also show lineages that lead to extinct taxa and so we define the reduced evolutionary tree to be the tree obtained from an evolutionary tree by (i) ignoring extinct lineages and (ii) not treating a lineage as a new species if it branches off a lineage that has no surviving taxa at the present. This is illustrated in figure 1 (the two trees on the lower right). Once again, ignoring edge lengths in a reduced evolutionary tree leads to a discrete phylogenetic tree, as illustrated in figure 1 (the circled vertices in the figure correspond to the root of the discrete phylogenetic tree).

2. Yule–Harding trees

The classical birth–death process has constant birth and death rates (λ and μ, respectively) and has a long history [1,6]. More general models were subsequently investigated (e.g. [7]) that allow for these rates to depend on time, or on the number of species present at a given time. In all such models, the tree either becomes extinct or the number of leaves tends to infinity with probability 1 as $t \to \infty$ (e.g. by a more general result due to Jagers [8]). Moreover, the shape of the resulting phylogenetic tree—even allowing these rates to vary as described—leads to the same distribution on phylogenetic tree shapes that depends solely on n (the number of leaves of the tree shape), the Yule–Harding (YH) distribution.

A very general model that leads to the YH distribution on tree shapes was described by Aldous [4, p.8]. Briefly stated, at any given time, the probability that the next event (speciation or extinction) is a speciation rather than an extinction is allowed to depend arbitrarily on the past. The only constraint required is that if the next event is a speciation (resp. extinction) then each species at that time is equally likely to speciate (resp. become extinct). We will refer to any such process of this type as a neutral model. A further analysis of models that lead to the YH distribution on tree shapes has been described by Lambert and Stadler [5].

The following simple discrete stochastic process generates YH tree shapes: start with the tree on two leaves and at each step, when the tree has $k \geq 2$ leaves, select one of these leaves uniformly at random and attach a new leaf to the midpoint of its incident edge to produce a binary tree with $k + 1$ leaves. If we continue until the tree has n leaves, then the resulting tree shape has the YH distribution.

For any tree shape τ with n leaves there are precisely $n!/2^{s(\tau)}$ distinct binary phylogenetic trees with this shape, where $s(\tau)$ is the number of symmetry vertices in τ (i.e. where the two descendent subtrees of the vertex have isomorphic shape). The probability $p(\tau)$ that the YH tree has a particular shape τ was described recursively by Harding [9]. However, it also has an explicit formulation, namely:
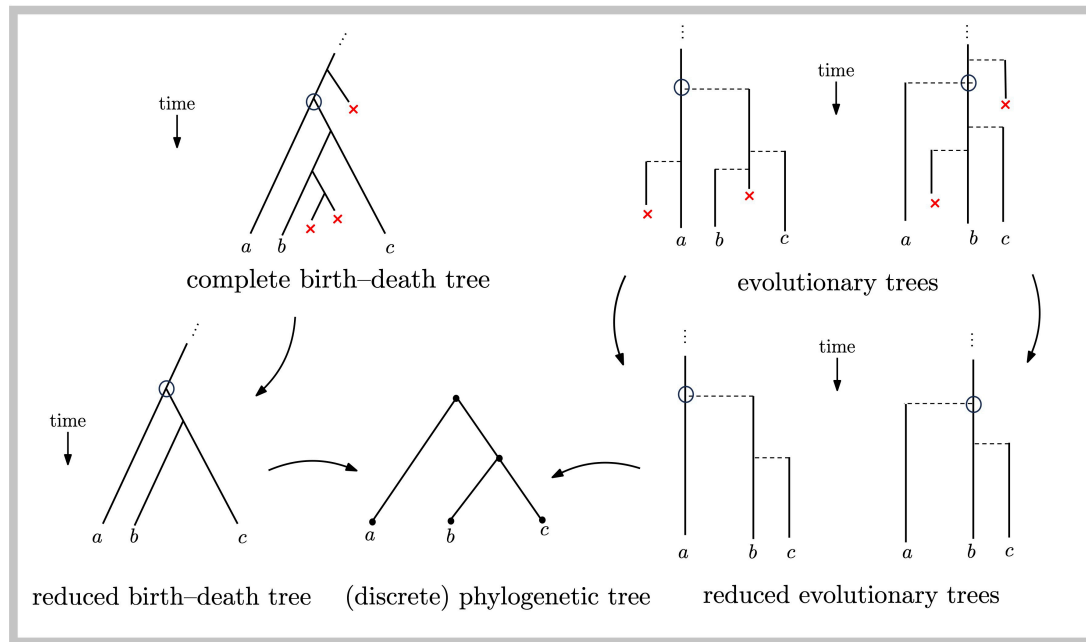
Figure 1. Classes of trees (see text for details).

$$p(\tau) = \frac{2^{n-1-s(\tau)}}{\prod_{v \in IV_\tau} n_\tau(v)},$$

where $s(\tau)$ is the number of symmetry vertices in τ (i.e. where the two descendent subtrees of the vertex have isomorphic shape), $n_\tau(v)$ is the number of leaves that are descendants of vertex v minus 1 and IV_τ is the set of interior vertices of τ. For example, the tree shape illustrated in figure 2 has probability $2^{5-1-3}/(4 \times 3 \times 1 \times 1) = 1/6$.

For any tree shape τ with n leaves there are precisely $n!/2^{s(\tau)}$ distinct binary phylogenetic trees with this shape. Thus, the YH distribution on tree shapes induces a distribution on binary phylogenetic trees by multiplying $p(\tau)$ by $2^{s(\tau)}/n!$. For example, the phylogenetic tree shown in figure 2 has $n = 5$ and $s(\tau) = 3$, and so has probability $1/90$.

It turns out that the YH distribution also describes the distribution of tree shapes that arise under a somewhat different process, namely, the Kingman coalescent (again assuming that the edge lengths are ignored).

Historical comment: E.F. Harding investigated the enumeration and probabilities of tree shapes and phylogenetic trees more than 50 years ago in his seminal paper [9]. The enumeration of tree shapes has a longer history, dating back to 1940, when Etherington [10] investigated the functional equation $f(x) = x + \frac{1}{2}(f^2(x) + f(x^2))$ that describes the generating function of these numbers. Harding also put forward a conjecture regarding the shape of the most probable phylogenetic trees on n leaves under this distribution. This conjecture was established a few years later [11]. However, formulating a conjecture regarding the most probable tree shape on n leaves (rather than the shape of a most probable phylogenetic tree, which has labelled leaves) eluded both Harding and the authors of [11], as the most probable shape exhibits an irregular behaviour as n grows.

(a) Depth of a random leaf in reduced trees

The depth of an extant leaf x in a reduced birth–death tree (or in a phylogenetic tree) is the number of edges between the root of the tree and leaf x. Similarly, the depth of an extant leaf x in a reduced evolutionary tree is the number of horizontal edges in the path from the root to x. Let D_n (resp. D'_n) denote the depth of a randomly chosen (extant) leaf in a reduced birth–death tree (resp. reduced evolutionary tree) with n leaves.

Notice that for the reduced birth–death tree (or the phylogenetic tree) in figure 1 the depth D_3 of a randomly chosen leaf equals 2 with probability 2/3 and equals 1 with probability 1/3. By contrast, in the reduced evolutionary trees in this figure there are two scenarios to consider. In the first (the left-hand reduced evolutionary tree in figure 1), the depth of a randomly chosen leaf is uniform on 0, 1, 2; in the other (the right-hand reduced evolutionary tree in figure 1) the depth is 0 with probability 1/3 and 1 with probability 2/3. Since the two scenarios are equally likely it follows that $D'_3 = 0, 1, 2$ with probabilities $\frac{1}{3}, \frac{1}{2}$ and $\frac{1}{6}$, respectively.

For neutral models (in which the reduced birth–death tree shapes follow the YH model), the distribution for D_n has the following exact expression:
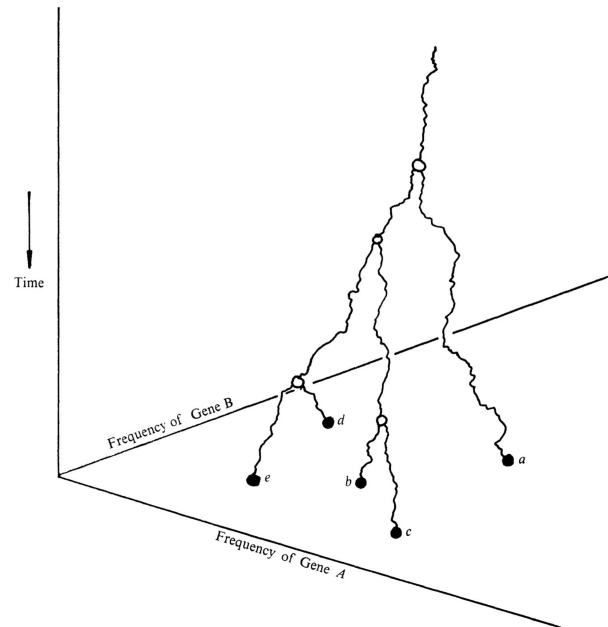
Figure 2. The binary phylogenetic tree from Harding's original 1971 paper [9]. After ignoring the edge lengths and leaf labels, there are three tree shapes on five leaves, and the tree shape shown has probability 1/6. The tree itself (with leaf labels) has probability 1/90.

$$\mathbb{P}(D_n = k) = 2^k \cdot \frac{c(n-1, k)}{n!}; \text{ for } k = 1, \ldots, n-1, \tag{2.1}$$

where $c(n-1, k)$ is the number of permutations on $n-1$ elements with k cycles (the (unsigned) Stirling numbers of the first kind) [12,13], and $\mathbb{P}(D_n = k) = 0$ for $k \geq n$ or $k = 0$.

By comparison, for D'_n under a neutral model we have the following result (the short proof is provided in the Appendix A.1):

Proposition 1 . *For $k = 0, \ldots, n-1$,*

$$\mathbb{P}(D'_n = k) = \frac{c(n, k+1)}{n!}, \tag{2.2}$$

and $\mathbb{P}(D'_n = k) = 0$ for $k \geq n$. Moreover, $\mathbb{E}[D'_n] = \frac{1}{2}\mathbb{E}[D_n] = \sum_{i=2}^{n} \frac{1}{i}$ and the total variational distance between D'_n and a Poisson distribution with mean $\ln(n) + \gamma - 1$ converges to 0, where γ is the Euler–Mascheroni constant.

(b) Yule–Harding trees versus uniform phylogenetic trees

We now list some of key features of the shape of Yule–Harding (YH) trees and how they differ from the shape of binary phylogenetic trees sampled uniformly at random, which is often called the *Proportional-to-Distinguishable-Arrangements* (PDA) model[1].

First, if T_n is a YH tree and we select uniformly at random one of the two subtrees incident with the degree-2 root of T_n then the number N of leaves in this subtree tree is uniformly distributed between 1 and $n-1$ (this follows by a simple Pólya-urn model argument). By contrast, the PDA model leads to a quite different distribution for N, since the probability that one of the two subtrees incident with the root consists of just one leaf equals $n/(2n-3)$, which converges to 1/2 as $n \to \infty$ (whereas the corresponding probability converges to 0 under the YH model).

Another aspect of the shape of a tree is how 'balanced' it is. There are many ways to measure tree balance (a recent book [14] provides an excellent survey). However, tree balance, roughly speaking, usually measures the extent to which (on average across the tree) the two subtrees below the interior vertices differ in size. Empirical phylogenetic trees tend to be less 'balanced' than the YH model predicts, but more balanced than expected in the PDA model [15].

A further difference between the YH and PDA models concerns the depth of a randomly chosen leaf (recall that depth refers to the number of edges between the leaf and the root of the tree). Suppose we select a leaf uniformly at random from a YH tree T_n; then the depth D_n of this leaf has an expected value of order $\ln(n)$ as $n \to \infty$ by equation (2.1). By contrast, for the PDA model, the corresponding expected value has order $\sqrt{n}$ [4].

Next, given a YH tree with n leaves, if we select uniformly at random a pair of leaves, then their most recent common ancestor (MRCA) vertex is likely to be 'close' to the root vertex of the tree. More precisely, the number of edges between the MRCA and the root converges in distribution (as n grows) to a geometric random variable G with $\mathbb{P}(G = g) = \frac{1}{3} \cdot (\frac{2}{3})^g$ for $g = 0, 1, \ldots$ [16]. Thus, the probability that the MRCA vertex of the chosen pair of leaves is the root of the YH tree converges to $\frac{1}{3}$

---

[1]The YH model and PDA models are special cases of the β-splitting model of Aldous [4] , with $\beta = 0$ and $\beta = -3/2$, respectively.

as n grows[2]. For the PDA model however, the probability that the MRCA of the chosen pair of leaves is the root of the tree tends to zero as $n \to \infty$.

We end this section by describing a recent result that exhibits both a difference and a similarity between the YH and PDA model. Let τ_k be the particular binary phylogenetic shape where all leaves have depth k (thus τ_k has 2^k leaves).

Claim 1: For each fixed k, the probability that the subtree of a YH tree T_n consisting of all vertices of depth at most k from the root has shape τ_k tends to 1 as $n \to \infty$.

This claim follows by a slight extension of the argument used to establish Proposition 5 in [18]. Moreover, since every rooted tree shape is a subtree of τ_k for some (sufficiently large) k it follows that the probability that a YH tree T_n contains a subtree of shape τ tends to 1 as $n \to \infty$. This last result also holds for the PDA model, but the proof is more involved (for details, see [18]) since Claim 1 (above) is easily shown to fail for the PDA model.

(c) Edge lengths of Yule trees

Next, consider the Yule process, a continuous-time Markov process on state space 1,2,3…, with a linear birth rate λ and a death rate of 0. Such a process corresponds to the growing number of lineages in a random rooted binary tree in which each lineage of the tree splits independently according to an exponential distribution with rate λ. This tree is commonly referred to as Yule tree and we adopt this terminology here, despite it being somewhat contrived (as noted by Amaury Lambert in this special issue).

Suppose we now sample a Yule tree just before the number of leaves of the tree changes from n to $n+1$. Ignoring the initial stem edge, select any other edge of this tree uniformly at random from either (i) the set of pendant edges or (ii) the set of interior edges. In either case, the length of the sampled edge is exponentially distributed with expected value $\frac{1}{2\lambda}$. Thus the notion that the pendant edges might tend to be shorter, on average, than the interior edges (because the pendant edges are in some sense 'still growing') is misleading. For further details, and extensions to birth–death trees see [19]. By contrast, if we select one of the two edges descended from the root of this Yule tree, then the expected length of this edge is different; it is precisely $\frac{1}{\lambda}\left(1 - \frac{1}{n}\right)$ [20].

We can also consider the sum of the lengths of all the edges of a Yule tree (again ignoring the stem edge). Let S_n denote this sum at the last moment when the tree has $n \geq 2$ leaves. Since there are $2n-2$ edges in a binary tree with n leaves, and since a randomly selected edge has expected length $1/2\lambda$, it follows that S_n has expected value $(n-1)/\lambda$. However, more can be said: S_n has a Gamma distribution with variance $(n-1)/\lambda^2$ and is asymptotically normally distributed as $n \to \infty$ [20].

Moving beyond Yule trees, the distributions of the interior and pendant edge lengths in birth–death trees were investigated in [19], with more recent results reported in [21].

(d) The longest pendant and shortest interior edges in birth–death trees

We now consider the length of the longest pendant edge or the shortest interior edges. The former are of interest to biologists, since they end at a species that is maximally distant from the other taxa. The latter are relevant to tree reconstruction, as they represent near-polytomies that are hard to resolve using genomic data. Assume $\lambda > \mu$. Let L_t denote the length of the longest pendant edge in a reduced birth–death tree at time t, and let N_t denote the number of leaves present at time t. The following result is from [22].

Proposition 2. *Conditional on $N_t > 0$, the ratio L_t/t converges in probability to $\frac{1}{2}$ and $\mathbb{E}\left[\frac{L_t}{t} | N_t > 0\right] \to \frac{1}{2}$ as $\lambda t \to \infty$.*

A noteworthy feature of Proposition 2 is that L_t/t converges to a constant that involves neither λ nor μ. By contrast, the longest pendant edge in the complete birth–death tree (i.e. including past extinctions) divided by t converges in probability to $\frac{1}{2}(1 - \mu/\lambda)$ as t grows [22]. A study across 114 mammal families (with increasing numbers of leaves) provided a reasonable fit to the value $t/2$ predicted by Proposition 2 [22], as did subsequent comparisons against other groups of taxa (e.g. birds) [23]. A heuristic argument for Proposition 2 was also described recently in [24], by considering ranked YH trees. Results for the lengths of the k-th longest pendant edge are also possible.

It is also of interest to consider the length S_t of the shortest interior edge, since resolving short edges from aligned DNA sequence data requires sequence lengths that (provably) grow with the inverse square of this length. Let S_t denote the length of the shortest interior edge in a Yule tree $\mathcal{T}_t$ grown for time t. Then from [22] (Proposition 4), we have:

$$\mathbb{P}\left(S_t \geq xe^{-\lambda t}/\lambda\right) \to \frac{1}{1+2x}, \text{ as } \lambda t \to \infty. \tag{2.3}$$

Less precisely, equation (2.3) with $x = \frac{1}{2}$ implies that, as t grows, there is an approximately 50% probability that the shortest interior edge will be smaller than the average interior edge length divided by the expected number of leaves.

---

[2]More generally, Sanderson [17] showed that for a randomly selected subset of leaves S of size k, the probability that the MRCA of S is the root of a YH tree converges to $1 - 2/(k+1)$ as $n \to \infty$.

3. Phylogenetic diversity

In the previous section, we considered the distribution of the sum of all the edge lengths in a Yule tree. This sum can be viewed as some measure of the total amount of evolution that has taken place in the tree. More generally, for any phylogenetic tree T on leaf set X and any subset Y of X, consider the sum of the lengths of the edges in the minimal subtree of T that connects the leaves in Y with the root of T, denoted here as $PD_{(T,\ell)}(Y)$. Formally, for an edge e of T, we let ℓ_e denote any (real-valued and positive) 'length' associated with e and we let ℓ denote the function $e \mapsto \ell_e$ over all edges of T. Then

$$PD_{(T,\ell)}(Y) = \sum_{e\,:\,C_T(e)\cap Y \neq \varnothing} \ell_e.$$

This quantity, referred to as the phylogenetic diversity of Y, has been advocated as a measure of biodiversity, beginning with a seminal paper by Daniel P. Faith [25]. The function $Y \mapsto PD_{(T,\ell)}(Y)$ satisfies a combinatorial strong exchange property that allows a subset of any given size that maximizes PD to be found via a linear-time greedy algorithm.

Notice that the ratio $PD_{(T,\ell)}(Y)/PD_{(T,\ell)}(X)$ measures, roughly speaking, the proportion of the total edge lengths of T that is spanned by Y, and is a quantity considered further in §3c. For birth–death trees we will assume that the edge lengths are proportional to time (and so ultrametric); however for fixed phylogenetic trees we allow arbitrary positive edge lengths.

(a) PD indices

Phylogenetic diversity provides a measure of diversity for subsets of species. However, conservation biologists are also interested in quantifying how much each species contributes to the total PD of a phylogenetic X-tree T. In other words, they are interested in functions $I_{(T,\ell)}: X \to \mathbb{R}^{\geq 0}$ for which

$$\sum_{x \in X} I_{(T,\ell)}(x) = PD_{(T,\ell)}(X).$$

We also require that $I_{(T,\ell)}(x)$ be a linear function of the edge lengths (thus, we can write $I_{(T,\ell)}(x) = \sum_e c_e(x)\ell_e$ for values $c_e(x)$) and, moreover, that $c_e(x) = 0$ if x is not a descendant of e, with $c_e(x)$ is strictly positive otherwise. The class of such functions can thus be written as

$$I_{(T,\ell)}(x) = \sum_{e \in P_T(x)} c_e(x)\ell_e \tag{3.1}$$

and they are referred to as PD indices.

The most popular PD index is fair proportion (FP) (also called 'evolutionary distinctiveness'), introduced in [26] and defined by putting $c_e(x) = 1/n_e$ in equation (3.1). In other words, FP shares out the length of each edge equally amongst the leaves of T that are descendants of that edge. This 'fair sharing' property is consistent with the mathematical result that FP is precisely the Shapley value of a cooperative game that is naturally associated with any phylogenetic tree with edge lengths [27]. Another PD index is equal splits (ES) defined by putting $c_e(x) = 1/N_e(x)$ in equation (3.1) where $N_e(x)$ is the product of the out-degrees of the interior vertices of T in the path from e to x. Other PD indices exist (e.g. by taking any convex combination of FP and ES).

Although a PD index satisfies the equation $\sum_{x \in X} I_{(T,\ell)}(x) = PD_{(T,\ell)}(X)$, if we consider a subset Y of X, the value $\sum_{x \in Y} I_{(T,\ell)}(x)$ does not generally equal $PD_{(T,\ell)}(Y)$, except in special cases (e.g. if T is a star tree). Nevertheless, a general inequality holds for any phylogenetic X-tree T, namely:

$$\sum_{x \in Y} I_{(T,\ell)}(x) \leq PD_{(T,\ell)}(Y), \text{ for all } Y \subseteq X.$$

This result is from [28] (Lemma 1.1), where the authors explored how large the difference $PD_{(T,\ell)}(Y) - \sum_{x \in Y} I_{(T,\ell)}(x)$ can be (over subsets of leaves Y of X) when $I_{(T,\ell)}$ is the FP or ES index and T is either fixed or free (and ℓ fixed or constrained). A further recent result concerning PD indices is that the rankings of species by FP can change (and even reverse) if some of the species were to disappear from the tree (for details, see [29,30]).

Turning next to random trees, let $FP_{\mathcal{T}_t}(x)$ be the fair proportion index of a leaf selected uniformly at random from a Yule tree $\mathcal{T}_t$ grown for time t. In this case,

$$\mathbb{E}[FP_{\mathcal{T}_t}(x)] = \frac{1}{\lambda}(1 + e^{-\lambda t}),$$

(from [22, Equation 32]) and so $FP_{\mathcal{T}_t}(x)$ converges in expectation to $\frac{1}{\lambda}$ as t grows. Since the expected length of the pendant edge incident with x converges to $\frac{1}{2\lambda}$ as t grows, one expects around half of the FP index for a random leaf to be contributed by its pendant edge, which is consistent with an early empirical observation from [26].

PD indices have also been recently defined and studied in a more general setting that includes unrooted phylogenetic trees, phylogenetic networks and set systems [31–33].

(b) The loss of PD under extinction at the present

Mass extinction events have played an important role in evolution [34] and we are currently entering a new (anthropogenic) one. The simplest process of extinction at the present is the field-of-bullets (FOB) model, in which species become extinct independently, and with the same probability ϵ. A more nuanced process—the generalized field-of-bullets model (gFOB) —assigns each species its own extinction probability ϵ_x while still assuming independence. In practice, species extinction
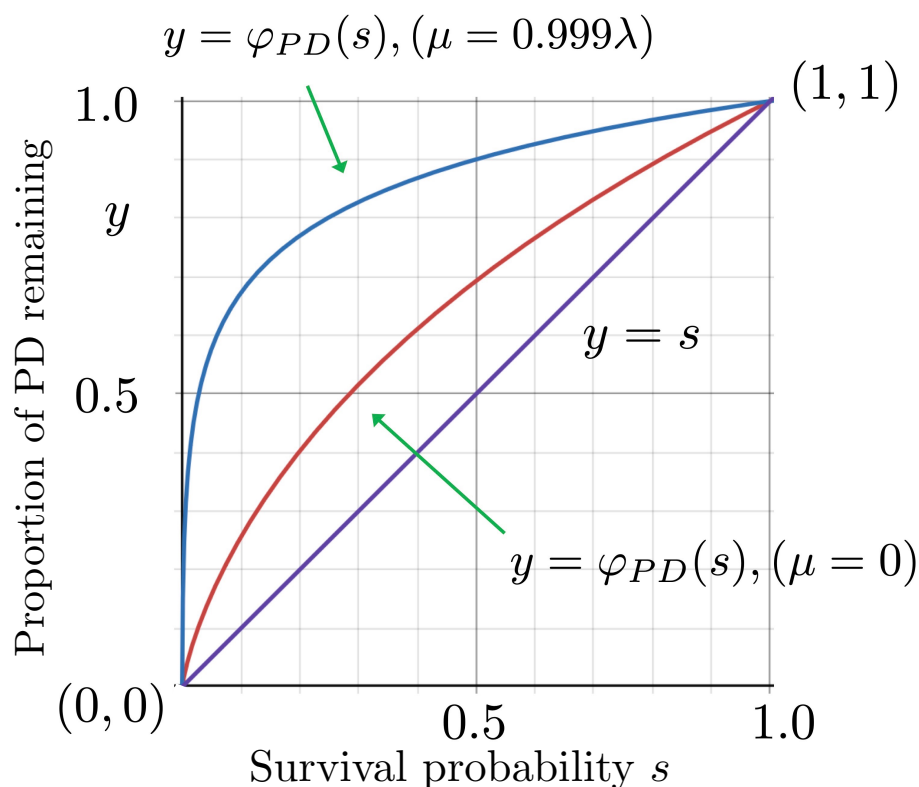
Figure 3. The proportion of PD remaining (*y*-axis) after a field of bullets extinction of variable intensity (*x*-axis is the survival probability *s*) at the tips of a birth–death tree as $t \to \infty$. Except when $s = 0,1$ this proportion always exceeds the proportion of species remaining (bottom line) and increases from the curve corresponding of a Yule tree (middle curve) to the unit step function as the birth–death process approaches criticality ($\mu/\lambda \to 1$).

probabilities can be estimated from their Red List status (from the International Union for Conservation of Nature) [35,36]. More realistic extinction processes would account for the fact that the extinction of some species may trigger the further extinction of other species (leading to 'extinction cascades'); alternatively, species that share common features may be correlated in their extinction risk if these features turn out to be deleterious. We do not consider these more complex models here.

For a given phylogenetic *X*-tree *T* with given edge lengths ℓ, the expected proportion of PD that is lost under a gFOB model is given by:

$$\frac{\sum_e \ell_e \prod_{x \in C_T(e)} \epsilon_x}{\sum_e \ell_e}. \tag{3.2}$$

To justify equation (3.2), simply observe that ℓ_e fails to contribute to the post-extinction PD score precisely if all species descended from *e* become extinct. Provided that the distribution of the edge lengths is not too extreme, and the ϵ_x values are bounded away from 0 and 1, the distribution of PD on *T* that is lost becomes asymptotically normally distributed as the number of leaves grows (for details, see [37], Theorem 3.1). Next we consider the loss of PD for birth–death trees (rather than for fixed trees).

(c) Loss of PD on birth–death trees under extinction at the present

Let $\mathcal{T}_t$ be a birth–death tree with $\lambda \geq \mu$, and apply the simple FOB model, with survival probability *s*, to the extant leaves of $\mathcal{T}_t$. What proportion of the tree remains? More precisely, let X_t denote the extant leaves of $\mathcal{T}_t$ and let Y_t denote the subset of X_t of the species that survive the FOB extinction event. What then can be said about the proportion of PD that survives, or, in other words, the ratio $PD_{\mathcal{T}_t}(Y_t)/PD_{\mathcal{T}_t}(X_t)$? From [38] the following convergence in probability holds as *t* becomes large:

$$\frac{PD_{\mathcal{T}_t}(Y_t)}{PD_{\mathcal{T}_t}(X_t)} \xrightarrow[t \to \infty]{P} \varphi_{PD}(s)$$

where

$$\varphi_{PD}(s) := \lim_{t \to \infty} \frac{\mathbb{E}[PD_{\mathcal{T}_t}(Y_t)]}{\mathbb{E}[PD_{\mathcal{T}_t}(X_t)]}.$$

When $\mu = 0$ (i.e. a Yule tree), $\varphi_{PD}(s)$ has a particularly simple form, namely:

$$\varphi_{PD}(s) = \frac{s \ln (1/s)}{1 - s}. \tag{3.3}$$

Otherwise, if $\mu > 0$, let $\rho = \mu/\lambda$, in which case:

$$\varphi_{PD}(s) := \begin{cases} \frac{\rho s}{\rho + s - 1} \cdot \frac{\ln(s/(1-\rho))}{\ln(1/(1-\rho))}, & \text{if } \rho = \mu/\lambda \neq 0, 1-s; \\ (1-s)/\ln(1/s), & \text{if } \rho = 1-s. \end{cases} \tag{3.4}$$

In all cases, $\varphi_{PD}(s)$ is a monotonically increasing concave function that satisfies $\varphi_{PD}(s) \geq s$ with equality if and only if $s \in \{0,1\}$. This is shown in figure 3, which also illustrates an interesting feature of $\varphi_{PD}(s)$: as μ approaches λ from below, the function $y = \varphi_{PD}(s)$ approaches the unit step function (which is zero at $s = 0$ and is 1 for $s \in (0,1]$). Roughly speaking, this is because the resulting critical trees, conditional on non-extinction, have very short pendant and near-pendant edges.

Moreover, for any value of μ and λ, as the survival probability (s) approaches 1 the function $y = \varphi_{PD}(s)$ has a slope equal to the expected proportion of the PD of the tree that is contributed by its pendant edges. The intuition behind this is that when s is close to 1, the only edges that disappear are isolated pendant edges (interior edges only start to disappear when two adjacent leaves disappear). Thus, for Yule trees, the slope of $y = \varphi_{PD}(s)$ at $s = 1$ is $\frac{1}{2}$, whereas for birth–death trees with $\mu > 0$, the slope at $s = 1$ is smaller.

> **Heuristic derivation of** φ_{PD}: To provide some intuition for equation (3.3), here we give a brief (but non-rigorous) derivation of it. First recall that constant-rate birth–death trees induce a tree shape that follows the YH distribution, and for this distribution, the expected number of edges that have $k \leq n - 1$ leaves as descendants is $2n/(k(k+1))$. Next, observe that the length of each such edge contributes to the PD score with probability $(1 - (1 - s)^k)$ (i.e. if at least one of these k leaves survives). Now for a Yule tree, the expected length of the (interior and pendant) edges is the same, and so $\varphi_{PD}(s)$ reduces to $\sim \sum_{k=1}^{n} \frac{1}{k(k+1)}(1 - (1-s)^k)$, which converges to $s \ln(1/s)/(1-s)$ as $n \to \infty$.

(d) Stochastic diversity indices for extinction at the present

Let X denote a set of extant species and consider the (random) subset $\mathcal{S}$ of species in X that will still be extant (i.e. still present) at some future time $t_* > t$ under some extinction model. Let (T, ℓ) be an associated phylogenetic tree with leaf set X. Focusing on any particular species $x \in X$, let $\mathcal{S}_x = \mathcal{S} \setminus \{x\}$ and let

$$\psi_x = \mathbb{E}[PD_{(T,\ell)}(\mathcal{S}_x \cup \{x\}) - PD_{(T,\ell)}(\mathcal{S}_x)].$$

Thus ψ_x measures the expected additional PD at time t_* if species x is present compared with if it is absent at time t_* (noting that other species may also be extinct at time t_*). If we now specialize the extinction model to the gFOB model, where each species s in X has probability ϵ_s of being extinct at time t_*, then:

$$\psi_x = \sum_{e \in P_T(x)} \ell_e \prod_{s \in C_T(e) \setminus \{x\}} \epsilon_s \tag{3.5}$$

Analogously, let

$$\psi'_x = \mathbb{E}[PD_{(T,\ell)}(\mathcal{S}_x \cup \{x\}) - PD_{(T,\ell)}(\mathcal{S}_x \cup \mathcal{I}_x)],$$

where $\mathcal{I}_x \in \{\emptyset, \{x\}\}$ and $\epsilon_x = \mathbb{P}(\mathcal{I}_x = \emptyset)$. Thus, ψ'_x is the expected additional PD at time t_* if species x is present at time t_* compared with if its state is uncertain at time t_*. Straightforward algebra leads to the following relationship between ψ_x and ψ'_x under the gFOB model:

$$\psi'_x = \psi_x \cdot \epsilon_x. \tag{3.6}$$

(see [39] for details). The indices ψ and ψ' linked by equation (3.6) are part of the recently proposed EDGE2 protocol [38] for assessing potential biodiversity loss in coming decades. Note that ψ and ψ' are not PD indices in the sense of §3a since summing either of them over all elements of X is not, in general, equal to $PD_{(T,\ell)}(X)$ (however this is the case for ψ when T is a star tree).

We end this section by noting a curious connection between the index ψ and the fair proportion index from §3a. Under the FOB model, suppose that the extinction probability ϵ is described by a uniform distribution on $[0,1]$. Then:

$$FP_{(T,\ell)}(x) = \mathbb{E}[\psi_x(\epsilon)], \text{ for all } x \in X, \tag{3.7}$$

where the expectation is with respect to this uniform prior on ϵ. The short proof of equation (3.7) is in the appendix A.2.

4. Feature diversity (FD)—a generalization of PD

The PD of a subset Y of extant species is sometimes assumed to be a proxy for the diversity of features (traits, characteristics, etc.) that are present among the species in Y. Under certain settings, this can be shown to be formally correct, as we describe shortly; however, under more realistic models of feature evolution it does not exactly apply. One reason for this which we model here, is that features are not only gained along lineages, but can also be lost (e.g. the evolution of flight in birds, which has subsequently been lost in some species). Indeed, a number of authors have pointed out that PD can be a poor proxy for feature diversity (FD) (e.g. [40,41]).

Features can be either discrete or continuous. Here, for simplicity, we consider just the discrete setting. Suppose that each species x in X has an associated set F_x of features (some or all of which may be shared by other species). Let $\mathbb{F} = (F_x : x \in X)$ denote the assignment of features to each species in X and suppose that feature f has an assigned strictly positive real-valued weight $w(f)$ (this may reflect the complexity or importance of this feature). For any subset Y of X, its feature diversity, denoted $FD_{\mathbb{F}}(Y)$, is given by:

$$FD_{\mathbb{F}}(Y) = \sum_{f \in \bigcup_{y \in Y} F_y} w(f);$$

(i.e. the sum of the $w(f)$ values across all features present in at least one species in Y). Thus, when $w(f) = 1$ for all features then $FD_{\mathbb{F}}(Y)$ is simply the number of features present among the species in Y. Notice that, unlike PD, FD does not involve any phylogenetic tree. Nevertheless, we can view PD as a special case of FD (but not conversely); as the following box explains.

> **PD as FD:** Given (T, ℓ), PD can be modelled exactly by FD as follows: Assign to each species x in X a feature f_e for each edge e on the path in T from the root to leaf x, and assign feature f_e weight $w(f_e) = \ell_e$. Then $PD_{(T, \ell)}(Y) = FD_{\mathbb{F}}(Y)$ for all sets $Y \subseteq X$. This abstract identity can be helpful. For example, the function $Y \longmapsto FD_{\mathbb{F}}(Y)$ (for any $\mathbb{F}$) satisfies the submodular inequality (which states that $\phi(Y_1 \cup Y_2) + \phi(Y_1 \cap Y_2) \leq \phi(Y_1) + \phi(Y_2)$) and so PD does too. However, FD cannot generally be represented by PD. A counter example is the feature sets for species x, y, z given by $F_x = \{f_1, f_2\}$, $F_y = \{f_1, f_3\}$ and $F_z = \{f_2, f_3\}$ with $w(f_i) = 1$. The non-nested nature of these sets implies that the resulting FD function cannot be realized as PD on any tree with edge lengths.

It turns out that FD can be represented by PD on some tree when the collection $\mathcal{C}_{\mathbb{F}}$ of subsets $X_f = \{x \in X : f \in F_x\}$ is nested (i.e. the condition that $A, B \in \mathcal{C}_{\mathbb{F}} \Rightarrow A \cap B \in \{A, B, \varnothing\}$ holds), by results in [42] (Proposition 1 and Theorem 1). Moreover, since a nested collection of subsets of X corresponds to a subset of the clusters of a phylogenetic tree, it is possible to describe the evolution of each character in $\mathbb{F}_X$ on that tree by a single gain (of the feature on one edge e) and no subsequent loss (on any descendant edge of e). A slightly stronger statement (from [42]) is the following:

Proposition 3. *For feature assignment $\mathbb{F}_X$, there is a phylogenetic X-tree T with non-negative edge lengths ℓ (not necessarily ultrametric) so that $FD(Y) = PD_{(T, \ell)}(Y)$ for all subsets Y of X if and only if the evolution of each feature in $\mathbb{F}_X$ can be described by an evolutionary scenario on T involving single gains and no loss.*

In fact, the 'only if' part holds even if we restrict the subsets Y of X to be of sizes of at most 3. The key phrase in Proposition 3 is 'can be described'; there is no requirement that this is how the feature actually evolved. For example, a feature that arises on an interior edge and is subsequently lost on one of the edge's two descendant edges can be described by a scenario that involves a single gain (on the other child edge) and no loss.

(a) A stochastic model of FD on fixed trees

Proposition 3 is purely combinatorial, so it is instructive to consider a stochastic model of feature evolution on a phylogenetic X-tree T (with positive edge length function $\ell = [\ell_e]$), where each new feature arises once in the tree and can subsequently be lost. This is illustrated in figure 4.

In this model,

- a new feature arises (on any edge) at rate $r > 0$ per unit time, and
- each feature (on any edge) is subject to an independent loss process with rate $\nu \geq 0$.

This simple model can be described by a classical (constant-birth, linear-death) stochastic process for which the following results hold (see e.g. [44, p. 461]). Consider a directed path in the tree T of length L (the sum of the ℓ_e values on the path). Conditional on there being k features present at the start of this path, let $N_L^{(k)}$ denote the number of features present at the end of this path. For $k = 0$, $N_L^{(0)}$ has a Poisson distribution with mean $\frac{r}{\nu}(1 - e^{-\nu L})$, and this mean is asymptotic to rL as $L \to 0$ and to r/ν as $L \to \infty$. More generally, if there are k features present at the start of the path, then $N_L^{(k)}$ is the sum of two independent random variables: one distributed as $N_L^{(0)}$, the other has a Gamma distribution with parameters $(k, \nu L)$. In particular, as $L \to \infty$, $N_L^{(k)}$ converges to a Poisson distribution with mean r/ν.

To simplify what follows, we will assume that $w(f) = w > 0$, though similar results hold more generally when $w(f)$ are i.i.d. random variables taking values in $[0, B]$ for some constant B. Given a phylogenetic X-tree T, the assignment $\mathbb{F}_X$ of features present at the leaves of T is now a random variable, and so is $FD(Y)$ for each subset Y of X. Let us then consider the expected FD value (i.e. $\mathbb{E}[FD_{(T, \ell)}(Y)]$), which depends on T, $\ell = [\ell_e]$, r and ν.

We can now ask whether the expected FD function $Y \mapsto \mathbb{E}[FD_{(T, \ell)}(Y)]$ coincides exactly with the PD function on T, with some appropriately chosen non-negative edge lengths (say ℓ'). It is easy to see that when the rate of loss ν equals zero, then the two functions always coincide (and with $\ell = \ell'$). Moreover, when $\nu > 0$ and T is one of the four tree shapes shown in figure 5, the expected FD function can be described exactly by PD for an appropriate choice of ℓ'. Note that these four classes of trees all have depth at most 2; however, not all trees of depth 2 fall into one of these four classes.
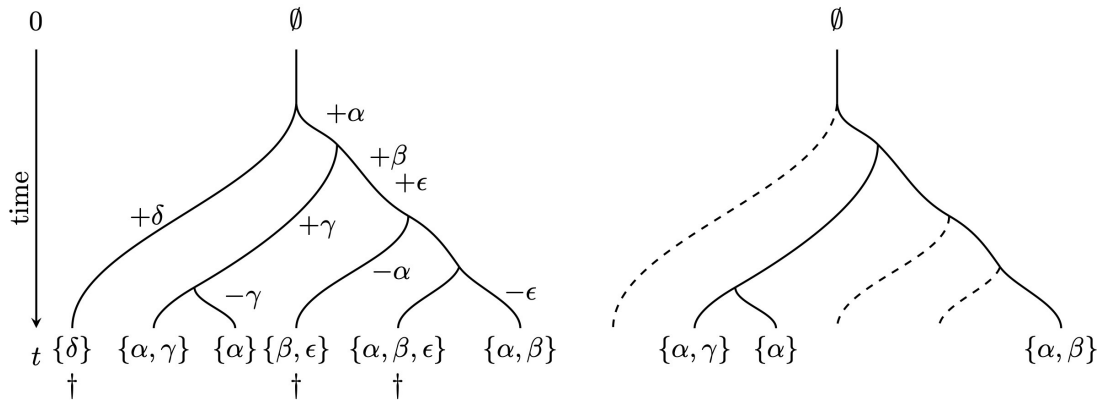
Figure 4. A simple example to illustrate the gain and loss of features on a tree and the FOB process (from [43]). Left: new features arise (indicated by +) and disappear (indicated by −) along edges of the tree. Each new feature arises only once in the tree, and for convenience, no features are present at time 0. Five features are present among the leaves of the tree at time t: $\{\alpha, \beta, \gamma, \delta, \epsilon\}$. A rapid extinction event at the present then results in the loss of three leaves of the tree (denoted by †). Right: three features (α, β, γ) remain among the leaves of the resulting tree. Thus, the ratio of surviving features to total features is $3/5 = 0.6$.
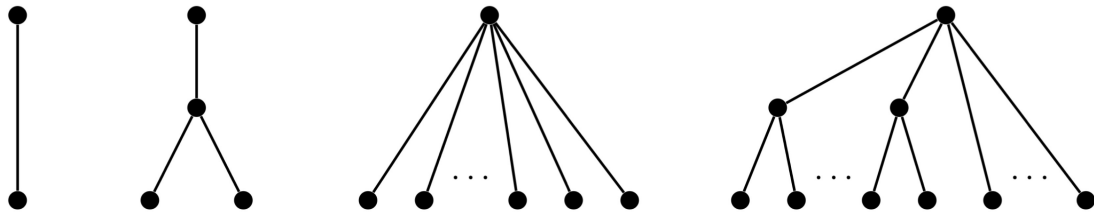

Figure 5. The (only) four shapes of trees for which the expected feature diversity can be exactly described by phylogenetic diversity (from [45]).

Apart from these special cases, the following result was recently established in [45] (Supplementary Material, Theorem 1.5). It can be regarded as a stochastic analogue of the combinatorial result in Proposition 3.

Proposition 4. *Suppose that $\nu > 0$, and T is a phylogenetic X-tree, with strictly positive edge lengths ℓ. Then there exists edge lengths ℓ' for T for which $\mathbb{E}[FD_{(T,\ell)}(Y)] = PD_{(T,\ell')}(Y)$ for all subsets Y of X if and only if T has the shape of one of the four trees shown in figure 5.*

As was the case for Proposition 3, the 'only if' part of Proposition 4 also holds if we restrict the subsets Y of X to be just those of size at most 3. This result remains true even if one allows the edge lengths ℓ and ℓ' to be non-ultrametric.

The function $Y \mapsto \mathbb{E}[FD_{(T,\ell)}(Y)]$ (and a related deterministic measure of the evolution of feature diversity) forms the basis of the recently proposed 'EvoHeritage' suite of biodiversity measures [45], in which the gain and loss of features provide a more nuanced measure of biodiversity than simpler PD-based measures.

(b) FD in birth–death trees with extinction at the present

In the previous section, we considered FD on a given tree under a stochastic model of feature gain and loss. In this section, we allow the tree to be random (a birth–death tree) and subject to the FOB model operating on the leaves at the present. Thus we have the following three-layer stochastic process:

- a birth–death tree $\mathcal{T}_t$ with parameters λ, μ grown for time t from time 0;
- a feature evolution model operating on the edges of $\mathcal{T}_t$ with gain and loss parameters r, ν (as in §4a);
- a FOB model operating at time t on the tips of $\mathcal{T}_t$, with survival probability s.

In this setting, we are interested in determining the proportion of features at the tips of the tree that survive the FOB model.

Next, we recall a standard result from birth–death theory. Consider a linear birth–death process X_t, with birth rate λ, death rate θ, starting from a single individual at time 0. At time t, sample the individuals present independently with sampling probability $s > 0$, let X_t ($t \geq 0$) denote the number of these sampled individuals, and let $R_t^s(\lambda, \theta) = \mathbb{P}(X_t > 0)$. In this case:

$$R_t^s(\lambda, \theta) = \begin{cases} \frac{s(\lambda - \theta)}{s\lambda + (\lambda(1-s) - \theta)e^{(\theta - \lambda)t}}, & \text{if } \lambda \neq \theta; \\ \frac{s}{1 + \lambda s t}, & \text{if } \lambda = \theta. \end{cases}$$

(4.1)

In particular, $R_t^s(\lambda, \theta)$ converges to 0 if $\lambda \leq \theta$ and converges to a strictly positive value, namely $1 - \theta/\lambda$, if $\lambda > \theta$ [7,46].

Let $\mathcal{T}_t^s$ denote the tree obtained from $\mathcal{T}_t$ by deleting all leaves at time t that do not survive the FOB model. The number of species at time t in $\mathcal{T}_t^s$ that have a copy of a particular feature that arose at some fixed time $t_0 \in (0, t)$ in $\mathcal{T}_t$ is described exactly by the birth–death process X_{t-t_0} with birth rate λ, death rate $\theta = \mu + \nu$ and extinction at the present (with survival probability s for each individual). It follows that as t becomes large, it becomes increasingly certain that none of the species at time t in $\mathcal{T}_t$

will contain a given feature f present at t_0 if $\lambda \le \mu + \nu$, whereas if $\lambda > \mu + \nu$, there is a positive limiting probability that f will be present in the extant leaves of $\mathcal{T}_t$ (even after pruning under the FOB model).

Let $\mathcal{F}_t$ (respectively $\mathcal{F}_t^s$) be the random variable that counts the number of features present among the leaves of $\mathcal{T}_t$ before (respectively, after) a FOB extinction event at the present (with a survival probability per species of s). Thus, $\mathbb{E}[\mathcal{F}_t] = \mathbb{E}[\mathcal{F}_t^1]$, and we have:

$$\mathbb{E}[\mathcal{F}_t^s] = \int_0^t re^{(\lambda - \mu)\tau} R_{t-\tau}^s(\lambda, \mu + \nu)d\tau + F_0 R_t^s(\lambda, \mu + \nu), \tag{4.2}$$

where F_0 is the number of features initially present at time 0 (i.e. at the top of the stem edge).

An intuitive (but not rigorous) justification of equation (4.2) is as follows. The expected number of tree lineages at time τ is $e^{(\lambda - \mu)\tau}$, so the expected rate at which new features arise at time τ is r times this function. We then multiply this expected rate by the probability that a given feature at time τ survives to the present and also avoids the FOB extinction (i.e. we multiply by $R_{t-\tau}^s(\lambda, \mu + \nu)$) and integrate this modified rate from 0 to t to obtain the first term in equation (4.2) (the second term handles features present at the top of the stem edge and is asymptotically negligible compared with the first term as $t \to \infty$). A more precise proof is given in [47]. Next, let

$$\varphi_{FD}(s) = \lim_{t \to \infty} \frac{\mathbb{E}[\mathcal{F}_t^s]}{\mathbb{E}[\mathcal{F}_t]}.$$

It can be shown that for $\lambda > \mu$ (and conditional on the non-extinction of $\mathcal{T}_t$), we have the following convergence in probability:

$$\frac{\mathcal{F}_t^s}{\mathcal{F}_t} \xrightarrow[t \to \infty]{P} \varphi_{FD}(s)$$

(for details, see [47], Theorem 5.1). Thus, $\varphi_{FD}(s)$ estimates the proportion of FD among the extant species of a large birth–death tree (just prior to the FOB extinction) that survives this extinction event.

It is easily verified that $\varphi_{FD}(s)$ is an increasing concave function that satisfies $1 \ge \varphi_{FD}(s) \ge s$ for all s. This is similar to the earlier story for PD (indeed when $\nu = 0$, $\varphi_{FD} = \varphi_{PD}$). However, by contrast to $\varphi_{PD}(s)$, there does not seem to be a simple explicit expression for $\varphi_{FD}(s)$ when $\nu > 0$. Nevertheless, $\varphi_{FD}(s)$ can be expressed as a ratio of two integrals as follows. We will assume that $\lambda > \mu$ (as we did for PD) and we let $\rho = (\mu + \nu)/\lambda$. From [47] we have:

$$\varphi_{FD}(s) = s\frac{I(s)}{I(1)}, \quad \text{where } I(s) = \int_0^\infty \frac{e^{-(\lambda - \mu)\tau}}{g_s(\tau)}d\tau, \tag{4.3}$$

$$\text{and } g_s(\tau) = \begin{cases} (1 - s - \rho)e^{-\lambda(1-\rho)\tau} + s, & \text{if } \rho \ne 1; \\ 1 + \lambda s\tau, & \text{if } \rho = 1. \end{cases}$$

Notice that φ_{FD} does not depend on the rate (r) at which features arise (indeed it involves just three parameters rather than the original five). It can also be checked that when ν is set to 0, these formulae for φ_{FD} reduce to the more explicit earlier formulae for the analogous PD ratios (equations (3.3) and (3.4)).

(c) The relationship between φ_{FD} and φ_{PD}

In §4, Propositions 3 and 4 showed that on a fixed tree, the FD function (either combinatorial or its expectation under a stochastic model) does not generally agree with the PD function, even if we are free to chose edge lengths for the PD calculation. Here we consider a stochastic analogue for birth–death trees, by asking how $\varphi_{FD}(s)$ compares with $\varphi_{PD}(s)$.

As noted earlier, $\varphi_{FD}(s)$ and $\varphi_{PD}(s)$ are increasing concave functions of s, greater or equal to s, and these two functions coincide when $\nu = 0$. At the other extreme, where $\nu \to \infty$, it is easily shown that $\varphi_{FD}(s)$ converges to s (this is because the only features that will be present at the leaves are those that arise near the end of pendant edges).

Numerical evaluations of $\varphi_{FD}(s)$ using equation (4.3) suggest that for $s \ne 0,1$ the value of $\varphi_{FD}(s)$ is strictly monotone decreasing with ν if the other parameters are fixed. This monotonicity has been formally established for the critical case where $\lambda = \mu + \nu$ (i.e. $\rho = 1$), since in that case the expression for $\varphi_{FD}(s)$ (described above) becomes more tractable [47]. If monotonicity holds generally, this would imply that the following inequality also holds (since $\varphi_{FD} = \varphi_{PD}$ when $\nu = 0$):

$$\varphi_{FD}(s) < \varphi_{PD}(s), \quad \text{for all } \lambda > \mu, s \ne 0,1 \text{ and } \nu > 0. \tag{4.4}$$

What can easily be established is that a variant of inequality (4.4) holds, even on fixed trees (rather than just for birth–death trees). In this variation the temporal length of each edge (in the calculation of *PD*) is replaced by the expected number of features that arise on the edge and that are also present in at least one species at the present (prior to the FOB model). More precisely, suppose that T is a phylogenetic X-tree T, with edge lengths $\ell = [\ell_e]$. To allow features to be present at the root of this tree, we add an additional stem edge above the root, and assume that no features are present at the start of this stem edge. For an edge e of T of length ℓ, let ℓ'_e be the expected number of features that arise on edge e and, in addition, are present in at least one of the leaves of T, and let $\mathcal{S}$ be the set of species present after the FOB extinction event. The following inequality then holds (a proof is provided in Appendix A.3).

Proposition 5. For all choice of (T, ℓ), $s \in [0,1]$ and $r, \nu > 0$ the following inequality:

$$\frac{\mathbb{E}[FD_{(T,\ell)}(\mathcal{S})]}{\mathbb{E}[FD_{(T,\ell)}(X)]} \le \frac{\mathbb{E}[PD_{(T,\ell')}(\mathcal{S})]}{PD_{(T,\ell')}(X)},$$

holds for all values of $s \in [0,1]$, *and all choices of* (T, ℓ), *and all* $\nu \ge 0$. *Moreover, the inequality is strict unless* $s \in \{0,1\}$ *or* T *is a star tree.*

(d) Concluding comments

If inequality (4.4) is true generally, it would provide a relevant message for biodiversity theory. Namely, using PD as a proxy to measure FD will, in expectation, provide a systematic underestimate of the true proportion of feature diversity that is expected to be lost in an extinction event in the near future, at least for the simple models considered in this paper. It would therefore be interesting to investigate the degree to which such underestimation holds on real phylogenetic trees (perhaps within the framework of EvoHeritage [45]). A mathematical proof (or disproof) of inequality (4.4) would also be helpful.

The results reviewed in this paper also suggest further questions and lines of inquiry. For example, how does the distribution of PD and FD change under phylogenetic models that generate trees that are less balanced (and so more typical of real phylogenies) than the YH model predicts? Also, how do the results concerning FD generalize to models of feature evolution that allow a feature to arise more than once on a given tree? Finally, what can be said about the predicted relative loss of PD and FD under more realistic models of extinction at the present that recognize dependencies between species (e.g. extinction cascades)?

Appendix

A.1. Proof of Proposition 1

Proof: We have $D'_1 = 0$ (with probability 1). For $n > 1$, a leaf x chosen uniformly at random in a reduced evolutionary tree with n leaves, was either (i) present when this reduced tree had $n-1$ species (with probability $1 - \frac{1}{n}$) or it was the last new species to form in this tree (with probability $\frac{1}{n}$). Thus,

$$D'_n - D'_{n-1} = \begin{cases} 1, & \text{with probability } \frac{1}{n}; \\ 0, & \text{with probability } 1 - \frac{1}{n}. \end{cases}$$

Now, the neutrality assumption on the process generating the evolutionary tree implies that $(D'_n, n \ge 1)$ is a Markov chain. Thus we can write,

$$1 + D'_n = Y_1 + Y_2 + \cdots + Y_n,$$

where the Y_i are independent Bernoulli variables, with $Y_i = 1$ with probability $\frac{1}{i}$ and $Y_i = 0$ otherwise. Equation (2.2) now follows from standard results concerning cycles in random permutations (see e.g. §3.1 of [48]), and the remaining claims follow by the linearity of expectation (and asymptotics) and Poisson approximation theory (see e.g. [49] or [50]). ∎

A.2. Proof of equation (3.7)

From equation (3.5), with $\epsilon_x = \epsilon$ for all $x \in X$, we have:

$$\mathbb{E}[\psi_x(\epsilon)] = \int_0^1 \sum_{e \in P_T(x)} \ell_e \epsilon^{|C_T(e)|-1} d\epsilon = \sum_{e \in P_T(x)} \ell_e \int_0^1 \epsilon^{|C_T(e)|-1} d\epsilon = \sum_{e \in P_T(x)} \ell_e \left[\frac{\epsilon^{|C_T(e)|}}{|C_T(e)|} \right]_0^1 = FP_{(T,\ell)}(x).$$

The penultimate equality is from $\int t^{k-1} dt = t^k/k$, and the last equality is from $\left[\frac{\epsilon^{|C_T(e)|}}{|C_T(e)|} \right]_0^1 = \frac{1}{|C_T(e)|} = \frac{1}{n_e}$, together with the definition of $FP_{(T,\ell)}(x)$. ∎

A.3. Proof of Proposition 5

Let $\widetilde{\ell}_e = \frac{r}{\nu}(1 - e^{-\nu \ell_e})$, and let $\phi_e^T(k)$ be the probability that a given feature present at the end of edge e is present in exactly k leaves of T. Then ℓ'_e (defined just before Proposition 5) equals $\widetilde{\ell}_e \cdot (1 - \phi_e^T(0))$, and:

$$\mathbb{E}[FD_{(T,\ell)}(\mathcal{S})] = \sum_e \widetilde{\ell}_e \cdot \left(\sum_{k=1}^{n_e} \phi_e^T(k)(1 - \epsilon^k) \right),$$

where $\epsilon = (1 - s)$ is the FOB extinction probability. Now, since $k \leq n_e$ we have: $\mathbb{E}[FD_{(T,\ell)}(\mathcal{S})] \leq \sum_e \widetilde{\ell}_e \cdot \left(\sum_{k=1}^{n_e} \phi_e^T(k) \right)(1 - \epsilon^{n_e})$, and so

$$\frac{\mathbb{E}[FD_{(T,\ell)}(\mathcal{S})]}{\mathbb{E}[FD_{(T,\ell)}(X)]} \leq \frac{\sum_e \widetilde{\ell}_e \cdot \left(\sum_{k=1}^{n_e} \phi_e^T(k) \right)(1 - \epsilon^{n_e})}{\sum_e \widetilde{\ell}_e \cdot \left(\sum_{k=1}^{n_e} \phi_e^T(k) \right)}. \tag{A 1}$$

In addition, since $\sum_{k=1}^{n_e} \phi_e^T(k) = 1 - \phi_e^T(0)$, and $\ell_e' = \widetilde{\ell}_e \cdot (1 - \phi_e^T(0))$ (by definition), Inequality (A 1) can be rewritten as

$$\frac{\mathbb{E}[FD_{(T,\ell)}(\mathcal{S})]}{\mathbb{E}[FD_{(T,\ell)}(X)]} \leq \frac{\sum_e \ell_e'(1 - \epsilon^{n_e})}{\sum_e \ell_e'} = \frac{\mathbb{E}[PD_{(T,\ell')}(\mathcal{S})]}{PD_{(T,\ell')}(X)},$$

which establishes the claimed inequality. Moreover, the inequality is strict when $s \neq 0,1$, $\nu > 0$, and T is not a star tree because in that case $\sum_{k=1}^{n_e} \phi_e^T(k)\epsilon^k > \sum_{k=1}^{n_e} \phi_e^T(k)\epsilon^{n_e}$ for any edge e that is not a pendant edge of T. ∎

References

1. Yule GU. 1925 A mathematical theory of evolution, based on the conclusions of Dr. J. C. Willis, F. R. S. *Phil. Trans. R. Soc. Lond. B* **213**, 21–87. (doi:10.1098/rstb.1925.0002)

2. Kubatko L, Knowles LL. 2023 *Species tree inference: a guide to methods and applications*. Princeton, NJ: Princeton University Press. (doi:10.2307/j.ctv2wr4wdf)

3. Roch S, Snir S. 2013 Recovering the Treelike Trend of Evolution Despite Extensive Lateral Genetic Transfer: A Probabilistic Analysis. *J. Comput. Biol.* **20**, 93–112. (doi:10.1089/cmb.2012.0234)

4. Aldous D. 1996 Probability Distributions on Cladograms. In *Random discrete structures the IMA volumes in mathematics and its applications*, pp. 1–18, vol. **76**. New York, NY: Springer. (doi:10.1007/978-1-4612-0719-1_1)

5. Lambert A, Stadler T. 2013 Birth–death models and coalescent point processes: The shape and probability of reconstructed phylogenies. *Theor. Popul. Biol.* **90**, 113–128. (doi:10.1016/j.tpb.2013.10.002)

6. Feller W. 1939 Die Grundlagen der Volterraschen Theorie des Kampfes ums Dasein in wahrscheinlichkeitstheoretischer Behandlung. *Acta Biotheor.* **5**, 11–40. (doi:10.1007/bf01602932)

7. Kendall DG. 1948 On the Generalized 'Birth-and-Death' Process. *Ann. Math. Stat.* **19**, 1–15. (doi:10.1214/aoms/1177730285)

8. Jagers P. 1992 Stabilities and instabilities in population dynamics. *J. Appl. Probab.* **29**, 770–780. (doi:10.1017/s0021900200043667)

9. Harding EF. 1971 The probabilities of rooted tree-shapes generated by random bifurcation. *Adv. Appl. Probab.* **3**, 44–77. (doi:10.1017/S0001867800037587)

10. Etherington IMH. 1940 XV.—On Non-Associative Combinations. *Proc. R. Soc. Edinb.* **59**, 153–162. (doi:10.1017/S0370164600012244)

11. Hammersley JM, Grimmett GR. 1974 Maximal solutions of the generalized subadditive inequality. In *Stochastic geometry (a tribute to the memory of Rollo Davidson)*, pp. 270–284. London, UK: Wiley.

12. Mahmoud HM. 1992 *Evolution of random search trees*. New York, NY: Wiley.

13. Lynch WC. 1965 More Combinatorial Properties of Certain Trees. *Comput. J.* **7**, 299–302. (doi:10.1093/comjnl/7.4.299)

14. Fischer M, Herbst L, Kersting SJ, Kühn AL, Wicke K. 2023 Extension of tree balance and imbalance indices and alternative ways of measuring tree balance. In *Tree balance indices, a comprehensive survey*, pp. 341–344. Cham, Switzerland: Springer International Publishing. (doi:10.1007/978-3-031-39800-1_25)

15. Aldous DJ. 2001 Stochastic models and descriptive statistics for phylogenetic trees, from Yule to today. *Stat. Sci.* **16**, 23–34. (doi:10.1214/ss/998929474)

16. Steel M, McKenzie A. 2001 Properties of phylogenetic trees generated by Yule-type speciation models. *Math. Biosci.* **170**, 91–112. (doi:10.1016/s0025-5564(00)00061-4)

17. Sanderson MJ. 1996 How Many Taxa Must be Sampled to Identify the Root Node of a Large Clade? *Syst. Biol.* **45**, 168–173. (doi:10.2307/2413613)

18. Bienvenu F, Steel M. 2024 0–1 Laws for Pattern Occurrences in Phylogenetic Trees and Networks. *Bull. Math. Biol.* **86**, 94. (doi:10.1007/s11538-024-01316-x)

19. Mooers A, Gascuel O, Stadler T, Li H, Steel M. 2012 Branch Lengths on Birth–Death Trees and the Expected Loss of Phylogenetic Diversity. *Syst. Biol.* **61**, 195–203. (doi:10.1093/sysbio/syr090)

20. Stadler T, Steel M. 2012 Distribution of branch lengths and phylogenetic diversity under homogeneous speciation models. *J. Theor. Biol.* **297**, 33–40. (doi:10.1016/j.jtbi.2011.11.019)

21. Dieselhorst T, Berg J. 2024 The distribution of interior branch lengths in phylogenetic trees under constant-rate birth-death models. *arXiv* https://arxiv.org/html/2407.13403v1

22. Bocharov S, Harris S, Kominek E, Mooers AØ, Steel M. 2023 Predicting Long Pendant Edges in Model Phylogenies, with Applications to Biodiversity and Tree Inference. *Syst. Biol.* **72**, 575–589. (doi:10.1093/sysbio/syac059)

23. Kominek 2023 E. Prioritizing evolutionarily isolated insects for zoo-based conservation. MSc Thesis, Simon Fraser University.

24. Disanto F, Fuchs M. 2023 Distribution of External Branch Lengths in Yule Histories. *Electron. J. Comb* **30**, 23. (doi:10.37236/11438)

25. Faith DP. 1992 Conservation evaluation and phylogenetic diversity. *Biol. Conserv.* **61**, 1–10. (doi:10.1016/0006-3207(92)91201-3)

26. Redding DW, Hartmann K, Mimoto A, Bokal D, Devos M, Mooers AØ. 2008 Evolutionarily distinctive species often capture more phylogenetic diversity than expected. *J. Theor. Biol.* **251**, 606–615. (doi:10.1016/j.jtbi.2007.12.006)

27. Fuchs M, Jin EY. 2015 Equality of Shapley value and fair proportion index in phylogenetic trees. *J. Math. Biol.* **71**, 1133–1147. (doi:10.1007/s00285-014-0853-0)

28. Bordewich M, Semple C. 2024 Quantifying the difference between phylogenetic diversity and diversity indices. *J. Math. Biol.* **88**, 40. (doi:10.1007/s00285-024-02059-y)

29. Fischer M, Francis A, Wicke K. 2023 Phylogenetic Diversity Rankings in the Face of Extinctions: The Robustness of the Fair Proportion Index. *Syst. Biol.* **72**, 606–615. (doi:10.1093/sysbio/syac030)

30. Manson K. 2024 The robustness of phylogenetic diversity indices to extinctions. *J. Math. Biol.* **89**, 5. (doi:10.1007/s00285-024-02098-5)

31. Wicke K, Steel M. 2020 Combinatorial properties of phylogenetic diversity indices. *J. Math. Biol.* **80**, 687–715. (doi:10.1007/s00285-019-01438-0)

32. Abhari N, Colijn C, Mooers A, Tupper P. 2024 Capturing diversity: Split systems and circular approximations for conservation. *J. Theor. Biol.* **578**, 111689. (doi:10.1016/j.jtbi.2023.111689)

33. Moulton V, Spillner A, Wicke K. 2024 Phylogenetic Diversity Indices from an Affine and Projective Viewpoint. Bull. Math. Biol. **86**, 103. (doi:10.1007/s11538-024-01332-x)

34. Raup D. 1991 Extinction: bad genes or bad luck? *New Sci*. **131**, 46–49.

35. Mooers AØ, Faith DP, Maddison WP. 2008 Converting Endangered Species Categories to Probabilities of Extinction for Phylogenetic Conservation Prioritization. *PLoS One* **3**, e3700. (doi:10.1371/journal.pone.0003700)

36. Gumbs R *et al*. 2023 The EDGE2 protocol: Advancing the prioritisation of Evolutionarily Distinct and Globally Endangered species for practical conservation action. *PLoS Biol*. **21**, e3001991. (doi:10.1371/journal.pbio.3001991)

37. Faller B, Pardi F, Steel M. 2008 Distribution of phylogenetic diversity under random extinction. *J. Theor. Biol*. **251**, 286–296. (doi:10.1016/j.jtbi.2007.11.034)

38. Lambert A, Steel M. 2013 Predicting the loss of phylogenetic diversity under non-stationary diversification models. *J. Theor. Biol*. **337**, 111–124. (doi:10.1016/j.jtbi.2013.08.009)

39. Steel M, Mimoto A, Mooers AØ. 2007 Hedging Our Bets: The Expected Contribution of Species to Future Phylogenetic Diversity. *Evol. Bioinform*. **3**, 117693430700300. (doi:10.1177/117693430700300024)

40. Mazel F *et al*. 2018 Prioritizing phylogenetic diversity captures functional diversity unreliably. *Nat. Commun*. **9**, 2888. (doi:10.1038/s41467-018-05126-3)

41. Kelly S, Grenyer R, Scotland RW. 2014 Phylogenetic trees do not reliably predict feature diversity. *Divers. Distrib*. **20**, 600–612. (doi:10.1111/ddi.12188)

42. Wicke K, Mooers A, Steel M. 2021 Formal Links between Feature Diversity and Phylogenetic Diversity. *Syst. Biol*. **70**, 480–490. (doi:10.1093/sysbio/syaa062)

43. Overwater M, Pelletier D, Steel M. 2023 The expected loss of feature diversity (versus phylogenetic diversity) following rapid extinction at the present. *J. Math. Biol*. **87**, 53. (doi:10.1007/s00285-023-01988-4)

44. Feller W. 1950 *An introduction to probability theory and its applications*. New York, NY: Wiley.

45. Rosindell J, Manson K, Gumbs R, Pearse WD, Steel M. 2024 Phylogenetic Biodiversity Metrics Should Account for Both Accumulation and Attrition of Evolutionary Heritage. *Syst. Biol*. **73**, 158–182. (doi:10.1093/sysbio/syad072)

46. Yang Z, Rannala B. 1997 Bayesian phylogenetic inference using DNA sequences: a Markov Chain Monte Carlo Method. *Mol. Biol. Evol*. **14**, 717–724. (doi:10.1093/oxfordjournals.molbev.a025811)

47. Overwater M. 2023 On the mathematical links between phylogenetic and feature diversity. MSc Thesis, (ed. J Picard), University of Canterbury.

48. Pitman J. 2006 *Combinatorial stochastic processes: école d'été de probabilités de saint-flour xxxii-2002*. Berlin, Heidelberg, Germany: Springer-Verlag.

49. Barbour AD, Holst L, Janson S. 1992 Poisson process approximation. In *Poisson approximation*, pp. 202–252. Oxford, UK: Oxford University Press. (doi:10.1093/oso/9780198522355.003.0010)

50. Arratia R, Tavare S. 1992 The Cycle Structure of Random Permutations. *Ann. Probab* **20**, 1567–1591. (doi:10.1214/aop/1176989707)