The expected loss of feature diversity (versus phylogenetic diversity) following rapid extinction at the present

Marcus Overwater and Mike Steel

October 14, 2022

Biomathematics Research Centre,
School of Mathematics and Statistics,
University of Canterbury, Christchurch, New Zealand

Abstract

The current rapid extinction of species leads not only to their loss but also the disappearance of the unique features they harbour, which have evolved along the branches of the underlying evolutionary tree. One proxy for estimating the feature diversity (FD) of a set S of species at the tips of a tree is 'phylogenetic diversity' (PD): the sum of the branch lengths of the subtree connecting the species in S. For a phylogenetic tree that evolves under a standard birth–death process, and which is then subject to a sudden extinction event at the present (the simple 'field of bullets' model with a survival probability of s per species) the proportion of the original PD that is retained after extinction at the present is known to converge quickly to a particular concave function $\varphi_{PD}(s)$ as t grows. To investigate how the loss of FD mirrors the loss of PD for a birth–death tree, we model FD by assuming that distinct discrete features arise randomly and independently along the branches of the tree at rate r and are lost at a constant rate ν. We derive an exact mathematical expression for the ratio $\varphi_{FD}(s)$ of the two expected feature diversities (prior to and following an extinction event at the present) as t becomes large. We find that although φ_{FD} has a similar behaviour to φ_{PD} (and coincides with it for $\nu = 0$), when $\nu > 0$, $\varphi_{FD}(s)$ is described by a function that is different from $\varphi_{PD}(s)$. We also derive an exact expression for the expected number of features that are present in precisely *one* extant species. Our paper begins by establishing some generic properties of FD in a more general (non-phylogenetic setting) and applies this to fixed trees, before considering the setting of random (birth–death) trees.

Keywords: Feature diversity, phylogenetic diversity, birth–death processes, extinction

1 Introduction

Phylogenetic trees provide a way to quantify biodiversity and the extent to which it might be lost in the face of the current mass extinction event. One such biodiversity measure is *phylogenetic diversity* (PD), which associates to each subset S of extant species, the sum of the branch lengths of the underlying evolutionary tree that connects (just) these species to the root of the tree. This measure, pioneered by Daniel P. Faith [2], provides a more complete measure of biodiversity than merely counting the number of species in S (see e.g. [11]). Moreover, if new features evolve along the branches of a tree and are never lost, then the resulting features present amongst the species in the subset S (the *feature diversity* (FD) of S) are directly correlated with the phylogenetic diversity of S [21]. However, when features are lost, it has recently been shown mathematically that under simple (deterministic or stochastic) models of feature gain and loss, FD necessarily deviates from PD except for very trivial types of phylogenetic trees (Theorems

2 and 3 of [16]). More generally, the question of the extent to which PD captures feature or functional diversity has been the subject of considerable debate in the biological literature (see [1, 7, 8, 9, 14, 19, 18]).

In this paper, we investigate a related question: under a standard phylogenetic diversification model and a simple stochastic process of feature gain and loss, what is the proportion of expected feature diversity that will be lost in a mass extinction event at the present? And how does this ratio compare with the expected phylogenetic diversity that will be lost? Although the latter ratio (for PD) has been determined in earlier work, here we provide a corresponding result for FD, and show how it differs from PD when the rate of feature loss is non-zero. Our results suggest that the relative loss of FD under a mass extinction event at the present is greater than the relative loss of PD. We also investigate the number of features that are expected to be found in just one species at the present.

We begin by considering the properties of FD in a more general setting based purely on the species themselves (i.e. not involving any underlying phylogenetic tree or network) and then consider FD on fixed phylogenetic trees before presenting the results for (random) birth–death trees. Some of the results of these earlier sections are applied in the later sections.

2 General Properties of Feature Diversity without reference to phylogenies

This section considers the generic properties of expected FD for sets of species, and thus, no underlying phylogenetic tree or stochastic process that generates a tree, or model of feature evolution is assumed. We mostly follow the notation from [21].

Definitions

Let X be a labelled set of species and F_x be a non-empty set of discrete features (e.g. traits) associated with each species $x \in X$.

- The collection of ordered pairs $\mathbb{F} = \{(x, F_x) : x \in X\}$ is called a *feature assignment* on X.
- For any subset A of X, let $\mathcal{F}(A) = \bigcup_{x \in A} F_x$ and let $\mu : \mathcal{F}(X) \to \mathbb{R}^{>0}$ be a function assigning some richness or novelty to a feature $f \in \mathcal{F}(X)$.
- The *feature diversity* of some subset A of a set of species X is defined as,

$$FD(A) = \sum_{f \in \mathcal{F}(A)} \mu(f). \tag{2.1}$$

A default option for μ is to set $\mu(f) = 1$, which simply counts the number of features present. Notice that for any subset A of X we have $FD(A) \leq \sum_{x \in A} FD(\{x\}) = \sum_{x \in A} \sum_{f \in F_x} \mu(f)$.

2.1 Feature diversity loss under a 'Field of Bullets' model of extinction at the present

Definition 2.1. Consider a sudden extinction event taking place across a set of species ([15]). In the *generalised field of bullets* (g-FOB) model, each species $x \in X$ either survives the extinction event (with probability s_x) or disappears (with probability $1 - s_x$), and these survival events are assumed to be independent among the species. We write $\mathbf{s}(X)$ (or, more briefly, $\mathbf{s}$) to denote the vector $(s_x : x \in X)$ and we let $\mathcal{X}$ denote the (random) subset of X corresponding to the species that survive the extinction event. If $s_x = s$ for all $x \in X$, then we have the simpler *field of bullets* (FOB) model.

We define the following quantity:

$$\varphi_{(\mathbb{F},\mathbf{s})} = \frac{FD(\mathcal{X})}{FD(X)},$$

which is the proportion of feature diversity that survives the extinction event. In the case of the FOB model, we denote this ratio by $\varphi_{(\mathbb{F},s)}$.

Proposition 2.1. For the g-FOB model,

$$\mathbb{E}[\varphi_{(\mathbb{F},\mathbf{s})}] = \sum_{f\in\mathcal{F}(X)} \tilde{\mu}(f) \left(1 - \prod_{x:f\in F_x} (1-s_x)\right),$$

where $\tilde{\mu}(f) = \frac{\mu(f)}{\sum_{f\in\mathcal{F}(X)}\mu(f)}$ are the normalised μ values (which sum to 1).

For the FOB model, the equation in Proposition 2.1 simplifies to

$$\mathbb{E}[\varphi_{(\mathbb{F},s)}] = \sum_{f\in\mathcal{F}(X)} \tilde{\mu}(f)\left(1-(1-s)^{n_f}\right), \tag{2.2}$$

where $n_f = |\{f : \exists x : f \in F_x\}|$, the number of species in X that possess feature f.

Proof. We have $\varphi_{(\mathbb{F},\mathbf{s})} = \sum_{f\in\mathcal{F}(X)} \tilde{\mu}(f)\cdot\mathbb{I}_f$ where $\mathbb{I}_f$ is the Bernoulli variable that takes the value 1 if at least one species in X with feature f survives the g-FOB extinction event, and 0 otherwise. Applying linearity of expectation and noting that $\mathbb{P}(\mathbb{I}_f = 1) = 1-(1-s)^{n_f}$ gives the result. □

Notice that $\mathbb{E}[\varphi_{(\mathbb{F},s)}] = s$ at $s = 0, 1$. The behaviour between these two extreme values of s is described next.

Proposition 2.2. Under the FOB model, the following hold:

(i) $\mathbb{E}[\varphi_{(\mathbb{F},s)}] \geq s$ for all $s \in [0,1]$.

(ii) $\mathbb{E}[\varphi_{(\mathbb{F},s)}] = s$ for a value of $s \in (0,1)$ if and only if the sets F_x in $\mathbb{F}$ are pairwise disjoint, in which case $\mathbb{E}[\varphi_{(\mathbb{F},s)}] = s$ for all $s \in [0,1]$.

(iii) If the sets F_x are not pairwise disjoint, then $\mathbb{E}[\varphi_{(\mathbb{F},s)}]$ is a strictly concave increasing function of s. In particular, for all $s \in (0,1)$ we have $\mathbb{E}[\varphi_{(\mathbb{F},s)}] > s$.

Proof. Part (i): Since $n_f \geq 1$ in Eqn. (2.2) and $1-(1-s)^{n_f} \geq s$ for all $n_f \geq 1$, the claimed inequality is immediate. Part (ii): $n_f = 1$ if and only if $1-(1-s)^{n_f} = s$ for all $s \in [0,1]$, and if $n_f > 1$ then $1-(1-s)^{n_f} > s$ for every $s \in (0,1)$. Part (iii): By Eqn. (2.2),

$$\frac{d}{ds}\mathbb{E}[\varphi_{(\mathbb{F},s)}] = \sum_{f\in\mathcal{F}(X)} \tilde{\mu}(f) n_f (1-s)^{n_f-1} \geq 0,$$

and

$$\frac{d^2}{ds^2}\mathbb{E}[\varphi_{(\mathbb{F},s)}] = -\sum_{f\in\mathcal{F}(X):n_f\geq 2} \tilde{\mu}(f) n_f (n_f-1)(1-s)^{n_f-2} < 0,$$

from which the claimed results follow. □

2.2 Approximating $\varphi_{(\mathbb{F},s)}$ by its expected value

For each $n \geq 1$, let X_n be a labelled set of n species with feature assignment $\mathbb{F}_n$. Let $\mathcal{X}_n$ denote the (random) set of species after a g-FOB extinction event with survival probability vector $\mathbf{s}(X_n)$, which assigns each $x \in X_n$ a corresponding survival probability $s_n(x)$. Note that we make no assumption regarding how the species in X_n and X_m are related (e.g. they may be disjoint, overlapping or nested), or any apriori relationship between $\mathbf{s}(X_n)$ and $\mathbf{s}(X_m)$.

In the following result, we provide a sufficient condition under which $\varphi_{(\mathbb{F}_n,\mathbf{s}_n)}$ is likely to be close to its expected value $\mathbb{E}[\varphi_{(\mathbb{F}_n,\mathbf{s}_n)}]$ (readily computed via Eqn. (2.1)) when the number of species n is large. This condition allows some species to contribute proportionately more FD than other species do on average, and this proportion can grow with n, provided that it does not grow too quickly.

Let $R_n = \sum_{x \in X_n} \left[\frac{FD(\{x\})}{FD(X_n)}\right]^2$ and let $\text{av}FD(X_n) = FD(X_n)/n$, which is the average contribution of each species to the total FD.

Proposition 2.3.

(a)
$$\mathbb{P}\left(\left|\varphi_{(\mathbb{F}_n,\mathbf{s}_n)} - \mathbb{E}[\varphi_{(\mathbb{F}_n,\mathbf{s}_n)}]\right| > \epsilon\right) \leq 2\exp\left(-\frac{2\epsilon^2}{R_n}\right).$$

(b) If $R_n \to 0$ as $n \to \infty$ then $\varphi_{(\mathbb{F}_n,\mathbf{s}_n)} - \mathbb{E}[\varphi_{(\mathbb{F}_n,\mathbf{s}_n)}] \xrightarrow{P} 0$.

(c) Suppose that for each $x \in X$, $FD(\{x\})/\text{av}FD(X_n) \leq B_n$, where and $B_n^2 = o(n)$ (e.g. $B_n = \sqrt[3]{n}$). We then have the following convergence in probability as $n \to \infty$:
$$\varphi_{(\mathbb{F}_n,\mathbf{s}_n)} - \mathbb{E}[\varphi_{(\mathbb{F}_n,\mathbf{s}_n)}] \xrightarrow{P} 0.$$

Proof. Let $\mathbf{Y_n} = \{Y_i : i \in [n]\}$ be a sequence of Bernoulli random variables where each Y_i takes the value of 1 if species x_i survives and 0 otherwise. For the g-FOB model, the random variables Y_i are independent. We can write $\varphi_{(\mathbb{F}_n,\mathbf{s}_n)} = h(\mathbf{Y_n})$ where $h(y_1, \ldots, y_n)$ is the ratio $\frac{FD(\{x_i \in X_n : y_i = 1\})}{FD(X_n)}$. We have:
$$|h(y_1, ..., y_i, ..., y_n) - h(y_1, ..., y_i', ..., y_n)| \leq \frac{FD(\{x_i\})}{FD(X_n)}.$$

We now apply McDiarmid's inequality [10] to obtain (for each $\epsilon > 0$):
$$\mathbb{P}(\left|\varphi_{(\mathbb{F}_n,\mathbf{s}_n)} - \mathbb{E}[\varphi_{(\mathbb{F}_n,\mathbf{s}_n)}]\right| \geq \epsilon) \leq 2\exp\left(\frac{-2\epsilon^2}{R_n}\right), \tag{2.3}$$

This establishes Part (a). Part (b) now follows immediately, and Part (c) follows from Part (b), since the condition in Part (c) implies that
$$R_n = \sum_x \left[\frac{FD(\{x\})}{n \cdot \text{av}FD(X_n)}\right]^2 \leq \sum_x \left(\frac{B_n}{n}\right)^2 = \sum_x \frac{B_n^2}{n^2} = B_n^2/n \to 0,$$
as $n \to \infty$.

□

Remark:. Note that Proposition 2.3(c) can fail when the condition stated in Part (c) does not hold, even for the simpler FOB model. We provide a simple example to demonstrate this.

Let $X_n = \{1, \ldots, n\}$ and let $F_i = \{f\}$ for $i = 1, \ldots, n-1$ and $F_n = \{g\}$ where f and g are distinct features with $\mu(f) = \mu(g) = 1$. In this case:

$$\varphi_{(\mathbb{F},\mathbf{s})} = \begin{cases} 0, & \text{w.p. } (1-s)^n; \\ \frac{1}{2}, & \text{w.p. } (1-s)^{n-1}s + (1-(1-s)^{n-1})(1-s); \\ 1, & \text{w.p. } (1-(1-s)^{n-1})s. \end{cases}$$

Therefore, $\varphi_{(\mathbb{F}_n,\mathbf{s}_n)} - \mathbb{E}[\varphi_{(\mathbb{F}_n,\mathbf{s}_n)}]$ does not converge in probability to 0 (or to any constant) as $n \to \infty$.

2.3 Consequences for Phylogenetic Diversity

Proposition 2.2 and Proposition 2.3 provide a simple way to derive certain results concerning phylogenetic diversity - both on rooted trees and also for rooted phylogenetic networks (specifically for the subNet diversity measure described in [20]). To each edge e of a rooted phylogenetic tree (or network), associate some unique feature f_e and give it the value $\mu(f_e) = \ell(e)$, where $\ell(e)$ is the length of edge e in the tree (or network). For any subset Y of X (the leaf set of T) we then have:

$$PD(Y) = FD(Y).$$

It follows that for the simple field of bullets models, PD satisfies the concavity properties described in Proposition 2.2, where the condition that the sets F_x are pairwise disjoint corresponds to the tree being a star tree. These results were established for rooted trees by specific tree-based arguments (see e.g. Section 5 of [6]), but they directly follow from the more general framework above, and extend beyond trees.

Using this same link between PD and FD, Proposition 2.3 provides a further application to any sequence of rooted phylogenetic trees T_n with n leaves and ultrametric edge lengths. For example, the ratio of surviving PD to original PD under the FOB model converges in probability to the expected value of this ratio for a sequence of trees T_n if the total PD of T_n grows at least as fast as nL/n^β, where L is the height of the tree and $0 < \beta < 1/2$. This condition holds, for example, for Yule trees [17].

3 The feature diversity ratio φ for a model of feature evolution on a phylogenetic tree

Consider a rooted binary phylogenetic tree T_n, in which each edge has a positive length that corresponds to a temporal duration, with the root ρ of T_n being placed at the top of a stem edge at time 0, and with each leaf in the leaf set $X_n = \{x_1, \ldots, x_n\}$ of T_n being placed at time t (as in Fig. 3.1). For convenience, we will assume in this section that $\mu(f) = 1$ for all features; however, this assumption can be relaxed (e.g. by allowing $\mu(f)$ to take values in a fixed interval $[a, b]$ where $a > 0$ according to some fixed distribution, and independently between features) without altering the results significantly.

On T_n, we apply a stochastic process in which (discrete) features arise independently along the branches of this tree at rate r, and each feature that arises is novel (i.e. it has not appeared earlier elsewhere in the tree). Once a feature arises, it is then carried forward in time along the branches of T_n (and is passed on to the two lineages arising at any speciation event). In addition, any feature can be lost from a lineage at any point according to a continuous-time pure-death process that operates at rate ν. This model was recently discussed in [16]. Under this process, each leaf x of T_n will have a (possibly empty) set of features (F_x). Fig. 3.1 illustrates the processes described. Note that $|F_x|$ (for any $x \in X_n$) and $FD(X_n)$ are now random variables.

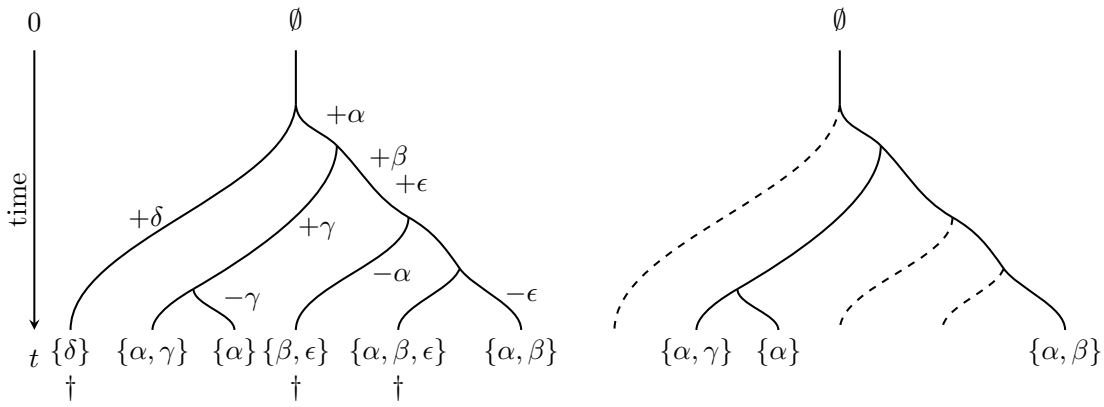

Figure 3.1 An example of feature gain and loss on a tree and the impact of extinction at the present. *Left:* New features arise (indicated by $+$) and disappear (indicated by $-$) along the branches of the tree. To simplify this example, no features are present at time 0; however, our results do not require this restriction. In total, there are five features present amongst the leaves of the tree at time t, namely $\{\alpha, \beta, \gamma, \delta, \epsilon\}$. *Right:* An extinction event at the present (denoted by †) results in the loss of three of the extant species, leaving just three features being present among the leaves of the resulting pruned tree, namely $\{\alpha, \beta, \gamma\}$. Thus, the ratio of surviving features to total features is $3/5 = 0.6$.

Let N_ℓ denote the number of features at the end of any path P in T_n that starts at time $t = 0$ and ends at time $t = \ell$. Then N_ℓ is described by a continuous-time Markov process that has a constant birth rate and a linear death rate. It is then a classical result [3] that N_ℓ has expected value $\frac{r}{\nu}(1 - e^{-\nu\ell}) + F_0 e^{-\nu t}$, where $F_0 = |F_\rho|$ is the number of features present at time $t = 0$ (i.e. at the top of the stem edge), and N_ℓ converges to a Poisson distribution with mean $\frac{r}{\nu}$ as ℓ grows. Moreover, if $N_0 = 0$ then N_ℓ has a Poisson distribution for any value of $\ell > 0$ ([3] p. 461), and so, regardless of the value of F_ρ, the random variable $|F_x \setminus F_\rho|$ has a Poisson distribution with mean $\frac{r}{\nu}(1 - e^{-\nu t})$. The (random) number of features at any leaf x of T_n (i.e. $|F_x|$) has the same distribution as N_t. Note that for distinct leaves x and y of T_n, the random variables $|F_x|$ and $|F_y|$ are not independent.

Notational convention: In this section, we will write $FD(T_n)$ in place of $FD(X_n)$ and we will also write $FD(T_n^s)$ in place of $FD(\mathcal{X}_n)$ when $\mathcal{X}_n$ is the subset of the set of leaves (X_n) of T_n that survive under a FOB model with a survival probability s. We will let F_ρ denote the set of features present at the root vertex ρ at the top of the stem edge.

The main result of this section is Theorem 3.1, and its proof relies on establishing a sequence of preliminary lemmas.

Lemma 3.1.

$$\mathbb{E}[FD(T_n)] \leq \frac{r}{\nu} \cdot n + |F_\rho|.$$

Proof. Let ℓ_x denote the length of the path from the end of the stem edge to the leaf x (these will all be the same if the edge lengths are ultrametric). In this case:

$$FD(T_n) = \left| \bigcup_{x \in X_n} F_x \right| \leq \sum_{x \in X_n} |F_x \setminus F_\rho| + |F_\rho|.$$

Now, for each $x \in X$, we have $\mathbb{E}[|F_x \setminus F_\rho|] = \frac{r}{\nu}(1 - e^{-\nu \ell_x})$ and thus:

$$\mathbb{E}[FD(T_n)] \leq \frac{r}{\nu} \cdot n + |F_\rho|.$$

□

Lemma 3.2. Let $\beta > 0$ be a fixed constant. Given a sequence T_n of trees with leaf set X_n, let $\mathcal{E}_n$ be the event that $|F_x| \leq n^\beta$ for every $x \in X_n$. Then $\lim_{n\to\infty} \mathbb{P}(\mathcal{E}_n) = 1$.

Proof. We combine the Bonferroni inequality with a standard right-tail probability bound for a Poisson variable. Firstly, observe that:

$$\mathbb{P}(\mathcal{E}_n) \geq 1 - \sum_{x\in X_n} \mathbb{P}(|F_x| > n^\beta) = 1 - n\mathbb{P}(|F_{x_1}| > n^\beta). \tag{3.1}$$

Here $|F_{x_1}|$ has a Poisson distribution with mean $m \leq r/\nu$, and so the Chernoff bound on the right hand tail of this Poisson variable ([12], p. 97) gives $\mathbb{P}(|F_{x_1}| > n^\beta) \leq \left(\frac{em}{n^\beta}\right)^{n^\beta} e^{-m}$, so $n\mathbb{P}(|F_{x_1}| > n^\beta) \to 0$ as n grows. Applying this to the inequality in (3.1) establishes the result.

□

The proof of the following result is given in the Appendix.

Lemma 3.3. Let $(T_n, n \geq 1)$ be a sequence of rooted binary trees, with T_n having leaf set X_n. Then for a constant C (independent of n) we have:

$$\mathrm{Var}[FD(T_n)] \leq Cn\sqrt{n}. \tag{3.2}$$

Lemma 3.4. Let $(T_n, n \geq 1)$ be a sequence of rooted binary trees, with T_n having leaf set X_n, and suppose that $\mathbb{E}[FD(T_n)] \geq cn$ for some constant $c > 0$. Then $\frac{FD(T_n)}{\mathbb{E}[FD(T_n)]}$ converges in probability to 1 as $n \to \infty$.

Proof. By Lemma 3.3 we have

$$\mathrm{Var}\left[\frac{FD(T_n)}{n}\right] \to 0. \tag{3.3}$$

By Chebyshev's inequality:

$$\mathbb{P}\left(\left|\frac{FD(T_n)}{\mathbb{E}[FD(T_n)]} - 1\right| \geq \epsilon\right) \leq \mathrm{Var}\left(\frac{FD(T_n)}{\mathbb{E}[FD(T_n)]}\right)/\epsilon^2, \tag{3.4}$$

and since

$$\frac{FD(T_n)}{\mathbb{E}[FD(T_n)]} = \frac{FD(T_n)}{n} \cdot \frac{n}{\mathbb{E}[FD(T_n)]} \leq \frac{FD(T_n)}{n} \cdot \frac{1}{c}$$

the term on the right of Eqn. (3.4) is bounded above by $\frac{1}{c^2}\mathrm{Var}\left(\frac{FD(T_n)}{n}\right)$, which converges to 0 as $n \to \infty$ by Eqn. (3.3).

□

Lemma 3.5. Let $(T_n, n \geq 1)$ be a sequence of rooted binary trees, with T_n having leaf set X_n, and suppose that $\mathbb{E}[FD(T_n)] \geq cn$ for some constant $c > 0$. Suppose that for each $x \in X_n$, $FD(\{x\}) \leq B_n$, where $B_n^2 = o(n)$. Then $\frac{FD(T_n^s)}{\mathbb{E}[FD(T_n^s)]} \xrightarrow{P} 1$ as $n \to \infty$.

Proof. Let W_n denote the random variable $\frac{FD(T_n^s)}{\mathbb{E}[FD(T_n^s)]}$. We can write $W_n = H(\mathbf{Y_n})/n$, where $\mathbf{Y_n} = \{Y_i : i \in [n]\}$ is the sequence of Bernoulli random variables with $Y_i = 1$ if species x_i survives the FOB extinction event and $Y_i = 0$ otherwise, and $H(y_1, \ldots, y_n)$ is the ratio $\frac{FD(\{x_i \in X_n : y_i = 1\})}{\mathbb{E}[FD(T_n^s)]}$, where the numerator is as defined in Eqn. (2.1) with $\mu(f) = 1$ for all f. We then have:

$$|H(y_1, ..., y_i, ..., y_n) - H(y_1, ..., y_i', ..., y_n)| \leq FD(\{x_i\})/\mathbb{E}[FD(T_n^s)] \leq B_n/\mathbb{E}[FD(T_n^s)].$$

We now apply McDiarmid's inequality to obtain (for each $\epsilon > 0$):

$$\mathbb{P}(|W_n - 1| \geq \epsilon) \leq 2\exp\left(\frac{-2\epsilon^2 \mathbb{E}[FD(T_n^s)]^2}{nB_n^2}\right). \tag{3.5}$$

Now, $\mathbb{E}[T_n^s] \geq s\mathbb{E}[T_n]$ by Proposition 2.1(ii) (taking $\mu(f) = 1$ for all f) and, by assumption, $\mathbb{E}[T_n] \geq cn$. Thus we obtain:

$$\mathbb{P}(|W_n - 1| \geq \epsilon) \leq 2\exp\left(\frac{-2\epsilon^2 c^2 n}{B_n^2}\right).$$

Therefore, $\mathbb{P}(|W_n - 1| \geq \epsilon) \to 0$ as $n \to \infty$. Since this holds for all $\epsilon > 0$, we obtain the claimed result.

□

We can now state the main result of this section.

Theorem 3.1. *Let $(T_n, n \geq 1)$ be a sequence of binary trees and let features evolve on T_n according to the stochastic feature evolution process described. If $\mathbb{E}[FD(T_n)] \geq cn$ for some constant $c > 0$, and if $\frac{\mathbb{E}[FD(T_n^s)]}{\mathbb{E}[FD(T_n)]}$ converges to a constant c_s as n grows we have:*

$$\frac{FD(T_n^s)}{FD(T_n)} \xrightarrow{P} c_s$$

as $n \to \infty$.

Proof. Let $A_\epsilon(n)$ be the event that $\left|\frac{FD(T_n^s)}{\mathbb{E}[FD(T_n^s)]} - 1\right| < \epsilon$, and let $\mathcal{E}_n$ be the event described in Lemma 3.2 with $\beta = \frac{1}{3}$. Then, $\lim_{n\to\infty} \mathbb{P}(A_\epsilon(n)|\mathcal{E}_n) = 1$, by Lemma 3.5. Now,

$$P(A_\epsilon(n)) = \mathbb{P}(A_\epsilon(n)|\mathcal{E}_n)\mathbb{P}(\mathcal{E}_n) + \mathbb{P}(A_\epsilon(n)|\overline{\mathcal{E}_n})\mathbb{P}(\overline{\mathcal{E}_n}),$$

and since $\mathbb{P}(\mathcal{E}_n) \to 1$ as $n \to \infty$ (by Lemma 3.2) we have $\lim_{n\to\infty} \mathbb{P}(A_\epsilon(n)) = 1$. Since this holds for all $\epsilon > 0$, it follows that $\frac{FD(T_n^s)}{\mathbb{E}[FD(T_n^s)]} \xrightarrow{P} 1$ as n grows.

Moreover, from Lemma 3.4, we have $\frac{FD(T_n)}{\mathbb{E}[FD(T_n)]} \xrightarrow{P} 1$. Now we can write $\frac{FD(T_n^s)}{FD(T_n)}$ as follows:

$$\frac{FD(T_n^s)}{FD(T_n)} = \frac{FD(T_n^s)}{\mathbb{E}[FD(T_n^s)]} \cdot \frac{\mathbb{E}[FD(T_n)]}{FD(T_n)} \cdot \frac{\mathbb{E}[FD(T_n^s)]}{\mathbb{E}[FD(T_n)]}$$

The first two terms on the right of this equation each converge in probability to 1 as $n \to \infty$, whereas the third (deterministic) term converges to c_s. This completes the proof.

□

Remarks: We will apply Theorem 3.1 in the next section to establish a result for FD loss on birth-death trees. A sufficient condition for the inequality $\mathbb{E}[FD(T_n)] \geq cn$ in Theorem 3.1 to hold is that for some $\epsilon > 0$ and $\delta > 0$ the proportion of pendant edges of T_n of length $\geq \epsilon$ is at least δ for all $n \geq 1$. Briefly, the reason for this is that the expected number of features arising on each of these pendant edges and surviving to the end of this edge is bounded away from 0 and the features associated with distinct pendant edges are always different from each other, and different from any other features arising in the tree.

4 Feature diversity ratios in birth–death trees

In this section, we continue to investigate the stochastic model of feature gain and loss, but rather than considering fixed trees, we will now allow the trees themselves to be stochastically generated, following the simple birth–death processes that are common in phylogenetics. Thus, there will now be *three* stochastic processes in play: the linear-birth/linear-death process that generates the tree, the constant-birth/linear-death process of feature gain and loss operating along the branches of the tree, and the simple FOB extinction event at the present.

4.1 Definitions

Let $\mathcal{T}_t$ denote a birth–death tree grown from a single lineage for time t with birth and death parameters λ and μ, respectively. We will assume throughout that $\lambda > \mu$ (since otherwise the tree is guaranteed to die out as t becomes large).

On $\mathcal{T}_t$, we impose the model of feature gain and loss from the previous section with parameters r and ν. We now apply the FOB model in which each extant species (i.e. leaves of $\mathcal{T}_t$ that are present at time t) has a probability $s > 0$ of surviving and $1 - s$ of becoming extinct (independently across the extant species), and we let $\mathcal{T}_t^s$ denote the tree obtained from $\mathcal{T}_t$ by removing the species at the present that did not survive this process. We refer to $\mathcal{T}_t^s$ as the *pruned* tree, and the leaves of $\mathcal{T}_t$ and $\mathcal{T}_t^s$ that are present at time t as the *extant species (or leaves)* of these trees (to contrast them from leaves of $\mathcal{T}_t$ that lie at the endpoints of any extinct lineages). If $s < 1$, there may now be fewer features present among the (probably reduced number of) extant species in $\mathcal{T}_t^s$ than there were in $\mathcal{T}_t$.

Let $FD(\mathcal{T}_t)$ be the discrete random variable that counts the number of features present in at least one of the extant leaves of $\mathcal{T}_t$, and let $FD(\mathcal{T}_t^s)$ denote the number of these features that are also present in at least one extant leaf in the pruned tree $\mathcal{T}_t^s$.

4.2 Expected feature diversity

Next, we consider the expected values of $FD(\mathcal{T}_t)$ and $FD(\mathcal{T}_t^s)$, where this expectation is across all three processes (the birth–death process that generates $\mathcal{T}_t$, feature gain and loss on this tree, and the species that survive the extinction event at the present under the FOB model). Of particular interest is the ratio of these expectations, and their limit as t becomes large. Specifically, let:

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

Note that $\varphi_{FD}(s)$ is a function of five parameters $(s, r, \lambda, \mu, \nu)$; however, we will show that it is just a function of s and two other parameters. Notice also that once these parameters are fixed, $\varphi_{FD}(t,s)$ and $\varphi_{FD}(s)$ are monotone increasing functions of s taking the value 0 at $s = 0$ and 1 at $s = 1$. Moreover, $\varphi_{FD}(t,s)$ and $\varphi_{FD}(s)$ are both independent of r (the rate at which features arise along any lineage) as we formally show shortly.

4.3 Relationship to Phylogenetic Diversity (PD)

Recall that for a rooted phylogenetic tree T with branch lengths, the PD value of a subset S of leaves ($PD(S,T)$) is the sum of the lengths of the edges of the subtree of T that connect S and the root of the tree.

In the special case where $\nu = 0$, and where no features are present at time 0, $FD(\mathcal{T}_t^s)$ (conditioned on $\mathcal{T}_t$), has a Poisson distribution with a mean of r times $PD(\mathcal{S}_t, \mathcal{T}_t)$, where $\mathcal{S}_t$ is the (random) set of leaves at time t that survive the extinction event at the present. Consequently, $\mathbb{E}[FD(\mathcal{T}_t^s)|\mathcal{T}_t^s] = \mathbb{E}[rPD(\mathcal{S}_t, \mathcal{T}_t)] = r\mathbb{E}[PD(\mathcal{S}_t, \mathcal{T}_t)]$. Similarly, $\mathbb{E}[FD(\mathcal{T}_t)|\mathcal{T}_t] = r\mathbb{E}[PD(\mathcal{L}_t, \mathcal{T}_t)]$, where $\mathcal{L}_t$ is the set of extant leaves of $\mathcal{T}_t$. Thus, in this special case we have $\varphi_{FD}(s) = \varphi_{PD}(s)$, where:

$$\varphi_{PD}(s) = \lim_{t\to\infty} \frac{\mathbb{E}[PD(\mathcal{S}_t, \mathcal{T}_t)]}{\mathbb{E}[PD(\mathcal{L}_t, \mathcal{T}_t)]}.$$

The function $\varphi_{PD}(s)$ was explicitly determined in [6, 13] as follows:

$$\varphi_{PD}(s) := \begin{cases} \frac{\rho s}{\rho+s-1} \cdot \frac{\ln(s/(1-\rho))}{\ln(1/(1-\rho))}, & \text{if } \rho = \mu/\lambda \neq 0, 1-s; \\ -s\ln(s)/(1-s), & \text{if } \rho = 0 \text{ (i.e. a Yule tree);} \\ (1-s)/\ln(1/s), & \text{if } \rho = 1-s. \end{cases} \tag{4.1}$$

5 Calculating $\varphi_{FD}(s)$

We first recall a standard result from birth–death theory. Let $\{X_t : t \geq 0\}$ denote the number of individuals at time t in a linear birth–death process (starting with a single individual at time 0), and with a birth rate λ, a death rate θ and a sampling probability at the present of $s > 0$. Let $R_t^s(\lambda, \theta)$ be the probability that $X_t > 0$. Then

$$R_t^s(\lambda,\theta) = \begin{cases} \frac{s(\lambda-\theta)}{s\lambda+(\lambda(1-s)-\theta)e^{(\theta-\lambda)t}}, & \text{if } \lambda \neq \theta; \\ \frac{s}{1+\lambda s t}, & \text{if } \lambda = \theta. \end{cases} \tag{5.1}$$

In particular, $R_t^s(\lambda, \theta)$ converges to 0 if $\lambda \leq \theta$ and converges to a strictly positive value $1 - \theta/\lambda$ if $\lambda > \theta$ [5, 22].

The number of species at time t in $\mathcal{T}_t^s$ that have a copy of particular feature f 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 parameters λ and $\theta = \mu+\nu$ and survival probability s at the present; it follows that as t becomes large, it becomes increasingly certain that none of the species at time t in the pruned tree will contain feature f if $\lambda \leq \mu + \nu$, whereas if $\lambda > \mu + \nu$, there is a positive limiting probability that f will be present in the extant leaves of the pruned tree.

Since $\varphi_{FD}(s) = \varphi_{PD}(s)$ (as given by Eqn. (4.1)) for all values of s when $\nu = 0$, in this section we will assume that $\nu > 0$ (in addition to our universal assumption that $\lambda > \mu$). Our main result provides an explicit formula for $\varphi_{FD}(s)$ in Part (a), and describes some of its key properties in Parts (b) and (c).

Theorem 5.1.
Given $\lambda > \mu$ and $\nu > 0$, let $\rho = \frac{\mu+\nu}{\lambda}$, $\beta = 1 - \frac{\lambda-\mu}{\nu}$. Then:

(a)

$$\varphi_{FD}(s) = s \cdot \frac{I(s)}{I(1)},$$

where

$$I(s) = \int_0^1 \frac{dx}{1 - s\left(\frac{1-x^\beta}{1-\rho}\right)}, \quad \textit{when } \rho \neq 1$$

and

$$I(s) = \int_0^1 \frac{dx}{\nu/\lambda - s\ln x}, \quad \textit{when } \rho = 1.$$

(b) Conditional on the non-extinction of $\mathcal{T}_t$, $\frac{FD(\mathcal{T}_t^s)}{FD(\mathcal{T}_t)}$ converges in probability to $\varphi_{FD}(s)$ as $t \to \infty$.

(c) $\varphi_{FD}(s)$ is an increasing concave function that satisfies $1 \geq \varphi_{FD}(s) \geq s$ for all s.

Remarks:

(i) Notice that although $\varphi_{FD}(s)$ depends on five parameters $(r, s, \lambda, \mu, \nu)$, Theorem 5.1(a) reveals that just three derived parameters suffice to determine $\varphi_{FD}(s)$, namely s and the ratios $\rho_1 = \mu/\lambda \in (0, 1)$ and $\rho_2 = \nu/\lambda$ (these determine ρ and β, since $\rho = \rho_1 + \rho_2$ and $\beta = 1 - 1/\rho_2 + \rho_1/\rho_2$). Notice also that $\rho = 1 \Leftrightarrow \beta = 0$ and $\rho > 1 \Leftrightarrow \beta > 0$.

(ii) Our proof relies on establishing the following exact expression for $\mathbb{E}[FD(\mathcal{T}_t^s)]$:

$$\mathbb{E}[FD(\mathcal{T}_t^s)] = re^{(\lambda-\mu)t} \int_0^t e^{-(\lambda-\mu)\tau} R_\tau^s(\lambda, \mu+\nu)d\tau + F_0 R_t^s(\lambda, \mu+\nu). \tag{5.2}$$

where F_0 is the number of expected features present at time $t = 0$.

In particular, this also provides an exact expression for $\varphi_{FD}(t,s)$. Notice that the ratio of the expected number of features present in the pruned tree, divided by the expected number of species in the pruned tree is $\frac{FD(\mathcal{T}_t^s)}{se^{(\lambda-\mu)t}}$, and this ratio converges to $\frac{r(1-\rho)}{\nu}\int_0^1 \frac{dx}{1-s-\rho+sx^\beta}$ as $t \to \infty$ when $\rho \neq 1$ (via a further analysis of Eqn. (5.2) in the Appendix).

(ii) We saw in Section 4.3 that when $\nu = 0$, $\varphi_{FD}(s) = \varphi_{PD}(s)$. At the other extreme, if λ and μ are fixed, and we let $\nu \to \infty$, then $\varphi_{FD}(s)$ converges to s (since $\rho \to -\infty$ and $\beta \to 1$ in Theorem 5.1(a)). Informally, when ν is large compared to λ, most of the features present among the extant leaves of $\mathcal{T}_t$ will have arisen near the end of the pendant edges incident with these extant leaves (a formalisation of this claim appears in [16]); if we now apply the FOB model then the expected proportion of these features that survive will be close to the expected proportion of leaves that survive, namely s.

5.1 Illustrative examples

First, consider a Yule tree (i.e. $\mu = 0$) grown for time t. Figure 5.1 (left) plots $\varphi_{FD}(s)$ for values of $\nu/\lambda \in \{0, 0.5, 1, 2, 10\}$. When $\nu = 0$, $\varphi_{FD}(s)$ describes the proportional loss of expected PD in the pruned tree, and as ν increases, $\varphi_{FD}(s)$ converges towards s (the expected proportion of leaves that survive extinction at the present). Figure 5.1 (right) plots $\varphi_{FD}(s)$ for birth–death trees with $\mu/\lambda = 0.8$, showing a similar trend, however with $\varphi_{FD}(s)$ ranging higher above the curve $y = s$.

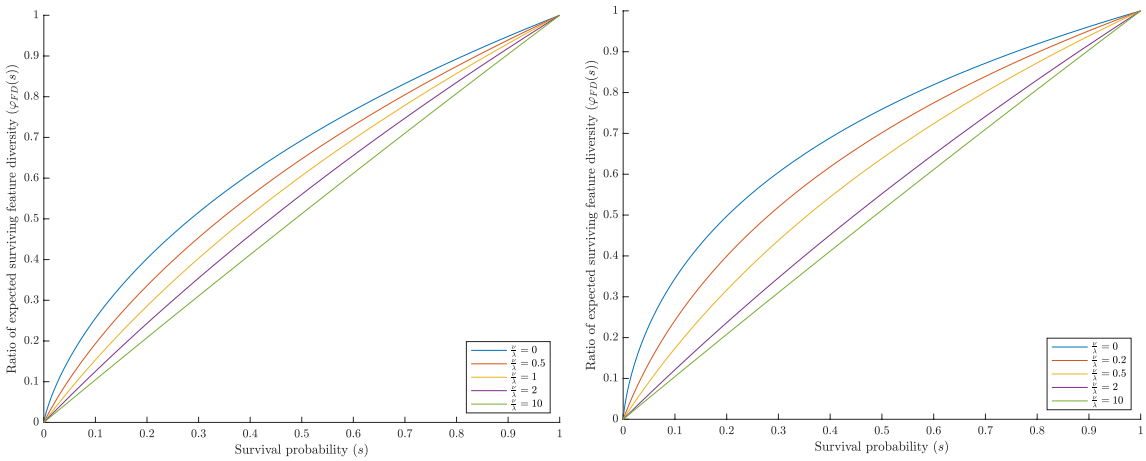

Figure 5.1 *Left:* The function $\varphi_{FD}(s)$ as a function of s for pure-birth Yule trees ($\mu = 0$). *Right:* The function $\varphi_{FD}(s)$ for birth–death trees with $\mu/\lambda = 0.8$. The curves within each graph show the effect of increasing ν, for values of ν/λ between 0 (top) and 10 (bottom). The top curve also corresponds to the phylogenetic diversity ratio $\varphi_{PD}(s)$.

5.2 Features that appear in only one extant species

Let $\mathcal{U}_t$ denote the number of features that that are present in precisely one species in $\mathcal{T}_t$, and let $U_t = \mathbb{E}[\mathcal{U}_t]$. The following result describes a simple relationship between U_t and $F_t = \mathbb{E}[FD(\mathcal{T}_t)]$.

Proposition 5.1.

$$\frac{dF_t}{dt} = re^{(\lambda-\mu)t} - (\mu+\nu)U_t. \tag{5.3}$$

Proof. Let $\mathcal{F}_t$ denote the number of features present at time t among the leaves of $\mathcal{T}_t$. Consider evolving $\mathcal{T}_t$ for an additional (short) period δ into the future. Then, conditional on $\mathcal{N}_t$ (the

number of leaves of $\mathcal{T}_t$ present at time t) and $\mathcal{U}_t$:

$$\mathcal{F}_{t+\delta} - \mathcal{F}_t = \begin{cases} 1, & \text{with probability } r\mathcal{N}_t\delta + o(\delta); \\ -1, & \text{with probability } (\mu+\nu)\delta \cdot \mathcal{U}_t + o(\delta); \\ 0, & \text{with probability } 1 - ((\mu+\nu)\mathcal{U}_t + r\mathcal{N}_t)\delta + o(\delta). \end{cases}$$

Applying the expectation operator (and using $\mathbb{E}[\mathcal{F}_{t+\delta}] = \mathbb{E}[\mathbb{E}[\mathcal{F}_{t+\delta}|\mathcal{N}_t, \mathcal{U}_t]]$) and letting $\delta \to 0$ leads to the equation stated.

□

More generally, let $\mathcal{U}_t^s$ denote the number of features that are present in precisely one species in $\mathcal{T}_t^s$, and let $U_t^s = \mathbb{E}[\mathcal{U}_t^s]$. For a birth–death process $\{X_t : t \geq 0\}$ with birth rate λ, death rate θ, and a sampling probability s, let $Q_t^s(\lambda, \theta)$ be the probability that $X_t = 1$, which equals $\frac{1}{s}R_t^s(\lambda,\theta)^2 e^{(\theta-\lambda)t}$ [5, 22]. Using similar arguments to those elsewhere in this paper, it can be shown that:

$$U_t^s = re^{(\lambda-\mu)t} \int_0^t e^{-(\lambda-\mu)\tau} Q_\tau^s(\lambda, \mu+\nu) d\tau + F_0 Q_t^s(\lambda, \mu+\nu).$$

6 Concluding comments

In this paper, we have considered, in order, three types of data to quantify the expected loss of feature diversity: sets of features across species (without any model of feature evolution or phylogeny), sets of features at the tips of a given phylogenetic tree, and sets of features at the tips of a random (birth–death) tree. The results of the earlier sections also proved helpful in establishing certain results in later sections.

In terms of wider significance to biodiversity conservation, our results and graphs in Section 5 suggest that the extent of relative feature diversity loss following extinction at the present is likely to be greater than that predicted by relative phylogenetic diversity loss for any given extinction rate ($1 - s \in (0, 1)$).

Of course, our results are based on simple models (of feature gain and loss, and extinction at the present) and so exploring how these results might extend to more complex and realistic biological models would be worthwhile for future work.

7 Acknowledgements

We thank François Bienvenu for helpful suggestions on an earlier draft of this manuscript. We also thank the New Zealand Marsden Fund (MFP-UOC2005) for supporting this research.

References

[1] V. Devictor, D. Mouillot, C. Meynard, F. Jiguet, W. Thuiller, and N. Mouquet. Spatial mismatch and congruence between taxonomic, phylogenetic and functional diversity: The need for integrative conservation strategies in a changing world. *Ecology Letters*, 13:1030–1040, June 2010.

[2] D. P. Faith. Conservation evaluation and phylogenetic diversity. *Biological Conservation*, 61(1):1–10, 1992.

[3] W. Feller. *An Introduction to Probability Theory and its Applications*, volume 1. Wiley, 3rd edition, 1950.

[4] P. Jagers. Stabilities and instabilities in population dynamics. *Journal of Applied Probability*, 29:770–780, 1992.

[5] D. G. Kendall. On the generalized birth-and-death process. *Annals of Mathematical Statistics*, 19:1–15, 1948.

[6] A. Lambert and M. Steel. Predicting the loss of phylogenetic diversity under non-stationary diversification models. *Journal of Theoretical Biology*, 337:111–124, 2013.

[7] F. Mazel, A. O. Mooers, G. V. D. Riva, and M. W. Pennell. Conserving phylogenetic diversity can be a poor strategy for conserving functional diversity. *Systematic Biology*, 66(6):1019–1027, June 2017.

[8] F. Mazel, 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. Prioritizing phylogenetic diversity captures functional diversity unreliably. *Nature Communications*, 9:2888, July 2018.

[9] F. Mazel, 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. Reply to: "Global conservation of phylogenetic diversity captures more than just functional diversity". *Nature Communications*, 10:858, Feb. 2019.

[10] C. McDiarmid. On the method of bounded differences. In *Surveys in combinatorics, London Mathematical Society Lecture Notes Series 141*, pages 148–188. Cambridge University Press, 1989.

[11] J. T. Miller, G. Jolley-Rogers, B. D. Mishler, and A. H. Thornhill. Phylogenetic diversity is a better measure of biodiversity than taxon counting. *Journal of Systematics and Evolution*, 56(6):663–667, 2018.

[12] M. Mitzenmacher and E. Upfal. *Probability and computing: Randomized algorithms and probabilistic analysis*. Cambridge University Press, 2005.

[13] A. Mooers, O. Gascuel, T. Stadler, H. Li, and M. Steel. Branch lengths on birth–death trees and the expected loss of phylogenetic diversity. *Systematic Biology*, 61(2):195–203, 2012.

[14] N. R. Owen, R. Gumbs, C. L. Gray, and D. P. Faith. Global conservation of phylogenetic diversity captures more than just functional diversity. *Nature Communications*, 10:859, Feb. 2019.

[15] D. M. Raup. *Extinction: Bad Genes or Bad Luck?* Oxford University Press, 1993.

[16] J. Rosindell, K. Manson, R. Gumbs, W. Pearse, and M. Steel. Phylogenetic biodiversity metrics should account for both accumulation and attrition of evolutionary heritage. Technical Report 2022.07.16.499419, BioRxiv, July 2022.

[17] T. Stadler and M. Steel. Distribution of branch lengths and phylogenetic diversity under homogeneous speciation models. *Journal of Theoretical Biology*, 297(2):33–40, 2012.

[18] C. M. Tucker, T. Aze, M. W. Cadotte, J. L. Cantalapiedra, C. Chisholm, and S. Díaz. Assessing the utility of conserving evolutionary history. *Biological Reviews*, 94:1740–1760, 2019.

[19] C. M. Tucker, T. J. Davies, M. W. Cadotte, and W. D. Pearse. On the relationship between phylogenetic diversity and trait diversity. *Ecology*, 99(6):1473–1479, 2018.

[20] K. Wicke and M. Fischer. Phylogenetic diversity and biodiversity indices on phylogenetic networks. *Mathematical Biosciences*, 298:80–90, 2018.

[21] K. Wicke, A. Mooers, and M. Steel. Formal links between feature diversity and phylogenetic diversity. *Systematic Biology*, 70:480–490, 2021.

[22] Z. Yang and B. Rannala. Bayesian phylogenetic inference using DNA sequences: a Markov chain Monte Carlo method. *Molecular Biology and Evolution*, 14(7):717–724, July 1997.

8 Appendix: Proof of Lemma 3.3 and Theorem 5.1

Proof of Lemma 3.3

We first establish Eqn (3.2) in the case where the each tree T_n has no features at the top of the stem edge (i.e. $F_\rho = \emptyset$) by using induction on n. We then extend the result to allow F_ρ to be an arbitrary finite set.

For the case $F_\rho = \emptyset$, Eqn. (3.2) clearly holds for $n = 1, 2$, so suppose that $n \geq 3$, and let T_n be a binary tree (with a stem edge) and n leaves with $F_\rho = \emptyset$. In this case:

$$FD(T_n) = FD(T^1) + FD(T^2) + G.$$

where the trees T^i (with $i = 1, 2$) are subtrees of T_n obtained by deleting the stem edge and its endpoints (and setting the set of features at the top of the stem edge of T^1 and T^2 equal to the empty set), and G is the number of features that arise on the stem edge of T_n and are present in at least one leaf of T_n.

Notice that $FD(T^1)$, $FD(T^2)$ and G are independent random variables, and so:

$$\mathrm{Var}[FD(T_n)] = \mathrm{Var}[FD(T^1)] + \mathrm{Var}[FD(T^2)] + \mathrm{Var}[G]. \tag{8.1}$$

Now, $G = \sum_{i=1}^{K} \mathbb{I}_i$, where K is the number of features that arise along the stem edge of T_n and $\mathbb{I}_i$ is a random variable taking the value 1 if feature i (on the stem edge) is present in at least one leaf of T_n and 0 otherwise. Thus G is the sum of a Poisson number of i.i.d. Bernoulli random variables, and so G has a Poisson distribution. Consequently:

$$\mathrm{Var}[G] = \mathbb{E}[G] = \frac{r}{\nu}(1 - e^{-\nu\ell})\mathbb{P}(\mathbb{I}_1 = 1) \leq \frac{r}{\nu}. \tag{8.2}$$

Each subtree T^i has $n_i < n$ leaves (since $n_1 + n_2 = n$ and $n_1, n_2 \geq 1$), and $\mathrm{Var}[G] \leq \frac{r}{\nu}$. We can now invoke the induction hypothesis and Eqns. (8.1) and (8.2) to obtain:

$$\mathrm{Var}[FD(T_n)] \leq C(n_1\sqrt{n_1} + n_2\sqrt{n_2}) + \frac{r}{\nu}.$$

Since $n_1\sqrt{n_1} + n_2\sqrt{n_2} \leq 1 + (n-1)\sqrt{n-1}$, we have:

$$\mathrm{Var}[FD(T_n)] \leq C\left(1 + (n-1)\sqrt{n-1}\right) + \frac{r}{\nu} \leq Cn\sqrt{n},$$

for $C > 3\frac{r}{\nu}$ (noting that $n \geq 3$). This establishes the induction step to justify Eqn. (3.2) when $F_\rho = \emptyset$.

Next, suppose that F_ρ not constrained to be $\emptyset$. Let H be the number of features of F_ρ that are present in at least one leaf of T_n. Then

$$FD(T_n) = FD(T_n | F_\rho = \emptyset) + H,$$

and since the two random variables on the right are independent, we have:

$$\mathrm{Var}[FD(T_n)] = \mathrm{Var}[FD(T_n) | F_\rho = \emptyset] + \mathrm{Var}[H].$$

Now, $0 \leq H \leq |F_\rho|$ with probability 1, and so $\mathrm{Var}[H] \leq |F_\rho|^2/4$. Thus, applying the case where $F_\rho = \emptyset$ we have:

$$\mathrm{Var}[FD(T_n)] \leq Cn\sqrt{n} + |F_\rho|^2/4n^2,$$

which again implies that Eqn. (3.2) holds. □

Proof of Theorem 5.1

Part (a): Consider a birth–death tree $\mathcal{T}_t$ with parameters λ, μ, a feature evolution model with parameters r, ν, and a survival probability s for leaves at the present. Let $\mathcal{F}_t^s = FD(\mathcal{T}_t^s)$, and let $\mathcal{G}_t^s$ be the random variable that has the same distribution as $\mathcal{F}_t^s$ with the initial condition $\mathcal{G}_0^s = 0$ (i.e. no features at the root of the tree). Let Δ_t^s be the number of features that are present at the root of $\mathcal{T}_t$ and also in the pruned tree $\mathcal{T}_t^s$. Then

$$\mathcal{F}_t^s = \mathcal{G}_t^s + \Delta_t^s.$$

Let $F_t^s = \mathbb{E}[\mathcal{F}_t^s]$ and $G_t^s = \mathbb{E}[\mathcal{G}_t^s]$. The two random variables $\mathcal{G}_t^s$ and Δ_t^s are not independent (they are linked by both the underlying tree $\mathcal{T}_t$ and the pruning event at the present) however, applying the expectation operator gives:

$$F_t^s = \mathbb{E}[\mathcal{G}_t^s] + \mathbb{E}[\Delta_t^s] = G_t^s + F_0 R_t^s(\lambda, \mu + \nu), \tag{8.3}$$

where F_0 is the expected number of features present at the root of $\mathcal{T}_t^s$.

Now, consider $\mathcal{G}_{t+\delta}^s$, and the events that can occur in the interval $[0, \delta)$.

(a) A new feature arises (with probability $r\delta + o(\delta)$);

(b) A speciation event occurs (with probability $\lambda\delta + o(\delta)$);

(c) The lineage (and hence the tree) dies (with probability $\mu\delta + o(\delta)$).

(d) None of the above occur (with probability $1 - (r + \lambda + \mu) + o(\delta)$)

Let X be the random variable taking values in $\{a, b, c, d\}$ which denotes which of these four events occurs. In Case (a), the new feature is also present in $\mathcal{T}_t^s$ with probability $R_t^s(\lambda, \mu + \nu)$ and so

$$\mathbb{E}[\mathcal{G}_{t+\delta}^s | X = a] = \mathbb{E}[\mathcal{G}_t^s] + R_t^s(\lambda, \mu + \nu) = G_t^s + R_t^s(\lambda, \mu + \nu).$$

For Case (b), $\mathbb{E}[\mathcal{G}_{t+\delta}^s | X = b] = \mathbb{E}[\mathcal{G}_t^s + \mathcal{H}_t^s]$, where $\mathcal{H}_t^s$ is an independent copy of $\mathcal{G}_t^s$. Thus,

$$\mathbb{E}[\mathcal{G}_{t+\delta}^s | X = b] = 2G_t^s.$$

For Cases (c) and (d), we have: $\mathbb{E}[\mathcal{G}_{t+\delta}^s | X = c] = 0$ and $\mathbb{E}[\mathcal{G}_{t+\delta}^s | X = d] = G_t^s$. Thus, by the law of total expectation,

$$G_{t+\delta}^s = \mathbb{E}[\mathbb{E}[\mathcal{G}_{t+\delta}^s | X]] = G_t^s + (rR_t^s(\lambda, \mu + \nu) + (\lambda - \mu)G_t^s)\delta + o(\delta).$$

Consequently, the function G_t^s satisfies the first-order linear differential equation:

$$\frac{d}{dt}G_t^s - (\lambda - \mu)G_t^s = rR_t^s(\lambda, \mu + \nu), \tag{8.4}$$

subject to the initial condition $G_0^s = 0$. Solving Eqn. (8.4) gives:

$$G_t^s = \int_0^t re^{(\lambda-\mu)\tau} R_{t-\tau}^s(\lambda, \mu + \nu) d\tau \tag{8.5}$$

By making a change of variable we can rewrite Eqn. (8.5) as:

$$G_t^s = re^{(\lambda-\mu)t}\int_0^t e^{-(\lambda-\mu)\tau}R_\tau^s(\lambda,\mu+\nu)d\tau. \tag{8.6}$$

Combining this equation and Eqn. (8.3) provides the explicit expression for F_t^s described earlier (Eqn. (5.2)).

We now substitute in the expression for $R_t^s(\lambda,\theta)$ from Eqn. (5.1) (with $t=\tau$ and $\theta=\mu+\nu$). For $\rho\neq 1$, we have:

$$G_t^s = re^{(\lambda-\mu)t}s(1-\rho)\int_0^t \frac{e^{(\mu-\lambda)\tau}}{(1-s-\rho)e^{-\lambda(1-\rho)\tau}+s}d\tau. \tag{8.7}$$

Multiplying the numerator and denominator of the integrand by $e^{\lambda(1-\rho)\tau}$ gives $\frac{e^{-\nu\tau}}{1-s-\rho+se^{\lambda(1-\rho)\tau}}$, and then by making the substitution $x=e^{-\nu\tau}$, we obtain:

$$G_t^s = \frac{re^{(\lambda-\mu)t}s(1-\rho)}{\nu}\cdot\int_{e^{-\nu t}}^1 \frac{dx}{1-s-\rho+sx^\beta}, \tag{8.8}$$

for the value of β described in the theorem. Combining Eqns. (8.3) and (8.8) gives:

$$F_t^s = \frac{re^{(\lambda-\mu)t}s(1-\rho)}{\nu}\cdot\int_{e^{-\nu t}}^1 \frac{dx}{1-s-\rho+sx^\beta} + F_0R_t^s(\lambda,\mu+\nu). \tag{8.9}$$

Thus, for $\rho\neq 1$,

$$\frac{F_t^s}{F_t^1} = \frac{s\int_{e^{-\nu t}}^1 \frac{dx}{1-s-\rho+sx^\beta}+o(1)}{\int_{e^{-\nu t}}^1 \frac{dx}{-\rho+x^\beta}+o(1)},$$

where the two terms of order $o(1)$ (which converge to zero as t grows) refer to the last term on the right of Eqn. (8.8) which is bounded above by the constant F_0 and so is asymptotically negligible in comparison to the term $e^{(\lambda-\mu)t}$ in Eqn. (8.8). This gives the limit for $\varphi_{FD}(s)$ as stated in Part (a) for fixed $\rho\neq 1$.

In the case where $\rho=1$ (which implies $\beta=0$), we use the corresponding expression for $R_t^s(\lambda,\theta)$ from Eqn. (5.1) with $t=\tau$ and $\theta=\mu+\nu$ to obtain:

$$G_t^s = rse^{(\lambda-\mu)t}\int_0^t \frac{e^{-(\lambda-\mu)\tau}}{1+\lambda s\tau}d\tau.$$

By a similar approach to the above we are led to the second equation in Part (a).

Part (b): Let $\mathcal{T}_t$ be a birth–death tree with rates λ,μ where $\lambda>\mu$, let $n(\mathcal{T}_t)$ denote the number of leaves of $\mathcal{T}_t$ present at time t, and let $\mathcal{E}'$ be the event that $n(\mathcal{T}_t)>0$ (i.e. the non-extinction of $\mathcal{T}_t$).

Conditional on the event $\mathcal{E}'$, the number of leaves in $\mathcal{T}_t$ tends to infinity (with probability 1) as $t\to\infty$ [4], and so we can define a sequence of trees $T_1,T_2,\ldots,T_n,\ldots$ by letting T_k denote the tree $\mathcal{T}_\tau$ at the first time $\tau=\tau(k)$ when $\mathcal{T}_\tau$ has k extant leaves (we ignore leaves of $\mathcal{T}_t$ that have already become extinct by time τ).

Next, we establish that $\mathbb{E}[FD(T_n)]\geq cn$ for a constant $c>0$ (in order to apply Theorem 3.1). The tree T_n has n extant pendant edges, and the length of a randomly selected pendant edge in T_n has a strictly positive probability p of having length at least $\kappa>0$ (dependent on μ and λ), by Theorem 3.1 of [17]. Now, $\mathbb{E}[FD(T_n)]$ is bounded below by the total number of features that arises on the n pendant edges and survive to the end of the edge (since all these features will necessarily be distinct from each other, and from other features that arise in the tree). Moreover, for each edge having length at least κ the expected number of features that arise on

this edge and survive to the end of the edge is at least $\frac{r}{\nu}(1-e^{-\kappa/\nu})$. Thus the expected number of features contributed by the pendant edges to $FD(T_n)$ is at least $n \cdot p\frac{r}{\nu}(1-e^{-\kappa/\nu})$.

Thus, we can now apply Theorem 3.1, since $c_s = \lim_{n\to\infty} \frac{\mathbb{E}[FD(T_n^s)]}{\mathbb{E}[FD(T_n)]}$ exists, and equals $\varphi_{FD}(s)$, so $\frac{FD(T_n^s)}{FD(T_n)}$ (and thus $\frac{FD(\mathcal{T}_t^s)}{FD(\mathcal{T}_t)}$) converges to $\varphi_{FD}(s)$ as n (respectively t) grows.

Part (c): We apply Proposition 2.2. By Part (i) of that result, and conditioning on $\mathcal{T}_t$ we obtain $\mathbb{E}[FD(\mathcal{T}_t^s)|\mathcal{T}_t] \geq sFD(\mathcal{T}_t)$, and so, taking expectation again (over the distribution of $\mathcal{T}_t$) gives: $\mathbb{E}[FD(\mathcal{T}_t^s)] \geq s\mathbb{E}[FD(\mathcal{T}_t)]$, and thus $\varphi_{FD}(s) = \lim_{t\to\infty} \frac{\mathbb{E}[FD(\mathcal{T}_t^s)]}{\mathbb{E}[FD(\mathcal{T}_t)]} \geq s$.

The inequality $\varphi_{FD}(s) \leq 1$ is clear since, for any choice of $\mathcal{T}_t$, we have $FD(\mathcal{T}_t^s) \leq FD(\mathcal{T}_t)$ with probability 1.

For concavity, Proposition 2.2 implies that the conditional expectation $\mathbb{E}\left[\frac{FD(\mathcal{T}_t^s)}{FD(\mathcal{T}_t)}|\mathcal{T}_t\right]$ is concave as a function of s, and (by taking expectation over the distribution of $\mathcal{T}_t$), $\mathbb{E}\left[\frac{FD(\mathcal{T}_t^s)}{FD(\mathcal{T}_t)}\right]$ is also concave as a function of s. Finally, by Part (a) of the current theorem, $\mathbb{E}\left[\frac{FD(\mathcal{T}_t^s)}{FD(\mathcal{T}_t)}\right]$ converges (deterministically) to $\varphi_{FD}(s)$ and so this function is also concave as a function of s. $\square$