Phylogenetic Biodiversity Metrics Should Account for Both Accumulation and Attrition of Evolutionary Heritage

James Rosindell[1,*], Kerry Manson[2], Rikki Gumbs[3], William D. Pearse[1] and Mike Steel[2]

[1] *Department of Life Sciences, Silwood Park Campus, Imperial College London, Buckhurst Road, Ascot, Berkshire, SL5 7PY, United Kingdom*

[2] *Biomathematics Research Centre, University of Canterbury, Christchurch, New Zealand*

[3] *EDGE of Existence Programme, Zoological Society of London, Regent's Park, London NW1 4RY, UK*

**Corresponding author j.rosindell@imperial.ac.uk*

Abstract

Phylogenetic metrics are essential tools in ecology, evolution and conservation. Phylogenetic diversity (PD) in particular is one of the most prominent measures of biodiversity, and is based on the idea that biological features accumulate along the branches of phylogenetic trees. We argue that PD and many other phylogenetic biodiversity metrics fail to capture an essential process that we term attrition. Attrition is the gradual loss of features through causes other than extinction, for example through natural selection. We introduce 'EvoHeritage', a generalisation of PD that is founded on the joint processes of accumulation and attrition of features, and can be applied to phylogenetic trees or more complex networks. Whilst PD measures evolutionary history, EvoHeritage is required to capture a more pertinent subset of evolutionary history including only features that have survived attrition. We show that EvoHeritage is not the same as PD on a tree with scaled branches; instead, accumulation and attrition interact in a more complex non-monphyletic way that cannot be captured by branch lengths alone. This leads us to speculate that existing phylogenetic trees and networks may be insufficiently flexible objects to capture the nuances of evolutionary processes. We derive a dimensionless measure of EvoHeritage that reproduces species richness and PD at opposite ends of a continuum based on the intensity of attrition. We suggest how the existing calculus of PD-based metrics and other phylogenetic biodiversity metrics could be recast in terms of EvoHeritage accumulation and attrition. We give three empirical applications that all rely on our new approach. The first is in ecology, evaluating EvoHeritage as a predictor of community productivity against species richness and PD. The second is in evolution, quantifying living fossils and resolving their associated controversy. The third is in conservation, using a partitioning of EvoHeritage to re-balance priorities at the broad scale of the complete tree of life.

Key words:

phylogenetic diversity, phylogenetic metric, diversity metric, biodiversity, feature loss, accumulation, attrition, EvoHeritage, evolutionary heritage, living fossil, community productivity, evolutionary distinctiveness

Introduction

Phylogenetic diversity (PD) is one of the most prominent measures of the variety of life, biodiversity. By taking phylogenetic branch lengths into account, PD goes beyond species richness and aims to capture feature diversity and more generally 'future options' for humanity (Faith, 1992). With biodiversity in crisis, PD was introduced by Faith (1992) to foster quantitative approaches to the problem of deciding where to focus limited conservation resources. PD has since found many practical applications in conservation (Isaac et al., 2007) and as a biodiversity indicator, for example as highlighted by the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (Brondizio et al., 2019). PD has become a mainstream feature of studies in ecology, evolution and conservation, and has been applied to taxa ranging from microbes (Bassett et al., 2015) to plants (Forest et al., 2007; Gonzalez-Orozco et al., 2016; Molina-Venegas et al., 2021) and vertebrates (Rosauer and Jetz, 2015; Gumbs et al., 2020). The PD measure has remained remarkably consistent with its original definition, which now underpins a large 'calculus' of related PD metrics (Faith, 2013), all derived from the same fundamental principles. The PD of a set of species on a tree is defined as the sum of the branch lengths needed to connect all those species (Faith, 1992) to the root of the tree (though explicitly not right back to the origin of life; Faith and Baker, 2006). PD is typically measured on dated trees and therefore in units of millions of years, but could be measured on trees that are not ultrametric or that have other units.

Alongside PD, other biodiversity metrics such as functional diversity (Tilman et al., 1997) and trait diversity have also been broadly applied, with a myriad of different approaches (Petchey and Gaston, 2006). The question of how well PD captures feature diversity (and, relatedly, functional diversity) has recently received increased attention from both data-driven (Mazel et al., 2017, 2018) and theory-focused (Wicke et al., 2021) studies. The answer undoubtedly depends on which features (e.g. traits or functional traits) are being considered and how they are being measured (Faith, 2018), leading to some debate (Owen et al., 2019; Mazel et al., 2019) and suggesting fertile ground for future research. However, there are also conceptual reasons (some newly highlighted here; some long-acknowledged) why PD might not capture feature diversity or future options for humanity. The rationale for defining PD as a sum of ancestral branch lengths is that biological features arise along these branches and are passed on to descendants (Faith, 2018). This is a cumulative process relying on an additive underlying mechanism assuming that features are never lost. Ancient features can, however, be lost through the process of evolution, especially if they are selected against (Behe, 2010; Carroll et al., 2008). We refer to this as 'attrition', encompassing the idea that features, future options, and all forms of biological information (biodiversity) may be lost along the paths of a phylogenetic tree without extinction occurring. A practical and well documented example of such attrition can be found at the genetic level, where the same areas of the genome can be overwritten multiple times. This leads to problems in the form of 'substitution saturation' when building phylogenetic trees: mutations are not precisely cumulative, which results in skewing the apparent length of some branches (Xia et al., 2003).

Homoplasy (convergent evolution) has been acknowledged from the start as a reason why PD may not always capture feature diversity (Faith, 1992); two species with divergent evolutionary histories sometimes have similar features. This may not be so troublesome for PD as one might assume because convergence may apply only to a (potentially superficial) subset of features. A blind snake has a remarkable resemblance to an earthworm, but closer inspection of their bodies would reveal the clear divergence of their less obvious features (Laver and Daza, 2021). Convergent evolution requires two different ancestor species to have evolved to become similar in certain respects. This may occur as a result of adding the same feature independently more than once, but can also

occur through the *loss* of a feature that once distinguished the two lineages. For example, the blind snake losing its sight is part of its path to convergence with the earthworm. Many examples of homoplasy may therefore be explained by attrition. In the cases where homoplasy occurs without attrition, the consequences for PD may also be less concerning because two (or more) independent, though convergent, solutions to the same adaptive need may indeed represent more underlying features than one solution. We therefore suggest that attrition, not homoplasy, is the real challenge for the classic PD narrative.

There are many further conceptual reasons pointing to the importance of attrition. Consider the lengths of the 'terminal branches' that connect each species to the rest of the tree and thus correspond to evolutionary steps that are only present in one living species. PD incorporates (as a minimum) the length of the terminal branch for any species under consideration, because that branch is required in a connection to any other place on the tree. It is clear that species with long terminal branch lengths contribute more unique features to biodiversity than others. But how much more? The tuatara (*Sphenodon punctatus*), an extraordinary reptile endemic to New Zealand, has a terminal branch length of 275 million years and many unique features. It seems plausible to us that it possesses more unique features than, say, the duck-billed platypus (*Ornithorhynchus anatinus*), with a terminal branch length of about 64 million years, but having over four times as many features, as PD would assert, seems disproportionate. More extreme examples are not hard to find; the Filasterea clade of just three species is sister to animals and choanoflagellates together (Shalchian-Tabrizi et al., 2008) and possibly over 900 million years old (Ferrer-Bonet and Ruiz-Trillo, 2017; OneZoom Core Team, 2021). This issue is mirrored by an idea in population genetics that net differences between groups will become saturated, given sufficient time, rather than grow indefinitely, even if evolution remains rapid (Crozier et al., 2005). Saturation was addressed at the stage of tree construction for Crozier's genetic diversity measure, where the branch lengths are bounded between 0 and 1, capturing the probabilities of a genetic difference (Crozier, 1992) meaning that branch lengths interacted in a multiplicative fashion, rather than additive as PD would assert. Framing in terms of attrition produces a similar effect because older features have more opportunities to be lost. As a result, deeper parts of the tree still add contemporary features, but have less weight than more recent parts of the tree. For terminal branches, these effects could be addressed within PD by re-scaling branch lengths appropriately. Indeed, branch lengths measured in years may not be the best choice for PD (Faith, 1994a) or other downstream analyses (Letten and Cornwell, 2015). For interior branches, however, the effects of attrition are more complex because features may be gained on one branch but lost on a different branch, resulting in a non-monophyletic pattern of those features.

Another consequence of PD's assumption that features only accumulate is that the survival of ancient features is guaranteed by the presence of just one living descendant. Consider the features that evolved along the stem branch of all birds that connects them to their common ancestor with crocodilians (explicitly not the additional features that evolved later along the descendant branches). For example, we expect flight to have evolved along this branch. Suppose now that all birds went extinct, except for just one species. According to PD, all the features from the stem remain captured by that one species, even if, say, that species was flightless (provided we have a non-avian species as well). This does not seem biologically reasonable. In conservation applications it could potentially undervalue the contribution of descendant species in capturing ancient features that have already been lost in other living descendants. This cannot be addressed by re-scaling branch lengths. In our example, the stem of birds would have to be shorter if we happened to be measuring a set of flightless birds, but not otherwise, so there isn't a single length that we could coherently give it to enable a classic PD-based analysis.

The same issue is apparent if we consider the PD of a mixed group of both extant and extinct species. The PD of all living archosaurs (birds and crocodilians together) is not

increased by the addition of an extinct species which sits on the branch connecting the birds to the crocodilians, such as the Archaeopteryx (or a close relative directly ancestral to birds). This is because the connecting branches of the living species already includes all their extinct direct ancestors. Yet, in contradiction, we would intuitively expect Archaeopteryx to add many features to the group that are not present in any living bird or crocodilian. Although extinct species may not seem relevant to some current applications of PD, paleontologists may benefit from applying the PD concept to extinct clades and fossil species; to the best of our knowledge, they have not done so thus far. Furthermore, inconsistency in a contrived example is likely to be a symptom of broader conceptual problems that are of more general concern. Incorporating attrition would change the role of extinct species because some of the features of an extinct species may be lost along all paths to their living descendants. The extinct species, if included in the set being measured, would therefore add the lost features and increase the diversity accordingly.

The idea that evolution may reverse a previous change has a long history in phylogenetic methods, for example in the Brownian motion model of trait evolution (Felsenstein, 1985). Such models generally have the aim of understanding the evolution of individual specified traits that have probably been measured empirically. The aim of PD is very different: to capture all features and future options for humanity, including traits we may not yet be aware of. It demands a different approach and a different body of theory. An alternative to PD incorporating feature loss has been briefly considered (Faith, 1994b), but there was no mathematical foundation to enable its practical use, and it was concluded that PD was more suitable because of its simplicity (Faith, 1994b).

In this manuscript, we generalise phylogenetic diversity and the associated PD calculus by introducing a suite of new measures to quantitatively incorporate the concept of 'attrition'. We adopt the term 'evolutionary heritage' that was first coined by Mooers et al. (2005) to describe PD-based measures with a narrative that is more appealing for conservation initiatives working at the national level. However, we use the term for a new purpose: to highlight an important conceptual distinction between evolutionary history and evolutionary heritage. We take evolutionary heritage (hereafter EvoHeritage) to mean features, traits, functions, genetic sequences, future options and biological information in the most general sense, but explicitly only the components that are heritable and present in a given set of organisms, that is the components that have survived attrition. In contrast, evolutionary history may mean the number of years over which lineages have been independently evolving, or all the historic features that ever existed, whether or not they survive attrition. We assert that EvoHeritage is conceptually preferable to evolutionary history as a proxy for features and future options for humanity and thus as a target for practical conservation.

We will derive an expression for the total EvoHeritage of a collection of taxa, species or populations, and show how the concepts of accumulation and attrition translate through to the rest of the PD calculus with selected examples. We leave provision for EvoHeritage to be applied to more general structures than species trees and show how in a special case, it directly captures an existing measure of genetic diversity (Crozier, 1992). Beyond providing a generalised alternative to the PD calculus, we show that EvoHeritage facilitates new applications. We introduce a new measure of EvoHeritage *difference*, in which the difference between two distinct groups becomes saturated over time rather than growing without a bound. This measure may serve as an alternative to branch lengths in other phylogenetic analyses.

The concept of EvoHeritage allows us to define a scaled (dimensionless) quantity, denoted ϕ_ρ, which can be calculated on any directed acyclic graph with dated vertices (a generalisation of a dated tree). The quantity ϕ_ρ can thus be swapped for PD in existing applications and, depending on the value of its associated parameter ρ, encompasses both PD (when $\rho \to 0$) and species richness (when $\rho \to \infty$) at opposite ends of a continuum. We

demonstrate the use of ϕ_ρ as a predictor of community productivity compared with species richness and PD, using data from the Cedar Creek grassland experiments (Tilman et al., 1996). We also introduce methods to consider the fate of just the EvoHeritage that evolved during a particular period in geological history. This enables us to identify so-called 'living fossils' (Grandcolas and Trewick, 2016) as concentrated groups preserving the rarest ancestral features in the face of attrition. We derive an expression for 'partitioned EvoHeritage' which gives a way to fairly share EvoHeritage between groups, akin to the 'evolutionary distinctiveness' (ED) measure from Isaac et al. (2007). We illustrate this with an example application in conservation, showing that attrition is essential for proportionate comparisons among species at the scale of the whole tree of life. We conclude by identifying future research that we hope EvoHeritage and the concept of attrition will facilitate.

Introducing EvoHeritage

We introduce EvoHeritage in the first instance with a glossary of the new terms required to understand it. The glossary is in logical (rather than alphabetic) order to facilitate easier reading. We provide a graphical summary of many principles from the glossary in Figure 1 and some simple illustrative examples are shown in Figure 4.

Glossary

DAG: A directed acyclic graph, 'DAG', is a standard mathematical object consisting of a set of vertices connected by directed edges in which there is no path along the edges from any vertex back to itself. A DAG is a generalisation of a rooted tree where leaves and nodes are represented by vertices and branches are represented by edges directed away from the root.

Root vertex: A vertex of a DAG which has no edges directed towards it, equivalently that means it has no ancestral edges.

Terminal vertex: A vertex of a DAG which has no edges directed away from it, equivalently that means it has no descendant edges. In a tree, a terminal vertex may be called a leaf.

Interior vertex: A vertex of a DAG that is neither a terminal vertex nor a root vertex.

Terminal edge: An edge of a DAG directed towards a terminal vertex.

Path: Any set of edges that define a directed route between two given vertices of a DAG.

$\hat{\boldsymbol{o}}$: A root vertex corresponding to the origin of life.

EvoHeritage: features, traits, functions, genetic sequences, future options and biological information in the most general sense, but explicitly only the components that are heritable and present in a given set of vertices. EvoHeritage may be measured as a continuous or discrete quantity.

Accumulation: A natural process by which novel EvoHeritage is generated along edges with copies being passed on to descendants.

Attrition: A natural process by which EvoHeritage is lost over time without extinction

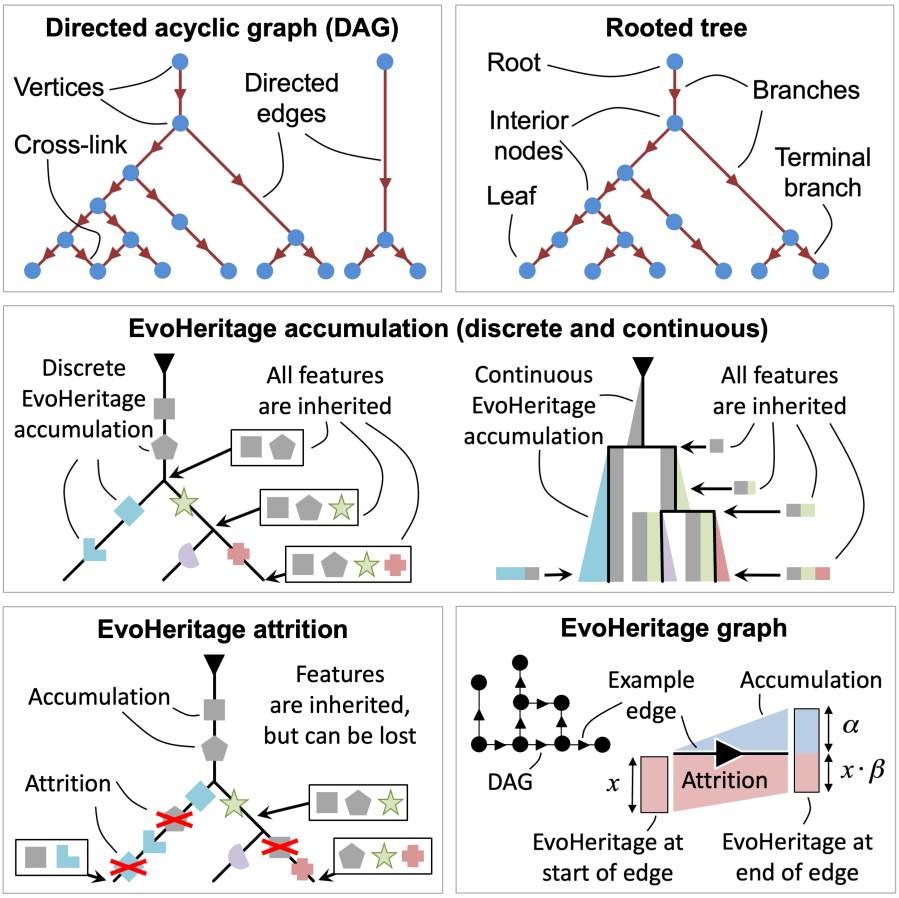

Fig. 1. An illustration of key terms and concepts from the glossary. Some concepts are clearer to visualise with discrete measurement of EvoHeritage and others are clearer with continuous EvoHeritage. In practice, if accumulation is sufficiently rapid compared to edge lengths, the distinction between discrete and continuous EvoHeritage will not be so important. This EvoHeritage graph shows only one form of EvoHeritage for simplicity, whereas the derivations permit a set of forms of EvoHeritage, each with its own rules for accumulation and attrition.

having occurred, for example, by not being passed on to any surviving descendants.

EvoHeritage graph: A DAG where the accumulation and attrition of EvoHeritage occurs in a specified way along each edge.

EvoHeritage form: EvoHeritage is considered to be of the same 'form' if it is subject to the same rules of accumulation and attrition on an EvoHeritage graph. An EvoHeritage graph may thus specify one or more forms of EvoHeritage, each with its own rules of accumulation and attrition.

Species: The species concept is optionally incorporated by identifying species as either a single vertex or as a set of vertices in an EvoHeritage graph.

ϕ: For a given set of vertices on an EvoHeritage graph, the total quantity of EvoHeritage present on those vertices.

Evolutionary History: For a given set of vertices on an EvoHeritage graph, the total quantity of EvoHeritage that has ever been present on those vertices or any of their ancestors. This is equivalent to the PD of the given set of vertices, including the root, which we take to be $\hat{o}$ by default.

Unique-$\boldsymbol{\phi}$: For two given sets of vertices on an EvoHeritage graph, this gives the amount of EvoHeritage that is present within the first set of vertices and absent from the second set of vertices.

Diff-$\boldsymbol{\phi}$: For two sets of vertices on an EvoHeritage graph, diff-ϕ is the proportion of total EvoHeritage that is only present within one of the two sets.

Partitioned-$\boldsymbol{\phi}$: A partitioning of EvoHeritage between vertices on an EvoHeritage graph, where the value of any part of EvoHeritage is divided equally among all copies of it.

Dated EvoHeritage graph: An EvoHeritage graph where all vertices also have dates attached.

Extant vertex: A terminal vertex with date $= 0$ on a dated EvoHeritage graph.

Extinct vertex: Any vertex with date > 0 on a dated EvoHeritage graph.

Ancestral EvoHeritage graph: This is constructed from a dated EvoHeritage graph by artificially augmenting EvoHeritage accumulation on certain edges in order to study the fate of targeted ancestral EvoHeritage. For example, removing EvoHeritage accumulation except at a key moment in geological history.

Living fossil: An extant vertex with a large share of rare ancestral EvoHeritage. After total EvoHeritage has been partitioned among the extant vertices of a dated ancestral EvoHeritage graph, those vertices with a high value are living fossils.

Standard conditions: For a DAG with additional edges lengths (akin to a tree with branch lengths), standard conditions generate an EvoHeritage graph according to a simplified set of rules. Specifically, (i) all EvoHeritage is of the same form and measured as a continuous quantity, (ii) EvoHeritage accumulates at a fixed deterministic rate of one unit of EvoHeritage per unit of edge length and (iii) attrition occurs at a fixed deterministic rate of ρ per unit of EvoHeritage per unit of edge length. The rate of attrition is thus proportionate to the total amount of EvoHeritage.

Standardised units: Under standard conditions (attrition at rate ρ), the total EvoHeritage at the end of a single disconnected edge with a given standard length. One unit of edge length is a convenient standard length to use by default. For dated EvoHeritage graphs this means using one million years. The standard length could be chosen less arbitrarily as the time back to $\hat{o}$, with $\hat{o}$ included as a vertex in the calculations. Alternatively, the median (or some other property) of the distribution of terminal edge lengths of species could be chosen in order to make standardised units relatable to the familiar species concept.

$\boldsymbol{\phi_\rho}$: A convenient notation for ϕ under standard conditions (attrition at rate ρ) and measured in standardised units on an EvoHeritage graph. It can thus be used in the same way as PD for any tree and a given $\rho \geqslant 0$. We may also write ϕ_ρ in place of ϕ for the other metrics involving ϕ. For example unique-ϕ_ρ.

Standard stochastic conditions: For an EvoHeritage graph with edge lengths, standard stochastic conditions consider one form of EvoHeritage measured in discrete units. EvoHeritage accumulation and attrition are defined in a probabilistic way so that the EvoHeritage of a set of vertices becomes a distribution of values rather than a single deterministic value. Specifically, (i) EvoHeritage accumulation occurs stochastically at a rate of one per unit of edge length so that the accumulation of EvoHeritage units corresponds to a Poisson process and (ii) attrition occurs independently for each EvoHeritage unit, with a fixed probability ρ per unit of edge length.

Rationale

We define EvoHeritage in general terms to support various future applications (please refer to the glossary). For greater generality, we use a directed acyclic graph (DAG) rather than a tree as the underlying structure and avoid referring to 'species' or 'higher taxa'; we imagine that ultimately species may equate to either a single vertex or a set of vertices in the DAG. Monotomies are permitted in a DAG and can be used to pinpoint ancestors that are significant for downstream analyses but are not most recent common ancestors and thus would not otherwise be identified as a vertex. Polytomies are permitted, though in most cases it would be better to resolve these (Kuhn et al., 2011) to avoid misinterpreting uncertainty as increased diversity. Multiple roots (vertices without any parent) are permitted and may, for example, correspond to independent origins of viruses coming from the tree of life, or multiple abiogenesis events. One can also interpret a posterior distribution of trees connected at their terminal vertices as a single DAG with multiple roots. Multiple independent routes through the DAG are permitted and may, for example, represent endosymbiosis, introgression, hybrid speciation or even gene flow between distinct populations of the same species.

We define an EvoHeritage graph as a DAG that specifies the accumulation and attrition of EvoHeritage explicitly, and thus contains all the data required to calculate EvoHeritage for any given set of vertices. EvoHeritage graphs are novel objects, so would have to be constructed in a bespoke manner, depending on the research objectives and available data. In most applications, a researcher would take an existing phylogeny with edge lengths and then specify (or estimate) a rate of attrition (ρ) to use the EvoHeritage calculus under standard conditions.

There may be different 'forms' of EvoHeritage potentially represented in an EvoHeritage graph, each with its own characteristics of accumulation and attrition. For example, we may have 'immortal' EvoHeritage as one form that is never lost to enable rare and truly irreversible evolutionary steps to be accounted for. Introducing the EvoHeritage graph concept enables a clear conceptual separation between the core EvoHeritage calculations (products of an EvoHeritage graph) and auxiliary details about the precise way that EvoHeritage is assumed to accumulate and become lost (e.g. building an EvoHeritage graph from a tree), something that is more likely to change in different applications. We leave open the possibility for EvoHeritage to be continuous or discrete, the latter going hand in hand with the possibility that EvoHeritage may be a stochastic distribution of values rather than a fixed value and thus may incorporate a mechanism for quantifying its own limits as a proxy for other biodiversity measures.

The quantity ϕ_ρ (derived later) is introduced as a special version of EvoHeritage (ϕ) that is dimensionless and can be calculated from a DAG (or tree) with edge lengths, and a value for parameter (ρ) representing attrition. It can therefore be directly swapped with PD in existing calculations on a tree with edge lengths for any given ρ. The idea was

inspired by Hill numbers that incorporate an 'order' or q parameter and recover a range of pre-existing metrics (or transformations of them) at special values of q (Hill, 1973; Jost, 2006). ϕ_ρ is measured in standardised units: multiples of the net amount of EvoHeritage that accumulates along a single unit of edge length, so formally, ϕ_ρ is dimensionless. We recommend measuring ϕ_ρ on a DAG that includes the origin of life as a vertex $\hat{o}$ (see supplementary material). This would enable groups with different crown ages to be compared without bias. Measuring ϕ_ρ in terms of standardised units enables it to converge precisely to a count of terminal edges (species richness in most cases) in the limit of infinite attrition ($\rho \to \infty$). Conversely, the value of ϕ_0 recovers PD. The value of $\phi_0(v)$ for any single extant vertex v on a dated tree, will give the total edge length to the root vertex. If the root is $\hat{o}$, then this is subtly different from PD (Faith and Baker, 2006) and is, in our view, more consistent with the concept of 'evolutionary history': any single extant species (or organism) arguably represents about four billion years of evolutionary history regardless of its position in the tree of life.

Derivation of EvoHeritage, ϕ, for a set of vertices on an EvoHeritage graph

In this section we give the formal definition of the EvoHeritage, ϕ, of a set of vertices (Eqn. 2 below). Throughout this manuscript, we describe an EvoHeritage graph mathematically by letting V represent the set of vertices, E be the set of directed edges, and I be the set of EvoHeritage forms. For a directed edge $e = (v_1, v_2)$ in E, let $d^V(e) = v_2$, the vertex descending from e, and let $a^V(e) = v_1$, the immediate ancestor vertex to e. For a vertex v in V, let $D^E(v)$ be the set of edges descending from v and let $A^E(v)$ be the set of ancestor edges to v. Thus $a^V(e) = v$ if and only if $e \in D^E(v)$, and similarly, $d^V(e) = v$ if and only if $e \in A^E(v)$.

For each form of EvoHeritage $i \in I$, let $\alpha_i(e)$ give the net amount of EvoHeritage of form i accumulated along each edge $e \in E$. Let $\beta_i(e)$ give the proportion of EvoHeritage of form i present at the start of the edge $e \in E$ that survives attrition along the edge to still be present at vertex $d^V(e)$.

To derive ϕ for a set A of vertices, we first iteratively calculate a term $p_i(v, A)$ for every vertex v in V. The term $p_i(v, A)$ represents the proportion of EvoHeritage of form $i \in I$ associated with the vertex v that survives to be counted on at least one vertex in the set A. Our Eqn. 1 gives the formula for calculating these values, where the iteration proceeds in a way such that $p_i(v, A)$ is calculated only after the values on all vertices descending from v have been calculated. We require the EvoHeritage graph to be a DAG (rather than any network which may have cycles) because this enables the iterative calculation of $p_i(v, A)$, and naturally reflects the unidirectional passage of time in evolution.

$$p_i(v, A) = \begin{cases} 1 & \text{when } v \in A; \\ 0 & \text{when } v \notin A \text{ and } D^E(v) = \emptyset; \\ 1 - \prod_{j \in D^E(v)} [1 - p_i(d^V(j), A) \cdot \beta_i(j)] & \text{otherwise.} \end{cases} \tag{1}$$

Should two or more copies of any EvoHeritage arrive at the same vertex (via multiple paths through the EvoHeritage graph), each copy is handled independently. This enables each vertex and each section of EvoHeritage to be treated independently. We are now in a position to write an expression for ϕ as a function of a set of vertices A.

$$\phi(A) = \sum_{i \in I} \sum_{e \in E} \alpha_i(e) \cdot p_i(d^V(e), A) \tag{2}$$

In words, ϕ of a set of vertices A is the sum, across all forms of EvoHeritage (first summation term) and across all edges (second summation term) in the EvoHeritage graph. Inside the sum is the EvoHeritage accumulated along each each edge $\alpha_i(e)$ multiplied by the proportion of that EvoHeritage surviving attrition to be counted in the set A.

Derivation of ϕ_ρ

In this section, we provide a simple formula for ϕ_ρ (Eqn. 8 below); however, to motivate and justify this definition, we require some preliminary theory. Let us first apply standard conditions to an EvoHeritage graph with specified edge lengths. We express this with an additional function $L(e)$ giving the length of each edge $e \in E$. Attrition under standard conditions occurs at rate $\rho \geqslant 0$ per unit of EvoHeritage per unit of edge length. Therefore, the proportion of EvoHeritage surviving attrition, $\mathrm{R}(l)$, as a function of edge length l, satisfies the differential equation:

$$\frac{d}{dl}\mathrm{R}(l) = -\rho \cdot \mathrm{R}(l), \quad \mathrm{R}(0) = 1. \tag{3}$$

Solving this equation yields:

$$\mathrm{R}(l) = e^{-\rho l}. \tag{4}$$

Note that here, $e = 2.71828\ldots$ and does not refer to an edge. As standard conditions include only one form of EvoHeritage, we exclude the EvoHeritage form subscripts here for convenience. We first derive $\beta(e)$, which gives the proportion of EvoHeritage that survives attrition along the edge $e \in E$. We use $l = L(e)$ in Eqn. 4 to write:

$$\beta(e) = e^{-\rho L(e)}. \tag{5}$$

We must now derive an expression for $\alpha(e)$ which includes EvoHeritage accumulation along the edge e but also attrition of the EvoHeritage generated earlier along the edge to give a net accumulation value. To do this, we integrate over edge e. This accounts for the EvoHeritage generated at every point along the edge, each having a different length over which it can be lost before reaching the end vertex. Given that EvoHeritage accumulation happens at a constant rate of 1 under standard conditions, we obtain:

$$\begin{aligned} \alpha(e) &= \int_0^{L(e)} 1 \cdot e^{-\rho l} dl \\ &= \left(\frac{e^{-\rho L(e)}}{-\rho} - \frac{e^{-\rho \cdot 0}}{-\rho} \right) \\ &= \frac{1 - e^{-\rho \cdot L(e)}}{\rho}. \end{aligned} \tag{6}$$

Thus for a DAG with edge lengths and a value of ρ, we can define an EvoHeritage graph by using Eqns. 5 and 6 for each edge of the DAG. We can then calculate ϕ via Eqns. 1 and 2 together with our EvoHeritage graph. The standardised unit S_u, based on an edge of unit length from Eqn. 6, can then be written as:

$$S_u = \frac{1 - e^{-\rho}}{\rho}. \tag{7}$$

Finally, for a set of vertices A, we define $\phi_\rho(A)$ as:

$$\phi_\rho(A) = \frac{\phi(A)}{S_u}. \tag{8}$$

Where $\phi(A)$ is calculated on an information graph constructed from a DAG with edge lengths using the value of ρ and Eqns 5 and 6. In the supplementary material we show that: (i) adding or taking away a 'monotomy' vertex does not affect the results of EvoHeritage calculations on the other vertices, and (ii) ϕ_ρ is independent of the fixed rate of EvoHeritage accumulation per unit of edge length, so there is no loss of generality when assuming a rate of 1, as under standard conditions.

Relationship between ϕ_ρ and PD

First, we consider how ϕ_ρ is related to PD and species richness. The extreme values of ρ, in the limit, give:

Theorem 1 Let $G = (V, E)$ be an EvoHeritage graph, and let $A \subseteq V$.

(i) $\lim_{\rho\to\infty} \phi_\rho(A) = \sum_{v\in A} |A^E(v)|$

(ii) $\lim_{\rho\to 0} \phi_\rho(A) = PD(A)$

In words: $\phi_\rho(A)$ converges to a count of edges that lead to vertices in A (species richness if the vertices represent species in a tree) when $\rho \to \infty$; and $\phi_\rho(A)$ converges to the PD of the set of vertices when $\rho \to 0$. The proof of the theorem is given in supplementary material.

Let us now consider whether EvoHeritage can be considered as a restatement of PD with appropriately chosen edge lengths. We say that PD corresponds to ϕ on a tree T if we can calculate some edge lengths $L(e) > 0$ on T so that, for every possible set A of terminal vertices, $PD(A)$, based on summation of the newly calculated edge lengths, gives the same value as $\phi(A)$ on T. It is clear that if $\beta(e) = 1$ everywhere on T, there is no attrition and ϕ corresponds to PD by setting $L(e) = \sum_{i\in I} \alpha_i(e)$. Supposing that the special case of zero attrition does not apply, we can consider which arrangements of edges and vertices in T would permit PD to correspond to ϕ.

Theorem 2 Let T be an EvoHeritage tree with terminal vertices given by $X \subset V$ and one EvoHeritage form with $\beta(e) < 1$ for every edge $e \in E$. In that case, $\phi(A) = PD(A)$ for all sets $A \subseteq X$ if and only if no edge is ancestral to more than two terminal vertices.

The proof of Theorem 2 follows as a consequence of a more general result proved in the supplementary material. In simpler terms, the theorem shows that PD can only correspond to ϕ on a tree when either attrition is explicitly turned off, or where the tree has a very simple shape: at most a polytomy at the root with each edge only branching at most once more. Such trees will not be encountered even in the simplest real-world applications. It can also be shown that a set (of a fixed number of vertices) chosen to maximise PD will not always capture the maximal ϕ_ρ (see supplementary material). An intuitive consequence of this section is that provided that attrition is small on the scale of the stem age of a tree, then PD remains a reasonable approximation for ϕ. Conversely, if attrition is large or we are studying an old tree with deeper splits, then the disparity between ϕ and PD is likely to be more pronounced and cannot be fully rectified by just re-scaling branch lengths. To make the intuition of this clearer, PD no longer corresponds to ϕ if it is possible for

accumulation and attrition to leave a non-monophyletic group that shares the same EvoHeritage (see Figure 2). These findings are consistent with Wicke et al. (2021) who highlight that PD with suitable branch lengths can correctly capture feature diversity only when features appear in a pattern that could have arisen without any feature loss.

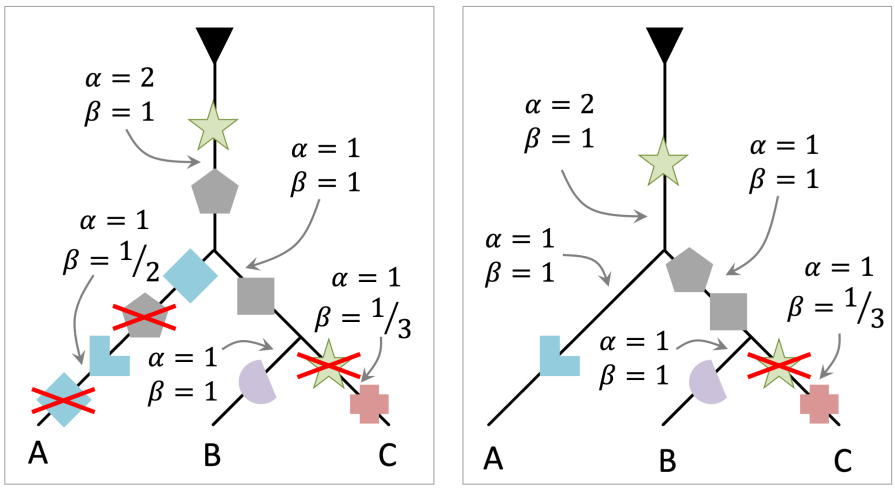

Fig. 2. Whilst neither PD nor EvoHeritage identifies individual features, it helps to consider the fate of individual features on a simple tree to appreciate why it is impossible to frame accumulation and attrition in terms of PD on a tree with scaled edge lengths. The left panel shows an example tree with features that are gained and lost on different branches. Certain types of feature loss can be absorbed by re-scaling edges. For instance, gain and loss of a feature along a single edge cancel out and shorten the effective length - this results in the α values on an EvoHeritage graph (e.g. the blue diamond on the terminal branch of vertex A). In other cases where feature gain and loss leaves a mono-phyletic group with that feature, we could express the gains and losses as a single gain at the stem of that group (e.g. the grey pentagon on the stem). The right panel shows a new equivalent tree where some of the attrition ($\beta < 1$) has been eliminated by re-scaling of α values on each edge. We cannot eliminate the attrition of the green star feature however because this forms a non-monophyletic group. If we try to add it separately to the terminal edges of both A and B then it presents as homoplasy and we may double count the same feature. If we try and compensate for the loss by reducing the α for the terminal edge of vertex C then we risk penalising for loss of a feature that survives on another vertex.

Relationship between EvoHeritage under standard stochastic conditions, feature diversity and PD

Our standard stochastic conditions consider one form of EvoHeritage measured in discrete units. These may be thought of as features (binary characters) which arise and are lost along the edges of a phylogenetic tree. The features exhibited by a vertex are those that still persist on it. Therefore, under standard stochastic conditions, ϕ can be thought of as a direct measure of feature diversity, under an explicit model of feature gain and feature loss. To distinguish standard stochastic conditions from the deterministic ones, we use new variables. Feature gains and losses are described by exponential distributions, with the respective rates $\lambda(e)$ and $\mu(e)$ along an edge e. We illustrate standard stochastic conditions on a tree (not a DAG) in order to facilitate a comparison with PD. We assume that the accumulation and loss of a feature is independent of the presence or absence of other features. We also require each feature to be distinct from all others. Thus the development of a new feature is not conditional on the number of features already arisen. We frame this as a 'constant birth' process. The rate at which features are lost does,

however, depend on how many features exist, so we have a 'linear death' process. We model these processes by a continuous time Markov chain, using constant birth and linear death rates, as shown in Fig. 3.

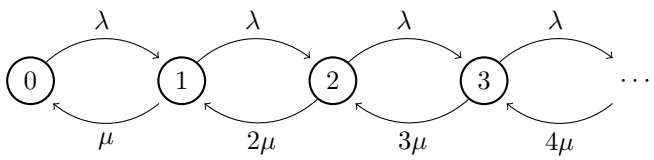

Fig. 3. A simple Markov chain showing the net number of discrete features gained (in the circles) and the stochastic rate of movement between states. The diagram illustrates a single edge of arbitrary length. The rate of EvoHeritage accumulation is λ per unit of time and the rate of attrition is μ per feature per unit of time. The overall rate of attrition per unit time thus increases proportionate to the number of features we have, hence the multiples of μ on the bottom row. The number of features that originate on a very long edge would thus increase and eventually fluctuate around a equilibrium distribution

Let T be an EvoHeritage tree (a tree with EvoHeritage accumulation and attrition specified akin to an EvoHeritage graph) and let $X \subseteq V$ represent the set of terminal vertices. Then ϕ of a set $A \subseteq X$ is the total quantity of EvoHeritage (features) present in A. This is directly equivalent to feature diversity (FD), measured as a count of features. To formalise this, let $\mathbb{F}_A = \bigcup_{x \in A} \mathbb{F}_x$ and let $\phi(A) = FD(A) = |\mathbb{F}_A|$. Similar to the deterministic setting, we find that PD rarely matches ϕ and FD. More precisely Theorem 3 shows that, even in the simple setting of an EvoHeritage tree, FD and ϕ cannot be described by PD for *any* choice of edge lengths on the tree, except in certain special cases.

Theorem 3 Let T be an EvoHeritage tree with vertex set V and edge set E where $X \subseteq V$ is the set of terminal vertices, and $r \in V$ is the root vertex. Let $L : E(T) \to \mathbb{R}_{>0}$ and $L' : E(T) \to \mathbb{R}_{>0}$ be two assignments of positive edge lengths for T.

(i) Suppose that at least one edge of T is ancestral to more than two terminal vertices. If $\mathbb{E}[FD_{T,L}(A)] = PD_{T,L'}(A)$ for every set of leaves $A \subseteq X$, then $\mu = 0$.

(ii) Conversely, if $\mu = 0$ and $L(e) = cL'(e)$ for every edge $e \in E(T)$, then $\mathbb{E}[FD_{T,L}(A)] = c\lambda PD_{T,L'}(A)$ for every set of leaves $A \subseteq X$, and any tree T.

The proof of this result is included in the supplementary material.

Developing a complete EvoHeritage calculus

PD refers not only to the original definition (the minimum spanning tree of a set of vertices) but also to a whole 'calculus' of derived measures based on the same underlying principles. To replicate the entirety of the PD calculus for EvoHeritage is beyond the scope of this work; however, we can begin the task and demonstrate that the rest will be achievable through future work. For worked examples of parts of the EvoHeritage calculus, refer to Figure 4.

Unique-ϕ

Unique EvoHeritage (unique-ϕ) is the EvoHeritage captured by a subset A of vertices in a given EvoHeritage graph, and not captured by any of the vertices in a second subset B of vertices in the graph. This is a generalised counterpart to 'PD loss' and 'PD

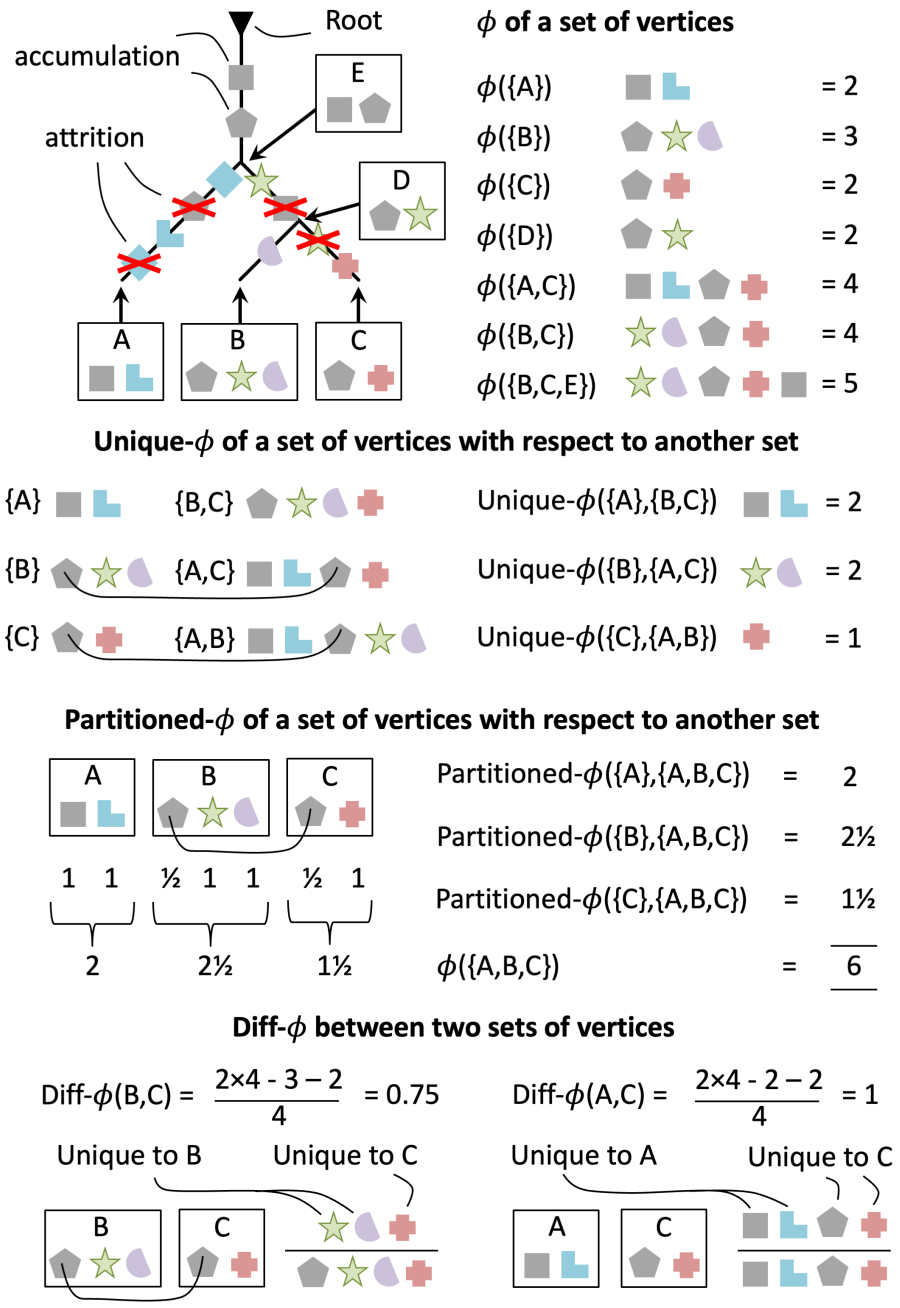

Fig. 4. Illustrative examples of EvoHeritage, ϕ, unique-ϕ, paritioned-ϕ and diff-ϕ. This example uses discrete EvoHeritage elements on the tree as this makes it easier to picture the underlying principles. The reader can imagine that if there were very large numbers of such discrete EvoHeritage elements, the result would be a system where EvoHeritage is effectively a continuous quantity.

endemism' (Faith, 2013), and is calculated in a similar manner. The need to define the set B more explicitly for unique-ϕ arises because extinct vertices are now potentially included. For most applications, we would expect researchers to choose that B contains all extant vertices that are not in A.

$$\text{unique-}\phi(A, B) = \phi(A \cup B) - \phi(B) \tag{9}$$

EvoHeritage difference, diff-ϕ

The EvoHeritage difference diff-ϕ between two subsets (A and B) of vertices is the EvoHeritage present within only one of the two sets, expressed as a proportion of the total EvoHeritage present in both subsets together. This can be written as follows:

$$\text{diff-}\phi(A, B) = \frac{\text{unique-}\phi(A, B) + \text{unique-}\phi(B, A)}{\phi(A \cup B)} \tag{10}$$

Note that diff-ϕ is symmetric so that diff-$\phi(A, B)$ =diff-$\phi(B, A)$. This metric is based on similar principles to the 'UniFrac' metric (Lozupone and Knight, 2005), which, in turn, is a phylogenetic version of the Jaccard index. Under standard conditions with $\rho \to 0$, diff-$\phi(A, B)$ will converge precisely to the same result as UniFrac. The Phylosor metric of phylogenetic difference (Bryant et al., 2008) is similar to UniFrac but is based on the Sørensen index; an EvoHeritage counterpart along similar lines to diff-ϕ would be easy to conceive.

In the supplementary material, we explore the relationship among diff-$\phi(\{A\}, \{B\})$, the difference between two individual vertices, and phylogenetic distance (the date of common ancestry between a pair of vertices). We also show there that under standard conditions, diff-ϕ between a pair of vertices saturates as the time since common ancestry increases. As $\rho \to 0$, we see that diff-$\phi(\{A\}, \{B\})$ converges to twice the date of common ancestry divided by a normalisation factor. In the other limit, as $\rho \to \infty$, we see that diff-$\phi(\{A\}, \{B\})$ converges to unity.

Relationship to genetic diversity

EvoHeritage has been constructed to avoid obligatory ties to the species concept and species trees. It therefore has many prerequisites to be applied in population genetics and for measuring both within- and between-species genetic diversity. There are many different measures of genetic diversity in general use (Hughes et al., 2008), and these may not necessarily fit the context of EvoHeritage. However, one early genetic diversity metric conceived by Crozier (1992) has a quantitative equivalent in terms of EvoHeritage. Crozier considered an unrooted tree with edge lengths between 0 and 1 corresponding to the probability of a difference being generated along the edge. This is directly comparable with our values of β (attrition) on each edge but without any values of α (accumulation). Crozier's genetic diversity is the probability that all vertices will not be identical. If we add a stem to the tree in any location, then the proportion of ancestral EvoHeritage that does not survive attrition to exist on every extant vertex is the same as Crozier's genetic diversity (see supplementary material). This may seem counter intuitive but attrition is interpreted as a multiplicative source of difference, and so not being inherited equally on all extant vertices is a source of variation and diversity. One could, perhaps, more closely equate this special case of EvoHeritage to 'genetic *history*' (Crozier et al., 2005) which is equivalent to the genetic diversity of Crozier (1992) but defined on a rooted tree.

Power set approach

For a set A, we let $\mathcal{P}(A)$ denote the power set of A (the set of all subsets of A). Here, we use the power set to introduce a new notation for EvoHeritage. We will see how this paves the way for a more elegant description of the broader EvoHeritage calculus, and for the use of Monte Carlo methods in numerical EvoHeritage calculations.

Let $G = (V, E)$ be an EvoHeritage graph with vertex set V and edge set E. Intuitively, the functions we introduce here will be used to provide a measure of connection. Specifically, the connection described is that from a single vertex $v \in V$, to some set H, utilising a particular set of edges, say $G \in \mathcal{P}(E)$. More formally, a function F of this type can be written as $F : V \times \mathcal{P}(E) \times H \to \mathbb{R}$, where the set H varies depending on what we are trying to calculate. For example, if we seek to calculate ϕ using this approach, $H = \mathcal{P}(V)$, so $A \in H$ is a subset of vertices for which we are calculating ϕ. We now define a function $Y(F, A)$ which calculates the percolation of EvoHeritage through the graph based on the measure of connection provided by F and a value of $A \in H$.

$$Y(F, A) = \sum_{i \in I} \sum_{G \in \mathcal{P}(E)} \left(\prod_{g \in G} \beta_i(g) \prod_{g \in E \setminus G} (1 - \beta_i(g)) \sum_{e \in E} \alpha_i(e) \cdot F(d^V(e), G, A) \right). \tag{11}$$

The calculation of $Y(F, A)$ involves a sum over all forms of EvoHeritage given in the set I. Each form of EvoHeritage is independent of the other forms, and this is expressed as the first sum in Eqn. 11. The next term in Eqn. 11 corresponds to a sum over all possible subsets of edges in the DAG. The broad idea is to view an EvoHeritage graph with attrition as a weighted distribution of DAGs without attrition. To see this in an example, consider a DAG on which only one edge e_1 has any attrition, so $\beta_i(e) = 1,\ \forall e \in E \setminus \{e_1\}$. We could express a calculation regarding this DAG as a weighted average across two possibilities: first, with weight $\beta(e_1)$, a calculation of F on the DAG with the edge e_1 included (for EvoHeritage not lost along that edge); second, with weight $1 - \beta(e_1)$, the same calculation of F on the DAG with the edge e_1 removed. Scaling this thinking up to a case where any possible subset of edges may be disconnected by attrition in different scenarios, we obtain the remainder of the terms given in Eqn. 11. These multiply factors of $\beta(e)$ and $1 - \beta(e)$ for different edges to weigh across the possible combinations of attrition. The format of the function Y shows how the sum over $\mathcal{P}(E)$ can be replaced with a Monte Carlo method for numerical approximation (see 'methods—empirical application').

We will now describe a specific choice of function for F so that $Y(F, A)$ returns $\phi(A)$ as derived earlier (Eqn. 2). To do this, we again use the concept of a path of edges that connect a pair of vertices. For a set of edges $G \in \mathcal{P}(E)$, we wish to count the number of different paths which use only edges from within G, and connect from a vertex $v \in V$ to any vertex a in set $A \subseteq V$. To do this, we introduce a function $c : V \times \mathcal{P}(E) \times \mathcal{P}(V) \to \mathbb{Z}_0^+$

$$c(v, G, A) = |\{g \in \mathcal{P}(G) : g \text{ is a path from } v \text{ to any } a \in A\}| \tag{12}$$

To differentiate between the sets of edges which contain at least one path from v to A and those that contain no such path, we introduce an 'indicator function', $\mathbb{I} : V \times \mathcal{P}(E) \times \mathcal{P}(V) \to \{0, 1\}$, defined by:

$$\mathbb{I}(v, G, A) = \begin{cases} 1 & \text{when } c(v, G, A) \geqslant 1; \\ 0 & \text{when } c(v, G, A) = 0. \end{cases} \tag{13}$$

We can now express ϕ in terms of Y and $\mathbb{I}$.

Theorem 4 For all subsets $A \subseteq V$, $\phi(A) = Y(\mathbb{I}, A)$.

For a proof of Theorem 4, refer to the supplementary material.

Partitioned-ϕ

PD can be partitioned with evolutionary distinctiveness (Isaac et al., 2007) or equal splits (Redding and Mooers, 2006). Evolutionary distinctiveness shares the length of each edge evenly between all extant descendant vertices regardless of tree topology (Isaac et al., 2007). Here, we provide an equivalent for EvoHeritage. We seek to share out (partition) the 'value' of all EvoHeritage present on a set of vertices B fairly among the subsets of vertices. Consider the subset $A \subseteq B$. The partitioned-ϕ of A with respect to B is the share of the value $\phi(B)$ which is attributed to the vertices in A. To calculate this, we need to account for the number of copies of EvoHeritage across all vertices of A and across all vertices of B (Fig. 4). This is especially complex in an EvoHeritage graph because it is possible for a single vertex to inherit more than one copy of the same EvoHeritage via cross-links in the underlying DAG (but it is a biological advantage, since it permits hybrids and other forms of gene-flow that are challening in a classical PD framework). Such vertices with multiple copies of the same EvoHeritage receive a greater proportion of the associated value. The power set approach provides an elegant solution because attrition is accounted for automatically by weighted averaging in the function Y. We only need to write a function to partition the value of the EvoHeritage from one vertex passing through a known subset of edges.

Suppose $A \subseteq B$. The function $R(d^V(e), G, A, B)$ gives the ratio of the number of unique paths from the vertex descending from the edge e to any vertex in the set B that are also paths to a vertex in the set A.

$$R(d^V(e), G, A, B) = \begin{cases} \dfrac{c(d^V(e), G, A)}{c(d^V(e), G, B)} & \text{when } c(d^V(e), G, B) \geqslant 1; \\ 0 & \text{when } c(d^V(e), G, B) = 0. \end{cases} \tag{14}$$

Applying Y to the function R gives us partitioned-$\phi(A, B)$.

$$\text{partitioned-}\phi(A, B) = Y\left(R, (A, B)\right) \tag{15}$$

Many applications of partitioned-ϕ will use it to assign a value to each individual extant vertex $b \in B$ so that the sum of all partitioned-ϕ values returns the total EvoHeritage of all extant vertices together $\phi(B) = \sum_{b \in B} \text{partitioned-}\phi(\{b\}, B)$ (see supplementary material). Under standard conditions, partitioned-ϕ_ρ of a single vertex within the context of all extant vertices will converge precisely to ED as $\rho \to 0$.

Definition of Hill-ϕ and relationship to previous work connecting PD with Hill numbers

Given a community of S species with relative species abundances of p_i, the Hill number ${}^qD$ of order q is given by:

$${}^qD = \left(\sum_{i=1}^{S} p_i^q \right)^{1/(1-q)} \tag{16}$$

with the particular case of $q = 1$ given by:

$$^{1}D = \lim_{q \to 1} {}^{q}D = \exp\left(-\sum_{i=1}^{S} p_i \log p_i \right). \tag{17}$$

Hill numbers (Hill, 1973) provide a true measure of community diversity (Jost, 2006). Depending on the order parameter q, Hill numbers can recover species richness (^{0}D), the exponential of Shannon entropy (^{1}D), and the inverse of the Gini–Simpson index (^{2}D). Previous work has combined PD with species abundance data in a new range of metrics (Cadotte et al., 2010), and some of this work specifically uses Hill numbers (Chao et al., 2010, 2014). The 'phylogenetic diversity of order q through t years ago' (Chao et al., 2010) (hereafter Hill PD) depends not only on q, which is analogous to the order parameter of the original Hill numbers, but also on t. The value of t gives a cutoff date before which segments of any edge do not contribute to diversity and are excluded from calculations. Hill PD, denoted $^{q}PD\,(t)$, is defined by:

$$^{q}PD\,(t) = \left\{ \sum_{i \in B_t} L_i \left(\frac{a_i}{t} \right)^{q} \right\}^{1/(1-q)}, \tag{18}$$

where B_t gives all edges younger than t, L_i gives the length of edge i after removing any portions older than t, and a_i gives the relative abundance of edge i, which is the sum of the relative abundances of all descendant extant terminal vertices. Hill PD effectively treats each unit length section of edge as a 'thing' that is copied to every extant individual of every descendant vertex. It calculates relative abundances of these things and from that, Hill numbers according to the original formula (Eqn. 16) (Chao et al., 2010, 2014). Note that t is effectively acting as a normalisation factor to ensure that:

$$\sum_{i \in B_t} L_i \left(\frac{a_i}{t} \right) = 1. \tag{19}$$

By removing t from the inner brackets, we can equivalently write a new expression:

$$^{q}PD\,(t) = t^{-q/(1-q)} \cdot \left\{ \sum_{i \in B_t} L_i \left(a_i \right)^{q} \right\}^{1/(1-q)}. \tag{20}$$

The value of t in Hill PD cuts away deep parts of very long edges and permits like-for-like comparisons between different clades (provided they are older than t). It thus mitigates some of our concerns with the original PD measure treating ancient, recent, and shared ancestry equivalently. Nevertheless, it is predicated on the same fundamental assumption as PD, namely that all EvoHeritage (on the edges under consideration) is copied to all descendants and there is no attrition.

Following the same rationale for Hill-ϕ, we replace a unit length section of edge with a unit of EvoHeritage that is generated at a rate of 1 per unit of edge length. We use the power set approach to obtain:

$$^{q}\text{Hill-}\phi(A) = Y\left(\left(\frac{C(d^V(e), G, A)}{Y\left(C(d^V(e), G, A)\right)} \right)^{q} \right)^{1/(1-q)}. \tag{21}$$

Equivalently:

$$Y\left(C(d^V(e), G, A)\right)^{-q/(1-q)} \cdot Y\left(\left(C(d^V(e), G, A)\right)^{q}\right)^{1/(1-q)}. \tag{22}$$

For the purpose of simpler notation, we consider abundances to manifest on an EvoHeritage graph as one vertex per organism, with all vertices corresponding to the same species (or taxa / group) connected to a single parent vertex via edges with $\alpha = 0$ and $\beta = 1$ (i.e. with no accumulation or attrition). If we are considering species abundances and a population-level structure is known we could go one step further and encode within-species diversity by changing the values of α and β that connect conspecifics accordingly.

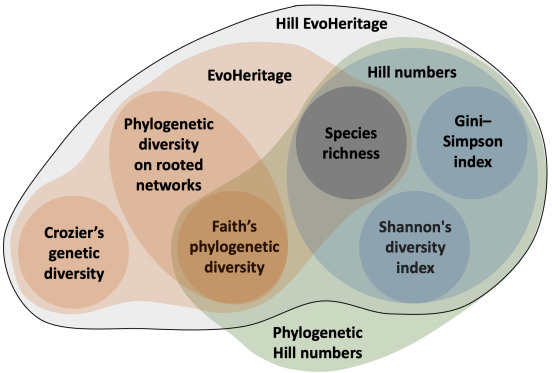

Fig. 5. Connections between overlapping concepts and generalisations. Blue corresponds to species abundance data; orange corresponds to phylogenetic data. Green corresponds to existing work connecting PD and Hill numbers.

Empirical applications –Methods

In this section, we will introduce three different empirical applications of our work that all rely on attrition: (i) an ecological application evaluating ϕ_ρ as a predictor of community productivity, (ii) an evolutionary application using the concepts of attrition and partitioned EvoHeritage to quantify living fossils, and (iii) a conservation prioritisation application comparing partitioned EvoHeritage and evolutionary distinctiveness across the tree of life. The results and discussion of each application will be given in later sections.

i) Evaluating ϕ_ρ as a predictor of community productivity

The field of biodiversity-ecosystem function (BEF) has yielded important insights for conservation, perhaps the most fundamental being that, in general and all other things being equal, greater biodiversity is associated with higher ecosystem function (Balvanera et al., 2006; Grace et al., 2016). A major contribution of phylogenetic conservation has been the general trend of phylogenetic diversity to correlate strongly with ecosystem productivity and, in notable cases, to do so more strongly than species richness or even functional diversity (Cadotte et al., 2009; Tucker et al., 2019). Perhaps the most influential study of these kinds of effects, by Cadotte et al. (2009), evaluated a range of other diversity metrics, including species richness, PD and functional diversity. Using biomass and species lists from each of the 13m×13m plots of the Cedar Creek Biodiversity Experiment (itself an iconic system in this strand of the literature; Tilman et al., 1996,

2001), the authors ranked 16 different individual predictors by their ability to explain plot-scale community productivity, measured as biomass accumulation. PD was second only to the presence of nitrogen fixers, and was an improvement over both species richness and a range of other functional diversity measures.

We repeat parts of the analyses of Cadotte et al. (2009), introducing ϕ_ρ alongside the PD and species richness predictors. One aim is to demonstrate our theoretical results (Theorem 1) with empirical data, showing that ϕ_ρ scales naturally between species richness and PD, depending on the strength of attrition.

We sourced the Cedar Creek Biodiversity Experiment data from (Tilman, 2021) to provide both plot biomass and a matrix of species presence by plot, and the phylogenetic tree directly from Cadotte et al. (2009). All analyses were performed in R (R Core Team, 2021). We used the chronos function within the R ape package (Paradis et al., 2004) to make the tree ultrametric, similar to the original study (Cadotte et al., 2009). The tree was rate-smoothed (Sanderson, 2002) setting the total tree depth to 1 for comparability with Cadotte et al. (2009). We developed an R function to iteratively calculate ϕ_ρ under standard conditions using Eqns. 2, 5 and 6 (which we release here and will release subsequently in *pez*; Pearse et al., 2015). We then applied this function to calculate ϕ_ρ for every one of the empirical plots. We repeated this process for a range of attrition parameters between $\rho = 10^{-4}$ and $\rho = 10^4$. To ensure we would see both the PD and species richness limits, and to help verify correctness of our new function, we also calculated PD and species richness for each plot directly using *pez* (Kembel et al., 2010; Pearse et al., 2015).

Given that the phylogeny of species present in the Cedar Creek plots has a single deep split dividing the true grasses *Poaceae* (8 species) from the *Pentapetalae* (10 species), we calculated the same diversity metrics for just the *Poaceae* and again for just the *Pentapetalae* as well as for both groups together. We calculated a Pearson correlation coefficient separately for plot biomass against every one of our plot diversity measures: PD, species richness and ϕ_ρ with each value of ρ, for all plants, for *Poaceae* only and for *Pentapetalae* only. We discuss our findings in the Results section.

ii) Applying attrition and partitioned-ϕ to quantify living fossils

The idea of 'living fossils' is contentious (Casane and Laurenti, 2013; Bennett et al., 2018; Lidgard and Love, 2021). Proponents of the concept say that living fossils are useful as motivation for a wide range of research questions (Lidgard and Love, 2018) and as a focus for conservation attention (Turner, 2019). One reason for the contention is perhaps the lack of a clear definition of a 'living fossil' and the wide range of different identifying characteristics that are mentioned across numerous studies (Bennett et al., 2018). Commonly claimed features of living fossils are slow rates of evolution, "gross similarity to an ancestral fossil" (Lidgard and Love, 2018, 2021) and long periods of morphological stability (Turner, 2019). Opponents argue that these characteristics are in conflict with mainstream evolutionary thinking (Casane and Laurenti, 2013) and reinforce misinterpretations about evolution because they appear to support the idea of a 'ladder of progress' (Omland et al., 2008), where some extant species can be considered more 'primitive'. It is also argued that even the most iconic living fossils, such as the coelacanths, do not really display morphological stability or slow rates of change at all (Casane and Laurenti, 2013). These are strong arguments, but, at the same time, it is hard to simply dismiss a large body of work on living fossils outright. It seems likely that it should be possible to distil some special property that connects all those species and clades that have been given the title.

Some attempt has been made to resolve the controversy surrounding living fossils

by recasting them as 'relict species' whilst giving warnings about misinterpretation (Grandcolas and Trewick, 2016). In contrast, Bennett et al. (2018) argue that a quantitative concept of living fossils is needed to underpin research and resolve some of the controversy. They went on to describe the 'evolutionary performance index', which, to the best of our knowledge, remains the only pre-existing quantitative index focused on living fossils. However, we note that evolutionary distinctiveness (ED) has also been proposed as a potential indicator for living fossils (Isaac et al., 2007). The evolutionary performance index focuses on monophyletic clades and consists of three components: species richness, clade age and the number of evolutionary and ecological changes since divergence. Bennett et al. (2018) were emphatic that their approach does not invoke a ladder of progress. However, for living fossils to be defined partly on the basis of having few morphological and ecological changes since divergence still feels reminiscent of the ladder's problematic characteristics. We will apply the idea of attrition and EvoHeritage to conceptualise living fossils and thereby show how the phenomenon of living fossils can be explained from tree topology alone without the need to invoke any heterogeneity in rates of genetic or morphological change.

We propose that the taxa (say, the extant vertices of a dated EvoHeritage graph) called living fossils do not necessarily possess more ancestral features *in total* than other extant vertices; rather, the features they have are *rarer*, or even *unique* to the living fossil among the extant vertices. To see this, consider EvoHeritage accumulation and attrition occurring together at an equal pace across every edge in a dated EvoHeritage graph. Now consider an ancient vertex with two descendant edges, one leading to a single extant vertex (a candidate living fossil) and the other leading to a diverse group of extant vertices. Because attrition is constant everywhere, all extant vertices will inherit an equal total amount of ancestral features. The candidate living fossil vertex, however, inherits an independent sample of ancestral features, meaning those features are more likely to be rare or unique. In contrast, the vertices from the diverse sister group share a greater amount of evolutionary history with each other, meaning many of the features they inherit will also tend to be shared and thus will not be rare.

Let us now explore how EvoHeritage concepts can be used to quantitatively identify living fossils from first principles on a dated EvoHeritage graph. One founding principle is that being a living fossil is not a binary categorisation; it is a continuous metric (Bennett et al., 2018; Turner, 2019). We use the term 'living-fossil-ness' first coined by Bennett et al. (2018). However, our approach beyond this is very different from that of Bennett et al. (2018).

In practice, clades containing more than one species may be identified as clades of living fossils (Bennett et al., 2018). For example, the coelacanths *Latimeria* consist of two extant species. For our purposes, we consider only individual vertices as the candidate entities. We suggest that although closely connected groups of vertices might happen to all be living fossils, this should be viewed as an emergent property from calculations that work at the level of a single vertex. EvoHeritage graphs intended for living fossil identification should thus be constructed so that all extant vertices are at comparable taxonomic levels in order to avoid bias in the results.

Fossils originate from many different periods in geological history and features originating at different times may be preserved in different patterns on a dated EvoHeritage graph. The concept of an ancestral EvoHeritage graph gives us the ability to select the precise periods in geological history that we want to focus on. These could be key moments such as a former mass extinction event, whole periods of geological history, or some more complex combination of time points corresponding to biases in the fossil record itself. An ancestral EvoHeritage graph is constructed from a dated EvoHeritage graph by artificially augmenting EvoHeritage accumulation (α) to focus on EvoHeritage generated at the time when the fossils of interest appeared in the fossil record. For example, by setting

EvoHeritage accumulation to zero along many edges and only retaining it on edges leading to the ancestral vertices of interest. We quantitatively define the 'living-fossil-ness' of an extant vertex v on an ancestral EvoHeritage graph as $\log_{10}(\text{partitioned-}\phi(v, A))$ where A is the set of all extant vertices and $v \in A$. Attrition will typically be an exponential process, so we take the log of partitioned-ϕ to make the comparisons more natural.

To illustrate our ideas, we will detect mammalian living fossils from two key points in geological history: the Triassic–Jurassic boundary 200 million years ago, and the Cretaceous–Paleogene boundary 66 million years ago. Our method requires no data beyond a phylogenetic tree with branch (edge) lengths. We sourced mammal trees as a posterior distribution from Upham et al. (2019) with edge length units in millions of years. To construct ancestral EvoHeritage graphs for each scenario, we first identify (or create) vertices $v \in K$, where $K \subseteq V$, so that $v \in K$ if and only if v is dated at the given key moment in geological history. For the Triassic–Jurassic boundary, we chose a K the consists of just the most recent common ancestor vertex for all mammals. For the Cretaceous–Paleogene boundary case, K contains a collection of vertices all dated at precisely 66 million years ago and newly generated as monotony vertices along every edge crossing that moment in history. We then set $\alpha(e)$ according to the following equation:

$$\alpha(e) = \begin{cases} 10^6 & \text{when } D^v(e) \in K \\ 0 & \text{otherwise.} \end{cases} \tag{23}$$

The value 10^6 is arbitrary, but affects all the final results equally while enabling the scores to fall in a convenient range. Attrition was calculated under standard conditions with a single EvoHeritage form and $\beta_i(e)$ was given by Eqn. 5. For each of our ancestral EvoHeritage graphs, we calculated partitioned-ϕ for every extant vertex using Monte Carlo methods. To see how, consider Eqn. 11: the first sum disappears because there is only one EvoHeritage form. The next sum over the power set $P(E)$, and both of the product terms involving functions of β will be replaced with a repeated stochastic process that considers each edge e of the EvoHeritage graph and disconnects it with probability $1 - \beta_i(e)$, then calculates the inner part of the expression from Eqn. 11. After substituting this and Eqn. 14 into Eqn. 15 we can write:

$$\text{partitioned-}\phi(v, A) = \sum_{T \in S} \left(\sum_{e \in E} \alpha_i(e) \cdot \frac{C(d^V(e), E(T), v)}{C(d^V(e), E(T), A)} \right). \tag{24}$$

In this equation, S is the set of EvoHeritage graphs (trees) resulting from repeated random removal of edges from the edge set E, and T is an example EvoHeritage graph (tree) from the set S, and $E(T)$ is the set of all edges remaining in T. Our ancestral EvoHeritage graph has $\alpha_i(e) = 0$ for most edges e and because EvoHeritage is only accumulated at a moment in time, no path through the EvoHeritage graph can pass more than one edge where $\alpha_i(e) \neq 0$. This simplifies calculation of the inner sum of Eqn. 24 to one that can be carried out with a single traversal of the edges to get partitioned-ϕ simultaneously for every extant vertex. We repeated the stochastic process 10^4 times and took the mean of those values. The whole calculation was repeated for 100 mammal trees from the posterior distribution of the trees provided by Upham et al. (2019). We then took the median of these 100 values to calculate the living-fossil-ness for each vertex because the distributions are often quite skewed. Thus, each living-fossil-ness value was obtained from 10^6 tree traversals. Everything was repeated for both the Triassic–Jurassic boundary case and the Cretaceous–Paleogene boundary case, as well as for both $\rho = 10^{-3}$ and $\rho = 10^{-2}$. We plotted the distribution of ranked living-fossil-ness across all mammals and showed box and whisker plots for the top 15 species, showing uncertainty from the 100 posterior trees.

iii) Conservation prioritisation across the whole tree of life

A motivator for PD and the PD calculus has been conservation (Faith, 1992), including the need to think quantitatively about conservation optimisation (Isaac et al., 2007) and to regard biodiversity as a broader concept than species richness (Brondizio et al., 2019). A flagship example is the prioritisation of species that are evolutionary distinct (ED) and globally endangered (GE) with 'EDGE lists' (Isaac et al., 2007), along with more recent developments of this idea (Steel et al., 2007). EDGE analyses used in real-world conservation applications are currently being carried out on relatively small clades such as mammals or birds in isolation (EDGE of Existence Programme, 2022). The opportunity to upscale this to include all described life is coming soon through the data synthesis of the Open Tree of Life (Hinchliff et al., 2015), methods for dating this tree (McTavish and Sanchez-Reyes, 2022) and methods (Wong and Rosindell, 2022) to automatically align it with the International Union for Conservation of Nature (IUCN) Red List of threatened species (IUCN, 2022). Although such an analysis is beyond the scope of our present work, it is germane to consider the weightings that a PD-based approach would give to selected species in an analysis of the whole tree of life and compare these with a EvoHeritage-based approach.

Our aim is to illustrate the difference between EvoHeritage and PD on the scale of the complete tree of life for a small number of selected species. Among the mammals we chose *Zaglossus attenboroughi* (Attenborough's long-beaked echidna) currently the (joint) top ranking EDGE mammal; *Solenodon paradoxus* (Hispaniolan solenodon), currently the EDGE mammal with the (joint) highest ED value (EDGE of Existence Programme, 2022); and *Mus musculus* (house mouse) as an example of a mammal with a very low ED score. We selected four other vertebrates that we expect to have high ED scores given their phylogenetic placement: *Sphenodon punctatus* (tuatara), *Neoceratodus forsteri* (Australian lungfish), *Latimeria chalumnae* (one of the two coelacanths) and *Rubicundus rubicundus* (a species of hagfish). Among the invertebrate animals, we chose *Xenoturbella churro* (a strange worm-like marine species) and *Trichoplax adhaerens* (a small and flat organ-less animal). We chose two single-celled eukaryotes, each with key phylogenetic positions: *Pygsuia biforma* and *Capsaspora owczarzaki* (Ferrer-Bonet and Ruiz-Trillo, 2017).

We used data from the OneZoom tree of life (Wong and Rosindell, 2022; OneZoom Core Team, 2021), which provides dates for key positions on the tree. Only a small number of vertices were not dated; in such cases we assumed that the vertices were evenly distanced along the path to the next dated vertex. We took the origin of life vertex as the crown of the tree and dated it at 4.025 billion years, which is the midpoint of the range of estimates from Dodd et al. (2017). We did not attempt to resolve the few polytomy vertices. For each species in our analyses, we extracted the length of each edge on a path from it to the origin of life, as well as the total number of descendants for each vertex along that path.

It was straightforward to calculate ED as well as the terminal branch length component of ED (Isaac et al., 2007). Unique-ϕ and partitioned-ϕ, unlike ED, depend, in part, on the structure of the whole tree, not only on the number of descendants shared by each vertex. Our calculations of both unique-ϕ and partitioned-ϕ are thus approximations that can be made with only the total number of descendants from each vertex. We applied these approximations in two different ways, making opposite assumptions about the structure of the unseen parts of the tree. In the first scenario, we assumed that all terminal vertices connect directly via an independent edge to the interior vertex, thus maximising the amount of independent evolution of the group. We refer to this as a *'star' tree*. In the second scenario, we assumed that all terminal vertices connect together in a polytomy with edges of zero length and then via a single long edge to the interior vertex. We refer to this as a *'stick' tree*.

We calculated α and β under standard conditions for each relevant edge using Eqns.

5 and 6. The value of α on the terminal edge to the focal vertex gave the component of partitioned-ϕ and unique-ϕ arising from the terminal edge. We estimated the total unique-ϕ by calculating, from first principles, the proportion of EvoHeritage of each ancestral edge that survives only to the focal vertex and to no other extant vertex. This is given by:

$$\sum_{x \in P(v,\hat{o})} \alpha(A^E(x)) \prod_{y \in P(v,x)} \beta(A^E(y)) \prod_{y \in C(x) \setminus P(v,x)} (1 - \beta(A^E(y))) \tag{25}$$

where $v \in V$ is the focal terminal vertex for which we are calculating unique-ϕ, $P(v,x) \subset V$ is the set of vertices along a path between vertex v and vertex x but excluding vertex x, $\hat{o}$ represents the origin of life vertex and $C(x)$ represents the set of all vertices descending from vertex x not including vertex x itself. In words, we perform a different calculation for each ancestral edge and sum the results. The α term gives the total EvoHeritage arising on an edge, and the product of various β and $1-\beta$ terms gives the proportion of this EvoHeritage that survives along the path to the focal vertex while also being cut off in all other descendants of all other interior vertices.

We estimated partitioned-ϕ via two different methods depending on whether we assumed 'star' or 'stick' trees for the other descendants from interior vertices. In all cases, we need to sum over the ancestral edges, calculate the EvoHeritage on each, and multiply by the proportion that survives to the focal vertex (the sum and first product terms of Eqn. 25). However, we need to keep track of the proportion of this EvoHeritage that is copied different numbers of times. In the case of a 'stick' tree, we can approximate number of copies with:

$$\sum_{x \in P(v,o)} \left(\alpha(A^E(x)) \left(\prod_{y \in P(v,x)} \beta(A^E(y)) \right) \left(1 + \sum_{y \in P(v,x)} (1 - \beta(G(y)))R(g(y)) \right)^{-1} \right) \tag{26}$$

where $G(y)$ gives the descendant edge from vertex y that is not an ancestor edge of the focal vertex, and $R(g(y))$ gives the total number of extant terminal vertices descended from edge $g(y)$. In words, the term that differs from Eqn. 25 is a reciprocal because we want to divide by the number of copies of EvoHeritage. As an approximation, we just take the mean number of such copies based on the proportion $(1 - \beta(G(y)))$ of surviving attrition along the stem of the 'stick tree' and the number of copies that would result in $R(g(y))$. In the case of a 'star tree', the number of copies from each interior vertex follows a binomial distribution with an independent trial for each descendant other than the focal vertex and a probability of success $(1 - \beta(G(y)))$. We applied a Monte Carlo approach, drawing the number of copies from the binomial distribution for each descendant vertex of each ancestral edge. We found the mean of 10^4 repeats of the Monte Carlo process.

Whilst IUCN Red List data are not available for most species in our analysis, we calculate EDGE scores for some species for comparison. To do this we used the formula from Isaac et al. (2007) :

$$EDGE = \ln(1 + \mathrm{ED}) + \mathrm{GE} \cdot \ln(2) \tag{27}$$

where $\mathrm{GE} \in \{0, 1, 2, 3, 4\}$ depends on IUCN Red List status with a higher score indicating a higher extinction risk.

Empirical applications –Results

i) Evaluating ϕ_ρ as a predictor of community productivity

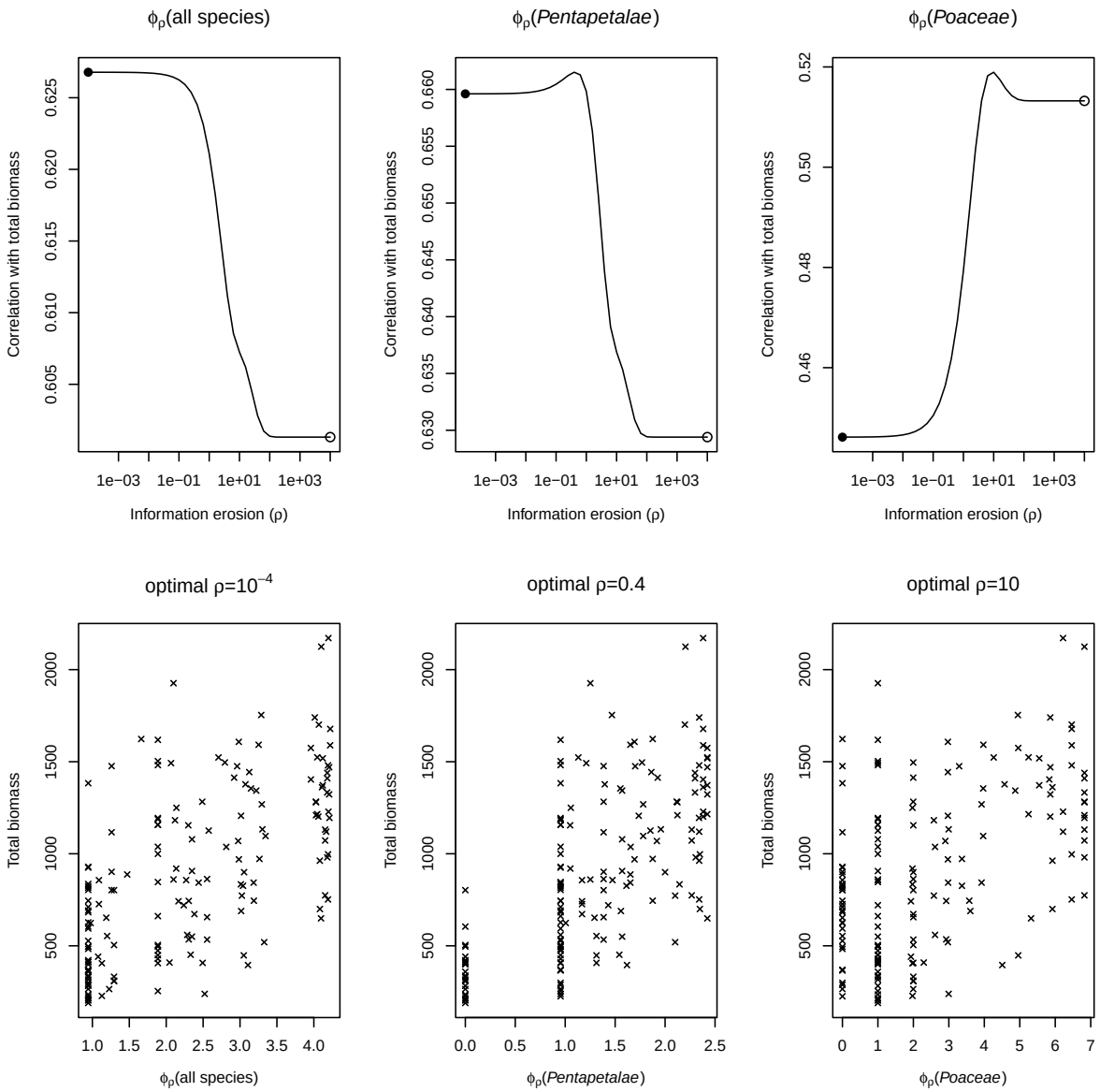

Fig. 6. Relationship between EvoHeritage ϕ_ρ and community productivity. The top row of panels show Pearson's correlation coefficient between ϕ_ρ and total biomass in the Cedar Creek plots. The horizontal axis shows attrition ρ. The solid circle gives the same correlation, only with PD (corresponding to ϕ_ρ with small ρ) and the open circle gives the correlation with species richness (corresponding to ϕ_ρ with large ρ). The bottom row of panels show scatterplots of biomass against ϕ_ρ for the value of ρ that gave the strongest correlation. The three columns correspond to correlations with measurements of: ϕ_ρ for all species (*Pentapetalae* and *Poaceae* together) (left), ϕ_ρ for *Pentapetalae* (centre), and ϕ_ρ for *Poaceae*, true grasses (right).

The strongest correlation of plot biomass against plot ϕ_ρ for all species was achieved for the smallest ρ, where ϕ_ρ converges to PD. The correlation coefficients of PD

and species richness coincide exactly with those of ϕ_ρ at each of its limits, as expected. Considering only the species richness of the *Pentapetalae*, ignoring the contribution of any of the *Poaceae* achieves a better correlation with biomass than any ϕ_ρ calculated for all species together. The PD for *Pentapetalae* correlated better with total biomass than the species richness of *Pentapetalae*. The strongest correlation with biomass was achieved with $\rho = 0.4$, calculated for *Pentapetalae* only. The diversity of *Poaceae* was the poorest predictor of biomass, regardless of how it was measured, but unlike the other cases, it performed better as species richness (*Poaceae*) than it did as PD(*Poaceae*). For *Poaceae*, ϕ_ρ was optimised when $\rho = 10$ where it performed better than either species richness or PD.

ii) Applying EvoHeritage attrition and partitioned-ϕ to quantify living fossils

Living-fossil-ness from the Triassic–Jurassic boundary yielded a stepped distribution of values with clear gaps between groups. The top species was *Ornithorhynchus anatinus*, the Duck-billed platypus. Next came all four species of echidna *Tachyglossidae* without much distinction between the four species. Then all seven species of shrew-opossum *Caenolestidae*, followed by *Dromiciops gliroides* (known as the monito del monte) and then various other opossums. Overall, the top 15 living fossils from the Triassic–Jurassic boundary are all Monotremes or South American marsupials. The rankings did not appear to be sensitive to the value of ρ, though the precise Living-fossil-ness numbers were dependent on ρ (Fig. 7). Living-fossil-ness from the Cretacous–Paleogene boundary yielded a smoother distribution of values, with noteworthy new entries to the top 15. The aardvark *Orycteropus afer* is top, followed by the three species of *Solenodon*. *Ornithorhynchus anatinus* came next, followed by another new entry, *Laonastes aenigmamus*, a rodent. After this, the top species are all *Tachyglossidae* and *Paucituberculata* species. Uncertainty in the placement of *Orycteropus afer* and *Laonastes aenigmamus*, both of which have long termal edges, leaves open the outside possibility for them to have much lower living-fossilness, but the median remains high.

iii) Conservation prioritisation across the whole tree of life

Of our 11 selected species, the one with the highest ED was *Pygsuia biforma* with a value of 1.125 billion years, mostly arising from a terminal edge length estimated at 900 million years. The second highest ED was for *Capsaspora owczarzaki*, with a value of 860 million years, of which 800 million years comes from its terminal edge. The next highest ED species, in order, were *Trichoplax adhaerens* (ED = 379), *Neoceratodus forsteri* (Australian lungfish) (ED = 346), *Sphenodon punctatus* (tuatara) (ED = 275), *Xenoturbella churro* (ED = 112), *Rubicundus rubicundus* (ED = 88) and *Latimeria chalumnae* (ED = 75).

Of our selected species, the three mammals had the lowest ED scores: *Solenodon paradoxus* (ED = 52), *Zaglossus attenboroughi* (ED = 49) and *Mus musculus* (ED = 4.7) (Fig. 8). The ED scores of *Solenodon paradoxus* and *Zaglossus attenboroughi*, published elsewhere, were 61.5 and 46.6, respectively (Gumbs et al., 2018). The EDGE score of the critically endangered *Zaglossus attenboroughi* was 6.68. The EDGE score of *Capsaspora owczarzaki* would be between 6.76 and 9.53 depending on its (unknown) GE score. Similarly, the EDGE score of *Pygsuia biforma* would be between 7.03 and 9.8.

The partitioned-ϕ and unique-ϕ with $\rho = 0$ were, as expected, identical to ED and terminal edge length, respectively (see supplementary material). The partitioned-ϕ and unique-ϕ with $\rho = 10^{-3}$ showed a somewhat similar pattern to ED, though the extremely high values for *Pygsuia biforma* and *Capsaspora owczarzaki* are somewhat dampened and almost all of their partitioned-ϕ score comes from the unique-ϕ component.

The partitioned-ϕ with $\rho = 10^{-2}$ was substantively different from ED. The

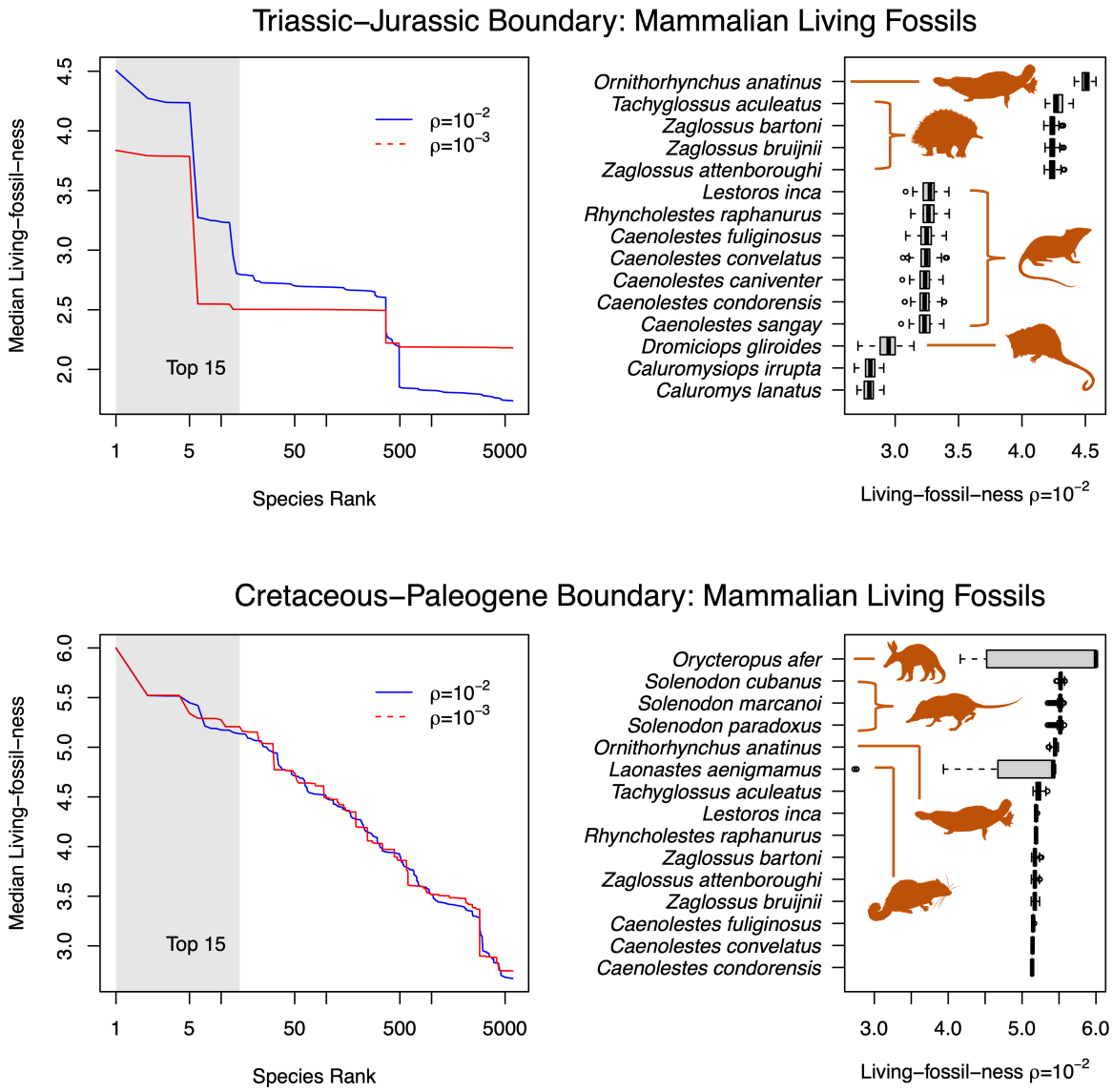

Fig. 7. Living-fossil-ness for mammal species calculated with reference to two different points in history t: the Triassic–Jurassic Boundary $t = 200$ Million years ago and the Cretaceous–Paleogene Boundary $t = 66$ Million years ago. The left panels show the living-fossil-ness with rank in log scale for two different values of attrition: $\rho = 10^{-2}$ (blue) and $\rho = 10^{-3}$ (red). The top 15 species are shaded. The box and whisker plots of the right focus on $\rho = 10^{-2}$ and top 15 species only. The uncertainty represented in the box and whisker plots is from 100 posterior trees. Each tree gives rise to one value in the distribution shown; each value is the result of 10^4 random draws of attrition via a Monte Carlo approach. Image credit: *Caenolestes* (representing shew opossums) and *Dromiciops* (monito del monte) Sarah Werning CC BY 3.0; *Laonastes* (Laotian rock rat) reprocessed version of original from creazilla.com; others are from the public domain.

partitioned-ϕ of *Mus musculus* was still by far the lowest, with 4.5 standardised units. The other species, all of which are considered to have high ED, now sit within a more compact range of between 31.1 and 100.5 standardised units, thus differing by a factor of about 3; ED scores for the same species differed by a factor of about 23. *Xenoturbella churro* and *Latimeria chalumnae* all fell disproportionately far and thus fell in the ranking of partitioned-ϕ with $\rho = 10^{-2}$ compared with their ranking under ED: both had a comparatively small terminal branch length (Fig. 8). In the limit as $\rho \to \infty$, both unique-ϕ and partitioned-ϕ converge to unity for all species (see supplementary material).

Discussion

i) Evaluating ϕ_ρ as a predictor of community productivity

One motivation for formulating EvoHeritage was to provide a process-based measure of diversity that scales continuously between species richness and PD. Our results showed that ϕ_ρ gives the same result as PD and species richness at its two extremes (Fig. 6 and Thm. 1). This approach changes the nature of debates in this field from a stark choice between species richness and PD (Mazel et al., 2018, 2019; Owen et al., 2019) to a question about the extent of attrition. Previous work has suggested a metric that will converge to species richness and PD in different limits, based on pruning parts of the tree older than a given cutoff (Chao et al., 2010). EvoHeritage is different because its approach is mechanistic and is based on the biological process of attrition. Attrition exists as a biological reality, occurring at a rate that is neither zero (PD) nor infinity (species richness); it should be somewhere in between. The incorporation of phylogeny into ecology has, in the past, been criticised on the grounds that phylogeny's role is relegated to a 'middleman' proxy for missing functional trait data (Swenson, 2013). Our approach allows us to move towards an eco-phylogenetic discipline where ecological processes give insight into evolutionary process of attrition of information and traits, allowing eco-phylogenetics to give mechanistic insight into two disciplines (Pearse et al., 2019; Davies, 2021).

We found that ϕ_ρ of all species present, for any value of ρ, could not improve on PD as a predictor of community productivity. This result is not a failure of ϕ_ρ but is rather an example of how it can be informative. Specifically, it suggests that the deepest branches of the tree are the most important predictors of productivity, because attrition, which can only reduce the influence of such branches, is optimal when set to zero. The phylogeny of the species we considered has a deep split between the true grasses *Poaceae* and the other species, all of which are *Pentapetalae*. PD of all species present in a plot will therefore primarily reflect the significant distinction between these two groups. This was confirmed by our finding that the species richness of *Pentapetalae* alone performed better as a predictor of biomass than any measure of total diversity. These results are probably the result of ecological factors such as species–species interactions and the special role of grasses in community productivity.

The best predictor of community productivity overall was the ϕ_ρ of *Pentapetalae* with an optimal ρ of 0.4 (see Fig. 6). This outperformed both species richness and PD and thus demonstrates the potential of ϕ_ρ as a general diversity measure. Although we found that the diversity of *Poaceae* is a poor predictor of community productivity, we highlight it as one example where species richness outperforms PD, though ϕ_ρ still outperforms both species richness and PD (Fig. 6).

The differences in the correlation coefficient are quite small across all our scenarios, which is consistent with the mixed results that comparisons between species richness and PD have so far yielded (Mazel et al., 2018). Furthermore, the best measure of overall diversity would not be one that completely ignores the presence of *Poaceae* as a component of diversity. These findings highlight the reasons why the best measure of overall diversity

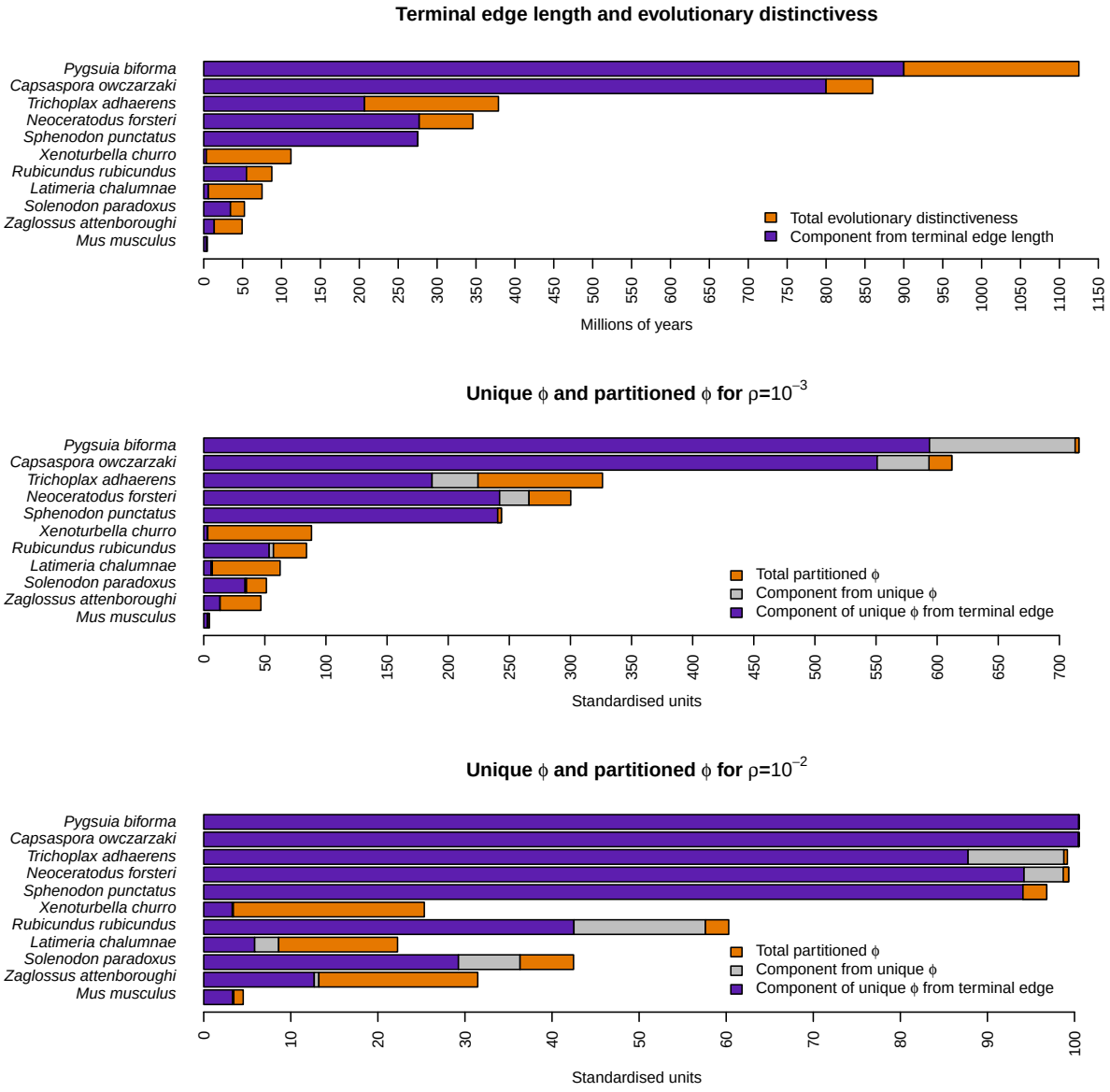

Fig. 8. Comparison of ED, terminal edge length, unique-ϕ and partitioned-ϕ for selected species, ordered by ED. The stacked bars of the top panel show ED separated into two components: the terminal edge length and the remainder, which is the contribution of the interior edges. The middle and bottom panels show the equivalent for ϕ with $\rho = 10^{-3}$ and $\rho = 10^{-2}$ respectively. The stacked bars of the EvoHeritage-based graphs show partitioned-ϕ separated out into three components: a contribution from unique-ϕ arising on the terminal edge, a contribution from unique-ϕ arising on interior edges but where the EvoHeritage is lost before reaching any other extant vertex, and a contribution from interior edges that is not unique. We are showing the midpoint of the 'stick' and 'star' cases for sister clades, but note that there was only a negligible difference between the two. All images of species were sourced from Wikimedia commons with credits as follows: *Capsaspora owczarzaki* by Arnau Sebé-Pedrós and Iñaki Ruiz-Trillo, *Trichoplax adhaerens* by Oliver Voigt, *Xenoturbella churro* by Greg Rouse, *Neoceratodus forsteri* by Mitch Ames and *Sphenodon punctatus* by Sid Mosdell from New Zealand

is not the same as the best predictor of community productivity; it can thus be misleading to compare PD with species richness on the basis of a single performance factor. Some debates around the role of PD have mirrored this and pointed to the conceptual foundations of PD in its defence (Owen et al., 2019). Our foundation for EvoHeritage as an alternative to PD is likewise based more on the conceptual reasons supporting its use, rather than its encouraging performance in this particular case.

ii) Applying EvoHeritage attrition and partitioned-ϕ to quantify living fossils

To the best of our knowledge, our method of using partitioned-ϕ on ancestral EvoHeritage graphs is the only attempt focused on quantifying living fossils, other than the work of Bennett et al. (2018). Although Bennett et al. (2018) provided inspiration for our quantification procedure, our approach is different in several important respects. First, we only calculated the 'living-fossil-ness' for individual vertices of a data EvoHeritage graph, whereas Bennett et al. (2018) primarily designated clades of species as living fossils. Second, we calculated the 'living-fossil-ness' differently depending on which period in geological history the fossils of interest came from. Third, our calculations are process-based, incorporating the concept of attrition. We note that ED (Isaac et al., 2007) has also been suggested as a metric of living fossils, though this was not its primary purpose (Bennett et al., 2018). Even though partitioned-ϕ converges to ED as a special case, our use of partitioned-ϕ is quite different because we only consider the fate of EvoHeritage created at a given moment in time. The lack of attrition in ED would mean there is nothing to choose between any extant descendants under these circumstances – all vertices would be equal.

Many of the top mammal species that we identified have a history of being referred to as living fossils in the primary literature, including the monotremes (Phillips et al., 2009), aardvark (Goździewska-Harłajczuk et al., 2018), solenodons (Wezel and Bender, 2002) and *Laonastes aenigmamus*, known locally as kha-nyou (Huchon et al., 2007). Other than *Dromiciops gliroides* (Bozinovic et al., 2004), we only found a mention of South American marsupisals as living fossils in the grey literature. Our measure of living-fossil-ness therefore demonstrates that the taxa historically identified as living fossils can be recovered from our EvoHeritage-based concept of living fossils that requires no heterogeneity in the rate of evolution (Fig. 7). The key to our proposed solution of the living fossil controversy comes from recognising that living fossils show ancestral features that are rare in living species, but need not show more ancestral features in total. Given that the identity of the living fossils themselves was originally established by a non-quantitative thought process, it is easy to imagine how the discovery of rare ancestral features in a living species would carry greater weight and be more worthy of commentary. In contrast, the discovery of ancestral features that are commonly seen in living species would be taken for granted. It also seems clear that rare ancestral features in a living species are more informative for reconstructing the features of extinct ancestors. Therefore, the phenomenon of living fossils as real and noteworthy objects of study might have been masked by the subtlety of rarity and a dearth of quantitative approaches. Our explanation still requires an imbalanced phylogenetic tree and does not seek to provide an explanation for why such imbalance occurs. It may be due to lower speciation or elevated extinction within a clade - if the latter it would make our living fossils 'relict species' in the spirit of Grandcolas et al. (2014). Regardless of the reason behind any phylogenetic imbalance, we have shown that the concept of living fossils does not need to invoke the 'ladder of progress' or be in conflict with mainstream thinking and can arise without any heterogenity in the rates of morphological or genetic change on a given phylogeny.

iii) Conservation prioritisation across the whole tree of life

Our illustrative example shows that the mammals, previously highlighted as conservation priorities by ED, would no longer be priorities when compared with the rest of non-mammalian life. This result is because of the disproportionate value given by PD (and ED) to extremely long branches. There is a large disparity, of a factor of about 23, in ED between species that could all be considered as evolutionary distinct. This highlights how the conceptual challenges with PD that we have identified can have an effect on real-world conservation applications. Indeed, we saw that the EDGE score of the critically endangered *Zaglossus attenboroughi* is less than the score that either *Pygsuia biforma* or *Capsaspora owczarzaki* would have even if they were least concern species. Despite their high scores, we note that *Capsaspora owczarzaki* and *Pygsuia biforma* are not technically 'EDGE species' until they are considered under threat of extinction according to the IUCN Red List (IUCN, 2022; Isaac et al., 2007; EDGE of Existence Programme, 2022). Of course, reconstructing such deep parts of the phylogeny is notoriously difficult, and it could be that the tree will rearrange around *Capsaspora owczarzaki* in some future work, radically changing its ED scores in the process. Alternatively, a new *Capsaspora* species may be discovered and this would also have a substantial effect. Indeed, the heterogeneous density of undescribed species around the tree of life is another source of bias that will require its own solution in future. In the context of our present work, uncertainty around deep parts of the tree of life is not a weakness of our analyses *per se*. It could instead be regarded as a weakness of the PD calculus to be so sensitive to rearrangements of the deep and unstable parts of the tree of life. The same is not true of EvoHeritage because the deeper parts of the tree are subject to more attrition and thus carry less influence for comparing extant vertices. Indeed, the difficulty of reconstructing ancient parts of the tree of life is perhaps prominent evidence that attrition is a real force that cannot be ignored in phylogenetic analyses at such scales.

We saw that the greater the value of attrition, the greater the departure of partitioned-ϕ from ED. For $\rho = 10^{-2}$, a relatively modest rate of attrition corresponding to a 1% loss of features over 1 million years, there were already substantive differences from ED. Firstly, there was less disparity between species that could be considered as having high ED. These scores ranged by a factor of about 3 instead of 23, as was the case with ED. Despite this, *Mus musculus*, included as an example of a low-ED species, remained at the low end of the distribution by a very substantial margin. Secondly, species with very short terminal edge lengths were disproportionately penalised by partitioned-ϕ. The EvoHeritage accumulated near the tips of the tree has greater value, as there has not yet been as much opportunity for it to be lost in the first place, so the cost of sharing this is high in terms of the final partitioned-ϕ score. Furthermore, having a long terminal edge provides the opportunity to preserve unique features from deeper ancestral edges that were lost along sister edges. This is particularly clear for species with a long terminal edge as well as a long interior edge shared by few other descendants. Previous work has used PD to argue for greater focus on conservation of microscopic organisms (Trewick and Morgan-Richards, 2016), our analyses qualitatively support this conclusion with microbes still being detected as important based on partitioned-ϕ. On a quantitative footing however, partitioned-ϕ provides a less extreme range of final values for downstream conservation prioritisation and continues to clearly identify the species with the least ED.

Our example was based on ED values and the corresponding EDGE scores (Isaac et al., 2007). More recent work extending the concept of EDGE has used mappings of IUCN Red List status onto the probability of extinction (Mooers et al., 2008) to optimise future PD with a metric known as HEDGE (Steel et al., 2007). One notable consequence of HEDGE is that the deeper edges of the tree have reduced influence because there is a small probability of all their descendants becoming extinct. The same mechanism attaches

a value to there being more descendants of a deep edge, as such edges are more likely to survive. HEDGE therefore mitigates some of the issues with PD that we have highlighted here, albeit as a byproduct of making the probability of extinction explicit. However, HEDGE is still based on PD and will not change the disproportionate value attached to terminal edges. Furthermore, one should not confuse the different underlying mechanisms and aims of HEDGE and EvoHeritage. We need a coherent system for attributing value (EvoHeritage accounting for attrition) as well as for assessing the risk of losing parts of that value (HEDGE); neither can be ignored and future work should amalgamate both.

There is another potential danger associated with application of PD maximisation to conservation. Work evaluating the rate of contemporary species extinction often compares this with background rates of extinction (and speciation) to establish that the present biodiversity crisis goes substantially beyond that expected from natural processes (De Vos et al., 2015). A comparison with the background rates cannot be meaningfully conducted for loss of PD. This is because every new year that passes adds over 2.2 million years of new PD to the tree of all described species of life (one year of PD for each described species). The background rate of overall PD generation across the tree of life, since the publication of PD (Faith, 1992), is already more than enough to 'pay off' the extinction of the duck-billed playtpus *Ornithorhynchus anatinus* with a terminal branch of 63 million years. Although such comparisons may seem unreasonable, they are correct within the strict context of PD maximisation. We worry that the certain generation of new PD may (explicitly or implicitly) end up perversely excusing some level of additional extinction in future analyses. This problem would be most prominent in large-scale quantitative analyses of present PD gain or loss compared with background rates of PD gain or loss. When we recast the problem in terms of EvoHeritage, however, we see that any generation of new EvoHeritage at the tips of the tree is already balanced against the natural process of attrition, just as speciation is balanced against natural extinction (which, we note, runs at a very low rate compared with extinction from anthropogenic causes). Future work may thus be able to meaningfully compare the background rate of EvoHeritage loss (through attrition and natural extinction together) with the current anthropogentic rate of EvoHeritage loss (including anthropogenic extinctions).

The future of EvoHeritage graphs and attrition

The foundations for EvoHeritage are based on a new concept of an EvoHeritage graph: a DAG where the processes of both EvoHeritage accumulation and attrition are defined along the edges. This creates an intentional separation between the question of useful metrics on a given EvoHeritage graph, and the question of how to acquire an EvoHeritage graph in the first place. It would be interesting to build EvoHeritage graphs directly from genetic data by adapting tree-building methods and we expect this will be possible in future work. Indeed, there have already been studies on building phylogenetic networks (Burbrink and Gehara, 2018; Huson and Scornavacca, 2011). It is also commonplace to express tree uncertainty with a posterior distribution of trees (Upham et al., 2019), the structure of which is technically already a DAG if we join the trees together by matching their terminal vertices.

A bigger step for building EvoHeritage graphs *de novo* is to incorporate attrition. We have shown through Theorem 2 that EvoHeritage accumulation and loss interact in a way that leads to non-monphyletic features and cannot be expressed with just a single dimension of data, such as an edge length. At least two dimensions are needed: one for EvoHeritage accumulation and the other for attrition. It may be that we use more than one EvoHeritage form, for example, so that the EvoHeritage within an endosymbiont has its own accumulation and loss that may be distinct from that of the host. We speculate

that attrition may be part of the reason why it is difficult to build trees and why, in practice, trees end up with so much uncertainty. Tree building may be like trying to fit a square peg into a round hole, not only because of well-documented reticulations (Burbrink and Gehara, 2018), but also because there is so much attrition going on at the genetic level, which can cause the tree signal to decay (Mossel and Steel, 2005). Yet the standard tool for capturing such data, the tree with just edge lengths, cannot possibly store the joint effects both accumulation and attrition.

To build an EvoHeritage graph instead of a tree would certainly be a break in convention. It would require a new data structure for storage and a new suite of methods. We would need to be prepared for different heterogeneity in EvoHeritage accumulation compared with attrition. The potential reward might be an object with less compromise and uncertainty, which would enable more of the source data to inform downstream analyses as well as new questions. For example how does the rate of attrition vary with body size? Did it change at important times of evolutionary pressure such as during past mass extinction events? Beyond this, future EvoHeritage graphs need not come only from adapted tree building procedures. For example, vertices may be individual organisms and the many edges may represent sexual reproduction with $\beta \approx 0.5$ because roughly equal EvoHeritage is inherited from each parent. This illustrates how the EvoHeritage graph is even more flexible than a phylogenetic network, and could span the bridge between unconnected fields such as population genetics and macroevolution.

In order to facilitate practical applications of the EvoHeritage calculus, we developed a model of exponential EvoHeritage decay that can be applied to a trees with branch lengths, given ρ, a rate of attrition. This has the useful property that many existing measures are recovered when we set $\rho = 0$. The metric ϕ_ρ in particular was necessary to provide something that could directly stand in for PD in future analyses. We can see several approaches to determining the most appropriate ρ in such studies. The approach we used here was to experiment with a range of values and make the effect of ρ a subject of study rather than an inconvenience. This can be a strength, such as in our application to community productivity where the continuum between species richness and PD challenges the conventional view that they are two discrete alternatives to be pitted against each other.

We see that in some cases, a fixed value of ρ will need to be decided upon, for example, in conservation applications. One approach is to use the value of ρ that optimises some payoff, such as a correlation with feature diversity or with community productivity, as we have done here. Another approach would be to use a rule of thumb. That may not seem ideal but if the alternative is to go to one extreme or the other (that is, to species richness or PD), then a rule of thumb is better than nothing. If we expect parts of the tree to be more recent than a few tens of millions of years to still behave like PD, and edges are measured in millions of years, then $\rho \ll 0.1$. Similarly, if we expect to see significant attrition over a billion years, then $\rho \gg 0.001$. Indeed, after investigating a wide range of ρ in our empirical studies for conservation prioritisation and living fossil detection, the interesting and intuitively correct results seem to be around $\rho \approx 0.01$. The optimal values of ρ from our community productivity application were 0.4 and 10 (see Fig 6), but these were calculated on a tree where the total edge length to the crown was set to 1, edges were not in units of millions of years, and the phylogeny was a non-random small subset of the tree of life (the vascular plants included in an experiment), and so, ρ values are not directly comparable. As a 'back of the envelope' comparison, the crown of our tree represented a common ancestor with *Amborella trichopoda*. This can be dated at 234.5 million years ago (Zanne et al., 2014), which would thus correspond to one unit of edge length on our tree. Scaling ρ accordingly gives values of 0.0017 and 0.043 with a midpoint of 0.022, all within the range given by our rule of thumb.

Conclusion

In this manuscript, we began by pointing out a range of conceptual challenges with PD. First, following PD, a single extant descendant of a deep edge is enough to guarantee the survival of all ancestral features on that edge. This does not seem biologically reasonable, for example, it would mean one species of flightless bird alone could represent all features of the stem edge of birds. Second, the PD of a set of vertices is not increased by resurrecting an extinct common ancestor. For example, PD ignores any value that the Archaeopteryx could add beyond that seen in living birds and crocodilians. Third, the features or future options for humanity cannot reasonably be directly proportionate to the edge length for extremely long edges (e.g. *Pygsuia biforma* is about than 23 times that of the top EDGE mammal *Zaglossus attenboroughi*). Whilst the last of the these challenges can be addressed for terminal edges by simply re-scaling longer branch lengths, the other challenges cannot. All these challenges are symptoms of one underlying cause: assuming that features accumulate and are never lost leading to non-monophyletic patterns of features that are not compatable with PD. A key realisation is that the same concerns are likely to apply to any other metric that is underpinned by the sums or means of inter-vertex distances in some way.

We introduced a new EvoHeritage calculus, which incorporates the process of 'attrition' capturing a gradual loss of features without extinction. Our mathematical results show that the independent forces of accumulation and attrition cannot be captured by with the mainstream concept of a single value for 'length' of each edge. We have provided standard conditions as a way to proceed with calculations using phylogenetic trees and accounting for attrition, until EvoHeritage graphs can be produced by other methods. Nevertheless, we have shown that it is impossible to frame EvoHeritage on a tree under standard conditions in terms of PD after a re-scaling of branch lengths.

We have introduced ϕ, unique-ϕ, partitioned-ϕ, diff-ϕ, and Hill-ϕ to illustrate EvoHeritage counterparts to wide ranging concepts from the PD calculus (Isaac et al., 2007; Lozupone and Knight, 2005; Chao et al., 2010). There are many more possible counterparts. For example, on an EvoHeritage graph with extant vertex range maps, we would like to see an 'EvoHeritage endemism' following the phylogenetic edemism concept of Rosauer et al. (2009). EvoHeritage endemism would spread the value of EvoHeritage equally across the area in space where that EvoHeritage exists. We also imagine a reinterpretation of tree balance in terms of EvoHeritage rather than of lineage numbers. Throughout the framing of EvoHeritage, we do not assume the underlying structure is a tree; EvoHeritage is therefore capable of being applied to phylogenetic networks (Huson and Bryant, 2006) and perhaps also to population genetics. We introduced standard conditions and ϕ_ρ to make the adoption of EvoHeritage easier as a straight swap for PD on any tree with edge lengths, given a value of ρ. To use the EvoHeritage calculus under standard conditions, one only needs a phylogenetic tree with branch lengths and a value of ρ (or a range of values). Our three applications of EvoHeritage all rely on its attrition component and they cover ecology, evolution and conservation. EvoHeritage can outperform PD and species richness as a predictor of community productivity; metrics on an ancestral EvoHeritage graph can detect living fossils and resolve their controversy; and conservation prioritisation at the scale of the whole tree of life is distorted without EvoHeritage attrition to rebalance it.

We are aware that there is already an apparent excess of phylogenetic metrics, many of them redundant (Tucker et al., 2017). Indeed, it has been argued that this 'jungle of indices' impedes, rather than facilitates, scientific progress and uptake in conservation applications (Pausas and Verdú, 2010).

We view the jungle in the different light of the process that we think created it. Researchers have qualitative questions that they wish to answer; doing so requires them to

have metrics with certain properties. It is easy to create metrics with the required properties from scratch and so researchers do so. The 'cost' of accidentally duplicating or reinventing something that already exists is very small for a metric compared with other research endeavours. The main message of this paper is thus not our individual metrics, but rather the conceptual importance of attrition and the flaws associated with just summing edge lengths, even if those edges have first been re-scaled in some way. We urge researchers to either account for attrition in the metrics they create, or stick to questions where the symptoms of attrition are unlikely to manifest, such as on trees with relatively recent crown ages, or where the quantity of interest is extrinsically justified as the units of edge length. One should thus use PD for a question explicitly about evolutionary history, the number of years things have been evolving for. However, it is hard to ignore attrition if we are making value judgements on the basis of something that accumulates, such as future options for humanity. To capture these things we need EvoHeritage and attrition. Whilst it will increase complexity, attrition will also make metrics more coherent and harder to manipulate. If we may summarise the original message of PD as 'species are not equal, their value depends on the branch lengths that represent their evolutionary histories', our message is 'unit lengths of branch are not equal, their value depends on both context and the evolution of their descendants'.

Acknowledgements

We thank (in alphabetic order) Ryan Chisholm, Dan Faith, Lucas Dias-Fernandes, Nisha Owen, Roseli Pellens, Sebastian Pippins, Laura Pollock and Yan Wong for useful discussions. We also thank the Rosindell lab group members, and members of the IUCN Phylogenetic Diversity Task Force. Through JR and WDP, this study is an output of the Georgina Mace Centre for the Living Planet at Imperial College London. KM and MS thank the New Zealand Marsden Fund (MFPUOC2005). JR thanks the Leverhulme Trust (RF-2022-497) for funding this research.

References

Balvanera, P., A. B. Pfisterer, N. Buchmann, J.-S. He, T. Nakashizuka, D. Raffaelli, and B. Schmid. 2006. Quantifying the evidence for biodiversity effects on ecosystem functioning and services. Ecology Letters 9:1146–1156.

Bassett, S. A., W. Young, M. P. Barnett, A. L. Cookson, W. C. McNabb, and N. C. Roy. 2015. Changes in composition of caecal microbiota associated with increased colon inflammation in interleukin-10 gene-deficient mice inoculated with *Enterococcus* species. Nutrients 7:1798–1816.

Behe, M. J. 2010. Experimental evolution, loss-of-function mutations, and "the first rule of adaptive evolution". The Quarterly Review of Biology 85:419–445.

Bennett, D. J., M. D. Sutton, and S. T. Turvey. 2018. Quantifying the living fossil concept. Palaeontologia Electronica 21.1. 14A.

Bozinovic, F., G. Ruiz, and M. Rosenmann. 2004. Energetics and torpor of a south american "living fossil", the microbiotheriid *Dromiciops gliroides*. Journal of Comparative Physiology B 174(4):293–297.

Brondizio, E. S., J. Settele, S. Díaz, and H. T. Ngo, eds. 2019. IPBES (2019): Global assessment report on biodiversity and ecosystem services of the Intergovernmental

Science-Policy Platform on Biodiversity and Ecosystem Services. IPBES Secretariat, Bonn, Germany.

Bryant, J., C. Lamanna, H. Morlon, A. Kerkhoff, B. Enquist, and J. Green. 2008. Microbes on mountainsides: contrasting elevational patterns of bacterial and plant diversity. Proceedings of the National Academy of Sciences 105:11505–11511.

Burbrink, F. and M. Gehara. 2018. The biogeography of deep time phylogenetic reticulation. Systematic Biology 67:743–755.

Cadotte, M., J. Cavender-Bares, D. Tilman, and T. Oakley. 2009. Using phylogenetic, functional and trait diversity to understand patterns of plant community productivity. PloS One 4(5):p.e5695.

Cadotte, M., T. Jonathan Davies, J. Regetz, S. Kembel, E. Cleland, and T. Oakley. 2010. Phylogenetic diversity metrics for ecological communities: integrating species richness, abundance and evolutionary history. Ecology Letters 13(1):96–105.

Carroll, S. B., B. Prud'homme, and N. Gompel. 2008. Regulating evolution. Scientific American 298:60–67.

Casane, D. and P. Laurenti. 2013. Why coelacanths are not 'living fossils' a review of molecular and morphological data. Bioessays 35(4):332–338.

Chao, A., C. Chiu, and L. Jost. 2010. Phylogenetic diversity measures based on hill numbers. Philosophical Transactions of the Royal Society B: Biological Sciences 365(1558):3599–3609.

Chao, A., C. Chiu, and L. Jost. 2014. Unifying species diversity, phylogenetic diversity, functional diversity, and related similarity and differentiation measures through hill numbers. Annual Review of Ecology, Evolution, and Systematics 45:297–324.

Crozier, R. H. 1992. Genetic diversity and the agony of choice. Biological Conservation 61:11–15.

Crozier, R. H., L. J. Dunnett, and P. Agapow. 2005. Phylogenetic biodiversity assessment based on systematic nomenclature. Evolutionary Bioinformatics Online 1:11–36.

Davies, T. J. 2021. Ecophylogenetics redux. Ecology Letters 24:1073–1088.

De Vos, J. M., L. N. Joppa, J. L. Gittleman, P. R. Stephens, and S. L. Pimm. 2015. Estimating the normal background rate of species extinction. Conservation biology 29:452–462.

Dodd, M., D. Papineau, T. Grenne, J. Slack, M. Rittner, F. Pirajno, J. O'Neil, and C. Little. 2017. Evidence for early life in earth's oldest hydrothermal vent precipitates. Nature 543(7643):60–64.

EDGE of Existence Programme. 2022. www.edgeofexistence.org .

Faith, D. 1994a. Biodiversity and systematics: the use and misuse of divergence information in assessing taxonomic diversity. Pacific Conservation Biology 1(1):53–57.

Faith, D. 1994b. Phylogenetic pattern and the quantification of organismal biodiversity. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences 345(1311):45–58.

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

Faith, D. P. 2013. Biodiversity and evolutionary history: useful extensions of the PD phylogenetic diversity assessment framework. Annals of the New York Academy of Sciences 1289:69–89.

Faith, D. P. 2018. Phylogenetic diversity and conservation evaluation: Perspectives on multiple values, indices, and scales of application. Pages 1–26 *in* Phylogenetic Diversity (R. Scherson and D. P. Faith, eds.). Springer.

Faith, D. P. and A. M. Baker. 2006. Phylogenetic diversity (PD) and biodiversity conservation: some bioinformatics challenges. Evolutionary bioinformatics 2:117693430600200007.

Felsenstein, J. 1985. Phylogenies and the comparative method. The American Naturalist 125(1):1–15.

Ferrer-Bonet, M. and I. Ruiz-Trillo. 2017. *Capsaspora owczarzaki*. Current Biology 27:R829–R830.

Forest, F., R. Grenyer, M. Rouget, T. Davies, R. Cowling, D. Faith, A. Balmford, J. Manning, S. Proches, M. van der Bank, and G. Reeves. 2007. Preserving the evolutionary potential of floras in biodiversity hotspots. Nature 445:757–760.

Gonzalez-Orozco, C. E., L. J. Pollock, A. H. Thornhill, B. D. Mishler, N. Knerr, S. W. Laffan, J. T. Miller, D. F. Rosauer, D. P. Faith, D. A. Nipperess, and H. Kujala. 2016. Phylogenetic approaches reveal biodiversity threats under climate change. Nature Climate Change 6:1110–1114.

Goździewska-Harłajczuk, K., J. Klećkowska-Nawrot, and K. Barszcz. 2018. Macroscopic and microscopic study of the tongue of the aardvark (*Orycteropus afer*, Orycteropodidae). Tissue and Cell 54:127–138.

Grace, J. B., T. M. Anderson, E. W. Seabloom, E. T. Borer, P. B. Adler, W. S. Harpole, Y. Hautier, H. Hillebrand, E. M. Lind, M. Pärtel, et al. 2016. Integrative modelling reveals mechanisms linking productivity and plant species richness. Nature 529:390–393.

Grandcolas, P., R. Nattier, and S. Trewick. 2014. Relict species: a relict concept? Trends in ecology and evolution 29(12):655–663.

Grandcolas, P. and S. A. Trewick. 2016. What is the meaning of extreme phylogenetic diversity? The case of phylogenetic relict species. Biodiversity Conservation and Phylogenetic Systematics 14:99–115.

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

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

Hill, M. O. 1973. Diversity and evenness: a unifying notation and its consequences. Ecology 54:427–432.

Hinchliff, C. E., S. A. Smith, J. F. Allman, J. G. Burleigh, R. Chaudhary, L. M. Coghill, K. A. Crandall, J. Deng, B. T. Drew, R. Gazis, and K. Gude. 2015. Synthesis of phylogeny and taxonomy into a comprehensive tree of life. Proceedings of the National Academy of Sciences 112:12764–12769.

Huchon, D., P. Chevret, U. Jordan, C. Kilpatrick, V. Ranwez, P. Jenkins, J. Brosius, and J. Schmitz. 2007. Multiple molecular evidences for a living mammalian fossil. Proceedings of the National Academy of Sciences 104(18):7495–7499.

Hughes, A. R., B. D. Inouye, M. T. Johnson, N. Underwood, and M. Vellend. 2008. Ecological consequences of genetic diversity. Ecology Letters 11:609–623.

Huson, D. and D. Bryant. 2006. Application of phylogenetic networks in evolutionary studies. Molecular biology and evolution 23:254–267.

Huson, D. and C. Scornavacca. 2011. A survey of combinatorial methods for phylogenetic networks. Genome Biology and Evolution 3:23–35.

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

IUCN. 2022. IUCN red list of threatened species .

Jost, L. 2006. Entropy and diversity. Oikos 113:363–375.

Kembel, S. W., P. D. Cowan, M. R. Helmus, W. K. Cornwell, H. Morlon, D. D. Ackerly, S. P. Blomberg, and C. O. Webb. 2010. Picante: R tools for integrating phylogenies and ecology. Bioinformatics 26:1463–1464.

Kuhn, T. S., A. O. Mooers, and G. H. Thomas. 2011. A simple polytomy resolver for dated phylogenies. Methods in Ecology and Evolution 2:427–436.

Laver, R. J. and J. D. Daza. 2021. Blind snakes beneath the surface: continuing the legacy of Richard Thomas. The Anatomical Record 304:2080–2084.

Letten, A. and W. Cornwell. 2015. Trees, branches and (square) roots: why evolutionary relatedness is not linearly related to functional distance. Methods in Ecology and Evolution 6(4):439–444.

Lidgard, S. and A. C. Love. 2018. Rethinking living fossils. BioScience 68:760–770.

Lidgard, S. and A. C. Love. 2021. The living fossil concept: reply to turner. Biology and Philosophy 36(2):1–16.

Lozupone, C. and R. Knight. 2005. Unifrac: a new phylogenetic method for comparing microbial communities. Applied and Environmental Microbiology 71:8228–8235.

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

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

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

McTavish, E. and L. Sanchez-Reyes. 2022. Chronosynth. https://github.com/OpenTreeOfLife/chronosynth.

Molina-Venegas, R., M. A. Rodríguez, M. Pardo-de Santayana, C. Ronquillo, and D. J. Mabberley. 2021. Maximum levels of global phylogenetic diversity efficiently capture plant services for humankind. Nature Ecology and Evolution 5:583–588.

Mooers, A., D. Faith, and W. Maddison. 2008. Converting endangered species categories to probabilities of extinction for phylogenetic conservation prioritization. PloS One 3(11):e3700.

Mooers, A., S. Heard, and E. Chrostowski. 2005. Evolutionary heritage as a metric for conservation. Phylogeny and conservation 8:120–138.

Mossel, E. and M. Steel. 2005. How much can evolved characters tell us about the tree that generated them? Pages 384–412 *in* Mathematics of Evolution and Phylogeny (O. Gascuel, ed.). Oxford University Press, Oxford.

Omland, K. E., L. G. Cook, and M. D. Crisp. 2008. Tree thinking for all biology: the problem with reading phylogenies as ladders of progress. BioEssays 30(9):854–867.

OneZoom Core Team. 2021. Onezoom tree of life explorer version 3.5 'chocolate chip starfish'. http://www.onezoom.org.

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

Paradis, E., J. Claude, and K. Strimmer. 2004. Ape: analyses of phylogenetics and evolution in R language. Bioinformatics 20:289–290.

Pausas, J. and M. Verdú. 2010. The jungle of methods for evaluating phenotypic and phylogenetic structure of communities. BioScience 60:614–625.

Pearse, W., M. Cadotte, J. Cavender-Bares, A. Ives, C. Tucker, S. Walker, and M. Helmus. 2015. Pez: Phylogenetics for the environmental sciences. Bioinformatics 31:2888–2890.

Pearse, W., P. Legendre, P. Peres-Neto, and T. Davies. 2019. The interaction of phylogeny and community structure: Linking the community composition and trait evolution of clades. Global Ecology and Biogeography 28:1499–1511.

Petchey, O. L. and K. J. Gaston. 2006. Functional diversity: back to basics and looking forward. Ecology Letters 9:741–758.

Phillips, M., T. Bennett, and M. Lee. 2009. Molecules, morphology, and ecology indicate a recent, amphibious ancestry for echidnas. Proceedings of the National Academy of Sciences 106(40):17089–17094.

R Core Team. 2021. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing , Vienna, Austria.

Redding, D. W. and A. O. Mooers. 2006. Incorporating evolutionary measures into conservation prioritization. Conservation Biology 20:1670–1678.

Rosauer, D. A. N., S. W. Laffan, M. D. Crisp, S. C. Donnellan, and L. G. Cook. 2009. Phylogenetic endemism: a new approach for identifying geographical concentrations of evolutionary history. Molecular Ecology 18:4061–4072.

Rosauer, D. F. and W. Jetz. 2015. Phylogenetic endemism in terrestrial mammals. Global Ecology and Biogeography 24:168–179.

Sanderson, M. J. 2002. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. Molecular Biology and Evolution 19:101–109.

Shalchian-Tabrizi, K., M. A. Minge, M. Espelund, R. Orr, T. Ruden, K. S. Jakobsen, and T. Cavalier-Smith. 2008. Multigene phylogeny of choanozoa and the origin of animals. PloS One 3:e2098.

Steel, M., A. Mimoto, and A. Mooers. 2007. Hedging our bets: the expected contribution of species to future phylogenetic diversity. Evolutionary Bioinformatics 3:117693430700300024.

Swenson, N. G. 2013. The assembly of tropical tree communities–the advances and shortcomings of phylogenetic and functional trait analyses. Ecography 36:264–276.

Tilman, D. 2021. Plant aboveground biomass data: Biodiversity II: Effects of Plant Biodiversity on Population and Ecosystem Processes version 10. Environmental Data Initiative .

Tilman, D., J. Knops, D. Wedin, P. Reich, M. Ritchie, and E. Siemann. 1997. The influence of functional diversity and composition on ecosystem processes. Science 277:1300–1302.

Tilman, D., P. Reich, J. Knops, D. Wedin, T. Mielke, and C. Lehman. 2001. Diversity and productivity in a long-term grassland experiment. Science 294(5543):843–845.

Tilman, D., D. Wedin, and J. Knops. 1996. Productivity and sustainability influenced by biodiversity in grassland ecosystems. Nature 379(6567):718–720.

Trewick, S. and M. Morgan-Richards. 2016. Phylogenetics and conservation in new zealand: the long and the short of it. Pages 81–97 *in* Biodiversity conservation and phylogenetic systematics (R. Pellens and P. Grandcolas, eds.). Springer.

Tucker, C., M. Cadotte, S. Carvalho, T. Davies, S. Ferrier, S. Fritz, R. Grenyer, M. Helmus, L. Jin, A. Mooers, and S. Pavoine. 2017. A guide to phylogenetic metrics for conservation, community ecology and macroecology. Biological Reviews 92:698–715.

Tucker, C. M., T. Aze, M. W. Cadotte, J. L. Cantalapiedra, C. Chisholm, S. Díaz, R. Grenyer, D. Huang, F. Mazel, W. D. Pearse, et al. 2019. Assessing the utility of conserving evolutionary history. Biological Reviews 94:1740–1760.

Turner, D. 2019. In defense of living fossils. Biology and Philosophy 34(2):1–22.

Upham, N. S., J. A. Esselstyn, and W. Jetz. 2019. Inferring the mammal tree: species-level sets of phylogenies for questions in ecology, evolution, and conservation. PLoS Biology 17(12):e3000494.

Wezel, A. and S. Bender. 2002. Agricultural land use in the coastal area of the Alexander von Humboldt National Park, Cuba and its implication for conservation and sustainability. GeoJournal 57(4):241–249.

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

Wong, Y. and J. Rosindell. 2022. Dynamic visualisation of million-tip trees: the OneZoom project. Methods in Ecology and Evolution 13:303–313.

Xia, X., Z. Xie, M. Salemi, L. Chen, and Y. Wang. 2003. An index of substitution saturation and its application. Molecular Phylogenetics and Evolution 26:1–7.

Zanne, A., D. Tank, W. Cornwell, J. Eastman, S. Smith, R. FitzJohn, D. McGlinn, B. O'Meara, A. Moles, P. Reich, and D. Royer. 2014. Three keys to the radiation of angiosperms into freezing environments. Nature 506:89–92.

Phylogenetic Biodiversity Metrics Should Account for Both Accumulation and Attrition of Evolutionary Heritage

September 8, 2022

James Rosindell[1,*], Kerry Manson[2], Rikki Gumbs[3], Will Pearse[1] and Mike Steel[2]

[1] *Department of Life Sciences, Silwood Park Campus, Imperial College London, Buckhurst Road, Ascot, Berkshire, SL5 7PY, United Kingdom*
[2] *Biomathematics Research Centre, University of Canterbury, Christchurch, New Zealand*
[3] *EDGE of Existence Programme, Zoological Society of London, Regent's Park, London NW1 4RY, UK*

**Corresponding author j.rosindell@imperial.ac.uk.*

Contents

1 Introducing EvoHeritage (ϕ) 3

1.1 Foundational results . . . 3

1.1.1 Proof: Generation of new vertices . . . 3

1.1.2 Proof: Suppression of vertices of degree 2 . . . 4

1.1.3 Proof: fixed EvoHeritage accumulation does not lose generality . . . 5

1.2 The role of the origin of life . . . 6

1.2.1 Why include the origin of life? . . . 6

1.2.2 Sensitivity to fluctuations in the origin of life . . . 6

1.3 Relationship to PD . . . 7

1.3.1 Proof: ϕ_ρ converges to a count of vertices (species richness) when $\rho \to \infty$ . . . 7

1.3.2 Proof: ϕ_ρ converges to PD when $\rho \to 0$ . . . 9

1.3.3 Conditions for ϕ to correspond to PD . . . 10

1.3.4 Proof: maximising PD does not always maximise ϕ_ρ . . . 20

1.4 EvoHeritage under stochastic conditions . . . 22

1.5 Sets that maximise PD do not necessarily maximise FD in our stochastic model . . . 26

2 Developing a complete EvoHeritage calculus 28

2.1 Relationship between generalised phylogenetic difference and classic phylogenetic distance for a pair of vertices . . . 28

2.2 Relationship to genetic diversity . . . 29

2.3 Proof: the power set based calculation of ϕ is equivalent to direct calculation . . . 30

2.4 Summing the partitioned-ϕ of individual vertices . . . 32

2.5 Values of partitioned-ϕ and unique-ϕ for extremal values of attrition 33

2.6 Supplementary figure illustrating ED and partitioned-ϕ . . . 35

1 Introducing EvoHeritage (ϕ)

1.1 Foundational results

1.1.1 Proof: Generation of new vertices

We show that any edge can be split into two edges to create an additional vertex without affecting the results of EvoHeritage calculations on existing vertices. Consider edge $e \in E$ with the associated values $\alpha_i(e)$ and $\beta_i(e)$ for each $i \in I$. Firstly $\alpha_i(e)$ gives the net amount of EvoHeritage of form i accumulated along each edge $e \in E$. Secondly $\beta_i(e)$ gives the proportion of EvoHeritage of form i present at the start of the edge e that survives attrition along the entire edge to still be present at its end vertex $d^V(e)$.

We will describe a method to insert a new vertex v^* along edge e, splitting the edge into two parts e_1 and e_2. Let $L(e)$ give the length of e. We describe the position of v^* by the equations:

$$L(e_1) = L(e) \cdot (1 - x) \tag{1}$$

$$L(e_2) = L(e) \cdot x \tag{2}$$

The EvoHeritage of form i accumulated along both new edges and surviving attrition to arrive at $d^V(e)$ must remain the same as it was before before the new vertex was added. Therefore

$$\alpha_i(e_1) \cdot \beta_i(e_2) + \alpha_i(e_2) = \alpha_i(e). \tag{3}$$

In addition, the total EvoHeritage of form i lost to attrition along both new edges must also remain unchanged and therefore

$$\beta_i(e_1) \cdot \beta_i(e_2) = \beta_i(e) \tag{4}$$

Provided that $\beta_i(e) \leq \beta_i(e_1) \leq 1$, $\beta_i(e) \leq \beta_i(e_2) \leq 1$, $0 \leq \alpha_i(e_2) \leq \alpha_i(e)$ and $0 \leq \alpha_i(e_1) \leq \frac{\alpha_i(e)}{\beta_i(e_2)}$ we can see numerous possible solutions for placement of the vertex that meet the condition that satisfy Eqns. 1 and 2. A general family of solutions is provided by taking a variable $0 \leq x \leq 1$ to define positioning of the new vertex. This leads to the natural solution below, which can be readily seen as having the required properties.

$$\beta_i(e_1) = \beta_i(e)^{1-x} \tag{5}$$

$$\beta_i(e_2) = \beta_i(e)^x \tag{6}$$

$$\alpha_i(e_2) = \alpha_i(e) \cdot x \tag{7}$$

$$\alpha_i(e_1) = \frac{\alpha_i(e) \cdot (1 - x)}{\beta_i(e)^x} \tag{8}$$

Under standard conditions, with attrition parameter ρ, we can set the following values:

$$\alpha_i(e_1) = \frac{1 - \exp(-\rho \cdot L(e_1))}{\rho} \tag{9}$$

$$\alpha_i(e_2) = \frac{1 - \exp(-\rho \cdot L(e_2))}{\rho} \tag{10}$$

$$\beta_i(e_1) = \exp(-\rho \cdot L(e_1)) \tag{11}$$

$$\beta_i(e_2) = \exp(-\rho \cdot L(e_2)) \tag{12}$$

Now, let us verify that these solutions are consistent with Eqns. 3 and 4.

$$\begin{aligned}
&\alpha_i(e_1) \cdot \beta_i(e_2) + \alpha_i(e_2) \\
=&\left(\frac{1 - \exp(-\rho \cdot L(e_1))}{\rho}\right) \cdot (\exp(-\rho \cdot L(e_2))) + \frac{1 - \exp(-\rho \cdot L(e_2))}{\rho} \\
=&\frac{\exp(-\rho \cdot L(e_2)) - \exp(-\rho \cdot (L(e_1) + L(e_2))) + 1 - \exp(-\rho \cdot L(e_2))}{\rho} \\
=&\frac{1 - \exp(-\rho \cdot L(e))}{\rho} = \alpha_i(e)
\end{aligned} \tag{13}$$

$$\beta_i(e_1) \cdot \beta_i(e_2) = \exp(-\rho \cdot (L(e_1) + L(e_2))) = \exp(-\rho \cdot L(e)) = \beta_i(e) \tag{14}$$

Hence for a new vertex added anywhere along edge e, consistent new values of α and β may be found for the newly created edges e_1 and e_2.

1.1.2 Proof: Suppression of vertices of degree 2

As a couterpoint to the generation of new vertices above, we now show that we can suppress vertices of degree two without impacting the calculation of ϕ. The values for EvoHeritage accumulation and attrition on the resultant edge are constructed from the original values of the merged edges. The construction is stated for a single form of EvoHeritage, with the understanding that if more than one form of EvoHeritage is present on an edge, independent constructions occur for each form.

Lemma 1.1 *Let $T = (V, E)$ be a rooted phylogenetic tree, and for some $n \geq 3$, let $P = v_1, v_2, \ldots, v_n$ be a path in T. Let $e_i = (v_i, v_{i+1})$ and suppose $v_2, \ldots, v_{n-1}$ are all vertices of degree 2. If e^* is the edge that results from suppressing $v_2, \ldots, v_{n-1}$, then setting $\alpha(e^*) = \sum_{i=1}^{n-1} \alpha(e_i) \prod_{j=i+1}^{n-1} \beta(e_j)$, and $\beta(e^*) = \prod_{i=1}^{n-1} \beta(e_i)$ does not alter ϕ values.*

Proof 1.2 *We prove that one vertex of degree 2 can be suppressed. By repeated application of this process we can extend the result to the entire path P.*

Consider v_1, v_2, v_3, a subpath of P, and suppose that v_1 is, of the three vertices, the closest to the root of T. Let R be the path from the root vertex to v_1, and let T' be the subtree descending from v_3. If $A \subseteq V(T')$, then

$$\begin{aligned}\phi(A) &= \sum_{e \in T'} \alpha(e) p(d^V(e), A) + \alpha(e_2) p(v_3, A) + \alpha(e_1) p(v_2, A) + \sum_{e \in R} \alpha(e) p(d^V(e), A) \\ &= \sum_{e \in T'} \alpha(e) p(d^V(e), A) + p(v_3, A)[\alpha(e_2) + \alpha(e_1)\beta(e_2) + \phi(\{v_1\})\beta(e_1)\beta(e_2)].\end{aligned}$$

Hence if we set $\alpha(e^*) = \alpha(e_2) + \alpha(e_1)\beta(e_2)$ and $\beta(e^*) = \beta(e_1)\beta(e_2)$ we do not change $\phi(A)$ after the suppression of v_2.

1.1.3 Proof: fixed EvoHeritage accumulation does not lose generality

Under standard conditions, EvoHeritage accumulates at a fixed deterministic rate of one unit of EvoHeritage per unit of edge length. This may seem like an arbitrary choice, so let us suppose that EvoHeritage accumulates instead at a fixed rate of x per unit of edge length. This would lead to a new value for $\alpha(e)$ (we omit the EvoHeritage form subscript as there is only a single EvoHeritage form under standard conditions) given by $\alpha^*(e)$, where

$$\begin{aligned}\alpha^*(e) &= \int_0^{L(e)} x \cdot e^{-\rho l}\, dl \\ &= \frac{x(1 - e^{-\rho L(e)})}{\rho} \\ &= x \cdot \alpha(e)\end{aligned} \tag{15}$$

Let u be a unit length edge. Then at an accumaltion rate of x per unit length a standardised unit changes from S_u to S_u^*, given by $S_u^* = \alpha^*(u) = x \cdot \alpha(u) = x \cdot S_u$. For a set of vertices A, let ϕ^* be the EvoHeritage of A with an accumulation rate of x. As a result the value of $\phi_\rho^*(A)$ is given by

$$\begin{aligned}\phi_\rho^*(A) = \frac{\phi^*(A)}{S_u^*} &= \frac{\sum_{e \in E} \alpha^*(e) \cdot p(d^V(e), A)}{S_u^*} \\ &= \frac{x \cdot \sum_{e \in E} \alpha(e) \cdot p(d^V(e), A)}{x \cdot S_u} \\ &= \frac{\phi(A)}{S_u} = \phi_\rho(A).\end{aligned} \tag{16}$$

This shows the independence of ϕ_ρ from the choice of x. Hence we can set the rate of EvoHeritage accumulation to one unit of EvoHeritage per unit of edge length without any loss of generality.

1.2 The role of the origin of life

1.2.1 Why include the origin of life?

As discussed in the main text, it is beneficial to include the origin of life $\hat{o}$ as a vertex in our formulation. One reason is to make EvoHeritage of groups with different crown ages comparable. Without including the origin of life the total PD of a group with a shallow crown (or stem) age would be unfairly penalised when it comes to a comparison with a group that has a deeper crown (or stem) age.

Another motivation is to enable a consistent definition of the EvoHeritage (or PD) of a single vertex (or species). Let us begin by considering classic PD of a single species on a tree. The distinction between unique-PD and PD is important here. The unique-PD of a single species should clearly be that species' terminal edge (branch) length. However, the total (not necessarily unique) PD of a single species has often been taken to also mean the terminal branch length of that species, if it is defined at all. We argue that a more logically consistent answer for the total PD of a single species should be the date back to the origin of life vertex $\hat{o}$, approximately 4 billion years (Dodd et al., 2017). This is because any living species is the result of that number of years of evolutionary history (some of which is unique to that species and some of which is not).

Set against these reasons for incorporating the origin of life $\hat{o}$ as a vertex, is the uncertainty in the value of its date, represented hereafter as τ. There may also be an intuition that the value of τ should not be relevant to measurement of contemporary diversity, as well as being a quantity that increases (albeit by a minute percentage of its total) with each further year that passes. So the value of τ measured now in 2022 will be 1 year shorter than the value measured in 2023. To resolve this we explore the sensitivity of our results to the value of τ.

1.2.2 Sensitivity to fluctuations in the origin of life

Consider the equation for a standardised unit based on the net EvoHeritage accumulation (the α value) of an edge with length τ:

$$S_u = \frac{1 - \exp(-\rho \cdot \tau)}{\rho}. \tag{17}$$

We can also write this as

$$S_u = \frac{1}{\rho}\left(1 - \frac{1}{\exp(\rho \cdot \tau)}\right) \tag{18}$$

It is now clear, given that τ is certainly very large, that unless ρ is extremely small $S_u \approx \frac{1}{\rho}$ and is independent of τ. We may also wish to constrain ρ from being so small as to prevent current uncertainty in τ from having any noticeable effect. Take $\tau_{min} \leq \tau \leq \tau_{max}$ to express the uncertainty in τ and suppose that we wish our resulting value of S_u to be accurate to within an error of N significant figures. We therefore require

$$\frac{\frac{1}{\rho}\left(1-\frac{1}{\exp(\rho\cdot\tau_{max})}\right)-\frac{1}{\rho}\left(1-\frac{1}{\exp(\rho\cdot\tau_{min})}\right)}{\frac{1}{\rho}\left(1-\frac{1}{\exp(\rho\cdot\tau_{min})}\right)} \leq 10^{-N} \tag{19}$$

Let us use the current best estimates of $\tau_{min} = 3.77 \times 10^9$ years and $\tau_{max} = 4.28 \times 10^9$ years following (Dodd et al., 2017). Let us take $N = 5$ for 5 significant figures, an accuracy greater than that used in many phylogenetic calculations. Equation 19 can now be solved numerically to give a value of ρ sufficient to certainly remove undesirable dependency on the uncertainty around the date of the origin of life. The result is $\rho \geq 10^{-8.525}$, thus we probably should not consider values of ρ below this value if we have opted to use the origin of life to define a standardized unit, and want to be certain that the origin of life uncertainty is not affecting our results.

1.3 Relationship to PD

1.3.1 Proof: ϕ_ρ converges to a count of vertices (species richness) when $\rho \to \infty$

We seek to show that $\lim_{\rho\to\infty} \phi_\rho(A) = \sum_{v\in A} |A^E(v)|$ where $A \subseteq V$. In words, the EvoHeritage of a set A, under standard conditions, converges to a count of edges that lead to vertices in A (species richness if vertices represent species in a tree) as attrition grows large. Let us begin by considering the values of $\alpha(e)$ (net EvoHeritage accumulation along edge e), $\beta(e)$ (proportion of EvoHeritage lost along edge e) and S_u (the size of a standardised unit) in this limit. For convenience, we repeat the definitions from the main text.

$$\alpha(e) = \frac{1-\exp(-\rho\cdot L(e))}{\rho} \tag{20}$$

$$\beta(e) = \exp(-\rho\cdot L(e)) \tag{21}$$

$$S_u = \frac{1-\exp(-\rho\cdot\tau)}{\rho} \tag{22}$$

Here we now have to assume that $L(e) > 0$. This is reasonable because one could collapse any zero branch lengths by merging the connected nodes, maybe resulting in a polytomy. Consequently $(\rho\cdot L(e)) \to \infty$ as $\rho \to \infty$, and so

$$\lim_{\rho\to\infty} \beta(e) = \lim_{\rho\to\infty} \exp(-\rho\cdot L(e)) = 0. \tag{23}$$

Let us now consider S_u.

$$\lim_{\rho\to\infty} S_u = \lim_{\rho\to\infty} \frac{1-\exp(-\rho\cdot\tau)}{\rho} = 0 \tag{24}$$

However, the value of ϕ_ρ will later be calculated as a fraction with S_u on the denominator, so we cannot take this limit and should instead write S_u as an approximation that will be valid in the limit of $\rho \to \infty$

$$S_u = \frac{1 - \exp(-\rho \cdot \tau)}{\rho} = \frac{1}{\rho} - \frac{1}{\rho \cdot \exp(\rho \cdot \tau)} \approx \frac{1}{\rho} \tag{25}$$

Following the same logic and again assuming $L(e) > 0$ we can say that $\alpha(e)$ is independent of $L(e)$ and can be approximated by

$$\alpha(e) \approx \frac{1}{\rho} \tag{26}$$

Now let us apply these values of α and β to give ϕ under standard conditions. Recall that under standard conditions there is only one form of EvoHeritage, and thus taking equation 1 from the main text we can see that

$$p(v, A) = \begin{cases} 1 & \text{when } v \in A \\ 1 - \prod_{j \in D^E(v)} (1 - p(D^V(j), A) \cdot \beta(j)) & \text{otherwise} \end{cases} \tag{27}$$

And so taking the limit for $\rho \to \infty$ we get $\beta(j) = 0$ and consequently that

$$p(v, A) = \begin{cases} 1 & \text{when } v \in A \\ 0 & \text{otherwise} \end{cases} \tag{28}$$

Now recall equation 2 from the main text.

$$\phi(A) = \sum_{i \in I} \sum_{e \in E} \alpha_i(e) \cdot p_i(D^V(e), A) \tag{29}$$

The outer sum is eliminated by again noting that there is only one form of EvoHeritage, and taking the limit for $\rho \to \infty$ again we can write

$$\phi(A) = \sum_{e \in E} \alpha(e) \cdot p(D^V(e), A) \approx \frac{1}{\rho} \cdot \left| \{ e \in E | D^V(e) \in A \} \right| = \frac{1}{\rho} \cdot \sum_{v \in A} |A^E(v)| \tag{30}$$

Now recall that

$$\phi_\rho(A) = \frac{\phi(A)}{S_u} = \frac{\frac{1}{\rho} \cdot \sum_{v \in A} |A^E(v)|}{\frac{1}{\rho}} = \sum_{v \in A} |A^E(v)| \tag{31}$$

Which is a count of the ancestral edges of vertices in A. If we consider a tree there would only exactly one edge leading towards each vertex and thus the result could equally be considered as a count of vertices (species richness if vertices represent species).

1.3.2 Proof: ϕ_ρ converges to PD when $\rho \to 0$

We aim to prove $\lim_{\rho\to 0} \phi_\rho(A) = PD(A)$. In words, $\phi_\rho(A)$ converges to classic PD of the set of vertices when $\rho \to 0$. Let us begin, as in section 1.3.1, by considering the values of $\alpha_1(e)$ (net EvoHeritage accumulation along edge e), $\beta_1(e)$ (proportion of EvoHeritage lost along edge e) and S_u (the size of a standardised unit) in this limit.

$$\alpha_1(e) = \frac{1 - \exp(-\rho \cdot L(e))}{\rho} \tag{32}$$

Taking the limit $\rho \to 0$ and being mindful of the fact that ρ appears on the denominator, our first step is to perform a Taylor expansion of the exponential term that will be perfect approximation in the limit as $\rho \to 0$

$$\alpha_1(e) = \frac{1 - \exp(-\rho \cdot L(e))}{\rho} \approx \frac{1 - (1 - \rho \cdot L(e))}{\rho} = L(e) \tag{33}$$

Similarly, in the limit as $\rho \to 0$ we can say

$$S_u = \frac{1 - \exp(-\rho)}{\rho} \approx \frac{1 - (1 - \rho)}{\rho} = 1 \tag{34}$$

And also

$$\beta_1(e) = \exp(-\rho \cdot L(e)) = 1 \tag{35}$$

Now following the derivation of ϕ

$$p_1(v, A) = \begin{cases} 1 & \text{when } v \in A \\ 1 - \prod_{j \in D^E(v)} (1 - p_1(D^V(j), A)) & \text{otherwise} \end{cases} \tag{36}$$

Which means that

$$p_1(D^V(e), A) = \begin{cases} 1 & \text{when at least one descendant of e is in set A} \\ 0 & \text{otherwise} \end{cases} \tag{37}$$

Which is equivalent to saying $p_1(D^V(e), A) = 1$ if and only if e is part of the minimum spanning set of edges for the vertices in A including the root vertex.

$$\phi(A) = \sum_{e \in E} \alpha_1(e) \cdot p_1(D^V(e), A) = \sum_{e \in E} L(e) \cdot p_1(D^V(e), A) \tag{38}$$

Which is classic PD of the set of vertices including the root vertex. To complete the proof,

$$\phi_\rho(A) = \frac{\phi(A)}{S_u} = \mathrm{PD}(A) \tag{39}$$

1.3.3 Conditions for ϕ to correspond to PD

In this section we aim to show which patterns of EvoHeritage attrition allow ϕ to correspond to PD. We will show that ϕ represents a true generalisation of PD. We consider only the case where one form of EvoHeritage is involved, and hence omit the EvoHeritage subscripts.

Let $T = (V, E)$ be a rooted phylogenetic tree. We say that ϕ corresponds to PD on T if for every set of leaves A we have $\phi(A) = PD(A)$. For each edge $e \in E(T)$, the EvoHeritage accumulation and attrition along e are, respectively, denoted by $\alpha(e)$ and $\beta(e)$. To achieve greater generality, we allow the α and β values of edges to be independent of each other (that is, not necessarily determined by a specific ρ value across an entire tree). The equation $\beta(e) = 1$ describes the situation where no EvoHeritage attrition occurs along e.

Furthermore, let $L(e)$ denote the length of e. Because we allow T to contain polytomies, we require that the length of every edge in $E(T)$ is strictly positive. We also make use of the terms parent and child vertex. If $v_1, v_2 \in V(T)$ share an edge, we call v_1 the *parent* of v_2 if v_1 is closer to the root of T. In this case we call v_2 a *child* of v_1. Additionally, we will use the term clade in the following specific sense. Let $e \in E(T)$ be an edge incident with the root of T. The *clade* of T determined by e is a subtree of T comprised of the root vertex, e, and the entire subtree descending from e.

Note that $\beta(e)$ is arbitrary if e is incident with the root vertex as no EvoHeritage precedes the root. If every edge exhibits no EvoHeritage attrition (ignoring for the moment those edges incident with the root), then we are able to make ϕ always equal PD by setting $\alpha(e) = L(e)$ for every edge e.

An example of a tree for which ϕ always corresponds to PD is given in Figure 1. Two edges close to the root exhibit EvoHeritage attrition. By careful selection of the accumulation values we can get the desired correspondence between ϕ and PD. These selections are given below. Note that we require $L(e) - 2L(z) \geq 0$ and $L(f) - 2L(z) \geq 0$ in order for $\alpha(e)$ and $\alpha(f)$ to be non-negative. Calculations of ϕ and PD for selected sets appear in Table 1. However, Theorem 1.4 illustrates that this particular example is the exception rather than the rule. For 'most' tree shapes we require every edge to exhibit no EvoHeritage attrition for ϕ to correspond to PD.

$\beta(a) = 1$	$\alpha(a) = L(a)$	$\beta(e) = \frac{1}{3}$	$\alpha(e) = L(e) - 2L(z)$
$\beta(b) = 1$	$\alpha(b) = L(b)$	$\beta(f) = \frac{1}{3}$	$\alpha(f) = L(f) - 2L(z)$
$\beta(c) = 1$	$\alpha(c) = L(c)$	$\beta(z) = 1$	$\alpha(z) = 9L(z)$
$\beta(d) = 1$	$\alpha(d) = L(d)$		

In light of Lemma 1.1 we consider phylogenetic trees with no monotomy vertices. After monotomy vertices are suppressed, the clades of a rooted phylogenetic tree may be categorised into three types. The types are differentiated by how many edges descend from the child of the root vertex.

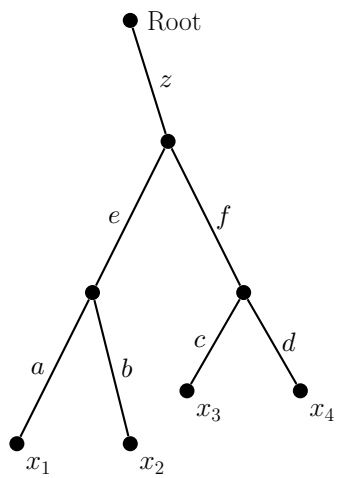

Figure 1: An example of a tree for which ϕ always corresponds to PD

Definition A *type-A clade* is one where the child of the root vertex has 3 or more descendant edges. A *type-B clade* is one where the child of the root vertex has exactly 2 descendant edges. If a type-B clade contains exactly two leaves we call it a *cherry*. A *type-C clade* is one which consists of a single edge.

Figure 2 shows a tree $T = (V, E)$ with one of each of the clade types. Triangles represent subtrees of T, possibly containing no more than the single vertex indicated. The dotted line divides $E(T)$ into two parts, according to the conditions of Theorem 1.4.

Now suppose EvoHeritage attrition occurs along every edge of a phylogenetic tree. If ϕ is to correspond to PD for every set of leaves of this tree, its structure is quite limited. Any descendant of the root can descend to at most two leaves. Figure 3 illustrates some of these limited structures, which contain only single-edge clades or Y-shaped ones. The results that follow formalise this idea.

Proposition 1.3 *Let $T = (V, E)$ be a rooted phylogenetic tree, on leaf set X, containing no monotomy vertices. Suppose $\beta(e) < 1$ for every edge $e \in E(T)$. Then $\phi(A) = PD(A)$ for all sets $A \subseteq X$ if and only if the clades of T are all cherries, single edges or a combination of both.*

The proof of Proposition 1.3 follows as a consequence of the following more general result.

Theorem 1.4 *Let $T = (V, E)$ be a rooted phylogenetic tree containing no monotomy vertices. Let X be the leaf set of T, and r denote the root vertex. Then $\phi(A) = PD(A)$ for every set of leaves $A \subseteq X$ if and only if each of the following conditions hold:*

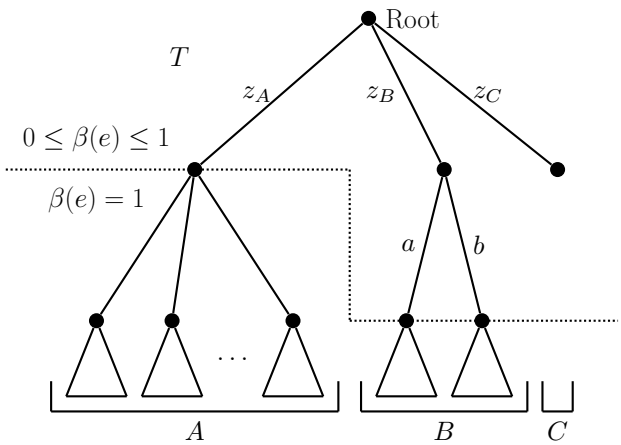

Figure 2: An example of the three clade types: A, B and C. Triangles represent sub trees. A clades of type B is considered a 'cherry' if it contains exactly two terminal vertices

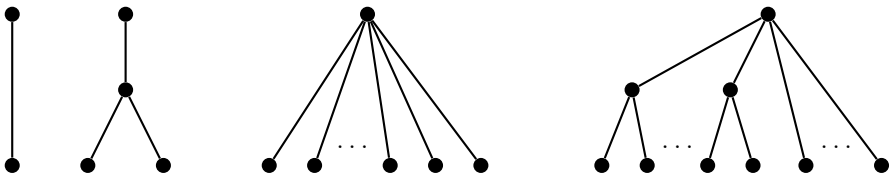

Figure 3: Examples of trees with EvoHeritage attrition occurring on every edge where ϕ may equal PD for every set of leaves: two trivial trees consisting of only a single edge or a cherry, a star tree, a tree composed entirely of cherries and single-edge clades.

A	$\phi(A)$	$PD(A)$
$\{x_1\}$	$\alpha(a)+\alpha(e)+\frac{1}{3}\alpha(z)$	$L(a)+L(e)+L(z)$
$\{x_2\}$	$\alpha(b)+\alpha(e)+\frac{1}{3}\alpha(z)$	$L(b)+L(e)+L(z)$
$\{x_3\}$	$\alpha(c)+\alpha(f)+\frac{1}{3}\alpha(z)$	$L(c)+L(f)+L(z)$
$\{x_1,x_2\}$	$\alpha(a)+\alpha(b)+\alpha(e)+\frac{1}{3}\alpha(z)$	$L(a)+L(b)+L(e)+L(z)$
$\{x_2,x_3\}$	$\alpha(b)+\alpha(c)$ $+\alpha(e)+\alpha(f)+\frac{5}{9}\alpha(z)$	$L(b)+L(c)$ $+L(e)+L(f)+L(z)$
$\{x_1,x_2,x_3\}$	$\alpha(a)+\alpha(b)+\alpha(c)$ $+\alpha(e)+\alpha(f)+\frac{5}{9}\alpha(z)$	$L(a)+L(b)+L(c)$ $+L(e)+L(f)+L(z)$
$\{x_1,x_2,x_3,x_4\}$	$\alpha(a)+\alpha(b)+\alpha(c)+\alpha(d)$ $+\alpha(e)+\alpha(f)+\frac{5}{9}\alpha(z)$	$L(a)+L(b)+L(c)+L(d)$ $+L(e)+L(f)+L(z)$

Table 1

- *In type-A clades:*
  1. *Every edge e not incident with the root has $\beta(e)=1$.*
  2. *For the edge z_A incident with the root, $0\leq\beta(z_A)\leq 1$.*
  3. *For every edge e in the clade, we have $\alpha(e)=L(e)$.*
- *In type-B clades:*
  1. *Every edge e not incident with the child of the root has $\beta(e)=1$.*
  2. *For the edge z_B incident with the root, and edges a,b immediate descendants of z_B, we allow $0\leq\beta(z_B),\beta(a),\beta(b)\leq 1$, provided that $\alpha(z_B)=\frac{L(z_B)}{\beta(a)\beta(b)}$, $\alpha(a)=L(a)+L(z_B)(1-\frac{1}{\beta(b)})$, and $\alpha(b)=L(b)+L(z_B)(1-\frac{1}{\beta(a)})$.*
  3. *For every edge $e\notin\{a,b,z_B\}$, we have $\alpha(e)=L(e)$.*
- *In type-C clades:*
  1. *For the edge z_C incident with the root, $0\leq\beta(z_C)\leq 1$.*
  2. *$\alpha(z_C)=L(z_C)$.*

Proof 1.5 *Let $\mathcal{T}=(V(\mathcal{T}),E(\mathcal{T}))$ be a clade of T containing the edge $z=(r,v')$. Observe that terms derived from distinct clades do not interact in either PD or ϕ calculations. As such, characterising the values of $\alpha(e)$ and $\beta(e)$ for each edge $e\in E(\mathcal{T})$ will characterise these for T as a whole.*

If $\mathcal{T}$ is a type-C clade, then it consists of only the edge z. The only set for which ϕ needs to correspond to PD is $\{v'\}$. Hence $\alpha(z) = \phi(\{v'\}) = PD(\{v'\}) = L(z)$. Otherwise, we suppose $\mathcal{T}$ contains at least 3 edges.

For a vertex v of $\mathcal{T}$, we say it satisfies the ($\star$) property if the following condition holds:

($\star$) Every edge e in the subtree descending from v has $\beta(e) = 1$.

Note that every terminal vertex of $\mathcal{T}$ satisfies ($\star$). The proof proceeds by showing that the ($\star$) property can be extended iteratively from the leaves to interior vertices apart from perhaps v'. Following this we prove that if v' has three or more descendant edges, then it too has the ($\star$) property. This means that we require v' to have exactly two descendant edges for some edge (other than z) to exhibit EvoHeritage attrition.

Let v be an interior vertex of $\mathcal{T}$ whose children all satisfy the ($\star$) property. If v is the root vertex, then the conditions of the theorem hold regardless of whether $\mathcal{T}$ is a type-A or type-B clade. Next suppose v is not the root, and let u be the parent vertex of v. We consider two cases: either u is the root vertex or it is not.

Case One: *Suppose u is not the root vertex.*

Since u is not the root, it has a child vertex w distinct from v. Select two leaves x_1, x_2 from the descendants of v such that the path from x_1 (resp. x_2) to v contains edge a (resp. b) incident to v, and $a \neq b$. Further, select terminal vertex x_3 descended from w. Figure 4(a) illustrates these choices.

In calculating the ϕ score of any subset of $\{x_1, x_2, x_3\}$, only those edges and paths shown can possibly make a non-zero contribution to the calculation. Hence we can use Lemma 1.1 to replace the path from v to x_1 (resp. x_2) with the edge a^* (resp. b^*), noting that $\beta(a) = 1$ if and only if $\beta(a^*) = 1$, and $\beta(b) = 1$ if and only if $\beta(b^*) = 1$. Similarly, we replace the path from vertex u to the root vertex (resp. x_3) with the edge z^* (resp. d^*). The result of these changes is shown in Figure 4(b). Table 2 shows ϕ and PD calculations for subsets of $\{x_1, x_2, x_3\}$, and a calculation of $\phi(\{u\})$.

Observe that

$$\phi(\{x_1\}) + \phi(\{x_3\}) - \phi(\{x_1, x_3\}) = \beta(a^*)\beta(c)\beta(d^*)\phi(\{u\}). \tag{40}$$

Now, under the assumption that ϕ corresponds to PD for these sets, and noting also that $PD(\{x_1\}) + PD(\{x_3\}) - PD(\{x_1, x_3\}) = L(z^*) > 0$, it follows that the left hand side of (40), and hence the right hand side also, must be strictly positive. This implies each factor is nonzero: $\beta(a^*), \beta(c), \beta(d^*), \phi(\{u\}) > 0$. Next, by expressing $L(a^*)$ in terms of ϕ, we will determine that $\beta(b^*) = 1$.

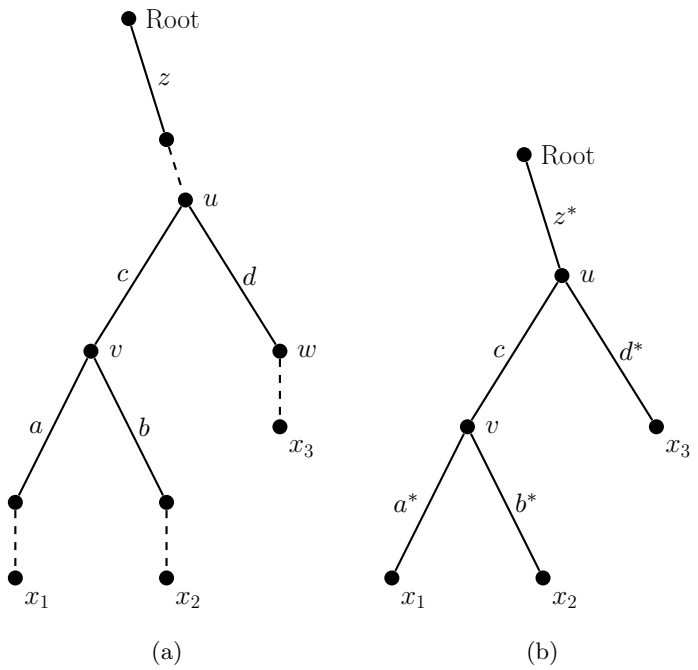

Figure 4: Selected parts of the tree T referred to in the main narrative

A	$\phi(A)$	$PD(A)$
$\{u\}$	$\alpha(z^*)$	
$\{x_1\}$	$\alpha(a^*) + \alpha(c)\beta(a^*) + \alpha(z^*)\beta(a^*)\beta(c)$	$L(a^*) + L(c) + L(z^*)$
$\{x_2\}$	$\alpha(b^*) + \alpha(c)\beta(b^*) + \alpha(z^*)\beta(b^*)\beta(c)$	$L(b^*) + L(c) + L(z^*)$
$\{x_3\}$	$\alpha(d^*) + \alpha(z^*)\beta(d^*)$	$L(d^*) + L(z^*)$
$\{x_1, x_2\}$	$\alpha(a^*) + \alpha(b^*) + \alpha(c)[\beta(a^*) + \beta(b^*) - \beta(a^*)\beta(b^*)]$ $+\alpha(z^*)[\beta(a^*) + \beta(b^*) - \beta(a^*)\beta(b^*)]\beta(c)$	$L(a^*) + L(b^*) + L(c) + L(z^*)$
$\{x_1, x_3\}$	$\alpha(a^*) + \alpha(d^*) + \alpha(c)\beta(a^*)$ $+\alpha(z^*)[\beta(d^*) + \beta(a^*)\beta(c) - \beta(a^*)\beta(c)\beta(d^*)]$	$L(a^*) + L(c) + L(d^*) + L(z^*)$
$\{x_2, x_3\}$	$\alpha(b^*) + \alpha(d^*) + \alpha(c)\beta(b^*)$ $+\alpha(z^*)[\beta(d^*) + \beta(b^*)\beta(c) - \beta(b^*)\beta(c)\beta(d^*)]$	$L(b^*) + L(c) + L(d^*) + L(z^*)$
$\{x_1, x_2, x_3\}$	$\alpha(a^*) + \alpha(b^*) + \alpha(d^*) + \alpha(c)[\beta(a^*) + \beta(b^*)$ $-\beta(a^*)\beta(b^*)] + \alpha(z^*)[\beta(d^*) + \beta(a^*)\beta(c)$ $+\beta(b^*)\beta(c) - \beta(a^*)\beta(b^*)\beta(c) - \beta(a^*)\beta(c)\beta(d^*)$ $-\beta(b^*)\beta(c)\beta(d^*) + \beta(a^*)\beta(b^*)\beta(c)\beta(d^*)]$	$L(a^*) + L(b^*) + L(c) + L(d^*) + L(z^*)$

Table 2: ϕ, PD values for Figure 4(b)

Using Table 2:

$$\begin{aligned} L(a^*) &= PD(\{x_1, x_2\}) - PD(\{x_2\}) \\ &= \phi(\{x_1, x_2\}) - \phi(\{x_2\}) \\ &= \alpha(a^*) + \alpha(c)[\beta(a^*) - \beta(a^*)\beta(b^*)] + \alpha(z^*)[\beta(a^*)\beta(c) - \beta(a^*)\beta(b^*)\beta(c)] \end{aligned}$$

and
$$\begin{aligned} L(a^*) &= PD(\{x_1, x_2, x_3\}) - PD(\{x_2, x_3\}) \\ &= \phi(\{x_1, x_2, x_3\}) - \phi(\{x_2, x_3\}) \\ &= \alpha(a^*) + \alpha(c)[\beta(a^*) - \beta(a^*)\beta(b^*)] + \alpha(z^*)[\beta(a^*)\beta(c) - \beta(a^*)\beta(b^*)\beta(c) \\ &\quad - \beta(a^*)\beta(c)\beta(d^*) + \beta(a^*)\beta(b^*)\beta(c)\beta(d^*)] \end{aligned}$$

Hence $\alpha(z^*)\beta(a^*)\beta(c)\beta(d^*) = \alpha(z^*)\beta(a^*)\beta(b^*)\beta(c)\beta(d^*)$. *As noted above, every factor on the left hand side of this equation is nonzero, therefore* $\beta(b^*) = 1$. *This in turn means we must have* $\beta(b) = 1$.

A symmetric argument shows that $\beta(a) = 1$, *and (repeating for other edges descending from* v *if necessary) that* v *has the* $(\star)$ *property. By iterating this procedure, we find that every vertex, other than possibly* v' *and the root, must satisfy the* $(\star)$ *property.*

Case Two: *Suppose the parent of* v *is the root vertex* ($v = v'$), *and* $\mathcal{T}$ *is a type-A clade.*

We select x_1, x_2, x_3, *three leaves of* $\mathcal{T}$, *such that the path from* x_1 *(resp.* x_2, x_3*) to* v *contains edge* a *(resp.* b, c*) incident to* v, *and* a, b, c *are all distinct. Figure 5(a) illustrates these choices. As in Case One, we use Lemma 1.1 to replace the path from* v *to* x_1 *(resp.* x_2, x_3*) with the edge* a^* *(resp.* b^*, c^**). The result of these changes is shown in Figure 5(b). Table 3 shows* ϕ *and PD calculations for subsets of* $\{x_1, x_2, x_3\}$, *and a calculation of* $\phi(\{u\})$.

Observe that

$$\phi(\{x_1\}) + \phi(\{x_3\}) - \phi(\{x_1, x_3\}) = \beta(a^*)\beta(c^*)\alpha(z). \tag{41}$$

Now, under the assumption that ϕ *corresponds to PD for these sets, and noting also that* $PD(\{x_1\}) + PD(\{x_3\}) - PD(\{x_1, x_3\}) = L(z) > 0$, *it follows that the left hand side of (41), and hence the right hand side also, must be strictly positive. This implies each factor is nonzero:* $\beta(a^*), \beta(c^*), \phi(\{v\}) > 0$. *Again we express* $L(a^*)$ *in terms of two* ϕ *expressions.*

$$\begin{aligned} L(a^*) &= PD(\{x_1, x_2\}) - PD(\{x_2\}) \\ &= \phi(\{x_1, x_2\}) - \phi(\{x_2\}) \\ &= \alpha(a^*) + \alpha(z)[\beta(a^*) - \beta(a^*)\beta(b^*)] \end{aligned}$$

and
$$\begin{aligned} L(a^*) &= PD(\{x_1, x_2, x_3\}) - PD(\{x_2, x_3\}) \\ &= \phi(\{x_1, x_2, x_3\}) - \phi(\{x_2, x_3\}) \\ &= \alpha(a^*) + \alpha(z)[\beta(a^*) - \beta(a^*)\beta(b^*) - \beta(a^*)\beta(c^*) + \beta(a^*)\beta(b^*)\beta(c^*)] \end{aligned}$$

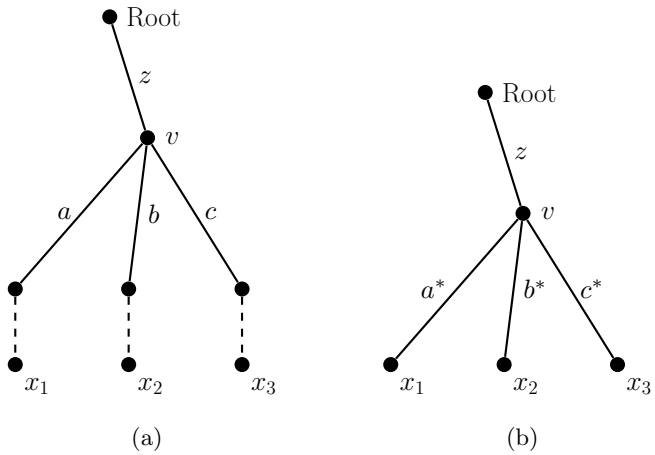

Figure 5: Diagram to show key edges and vertices referred to in the main narrative

A	$\phi(A)$	$PD(A)$
$\{v\}$	$\alpha(z)$	
$\{x_1\}$	$\alpha(a^*) + \alpha(z)\beta(a^*)$	$L(a^*) + L(z)$
$\{x_2\}$	$\alpha(b^*) + \alpha(z)\beta(b^*)$	$L(b^*) + L(z)$
$\{x_3\}$	$\alpha(c^*) + \alpha(z)\beta(c^*)$	$L(c^*) + L(z)$
$\{x_1, x_2\}$	$\alpha(a^*) + \alpha(b^*) + \alpha(z)[\beta(a^*) + \beta(b^*) - \beta(a^*)\beta(b^*)]$	$L(a^*) + L(b^*) + L(z)$
$\{x_1, x_3\}$	$\alpha(a^*) + \alpha(c^*) + \alpha(z)[\beta(a^*) + \beta(c^*) - \beta(a^*)\beta(c^*)]$	$L(a^*) + L(c^*) + L(z)$
$\{x_2, x_3\}$	$\alpha(b^*) + \alpha(c^*) + \alpha(z)[\beta(b^*) + \beta(c^*) - \beta(b^*)\beta(c^*)]$	$L(b^*) + L(c^*) + L(z)$
$\{x_1, x_2, x_3\}$	$\alpha(a^*) + \alpha(b^*) + \alpha(c^*) + \alpha(z)[\beta(a^*) + \beta(b^*) + \beta(c^*)$ $-\beta(a^*)\beta(b^*) - \beta(a^*)\beta(c^*) - \beta(b^*)\beta(c^*)$ $+\beta(a^*)\beta(b^*)\beta(c^*)]$	$L(a^*) + L(b^*) + L(c^*) + L(z)$

Table 3: ϕ, PD values for Figure 5(b)

Hence $\alpha(z)\beta(a^*)\beta(c^*) = \alpha(z)\beta(a^*)\beta(b^*)\beta(c^*)$. *Since every factor on the left hand side of the equation is nonzero, we get* $\beta(b^*) = 1$. *Symmetric arguments show that no EvoHeritage attrition occurs on any edge descending from* v. *Therefore every edge* $e \in E(\mathcal{T})\backslash\{z\}$ *has* $\beta(e) = 1$ *when* $\mathcal{T}$ *is a type-A clade.*

Accordingly, $L(a^*) = \alpha(a^*) + \alpha(z)[\beta(a^*) - \beta(a^*)\beta(b^*)] = \alpha(a^*)$. *Hence* $\alpha(a) = L(a)$ *(by Lemma 1.1). Again, by symmetry* $\alpha(e) = L(e)$ *for every edge* $e \in E(\mathcal{T})\backslash\{z\}$. *Finally, taking* $A = \{x_1\}$ *in Table 3 we obtain* $\alpha(z) = L(z)$.

Case Three: *Suppose the parent of* v *is the root vertex* ($v = v'$), *and* $\mathcal{T}$ *is a type-B clade.*

We proceed as in Case Two, using Figure 5 and Table 3 as if $L(c^*) = \alpha(c^*) = \beta(c^*) = 0$. *By Case One and Lemma 1.1,* $\beta(a^*) = \beta(a)$, *and* $\beta(b^*) = \beta(b)$. *Then*

$$\begin{aligned}
\phi(\{x_1\}) + \phi(\{x_2\}) - \phi(\{x_1, x_2\}) &= PD(\{x_1\}) + PD(\{x_2\}) - PD(\{x_1, x_2\}) \\
\alpha(z)\beta(a^*)\beta(b^*) &= L(z) \\
\alpha(z) &= \frac{L(z)}{\beta(a)\beta(b)}
\end{aligned}$$

Furthermore,

$$\begin{aligned}
\phi(\{x_1, x_2\}) - \phi(\{x_2\}) &= PD(\{x_1, x_2\}) - PD(\{x_2\}) \\
\alpha(a^*) + \frac{L(z)}{\beta(a)\beta(b)}(\beta(a) - \beta(a)\beta(b)) &= L(a^*) \\
\alpha(a^*) &= L(a^*) + L(z)\left(1 - \frac{1}{\beta(b)}\right) \\
\alpha(a) &= L(a) + L(z)\left(1 - \frac{1}{\beta(b)}\right) \qquad \text{[by Lemma 1.1]}
\end{aligned}$$

A symmetric argument gives the required value of $\alpha(b)$.

Conversely, it can be shown by direct calculation that if the conditions listed in the theorem hold, then $\phi(A) = PD(A)$ *for every set* $A \subseteq X$. *For type-A and type-C clades this is immediate, since the* ϕ *formula derived in the main body of the paper uses the value of neither* $\beta(z_A)$ *nor* $\beta(z_C)$. *There being no EvoHeritage preceding the root, there is no need to account for the rate at which it could be lost.*

We introduce some shorthand for the type-B case. Suppose A *is a set of leaves within a type-B clade. Let* Σ_a *denote that portion of* $PD(A)$ *which is derived from edges descended from* a. *We write* Σ_b *similarly for edges descended from* b.

Firstly, if A is entirely descended from a:

$$\begin{aligned}\phi(A) &= \Sigma_a + \alpha(a) + \alpha(z_B)\beta(a) \\ &= \Sigma_a + L(a) + L(z_B) - \frac{L(z_B)}{\beta(b)} + \frac{L(z_B)}{\beta(a)\beta(b)}\beta(a) \\ &= \Sigma_a + L(a) + L(z_B) \\ &= PD(A)\end{aligned}$$

A symmetric argument holds when A is entirely descended from b. Finally, if A is descended from both a and b:

$$\begin{aligned}\phi(A) &= \Sigma_a + \Sigma_b + \alpha(a) + \alpha(b) + \alpha(z_B)(\beta(a) + \beta(b) - \beta(a)\beta(b)) \\ &= \Sigma_a + \Sigma_b + L(a) + L(z_B) - \frac{L(z_B)}{\beta(b)} + L(b) + L(z_B) - \frac{L(z_B)}{\beta(a)} \\ &\quad + \frac{L(z_B)}{\beta(a)\beta(b)}(\beta(a) + \beta(b) - \beta(a)\beta(b)) \\ &= \Sigma_a + \Sigma_b + L(a) + L(b) + L(z_B) \\ &= PD(A)\end{aligned}$$

1.3.4 Proof: maximising PD does not always maximise ϕ_ρ

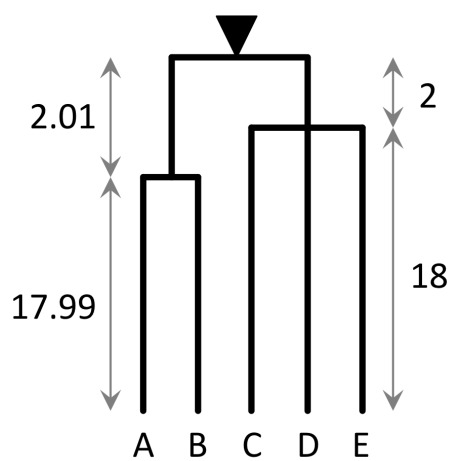

Figure 6: Example rooted tree with edge lengths, and leaves marked A-E. The downward pointing triangle illustrates the root.

The tree in Figure 6 provides an example where a set that maximises PD does not maximise ϕ_ρ, for $\rho = 10^{-3}$. Consider choosing a set of four out of five species in Figure 6. If we wish to maximise PD, the one species we would exclude to leave a set of four would clearly be either A or B. This gives a total PD of

$$\mathrm{PD}(\{A, C, D, E\}) = 17.99 + 2.01 + 2 + 3 \cdot 18 = 76 \tag{42}$$

By comparison, if we were to exclude either C, D or E, we would be left with a PD of

$$\mathrm{PD}(\{A, B, C, D\}) = 2 \cdot 17.99 + 2.01 + 2 + 2 \cdot 18 = 75.99 \tag{43}$$

Not let us consider the ϕ under standard conditions of the same two scenarios. In this case we do not add the origin of life or normalise into standardised units as neither of these steps would change the comparison between an $\{A, B, C, D\}$ scenario or an $\{A, C, D, E\}$ scenario.

$$\begin{aligned}\phi(\{A, B, C, D\}) = 2 \cdot \frac{1 - \exp(-\rho \cdot 17.99)}{\rho} + 2 \cdot \frac{1 - \exp(-\rho \cdot 18)}{\rho} + \\ \frac{1 - e^{-2.01\rho}}{\rho} \cdot \big(1 - (1 - \exp(-\rho \cdot 17.99))^2\big) + \\ \frac{1 - e^{-2\rho}}{\rho} \cdot \big(1 - (1 - \exp(-\rho \cdot 18))^2\big)\end{aligned} \tag{44}$$

In contrast

$$\begin{aligned}\phi(\{A, C, D, E\}) = \frac{1 - \exp(-\rho \cdot 17.99)}{\rho} + 3 \cdot \frac{1 - \exp(-\rho \cdot 18)}{\rho} + \\ \frac{1 - e^{-2.01\rho}}{\rho} \cdot \big(1 - (1 - \exp(-\rho \cdot 17.99))\big) + \\ \frac{1 - e^{-2\rho}}{\rho} \cdot \big(1 - (1 - \exp(-\rho \cdot 18))^3\big).\end{aligned} \tag{45}$$

Now let us take $\rho = 10^{-3}$ to obtain:

$$\phi(\{A, B, C, D\}) = 75.34094. \tag{46}$$

$$\phi(\{A, C, D, E\}) = 75.31622. \tag{47}$$

and so $\phi(\{A, B, C, D\}) > \phi(\{A, C, D, E\})$. However, we saw that $\mathrm{PD}(\{A, B, C, D\}) < \mathrm{PD}(\{A, C, D, E\})$ so the set of nodes that maximises PD is not the same as that which maximises ϕ. Of course, this is a contrived example, and even so the set of vertices that maximise PD may often also maximise ϕ_ρ depending on the value of ρ. What matters in practice, however, are the relative diversity values of a wide range of different sets of vertices (or species), and in terms of this we expect ϕ_ρ to behave very differently to PD. If we are performing an analysis based on more than just the tree, for example incorporating extinction risk as an additional distinguishing factor in a conservation prioritisation analysis, the the effect of changing to ϕ would be expected to be more substantial.

1.4 EvoHeritage under stochastic conditions

Our stochastic model considers one form of EvoHeritage measured in discrete units. These discrete EvoHeritage units may be thought of as features (binary characters) which arise and are lost along the edges of a phylogenetic tree. The features exhibited by a species are precisely those that persist until the terminal vertex representing it.

To separate stochastic conditions from the deterministic ones, we use new variables. Feature gains and losses will be described by exponential distributions, with respective rates $\lambda(e)$ and $\mu(e)$ along edge e. Here we consider the case where these rates have the same values for every edge in a tree, namely λ and μ.

We assume that the gain and loss of a feature is independent of the presence or absence of other features. We also require that a feature be gained before it can be lost, and that each feature is distinct from all others.

Thus the development of a new feature is not conditional on the number of features already arisen. We phrase this as a 'constant birth' process. But the attrition of features does depend on how many features exist, so we have a 'linear death' process. We model these processes by a continuous time Markov chain, using constant birth and linear death rates, as shown in Figure 7.

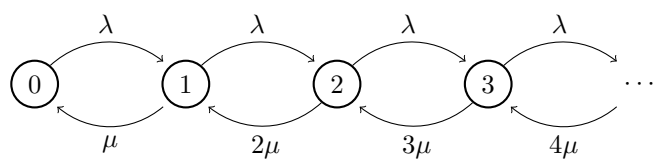

Figure 7

We note that this model is described in Queueing Theory as an infinite server queue, or $M/M/\infty$-queue. This connection provides a limiting distribution for the model, given by $p_j = e^{-\frac{\lambda}{\mu}} \frac{(\frac{\lambda}{\mu})^j}{j!}$, for $j \geq 0$.

Let $T = (V, E)$ be a rooted phylogenetic tree, where each edge $e \in E(T)$ has a corresponding positive length $L(e)$. For a vertex $v \in V$, we denote the set of features present at v by $\mathbb{F}_v$. This set consists of both new features, those which have been gained on the incident ancestral edge, and persistent features which accumulated along earlier edges and survived the attrition process.

Let $N_e(t)$ be the net number of new features present on edge e at distance t from the start of e. Then $N_e(0) = 0$ and $N_e(t)$ is a Poisson random variable with mean $m_t = \int_0^t \lambda e^{-\mu s} ds = \frac{\lambda}{\mu} e^{-\mu s}\Big|_t^0 = \frac{\lambda}{\mu}(1 - e^{-\mu t})$, i.e. $N_e(t) \sim P_0(m_t)$. Hence $\mathbb{E}[N_e(t)] = m_t$.

We can now derive an expression for the expected number of features which arise along edge e and are still present at the end of e.

$$\mathbb{E}[N_e(L(e))] = m_{L(e)} = \frac{\lambda}{\mu}(1 - e^{-\mu L(e)})$$

Next, let $v, v' \in V(T)$, where v' is a child of v and $(v, v') = e \in E(T)$. We write F_v for the random variable that counts the number of features present at vertex v. Then we define $P_{v'}$ to be the number of these F_v features which persist along the entirety of edge e to be present at v'. Conditional on F_v, $P_{v'}$ is described by a binomial distribution $P_{v'} \sim \text{Bin}(F_v, e^{-\mu L(e)})$, hence $\mathbb{E}[P_{v'}|F_v] = F_v e^{-\mu L(e)}$.

Finally, we can write $\mathbb{E}[F_{v'}|F_v] = F_v e^{-\mu L(e)} + \frac{\lambda}{\mu}(1 - e^{-\mu L(e)})$. When $\mu = 0$, we define $\mathbb{E}[F_{v'}|F_v]$ as $\lim_{\mu \to 0} \mathbb{E}[F_{v'}|F_v] = F_v + \lambda L(e)$.

Lemma 1.6 *Let T be a rooted phylogenetic tree, and let P be a path in T, beginning at v and ending at v'. Then $\mathbb{E}[F_{v'}|F_v]$ may be calculated as if v and v' were connected by a single edge whose length is the length of P as a whole.*

Proof 1.7 *Let $u, v, w \in V(T)$ and $e, f \in E(T)$. Suppose that $e = (u, v), f = (v, w)$. Then*

$$\begin{aligned}
\mathbb{E}[F_w|F_u] &= \mathbb{E}[F_v|F_u]e^{-\mu L(f)} + \frac{\lambda}{\mu}(1 - e^{-\mu L(f)}) \\
&= [F_u e^{-\mu L(e)} + \frac{\lambda}{\mu}(1 - e^{-\mu L(e)})]e^{-\mu L(f)} + \frac{\lambda}{\mu}(1 - e^{-\mu L(f)}) \\
&= F_u e^{-\mu(L(e)+L(f))} + \frac{\lambda}{\mu}(1 - e^{-\mu(L(e)+L(f))})
\end{aligned}$$

So the expected number of features is the same if we take the entire path as if it were composed of a single edge of length $L(e) + L(f)$. This observation can be applied iteratively to extend to paths of any finite length.

Definition: Let T be a rooted phylogenetic tree with leaf set X. The *feature diversity* (FD) of a set $A \subseteq X$ is the size of the set of features present among the taxa in A. Let $\mathbb{F}_A = \bigcup_{x \in A} \mathbb{F}_x$ and let $FD(A) = |\mathbb{F}_A|$.

Similar to the deterministic setting, we find that FD does not always match Phylogenetic Diversity (PD). Proposition 1.8 gives conditions for when the two measures do align. If T does contain monotomy vertices, we suppress these first using Lemma 1.6.

Theorem 1.8 *Let $T = (V, E)$ be a rooted phylogenetic tree containing no monotomy vertices. Let X be the leaf set of T, and r denote the root vertex. Let $L : E(T) \to \mathbb{R}$ and $L' : E(T) \to \mathbb{R}$ be two edge length assignments for T. Suppose at least one clade of T contains more than 3 edges. If $\mathbb{E}[FD_{T,L}(A)] = PD_{T,L'}(A)$ for every set of leaves $A \subseteq X$ then $\mu = 0$.*

Conversely, if $\mu = 0$ and $L(e) = cL'(e)$ for every edge $e \in E(T)$ then $\mathbb{E}[FD_{T,L}(A)] = c\lambda PD_{T,L'}(A)$ for every set of leaves $A \subseteq X$.

Proof 1.9 *Let $\mathcal{T} = (V(\mathcal{T}), E(\mathcal{T}))$ be a clade of T containing the edge $z = (r, v')$. We will determine that $\mu = 0$ or v' has less than 3 descendant edges. Since terms derived from distinct clades do not interact in either PD or FD*

calculations, this will prove the result for T as a whole. For simplicity's sake we consider the most natural scenario where the root vertex begins with no features. However this assumption may be relaxed without impacting the results shown below.

Let v be an interior vertex of $\mathcal{T}$, and let u be the parent vertex of v. We consider two cases: either u is the root vertex or it is not.

Case One: *Suppose u is not the root vertex.*

Since u is not the root, it has a child vertex w distinct from v. Select two leaves x_1, x_2 from the descendants of v such that the path from x_1 (resp. x_2) to v contains edge a (resp. b) incident to v, and $a \neq b$. Further, select terminal vertex x_3 descended from w. Figure 4(a) illustrates these choices.

In calculating FD of any subset of $\{x_1, x_2, x_3\}$, only those edges and paths shown can possibly make a non-zero contribution to the calculation. Hence we can use Lemma 1.6 to replace the path from v to x_1 (resp. x_2, x_3, r) with the edge a^* (resp b^*, d^*, z^*). The result of these changes is shown in Figure 4(b).

Writing $\mathbb{F}_i$ as shorthand in place of $\mathbb{F}_{x_i}$ for $i \in \{1, 2, 3\}$, Table 4 shows calculations of expected FD for selected sets. The expected FD for $|\mathbb{F}_1 \cap \mathbb{F}_3|$ is arrived at by noticing that any feature in this intersection must arise along z^*. Then it must survive attrition along edges c and a^* to be included in $\mathbb{F}_1$ and also along d^* to be included in $\mathbb{F}_3$. The expression for $|\mathbb{F}_1 \cap \mathbb{F}_2 \cap \mathbb{F}_3|$ is obtained similarly. If FD corresponds to PD then the last two lines of Table 6 must be equal, because

$$PD(\{x_1, x_2\}) - PD(\{x_2\}) = L'(a^*) = PD(\{x_1, x_2, x_3\}) - PD(\{x_2, x_3\}).$$

Hence,

$$|\mathbb{F}_1 \cap \mathbb{F}_3| = |\mathbb{F}_1 \cap \mathbb{F}_2 \cap \mathbb{F}_3|$$
$$e^{-\mu(L(a^*)+L(c)+L(d^*))}(1 - e^{-\mu L(z^*)}) = e^{-\mu(L(a^*)+L(b^*)+L(c)+L(d^*))}(1 - e^{-\mu L(z^*)}).$$

Suppose $\mu \neq 0$. Then $1 = e^{-\mu L(b^*)}$, giving $L(b^*) = 0$ which contradicts our earlier stipulation that all edge lengths be strictly positive. Therefore $\mu = 0$ in this case.

Case Two: *Suppose u is the root vertex.*

Assume v has at least three descendant edges. We select x_1, x_2, x_3, three leaves of $\mathcal{T}$, such that the path from x_1 (resp. x_2, x_3) to v contains edge a (resp. b, c) incident to v, and a, b, c are all distinct. Figure 5(a) illustrates these choices, and all edges relevant to the calculation of FD and PD appear in this figure. We use Lemma 1.6 to replace the path from v to x_1 (resp. x_2, x_3) with the edge a^* (resp. b^*, c^*). The result of these changes is shown in Figure 5(b). As in Case One, we have

$$|\mathbb{F}_1 \cap \mathbb{F}_3| = |\mathbb{F}_1 \cap \mathbb{F}_2 \cap \mathbb{F}_3|$$
$$e^{-\mu(L(a^*)+L(c^*))}(1 - e^{-\mu L(z)}) = e^{-\mu(L(a^*)+L(b^*)+L(c^*))}(1 - e^{-\mu L(z)}).$$

A	Expected FD of A
$\lvert\mathbb{F}_1\rvert$	$\frac{\lambda}{\mu}(1 - e^{-\mu(L(a^*)+L(c)+L(z^*))})$
$\lvert\mathbb{F}_2\rvert$	$\frac{\lambda}{\mu}(1 - e^{-\mu(L(b^*)+L(c)+L(z^*))})$
$\lvert\mathbb{F}_3\rvert$	$\frac{\lambda}{\mu}(1 - e^{-\mu(L(d^*)+L(z^*))})$
$\lvert\mathbb{F}_1 \cap \mathbb{F}_3\rvert$	$\frac{\lambda}{\mu}e^{-\mu(L(a^*)+L(c)+L(d^*))}(1 - e^{-\mu L(z^*)})$
$\lvert\mathbb{F}_1 \cap \mathbb{F}_2 \cap \mathbb{F}_3\rvert$	$\frac{\lambda}{\mu}e^{-\mu(L(a^*)+L(b^*)+L(c)+L(d^*))}(1 - e^{-\mu L(z^*)})$

Table 4: Expected FD for sets in Figure 4(b)

A	Expected FD of A
$\lvert\mathbb{F}_1\rvert$	$\frac{\lambda}{\mu}(1 - e^{-\mu(L(a^*)+L(z))})$
$\lvert\mathbb{F}_2\rvert$	$\frac{\lambda}{\mu}(1 - e^{-\mu(L(b^*)+L(z))})$
$\lvert\mathbb{F}_3\rvert$	$\frac{\lambda}{\mu}(1 - e^{-\mu(L(c^*)+L(z))})$
$\lvert\mathbb{F}_1 \cap \mathbb{F}_3\rvert$	$\frac{\lambda}{\mu}e^{-\mu(L(a^*)+L(c^*))}(1 - e^{-\mu L(z)})$
$\lvert\mathbb{F}_1 \cap \mathbb{F}_2 \cap \mathbb{F}_3\rvert$	$\frac{\lambda}{\mu}e^{-\mu(L(a^*)+L(b^*)+L(c^*))}(1 - e^{-\mu L(z)})$

Table 5: Expected FD for sets in Figure 5(b)

$\lvert\mathbb{F}_1 \cup \mathbb{F}_2\rvert$	$\lvert\mathbb{F}_1\rvert + \lvert\mathbb{F}_2\rvert - \lvert\mathbb{F}_1 \cap \mathbb{F}_2\rvert$
$\lvert\mathbb{F}_2 \cup \mathbb{F}_3\rvert$	$\lvert\mathbb{F}_2\rvert + \lvert\mathbb{F}_3\rvert - \lvert\mathbb{F}_2 \cap \mathbb{F}_3\rvert$
$\lvert\mathbb{F}_1 \cup \mathbb{F}_2 \cup \mathbb{F}_3\rvert$	$\lvert\mathbb{F}_1\rvert + \lvert\mathbb{F}_2\rvert + \lvert\mathbb{F}_3\rvert - \lvert\mathbb{F}_1 \cap \mathbb{F}_2\rvert - \lvert\mathbb{F}_1 \cap \mathbb{F}_3\rvert - \lvert\mathbb{F}_2 \cap \mathbb{F}_3\rvert + \lvert\mathbb{F}_1 \cap \mathbb{F}_2 \cap \mathbb{F}_3\rvert$
$\lvert\mathbb{F}_1 \cup \mathbb{F}_2\rvert - \lvert\mathbb{F}_2\rvert$	$\lvert\mathbb{F}_1\rvert - \lvert\mathbb{F}_1 \cap \mathbb{F}_2\rvert$
$\lvert\mathbb{F}_1 \cup \mathbb{F}_2 \cup \mathbb{F}_3\rvert - \lvert\mathbb{F}_2 \cup \mathbb{F}_3\rvert$	$\lvert\mathbb{F}_1\rvert - \lvert\mathbb{F}_1 \cap \mathbb{F}_2\rvert - \lvert\mathbb{F}_1 \cap \mathbb{F}_3\rvert + \lvert\mathbb{F}_1 \cap \mathbb{F}_2 \cap \mathbb{F}_3\rvert$

Table 6: Equal-sized feature sets

Again, the supposition that $\mu \neq 0$ leads to the contradiction that $L(b^*) = 0$. So $\mu = 0$ when v has at least three descendant edges.

The remaining possibility for T is that every clade consists of a single edge, or an edge incident with the root and exactly two descending edges.

Now we prove the second statement of the theorem. Suppose that $\mu = 0$, and for some constant c, that $L(e) = cL'(e)$ for every $e \in E(T)$. Let $A = \{x_1, \ldots, x_k\}$ be a set of leaves from T. Let p_i denote the path from the root vertex r to terminal vertex x_i, and P_A be the set of edges common to all p_i, for $1 \leq i \leq k$. Finally, let $Q_A = \bigcup p_i - P_A$, i.e. the set of edges contained in at least one of the paths p_i, but not all of them. Then

$$\mathbb{E}\left[|\bigcap_{x_i \in A} \mathbb{F}_{x_i}|\right] = \lim_{\mu \to 0} \frac{\lambda}{\mu} e^{-\mu \sum_{e \in Q_A} L(e)} \left(1 - e^{-\mu \sum_{e \in P_A} L(e)}\right) \tag{48}$$

$$= \lambda \sum_{e \in P_A} L(e) \tag{49}$$

For tree T and edge length assignment L,

$$\begin{aligned}
\mathbb{E}[FD(A)] &= \mathbb{E}[|\mathbb{F}_A|] \\
&= \mathbb{E}[|\mathbb{F}_{x_1} \cup \cdots \cup \mathbb{F}_{x_k}|] \\
&= \mathbb{E}[|\mathbb{F}_{x_1}|] + \cdots + \mathbb{E}[|\mathbb{F}_{x_k}|] \\
&\quad - \mathbb{E}[|\mathbb{F}_{x_1} \cap \mathbb{F}_{x_2}|] - \mathbb{E}[|\mathbb{F}_{x_1} \cap \mathbb{F}_{x_3}|] - \cdots - \mathbb{E}[|\mathbb{F}_{x_{k-1}} \cap \mathbb{F}_{x_k}|] \\
&\quad + \ldots \\
&\quad \pm \mathbb{E}[|\mathbb{F}_{x_1} \cap \cdots \cap \mathbb{F}_{x_k}|]
\end{aligned}$$

Now let e be an edge in $\bigcup p_i$, from which m members of A descend. When the final expression above is evaluated using Eqn. 49, then $L(e)$ is added a net $\binom{m}{1} - \binom{m}{2} + \binom{m}{3} - \cdots \pm \binom{m}{m}$ times. For all $m \geq 1$, we have $\sum_{j=0}^{m} (-1)^j \binom{m}{j} = 0$, and $\binom{m}{0} = 1$. Therefore $L(e)$ is added exactly once more than it is subtracted.

Hence $\mathbb{E}[FD_{T,L}(A)] = \lambda \sum_{e \in \cup p_i} L(e) = \lambda PD_{T,L}(A) = c\lambda PD_{T,L'}(A)$.

1.5 Sets that maximise PD do not necessarily maximise FD in our stochastic model

We give an example of a phylogenetic tree for which the sets that maximise phylogenetic diversity do not always correspond to those that maximise feature diversity under our stochastic model. The phylogenetic tree in Fig. 8 is adapted from the earlier figure that was used to illustrate the difference between maximal PD and maximal ϕ sets.

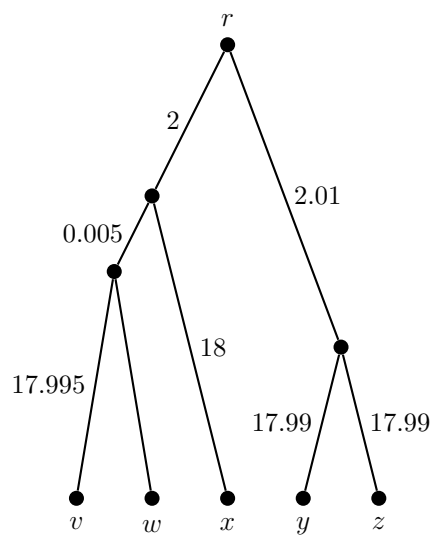

Figure 8: Binary phylogenetic tree, with edge lengths marked. A set of four leaves that maximises the PD score, $\{v, w, x, y\}$, does not maximise feature diversity under the stochastic model when $\mu = 10^{-3}$.

Firstly we note that $\{v, w, x, y\}$ is a set that maximises PD among sets of four leaves. For the stochastic comparison, we have no features present at the root vertex r. We calculate the feature diversity of the sets $\{v, w, x, y\}$ and $\{w, x, y, z\}$.

$$
\begin{aligned}
&FD(\{v, w, x, y\}) \\
=&FD(v) + FD(w) + FD(x) + FD(y) \\
&\quad - FD(v \cap x) - FD(w \cap x) - FD(v \cap w) + FD(v \cap x \cap x) \\
=&4 \cdot \frac{\lambda}{\mu}(1 - e^{-20\mu}) - 2 \cdot \frac{\lambda}{\mu}e^{-36\mu}(1 - e^{-2\mu}) - \frac{\lambda}{\mu}e^{-35.99\mu}(1 - e^{-2.005\mu}) + \frac{\lambda}{\mu}e^{-53.995\mu}(1 - e^{-2\mu})
\end{aligned}
$$

$$
\begin{aligned}
&FD(\{w, x, y, z\}) \\
=&FD(w) + FD(x) + FD(y) + FD(z) \\
&\quad - FD(x \cap w) - FD(y \cap z) \\
=&4 \cdot \frac{\lambda}{\mu}(1 - e^{-20\mu}) - \frac{\lambda}{\mu}e^{-35.98\mu}(1 - e^{-2.01\mu}) - \frac{\lambda}{\mu}e^{-36\mu}(1 - e^{-2\mu})
\end{aligned}
$$

Take $\mu = \frac{1}{1000}$. Then, considering the difference between these two FD scores we get

$$
\begin{aligned}
&FD(\{v,w,x,y\}) - FD(\{w,x,y,z\}) \\
=&1000\lambda\left(-e^{-0.03598}(1-e^{-0.00201}) + e^{-0.036}(1-e^{-0.002}) + e^{-0.03599}(1-e^{-0.002005}) - e^{-0.053995}(1-e^{-0.002})\right) \\
\approx&1000\lambda\left(-0.964660\cdot 0.002008 + 0.964640\cdot 0.001998 + 0.964650\cdot 0.002003 - 0.947437\cdot 0.001998\right) \\
\approx&1000\lambda\left(-0.001937 + 0.001927 + 0.001932 - 0.001893\right) \\
=&0.029\lambda
\end{aligned}
$$

Thus, since λ is always positive, the set $\{w,x,y,z\}$ does not maximise FD among sets of size four for this value of μ.

2 Developing a complete EvoHeritage calculus

2.1 Relationship between generalised phylogenetic difference and classic phylogenetic distance for a pair of vertices

Suppose we measure phylogenetic distance as the date D of common ancestry between two species a and b on a phylogenetic tree. We seek to calculate EvoHeritage difference, diff-ϕ, between these species under standard conditions with attrition at rate ρ. First, we express the phylogenetic tree as an EvoHeritage graph under standard conditions. Next we collapse all vertices except those corresponding to the two species, their common ancestor, and the root vertex at age $x \geq D$. By setting $A = \{a\}$ and $B = \{b\}$ we are left with the following calculation for diff-ϕ.

$$
\text{diff-}\phi(\{a\},\{b\}) = \frac{2\cdot\phi(\{a,b\}) - \phi(\{a\}) - \phi(\{b\})}{\phi(\{a,b\})} \tag{50}
$$

Note that we could measure ϕ in standardised units but as we are taking a ratio of different ϕ measures any effect of units would immediately cancel out. We determine expressions for the terms appearing on the right hand side of the above equation.

$$
\phi(\{a\}) = \phi(\{b\}) = \frac{1-\exp(-\rho\cdot x)}{\rho} \tag{51}
$$

$$
\begin{aligned}
\phi(\{a,b\}) &= 2\cdot\frac{1-\exp(-\rho\cdot D)}{\rho} + \frac{1-\exp(-\rho\cdot(x-D))}{\rho}\cdot(1-(1-\exp(-\rho\cdot D))^2) \\
&= \frac{2\cdot(1-\exp(-\rho\cdot D)) + (1-\exp(-\rho\cdot(x-D)))\cdot(2\cdot\exp(-\rho\cdot D) - \exp(-2\rho\cdot D))}{\rho} \\
&= \frac{2-\exp(-2\rho\cdot D) - 2\cdot\exp(-\rho\cdot x) + \exp(-\rho(x+D))}{\rho}
\end{aligned} \tag{52}
$$

Substituting these expressions into (50):

$$\text{diff-}\phi = \frac{4 - 2\cdot\exp(-2\rho\cdot D) - 4\cdot\exp(-\rho\cdot x) + 2\cdot\exp(-\rho(x+D)) - 2 + 2\cdot\exp(-\rho\cdot x)}{2 - \exp(-2\rho\cdot D) - 2\exp(-\rho\cdot x) + \exp(-\rho(x+D))} \tag{53}$$

$$= \frac{2 - 2\cdot\exp(-2\rho\cdot D) - 2\cdot\exp(-\rho\cdot x) + 2\cdot\exp(-\rho(x+D))}{2 - \exp(-2\rho\cdot D) - 2\cdot\exp(-\rho\cdot x) + \exp(-\rho(x+D))} \tag{54}$$

Let us consider what happens in the limit as $\rho \to \infty$. All of the exponential terms tend to 0 and therefore

$$\lim_{\rho\to\infty} \text{diff-}\phi = \frac{2}{2} = 1. \tag{55}$$

Recall that diff-ϕ is the proportion of total EvoHeritage across two sets which is unique to one set or the other. The above limit conforms with the intuition that with infinite EvoHeritage attrition, these two species will be maximally distant from one another.

Now, let us consider what happens when $\rho \to 0$. In this case substituting $\rho = 0$ into equation 54 gives an indeterminate expression ($\frac{0}{0}$). We use L'Hôpital's Rule to calculate this limit.

$$\lim_{\rho\to 0} \text{diff-}\phi = \lim_{\rho\to 0} \frac{4D\exp(-2\rho D) + 2x\exp(-\rho x) - 2(x+D)\exp(-\rho(x+D)}{2D\exp(-2\rho D) + 2x\exp(-\rho x) - (x+D)\exp(-\rho(x+D)} \tag{56}$$

$$= \frac{2D}{x+D} \tag{57}$$

2.2 Relationship to genetic diversity

We focus on the measure introduced by (Crozier, 1992) which measures the probability of there being two or more character states across all terminal vertices. The calculation is on an unrooted tree with all edge lengths bounded between 0 and 1 corresponding to the probability of a difference in character state between the vertices at either end of the edge. The measure is given by equation 3 in (Crozier, 1992)

$$P(\geq 2) = 1 - \prod_{k=1}^{2n-3} (1 - b_k) \tag{58}$$

Where there are n terminal vertices and the unrooted tree is composed of only bifurcations so that the total number of vertices is always given by $2n-3$. b_k gives the probability of a difference along a branch so $1-b_k$ gives the probability

that they are the same. The product term overall gives the probability that all vertices are the same.

To see how this corresponds to a measure in the ϕ world first add a stem and root to the tree from any interior vertex and direct all edges away from the root. Let us refer to this new edge as having index $2n-2$ so that we can write the edge as e_{2n-2} with all the other edges being indexed as before. Now set

$$\alpha(e_i) = \begin{cases} 1 & \text{when } i = 2n-2; \\ 0 & \text{otherwise.} \end{cases} \tag{59}$$

$$\beta(e_i) = \begin{cases} 0 & \text{when } i = 2n-2; \\ b_i & \text{otherwise.} \end{cases} \tag{60}$$

The proportion of ancestral EvoHeritage (from edge e_{2n-2}) that *does* survive to all terminal vertices is the amount that survives EvoHeritage attrition on all branches other than the stem (which is where the EvoHeritage is generated). This is given by

$$\prod_{k=1}^{2n-3} (1 - \beta(e_k)) \tag{61}$$

The proportion that *does not* survive is just one minus that giving

$$1 - \prod_{k=1}^{2n-3} (1 - \beta(e_k)) = 1 - \prod_{k=1}^{2n-3} (1 - b_k) = P(\geq 2) \tag{62}$$

And hence the relationship to genetic diversity from the main text is proven.

2.3 Proof: the power set based calculation of ϕ is equivalent to direct calculation

We will make use of the following lemma in the proof of Theorem 4 from the main text

Lemma 2.1 *For a set A, and function $\beta : A \to \mathbb{R}$:*

$$\sum_{G \in \mathcal{P}(A)} \prod_{g \in G} \beta(g) \prod_{g \in A \setminus G} (1 - \beta(g)) = 1$$

Proof 2.2 $1 = \prod_{g \in A} (\beta(g) + (1 - \beta(g))) = \sum_{G \in \mathcal{P}(A)} \prod_{g \in G} \beta(g) \prod_{g \in A \setminus G} (1 - \beta(g))$

To prove Theorem 4 from the main text, we first recall the recursive formulation of ϕ.

$$\phi(A) = \sum_{i \in I} \sum_{e \in E} \alpha_i(e) \cdot p_i(d^V(e), A), \tag{63}$$

where $p_i(v, A)$ is given by

$$p_i(v, A) = \begin{cases} 1 & \text{when } v \in A \\ 0 & \text{when } v \notin A \text{ and } D^E(v) = \emptyset \\ 1 - \prod_{j \in D^E(v)} [1 - p_i(d^V(j), A) \cdot \beta_i(j)] & \text{otherwise.} \end{cases} \tag{64}$$

On the other hand, starting with the expression for $Y(\mathbb{I}, A)$ and rearranging the order of summation we get:

$$\begin{aligned} Y(\mathbb{I}, A) &= \sum_{i \in I} \sum_{G \in \mathcal{P}(E)} \left(\prod_{g \in G} \beta_i(g) \prod_{g \in E \setminus G} (1 - \beta_i(g)) \sum_{e \in E} \alpha_i(e) \cdot \mathbb{I}(d^V(e), G, A) \right) \\ &= \sum_{i \in I} \sum_{e \in E} \alpha_i(e) \cdot \sum_{G \in \mathcal{P}(E)} \left(\prod_{g \in G} \beta_i(g) \prod_{g \in E \setminus G} (1 - \beta_i(g)) \cdot \mathbb{I}(d^V(e), G, A) \right) \end{aligned}$$

Hence, the theorem holds provided, for every edge $e \in E$, we can show

$$p_i(d^V(e), A) = \sum_{G \in \mathcal{P}(E)} \left(\prod_{g \in G} \beta_i(g) \prod_{g \in E \setminus G} (1 - \beta_i(g)) \cdot \mathbb{I}(d^V(e), G, A) \right). \tag{65}$$

The proof proceeds by induction on the number of edges in E. As the base case we consider the DAG with one edge, which we call e. Suppose $v = d^V(e)$. Note that $\mathbb{I}(v, G, A) = 1$ if and only if $v \in A$, for both $G = \emptyset$ and $G = \{e\}$. So in this case, beginning from the right hand side of Eqn. 65:

$$\sum_{G \in \mathcal{P}(E)} \left(\prod_{g \in G} \beta_i(g) \prod_{g \in E \setminus G} (1 - \beta_i(g)) \cdot \mathbb{I}(d^V(e), G, A) \right) \tag{66}$$

$$= \prod_{g \in \emptyset} \beta_i(g) \prod_{g \in \{e\}} (1 - \beta_i(g)) \cdot \mathbb{I}(v, \emptyset, A) + \prod_{g \in \{e\}} \beta_i(g) \prod_{g \in \emptyset} (1 - \beta_i(g)) \cdot \mathbb{I}(v, \{e\}, A) \tag{67}$$

$$= (1 - \beta_i(e)) \cdot \mathbb{I}(v, \emptyset, A) + \beta_i(e) \cdot \mathbb{I}(v, \{e\}, A) \tag{68}$$

$$= \begin{cases} 1 & \text{when } v \in A \\ 0 & \text{when } v \notin A \end{cases} \tag{69}$$

$$= p_i(v, A), \tag{70}$$

and the base case holds.

Now suppose that $D = (V, E)$ is a DAG, and that Eqn. 65 holds for all DAGs with less than $|E|$ edges.

We select an edge $e \in E$ and split up the sum on the right hand side of Eqn. 65 into two parts. The first part consists of subsets of E that contain e, and the

second consists of those subsets of E which exclude e. In this way, we create the following expression:

$$(1-\beta_i(e)) \cdot \sum_{G \in \mathcal{P}(E \setminus \{e\})} \left(\prod_{g \in G} \beta_i(g) \prod_{g \in E \setminus G} (1-\beta_i(g)) \cdot \mathbb{I}(d^V(e), G, A) \right)$$
$$+ \beta_i(e) \cdot \sum_{G \in \mathcal{P}(E \setminus \{e\})} \left(\prod_{g \in G} \beta_i(g) \prod_{g \in E \setminus G} (1-\beta_i(g)) \cdot \mathbb{I}(d^V(e), G \cup \{e\}, A) \right). \tag{71}$$

We consider two cases: either $d^V(e) \in A$ or $d^V(e) \notin A$.

Firstly, assume $d^V(e) \in A$. Then $\mathbb{I}(d^V(e), G, A) = \mathbb{I}(d^V(e), G \cup \{e\}, A) = 1$ because we consider the path (with no edges) from $d^V(e)$ to itself to be a valid path. Then the expression (71) is equal to 1 by Lemma 2.1. Since $p_i(d^V(e), A) = 1$ whenever $d^V(e) \in A$, we have shown that Eqn. 65 holds in this case.

Now assume $d^V(e) \notin A$. Then for every set G of edges, $\mathbb{I}(d^V(e), G, A) = \mathbb{I}(d^V(e), G \cup \{e\}, A)$ because e doesn't help us to connect to any vertex in A which was not already reachable via edges in G. Then the expression (71) becomes

$$\sum_{G \in \mathcal{P}(E \setminus \{e\})} \left(\prod_{g \in G} \beta_i(g) \prod_{g \in E \setminus G} (1-\beta_i(g)) \cdot \mathbb{I}(d^V(e), G, A) \right). \tag{72}$$

By the induction assumption this expression is equal to $p_i(d^V(e), A)$, and Eqn. 65 holds.

2.4 Summing the partitioned-ϕ of individual vertices

Many practical applications will find it useful to partition the EvoHeritage of all extant vertices between each individual extant vertex, however, we could use the same solution to partition between any mixture of extant and extinct vertices. This is not made explicit when partitioning classic PD, for example with the Evolutionary Distinctiveness (ED), or Equal Splits (ES) methods, because extinct species are never considered.

The sum of all partitioned-ϕ values returns the total ϕ of all extant vertices together: $\phi(B) = \sum_{b \in B} \text{partitioned-}\phi(\{b\}, B)$. To see this, note that the total number of paths $c(v, G, B)$ can be found by calculating the number of paths from v to each distinct b in B, then summing these values together. In symbols:

$$c(v, G, B) = \sum_{b \in B} c(v, G\{b\})$$

.

$$\sum_{b\in B} \text{partitioned-}\phi(\{b\}, B) \tag{73}$$

$$= \sum_{b\in B} Y(R, (\{b\}, B) \tag{74}$$

$$= \sum_{b\in B} \sum_{i\in I} \sum_{G\in\mathcal{P}(E)} \left(\prod_{g\in G} \beta_i(g) \prod_{g\in E\setminus G} (1-\beta_i(g)) \sum_{e\in E} \alpha_i(e) \cdot R(d^V(e), G, \{b\}, B) \right) \tag{75}$$

$$= \sum_{i\in I} \sum_{G\in\mathcal{P}(E)} \left(\prod_{g\in G} \beta_i(g) \prod_{g\in E\setminus G} (1-\beta_i(g)) \sum_{e\in E} \alpha_i(e) \cdot \frac{\sum_{b\in B} c(d^V(e), G, \{b\})}{c(d^V(e), G, B)} \right) \tag{76}$$

$$= \sum_{i\in I} \sum_{G\in\mathcal{P}(E)} \left(\prod_{g\in G} \beta_i(g) \prod_{g\in E\setminus G} (1-\beta_i(g)) \sum_{e\in E} \alpha_i(e) \cdot \mathbb{I}(d^V(e), G, B) \right) \tag{77}$$

$$= \phi(B) \tag{78}$$

2.5 Values of partitioned-ϕ and unique-ϕ for extremal values of attrition

Let $T = (V, E)$ be an EvoHeritage tree, and for each edge $e \in E$, let $L(e)$ give the length of e. Let $A \subset V$ be the set of terminal vertices of T. Suppose a is a terminal vertex of T, and that e_a is its corresponding terminal edge. We write $ED(a)$ for the evolutionary distinctiveness score of a in the tree T, and $\mathcal{D}(e)$ for the set of terminal vertices descended from edge e.

Proposition 2.3 *Let T be an EvoHeritage tree under standard conditions, with attrition parameter ρ. The extremal values of partitioned-ϕ_ρ and unique-ϕ_ρ are given by:*

- partitioned-$\phi_0(\{a\}, A) = ED(a)$
- unique-$\phi_0(\{a\}, A - \{a\}) = L(e_a)$
- partitioned-$\phi_\infty(\{a\}, A) = 1$
- unique-$\phi_\infty(\{a\}, A - \{a\}) = 1$

Recall that $\phi(A) = Y(\mathbb{I}, A)$ and partitioned-$\phi(A, B) = Y(R, (A, B))$. We begin by proving the statements where $\rho = 0$. From section 1.3.2 we have $\alpha(e) = L(e)$ and $\beta(e) = 1$ for all $e \in E$. Also, note that in this case $S_u = 1$.

Whenever $G \neq E$, the product $\prod_{g\in E\setminus G} (1-\beta(g))$ is nonempty, and equals zero, making these terms vanish in the relevant Y equations. For $G = E$, the function

$c(d^V(e), G, A)$ equals the number of terminal vertices in A that descend from e, that is: $|\mathcal{D}(e)|$. In particular, $c(d^V(e), G, \{a\})$ equals one when a is descended from e, and zero for all other edges. Therefore, dividing through by S_u to get a standardised value,

$$\begin{aligned}
\frac{\text{partitioned-}\phi_0(\{a\}, A)}{S_u} &= \frac{Y(R, (\{a\}, A))}{S_u} \\
&= \prod_{e \in E} 1 \prod_{e \in \emptyset} 0 \sum_{e \in E} L(e) R(d^V(e), E, \{a\}, A) \\
&= \sum_{\substack{e \in E, \\ a \in \mathcal{D}(e)}} L(e) \frac{1}{|\mathcal{D}(e)|} \\
&= ED(a).
\end{aligned} \tag{79}$$

Let $\bar{E}(a)$ be the set of edges that have a descendant terminal vertex distinct from a. Then, noting that $\{a\} \cup (A - \{a\}) = A$,

$$\begin{aligned}
\frac{\text{unique-}\phi_0(\{a\}, A - \{a\})}{S_u} &= \frac{\phi_0(A) - \phi_0(A - \{a\})}{S_u} \\
&= Y(\mathbb{I}, A) - Y(\mathbb{I}, A - \{a\}) \\
&= \sum_{e \in E} L(e) - \sum_{e \in \bar{E}(a)} L(e) \\
&= L(e_a).
\end{aligned} \tag{80}$$

Next, we prove the statements where $\rho \to \infty$. From section 1.3.1, as ρ increases we have $\alpha(e) \approx \frac{1}{\rho}$ and $\beta(e) \to 0$ for all $e \in E$. We use the approximation $S_u \approx \frac{1}{\rho}$, valid for large values of ρ.

Whenever $G \neq \emptyset$, the product $\prod_{g \in E \setminus G} (\beta(g))$ is nonempty, and equals zero, making these terms vanish in the relevant Y equations. For $G = \emptyset$, the function $c(d^V(e), G, A)$ equals one if and only if $d^V(e) \in A$, and equals zero otherwise. Therefore, dividing through by S_u to get a standardised value,

$$\begin{aligned}
\frac{\text{partitioned-}\phi_\infty(\{a\}, A)}{S_u} &= \frac{Y(R, (\{a\}, A))}{S_u} \\
&= \frac{\prod_{e \in \emptyset} 0 \prod_{e \in E} 1 \sum_{e \in E} \alpha(e) R(d^V(e), \emptyset, \{a\}, A)}{S_u} \\
&= \frac{\alpha(e_a)}{S_u} \\
&\approx \frac{\frac{1}{\rho}}{\frac{1}{\rho}} = 1
\end{aligned} \tag{81}$$

and

$$\begin{aligned}
\frac{\text{unique-}\phi_0(\{a\}, A - \{a\})}{S_u} &= \frac{\phi_0(A) - \phi_0(A - \{a\})}{S_u} \\
&= \frac{Y(\mathbb{I}, A) - Y(\mathbb{I}, A - \{a\})}{S_u} \\
&= \frac{\sum_{e \in E} \alpha(e) \mathbb{I}(d^V(e), \emptyset, A) - \sum_{e \in E} \alpha(e) \mathbb{I}(d^V(e), \emptyset, A - \{a\})}{S_u} \\
&= \frac{\alpha(e_a)}{S_u} \\
&\approx \frac{\frac{1}{\rho}}{\frac{1}{\rho}} = 1.
\end{aligned} \tag{82}$$

2.6 Supplementary figure illustrating ED and partitioned-ϕ

Please refer to figure 9.

References

Crozier, R. H. 1992. Genetic diversity and the agony of choice. Biological conservation 61:11–15.

Dodd, M. S., D. Papineau, T. Grenne, J. F. Slack, M. Rittner, F. Pirajno, J. O'Neil, and C. T. Little. 2017. Evidence for early life in earth's oldest hydrothermal vent precipitates. Nature 543:60–64.

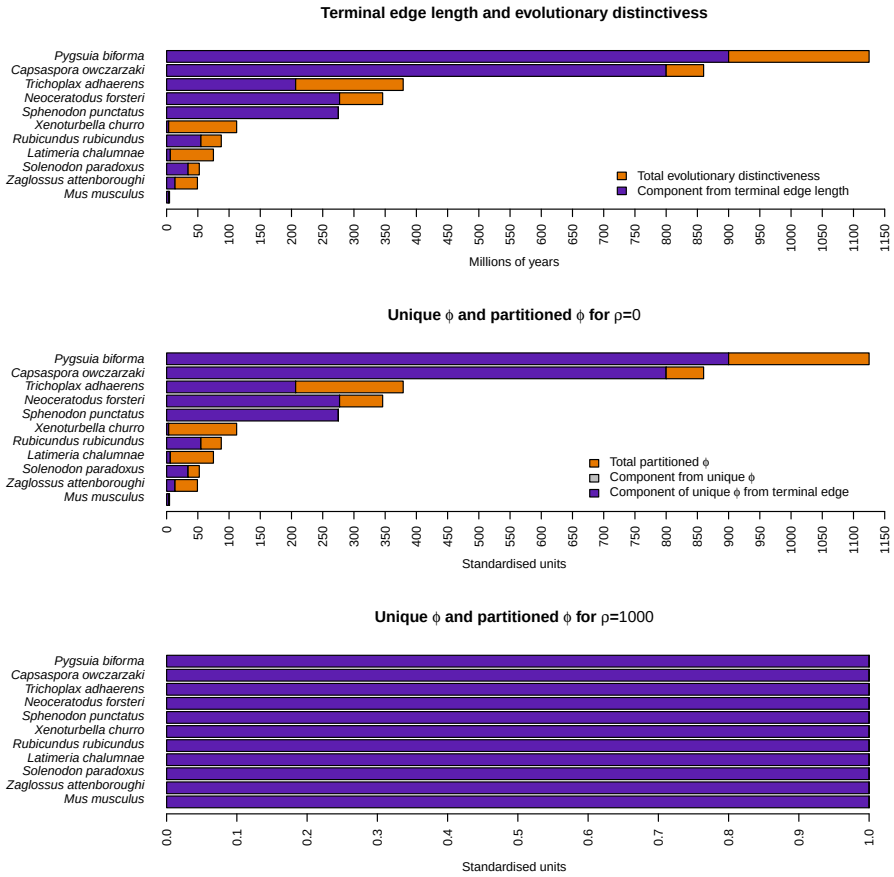

Figure 9: Comparison of ED, terminal edge length, unique-ϕ and partitioned-ϕ for selected species ordered by ED. The stacked bars of the top panel show ED separated into two components: the terminal edge legth, and the remainder which is a contribution from interior edges. The middle and bottom panel show an equivalent for ϕ with $\rho = 0$ and $\rho = 1000$ respectively. The results show unique-ϕ and partitioned-ϕ converge to terminal edge length and ED for $\rho \to 0$ and to unity for large ρ. We are showing the midpoint of the 'stick' and 'star' cases for sister clades but note there was only negligible difference between the two.