Journal of Theoretical Biology 337 (2013) 111–124

Contents lists available at ScienceDirect

Journal of Theoretical Biology

journal homepage: www.elsevier.com/locate/yjtbi

Predicting the loss of phylogenetic diversity under non-stationary diversification models

Amaury Lambert [a], Mike Steel [b,*]

[a] Collège de France, Center for Interdisciplinary Research in Biology CNRS UMR 7241, Paris, France
[b] Biomathematics Research Centre, University of Canterbury, Private Bag 4800, Christchurch, New Zealand

HIGHLIGHTS

- We study the random loss of phylogenetic diversity due to extinction at the present.
- We provide exact formulae for this loss under a range of extinction and speciation models.
- We obtain exact convergence results as the number of taxa or the depth of the tree grows.

ARTICLE INFO

Article history:
Received 12 June 2013
Received in revised form
10 August 2013
Accepted 13 August 2013
Available online 20 August 2013

Keywords:
Phylogenetic tree
Diversification process
Field of bullets model

ABSTRACT

For many species, the current high rates of extinction are likely to result in a significant loss of biodiversity. The evolutionary heritage of biodiversity is frequently quantified by a measure called phylogenetic diversity (PD). We predict the loss of PD under a wide class of phylogenetic tree models, where speciation rates and extinction rates may be time-dependent, and assuming independent random species extinctions at the present. We study the loss of PD when K contemporary species are selected uniformly at random from the N extant species as the surviving species, while the remaining $N-K$ become extinct (N and K being random variables). We consider two models of species sampling, the so-called field of bullets model, where each species independently survives the extinction event at the present with probability p, and a model for which the number of surviving species is fixed.

We provide explicit formulae for the expected remaining PD in both models, conditional on $N=n$, conditional on $K=k$, or conditional on both events. When $N=n$ is fixed, we show the convergence to an explicit deterministic limit of the ratio of new to initial PD, as $n\to\infty$, both under the field of bullets model, and when $K=k_n$ is fixed and depends on n in such a way that k_n/n converges to p. We also prove the convergence of this ratio as $T\to\infty$ in the supercritical, time-homogeneous case, where N simultaneously goes to ∞, thereby strengthening previous results of Mooers et al. (2012).

1. Introduction

1.1. Phylogenetic diversity

A typical question arising in biodiversity conservation is the following:

> If a random 10% of species from some clade were to disappear in the next 100 years due to current high rates of extinction, how much evolutionary heritage would be lost?

The answer depends on many factors, the first of which is how one measures evolutionary heritage. Here, we adopt phylogenetic diversity (PD) for this purpose – it assigns to any (surviving) subset of species the sum of the branch lengths of the evolutionary tree that span those species and the root of the tree (Faith, 1992). Thus, one can consider the ratio of the PD after a rapid mass extinction event (the ‘surviving PD’ score) to the initial PD score as a measure of the relative PD loss.

A second important factor in answering this question is the interplay of the tree shape and the process of extinction at the tips. For example, the extinction of a species at the end of a long pendant edge of the tree will lead to greater PD loss than the extinction of a species on a short pendant edge. However, this is just a part of the story, as interaction effects also occur – for instance, the extinction of two closely related species on short pendant edges that form a cherry in the tree at the end of a very long interior edge will lead to far more

---

* Corresponding author. Tel.: +64 21329705.
E-mail addresses: amaury.lambert@upmc.fr (A. Lambert), mike.steel@canterbury.ac.nz, mathmomike@gmail.com (M. Steel).

0022-5193/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.jtbi.2013.08.009

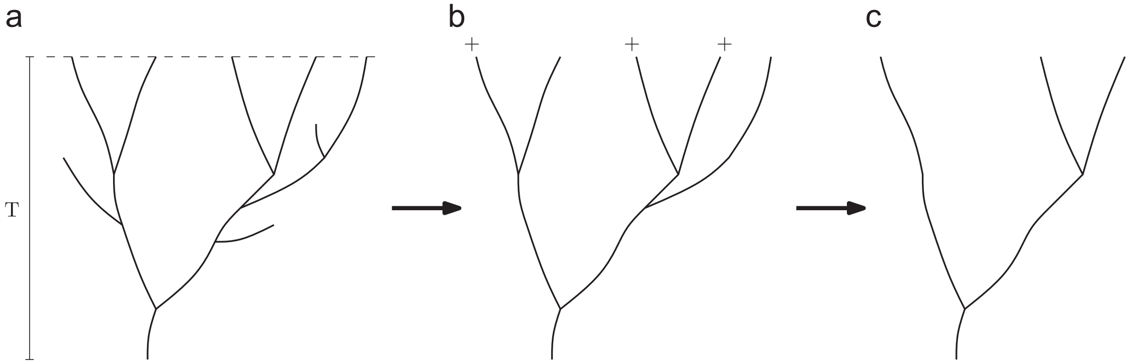

Fig. 1. (a) An evolutionary tree as it arises under a continuous model of speciation and extinction, observed at the present time T. (b) The reconstructed tree obtained by deleting lineages that do not survive to the present. The leaves are now subject to a further extinction event at the present (e.g. under a 'field of bullets' model). The tree connecting the surviving leaves (indicated by $+$ in (b)) is shown in (c).

PD loss than the extinction of two species with short to moderate pendant edge lengths that do not form a cherry.

This interplay of tree shape and possible species extinction scenarios will vary from data-set to data-set, and will generally depend on a large number of parameters (related to the tree, its branch lengths, the extinction risks of different species and how they are correlated), some of which are often not known with any precision.

In this paper, we establish general results and properties concerning relative PD loss, by showing how it can be estimated by closed-form formulae based on stochastic diversification models that describe how phylogenetic trees arise under speciation and extinction models (Aldous and Popovic, 2005; Aldous et al., 2011; Morlon et al., 2010; Purvis et al., 2000; Rabosky and Lovette, 2008), together with a simple 'field of bullets' model of random, instantaneous extinctions at the present. Thus, our approach is in a similar spirit to Nee and May (1997) and the more recent paper by Mooers et al. (2012), but our results generalize and strengthen these earlier results in some important ways:

- Most of our results allow the speciation rate $b(t)$ for a lineage to depend on time t, and the extinction rate $d(t;x)$ for a lineage to depend on time t and/or on a non-heritable trait x (i.e. a discrete or continuous trait changing in the same way in all species, for example the age of the species); this generalizes the classical (constant rate) birth–death model where $b(t)=b$ and $d(t;x)=d$ for constants $b,d\geq 0$, thereby allowing a greater biological realism.
- Rather than studying the limiting ratio of expected surviving PD to expected initial PD (as in Mooers et al., 2012), we analyze the actual ratio of new to initial PD and establish its convergence to explicitly computable functions under each of two limiting processes (increasing number of species and increasing time). This provides for statements with a greater statistical precision.
- We also present explicit exact formulae for the expected surviving PD (and the expected loss of PD), given a fixed initial number of species and the depth of the tree under sudden random mass extinctions at the present. We also provide a formula for when we explicitly condition on the number of species that survive this sudden extinction event.

Before proceeding to describe our results, we summarize some standard terms in probability theory that will be used throughout this paper.

1.2. Terminology from probability theory

Recall first that a *Bernoulli random variable* has just two outcomes (0,1), with 1 referred to as a 'success'. Given a sequence of independent and identically distributed (i.i.d.) Bernoulli random variables $X_1, X_2, \ldots$, where X_i has success probability p, the random variable J that specifies the first value $j\geq 1$ for which $X_j=1$ is a (shifted) *geometric random variable* with success probability p; its distribution is easily seen to be $\mathbb{P}(J=j)=(1-p)^{j-1}p$, for $j=1,2,\ldots$. For example, the number of rolls of a fair die until the number 4 first appears is a geometric random variable with success probability $p=1/6$.

A sequence of random variables X_n *converges in probability* to some constant value c if the probability that X_n differs from c by every given positive value ϵ tends to zero as $n\to\infty$. For example the proportion of tosses of a fair coin that result in a head converges in probability to $c=1/2$ (by the weak law of large numbers, or by the central limit theorem). A stronger notion is to say that X_n converges *almost surely* to some constant c, which means that any realization (e.g. numerical simulation) of the sequence (X_n) converges to c with probability 1. This actually holds also for the coin-tossing example (by the strong law of large numbers). Almost sure convergence implies convergence in probability but the converse need not hold.

1.3. Summary of results

We use a 'coalescent point process' characterization of the reconstructed tree (described below) in which the total number of extant leaves N at a given time T has a geometric distribution (with a parameter denoted $a=a_T$) and we study the loss of phylogenetic diversity when contemporary species are randomly removed from the standing species set. For any time-calibrated phylogeny, the *phylogenetic diversity* (PD) is the total sum of branch lengths, also called the total length of the tree. We consider two models of random species removal.

The first model, called the 'field of bullets' (Raup, 1992; Nee and May, 1997; Purvis et al., 2000), makes the assumption that every contemporary species (i.e. every tip of the phylogeny) is independently removed with probability $1-p$, where p will be called the *sampling probability*. We will denote by K the total number of sampled and so retained species, by $S_N(p)$ the remaining PD after the passage of the field of bullets, and by K_n and $S_n(p)$, respectively, the same quantities when conditioning the initial number N of species (i.e. before sampling) to equal n.

The second model consists in fixing the number of sampled species, to k say, and to sample these k species uniformly at random, provided that $N\geq k$. We will denote by $S_{N,k}$ the remaining PD after sampling these k species (and removing all others); similarly, $S_{n,k}$ denotes the same quantity when conditioning on the initial number of species to equal n.

Fig. 1 provides an overview of the three steps in the processes we consider: (a) the phylogenetic tree generated by a speciation