Information Accumulation and Loss as a Foundation for Sound Phylogenetic Biodiversity Metrics

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, and Phylogenetic Diversity (PD) is one of the most prominent measures of biodiversity. PD is based on the idea that biological features accumulate along the branches of phylogenetic trees, and that these features are of biological importance. We argue that PD, and other phylogenetic biodiversity metrics, fail to capture an essential biological process: 'information loss', whereby features, and biological information in general, can be lost through the process of evolution. We introduce Generalised Phylogeny Diversity (GPD), which is founded on the joint processes of information accumulation and loss and can be applied to phylogenetic trees or more complex networks. We derive a dimensionless measure ϕ from GPD that reproduces species richness and PD at opposite ends of a continuum of information loss intensity. We suggest how the existing calculus of PD-based metrics, and other phylogenetic biodiversity metrics, can be recast to incorporate information loss. To illustrate our approach, we give three empirical applications all reliant on new measures using the information loss concept. First, in ecology, evaluating ϕ as a predictor of community productivity against species richness and PD. Second, in evolution, quantifying living fossils and resolving their associated controversy. Third, in conservation, using a partitioning of GPD to re-balance our priorities at the broad scale of the complete tree of life.

Key words:

phylogenetic diversity, phylogenetic metric, diversity metric, biodiversity, feature loss, information accumulation, information loss, 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 associated '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 (IPBES) (Brondizio et al., 2019). PD has become a mainstream feature of studies in ecology, evolution and conservation 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 necessary 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, using 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 with 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 that effectively assumes 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 'information loss', encompassing the idea that features, traits, genetic sequences, and all forms of biological information (biodiversity) may be both gained and lost along the paths of a phylogenetic tree without extinction occurring. A practical and well documented example of information loss 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, 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 a 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 information loss. In the cases where homoplasy occurs without information loss, 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 information loss, and not homoplasy, is the real challenge for the classic PD narrative.

There are many further conceptual reasons pointing to the importance of information loss. 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 do 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 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 likely 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 saturate given sufficient time, rather than grow indefinitely, even if evolution remains rapid (Crozier et al., 2005). It was addressed at the stage of tree construction for Crozier's genetic diversity measure, where branch lengths are bounded between 0 and 1, capturing probabilities of a genetic difference (Crozier, 1992). Information loss would introduce 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 with less weight than more recent parts of the tree.

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 descendent. 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 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. This does not seem biologically reasonable. In conservation applications it could potentially undervalue the contribution of descendent species in capturing ancient features that have already been lost in other living descendants.

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, for example the Archaeopteryx (or a close relative directly ancestral to birds). This is because the connecting branches of the living species already include 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. Whilst 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 so far not done so. Furthermore, inconsistency in a contrived example is likely a symptom of broader conceptual problems that are of more general concern. Incorporating information loss would change the role of

extinct species because some of the features possessed by 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 back the lost features and increase diversity accordingly.

In this manuscript, we introduce 'Generalised Phylogenetic Diversity' (GPD), which quantitatively incorporates not only the accumulation of features and common ancestry as PD does, but also the concept of 'information loss'. We derive an expression for the total GPD value of a collection of taxa, species or populations, and show how the concepts translate through to the rest of the PD calculus with selected examples. We leave provision for GPD to be applied to structures more general 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 GPD facilitates entirely new applications. We introduce a new measure of 'generalised phylogentic *difference*' in which the difference between two distinct groups saturates over time rather than grows without bound. This measure may serve as an alternative to branch lengths in other phylogenetic analyses.

The concept of GPD allows us to define a scaled (dimensionless) quantity that we denote by ϕ_ρ, and 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 ϕ_ρ as a predictor of community productivity compared to species richness and PD, using data from the Cedar Creek grassland experiments (Tilman et al., 1996). GPD can also consider the fate of just the features evolving during a particular period in geological history. This enables GPD to identify so-called 'living fossils' (Grandcolas and Trewick, 2016) as concentrated groups preserving the most rare ancestral features in the face of information loss. We derive an expression for 'Partitioned GPD' which gives a way to fairly share GPD between groups akin to the Evolutionary Distinctiveness measure from (Isaac et al., 2007). We illustrate this with an example conservation application, showing that information loss is essential for proportionate comparisons between species at the scale of the whole tree of life. We conclude by identifying future research and conceptual advances that we hope GPD will facilitate.

Introducing Generalised Phylogenetic Diversity

We introduce GPD in the first instance with a glossary of the new terms required to understand it. The glossary is in logical (rather than alphabetic) 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 3.

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, that is no ancestral edges.

Terminal vertex: a vertex of a DAG which has no edges directed away from it, that is,

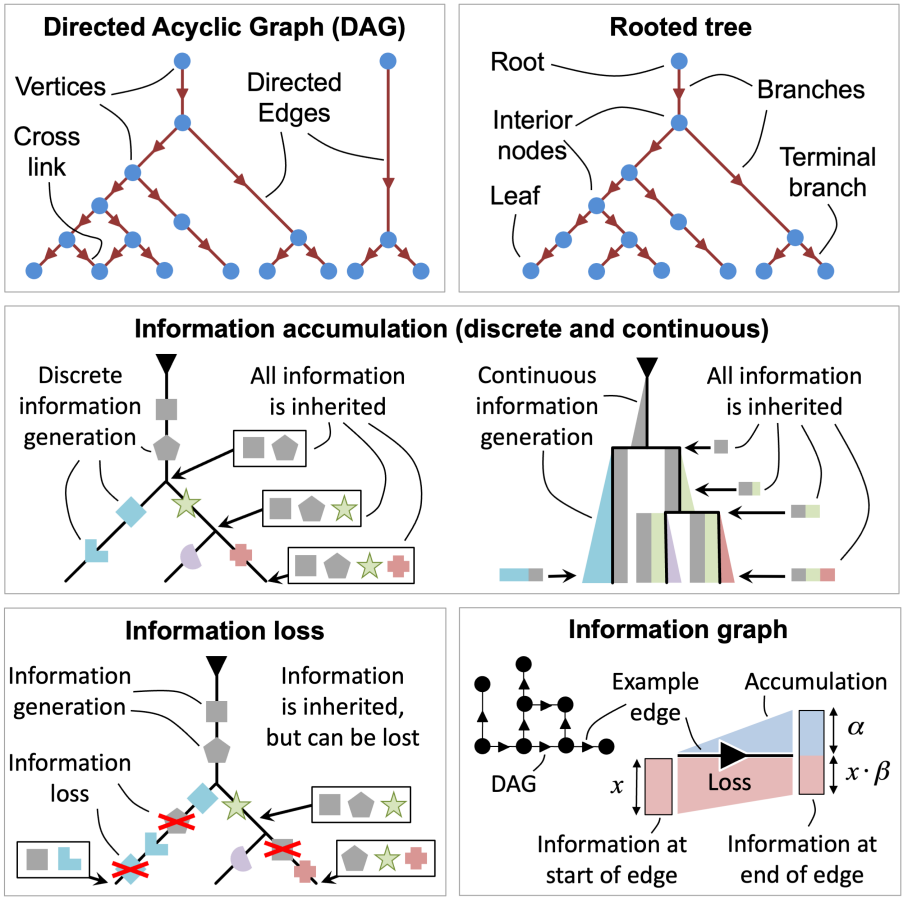

Fig. 1. An illustration of key terms and concepts from the glossary. Some concepts are clearer to visualise with discrete information and others are clearer with continuous information. In practice, if information accumulation is sufficiently rapid compared to edge lengths, the distinction between discrete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

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.

Information: biological 'information' in the most general sense including features, genetic sequences and any other sources of variation. Information may be measured as a continuous or discrete quantity.

Information accumulation: a natural process by which novel information is generated along edges with copies being passed on to descendants.

Information loss: a natural process by which information is lost over time without extinction having occurred, for example by not being passed on to surviving descendants.

Information graph: a DAG where the accumulation and loss of information occurs in a specified way along each edge.

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

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

GPD: Generalised Phylogenetic Diversity. For a given set of vertices on an information graph, the total quantity of information present on those vertices.

Unique GPD: for two given sets of vertices on an information graph, gives the quantity of information that is present within the first set of vertices and absent from the second set of vertices.

$\mathbf{GP}_{\text{diff}}$: Generalised Phylogenetic Difference. For two sets of vertices on an information graph, the proportion of total information that is only present within one of the two sets.

Partitioned GPD: a partitioning of GPD between vertices on an information graph, where the value of any information is divided equally between all copies of that information.

Dated information graph: an information graph where all vertices also have dates attached.

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

Extinct vertex: any vertex with date > 0 on a dated information graph.

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

Living fossil: an extant vertex with a large share of rare ancestral information. After partitioning GPD among extant vertices of a dated ancestral information graph, those vertices with 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 information graph according to a simplified set of rules. Specifically, i) all information is of the same form and measured as a continuous quantity; ii) information accumulates at a fixed deterministic rate of one unit of information per unit of edge length; iii) information loss occurs at a fixed deterministic rate of ρ per unit of information per unit of edge length. The rate of information loss is thus proportionate to the total amount of information.

$\hat{\boldsymbol{o}}$: a root vertex corresponding to the origin of life on an information graph.

Standardised units: under standard conditions (information loss at rate ρ), the total information at the end of a single disconnected edge with a given standard length. One unit of edge is a convenient standard length to use by default. For dated information graphs this means using one million years, however, the standard length could be chosen less arbitrarily as the time back to $\hat{\boldsymbol{o}}$ with $\hat{o}$ included as a vertex in the calculations.

$\boldsymbol{\phi_\rho}$: a convenient notation for GPD under standard conditions and measured in standardised units on an information graph. It can thus be used in the same way as PD for any given $\rho \geqslant 0$.

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

Rationale

We define GPD in general terms to support varied future applications (please refer to the glossary). For greater generality, we chose to 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 which are not most recent common ancestors and thus would not otherwise be identified as a vertex. Polytomies are permitted, though in most cases they would be better to be resolved (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 information graph as a DAG that specifies the accumulation and loss of information explicitly, and thus contains all the data required to calculate GPD for any given set of vertices. Information graphs are novel objects so would have to be constructed in a bespoke manner, based on the research objectives and available data. There may be different 'forms' of information potentially represented in an information graph, each with their own characteristics of accumulation and loss. For example, we may have 'immortal' information as one form that is never lost to enable rare and truly irreversible evolutionary steps to be accounted for. Introducing the information graph concept enables a clear conceptual separation between the core GPD calculations (products of an information graph) and auxiliary details about the precise way that information is assumed to accumulate and get lost (e.g. building an information graph from a tree), something more likely to change in different applications. We leave open the possibility for information to

be continuous or discrete, the latter going hand in hand with the possibility that GPD may be a stochastic distribution of values rather than a fixed value and thus incorporate a mechanism for quantifying its own limits as a proxy for other biodiversity measures.

The quantity ϕ_ρ is introduced as a special version of GPD (derived later) that is dimensionless and can be calculated given dates on each vertex of a DAG (or tree) and a value for its associated parameter (ρ) representing information loss. It can therefore be directly swapped with PD in existing calculations on a dated tree for any given ρ. The idea was inspired by Hill numbers that incorporate an 'order' or q parameter and recover a range of preexisting metrics (or transformations of them) at special values of q (Hill, 1973; Jost, 2006). One could study ϕ_ρ across a complete spectrum of ρ values, or fix ρ based on biologically reasoned arguments as we discuss later. Unlike GPD, ϕ_ρ is measured in standardised units: multiples of the net amount of information that accumulates along a single unit of edge, so formally ϕ_ρ is dimensionless. We recommend that ϕ_ρ be measured on a DAG that includes the origin of life as a vertex $\hat{o}$ (see supplementary information). 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 information loss ($\rho \to \infty$). Conversely, the value of ϕ_0 is comparable to 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 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 GPD

In this section we give the formal definition of the Generalised Phylogenetic Diversity of a set of vertices (Eqn. (2) below). Throughout this manuscript, we describe an information graph mathematically by letting V represent the set of vertices, E the set of directed edges, and I the set of information 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 information $i \in I$, let $\alpha_i(e)$ give the net amount of information of form i accumulated along each edge $e \in E$. Let $\beta_i(e)$ give the proportion of information of form i present at the start of the edge $e \in E$ that survives information loss along the edge to still be present at vertex $d^V(e)$.

To derive the GPD of 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 information of form $i \in I$ associated with the vertex v that survives to be counted on at least one vertex in the set A. 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 information 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 information arrive at the same vertex (via multiple paths through the GPD graph), each copy is lost independently. This enables each vertex and each section of information to be treated independently. Now we are in a position to write an expression for the GPD value of a set of vertices A.

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

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

Derivation of ϕ_ρ

In this section, we provide the 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 information graph with specified edge lengths. We express this with an additional function $L(e)$ giving the length of each edge $e \in E$. Information loss under standard conditions occurs at rate $\rho \geqslant 0$ per unit of information per unit of edge length. Therefore the proportion of information surviving loss, $\mathrm{I}(l)$, as a function of edge length l, satisfies the differential equation:

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

Solving this equation yields

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

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

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

Now we must derive an expression for $\alpha(e)$ which includes information accumulation along the edge but also loss of the information generated earlier along the edge to give a net accumulation value. To do this, we integrate over edge e. This accounts for the information generated at every point along the edge, each having a different length over which it can be lost before reaching the end vertex. Noting that information 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 information graph using Eqns. (5) and (6) for each edge of the DAG. We can then calculate GPD using Eqns. (1) and (2) together with our information graph. Finally, we can write S_u, the standardised unit, based on an edge of unit length from Eqn. (6) as

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

and for a set of vertices A, define $\phi_\rho(A)$ as

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

In supplementary information we show: i) adding or taking away a 'monotomy' vertex does not affect the results of GPD calculations on other vertices, ii) ϕ_ρ is independent of the fixed rate of information accumulation per unit of edge, so there is no loss of generality by assuming a rate of 1 as is done under standard conditions.

Relationship between GPD and PD

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

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

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

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

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

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

Theorem 2 Let T be an information tree with terminal vertices given by $X \subset V$ and one information form with $\beta(e) < 1$ for every edge $e \in E$. In that case, $GPD(A) = PD(A)$ 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 GPD on a tree when information loss is explicitly turned off, or in cases where the tree has a very simple shape at most it would be 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 maximal ϕ_ρ (see supplementary material). An intuitive consequence of this section is that provided information loss is small on the scale of the stem age of a tree, then PD remains a reasonable approximation for GPD. Conversely, if information loss is large or we are studying an old tree with deeper splits, then the disparity between GPD and PD will likely be more pronounced.

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

Our standard stochastic conditions consider one form of information 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. GPD under standard stochastic conditions can therefore 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 respective rates $\lambda(e)$ and $\mu(e)$ along edge e. We illustrate standard stochastic conditions on a tree (not a DAG) in order to facilitate 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 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 frame this as a 'constant birth' process. The rate that 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. 2.

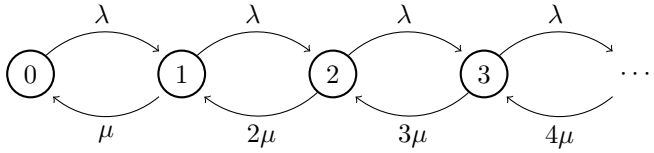

Fig. 2. A simple Markov chain showing the net number of discrete features gained (in the circles) and the stochastic rate of movement between states. The diagram illustrates a single edge of arbitrary length. The rate of information accumulation is λ per unit time and the rate of information loss is μ per feature per unit time.

Let T be an information tree (a tree with information accumulation and loss specified akin to an information graph) and let $X \subseteq V$ represent the set of terminal vertices. The *GPD* of a set $A \subseteq X$ is the total quantity of information (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 $GPD(A) = FD(A) = |\mathbb{F}_A|$. Similar to the deterministic setting, we find that PD rarely matches GPD and FD. Theorem 3 provides

the precise conditions for when the two measures do align.

Theorem 3 Let T be an information tree with vertex set V, edge set E, $X \subseteq V$ the set of terminal vertices, and $r \in V$ 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 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 supplementary material.

Developing a complete GPD 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 GPD is beyond the scope of this work, however, we can begin the task and demonstrate that the rest will be achievable as future work. For worked examples of parts of the GPD calculus refer to figure 3.

Unique GPD

The Unique GPD captured by a subset A of vertices in the graph, but which is 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:endemism' (Faith, 2013) and is calculated in a similar manner. The need to define the set B more explicitly for GPD arises because extinct vertices are now potentially included. For most applications, we would expect researchers to choose that B contain all extant vertices that are not in A.

$$\text{Unique GPD}(A, B) = \text{GPD}(A \cup B) - \text{GPD}(B) \tag{9}$$

Generalised phylogenetic difference

The generalised phylogenetic difference GP_{diff} between two subsets (A and B) of vertices is information present within only one of the two sets, expressed as a proportion of the total information present in both subsets together. This can be written as follows.

$$\text{GP}_{\text{diff}}(A, B) = \frac{\text{Unique GPD}(A, B) + \text{Unique GPD}(B, A)}{\text{GPD}(A \cup B)} \tag{10}$$

Note that GP_{diff} is symmetric so that $\text{GP}_{\text{diff}}(A, B) = \text{GP}_{\text{diff}}(B, A)$. 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 $\rho \to 0$, $\text{GP}_{\text{diff}}(A, B)$ will converge precisely to the same result as UniFrac. The Phylosor metric of phylogenetic difference (Bryant et al., 2008) is similar to UniFrac, but based on the Sørensen index, a GPD counterpart along similar lines to $\text{GP}_{\text{diff}}(A, B)$ would be easy to conceive.

In supplementary material we explore the relationship between $\text{GP}_{\text{diff}}(\{a\}, \{b\})$, 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

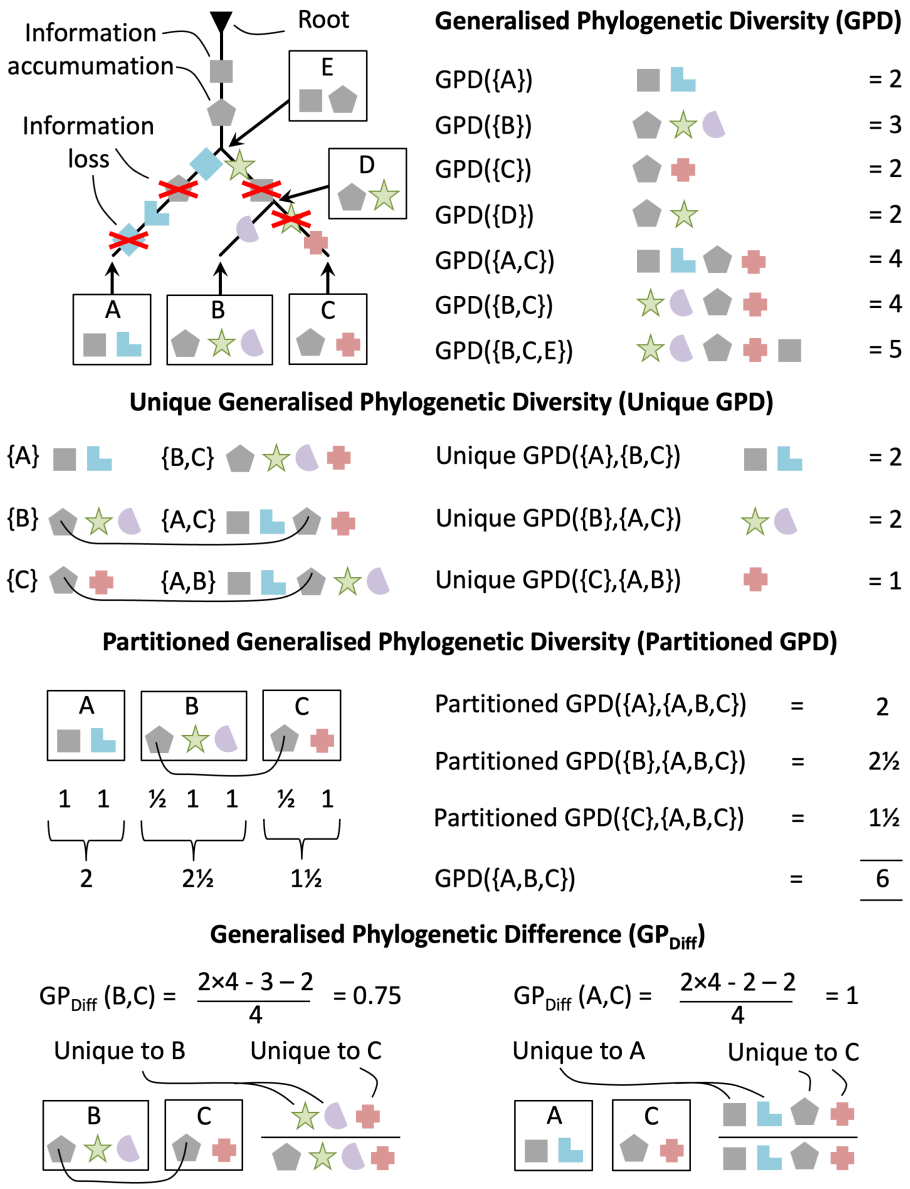

Fig. 3. Illustrative examples of GPD, Unique GPD, paritioned GPD and GP_{diff}. This example uses discrete information on the tree as this makes it easier to picture the underlying principles. The reader can imagine that if there were very large numbers of such discrete information elements, the result would be a system where information is effectively a continuous quantity.

$\mathrm{GP}_{\mathrm{diff}}$ between a pair of vertices saturates as the time since common ancestry increases. As $\rho \to 0$ we see that $\mathrm{GP}_{\mathrm{diff}}(\{a\},\{b\})$ converges to twice the date of common ancestry divided by a normalisation factor. In the other limit, as $\rho \to \infty$ we see that $\mathrm{GP}_{\mathrm{diff}}(\{a\},\{b\})$ converges to unity.

Relationship to genetic diversity

GPD has been constructed to avoid obligatory ties to the species concept and species trees. It therefore has many of the 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 GPD. However, one early genetic diversity metric conceived by Crozier (1992) does have a quantitative equivalent in GPD. Crozier considers 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 to our values of β (information loss) on each edge but without any values of α (information 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 information that does not survive information loss to exist on every extant vertex is the same as genetic diversity (see supplementary material). This may seem counter intuitive but information loss is interpreted as a source of multiplicative 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 GPD 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 GPD. We will see how this paves the way for a more elegant description of the broader GPD calculus, and for the use of Monte Carlo methods in numerical GPD calculations.

Let $G = (V, E)$ be an information 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 GPD, $H = \mathcal{P}(V)$ so that $A \in H$ is a subset of vertices for which we are calculating GPD. We now define a function $Y(F, A)$ which calculates the percolation of information through an information 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 information given in the set I. Each form of information is independent of 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 information graph with information loss as a weighted distribution of DAGs without information loss. To see this in an example, consider a DAG on which only one edge e_1 has any information loss, 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 edge e_1 included (for information not lost along that edge), second, with weight $1 - \beta(e_1)$, the same calculation of F on the DAG with edge e_1 removed. Scaling this thinking up to a case where any possible subset of edges may be disconnected by information loss in different scenarios, we get the remainder of the terms given in Eqn. 11. These multiply up factors of $\beta(e)$ and $1 - \beta(e)$ for different edges to weigh across the possible combinations of information loss. 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).

We will now describe a specific choice of function for F so that $Y(F, A)$ returns $GPD(A)$ 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 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

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

Now we can express GPD in terms of Y and $\mathbb{I}$.

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

For a proof of Theorem 4 refer to supplementary material.

Partitioned GPD

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 descendent vertices regardless of tree topology (Isaac et al., 2007). Here we provide an equivalent for GPD. We seek to share out (partition) the 'value' of all information present on a set of vertices B fairly between subsets of vertices. Consider the subset $A \subseteq B$. The Partitioned GPD of A with respect to B is the share of the value GPD(B) which is attributed to vertices in A. To calculate this, we need to take into account the number of copies of information across all vertices of A and across all vertices of B (Figure 3). This is especially complex on an information graph because it is possible for a single vertex to inherit more than one copy of the same information via cross links in the underlying DAG. Such vertices with multiple copies of the same information receive a greater proportion of the associated value. The power set approach provides an elegant solution because information loss is accounted for automatically by weighted averaging in the function Y. We only need to write a function to partition value of the information from one vertex through known edges.

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

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

Then applying Y to the function F gives us Partitioned GPD(A, B).

$$\text{Partitioned GPD}(A, B) = Y\left(F, (A, B)\right) \tag{15}$$

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

Definition of Hill GPD 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

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

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 ${}^0D$, the exponential of Shannon entropy ${}^1D$, and the inverse of the Gini-Simpson index ${}^2D$. 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 edge do not contribute to diversity and are excluded from calculations. Hill PD, denoted ${}^qPD\,(t)$, is defined by

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

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 descendent 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

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

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. Nevertheless it is predicated on the same fundamental assumption as PD, that all information (on edges under consideration) is copied to all descendants and there is no information loss.

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

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

Or 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 information 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 loss. If we are considering species abundances and some 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.

Empirical applications - Methods

In this section we will introduce three different empirical applications of our work that all rely on information loss: i) an ecological application evaluating ϕ_ρ as a predictor of community productivity; ii) an evolutionary application using the concepts of information loss and Partitioned GPD to quantify living fossils; iii) a conservation prioritisation application application comparing Partitioned GPD 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 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 pattern of phylogenetic diversity to also 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

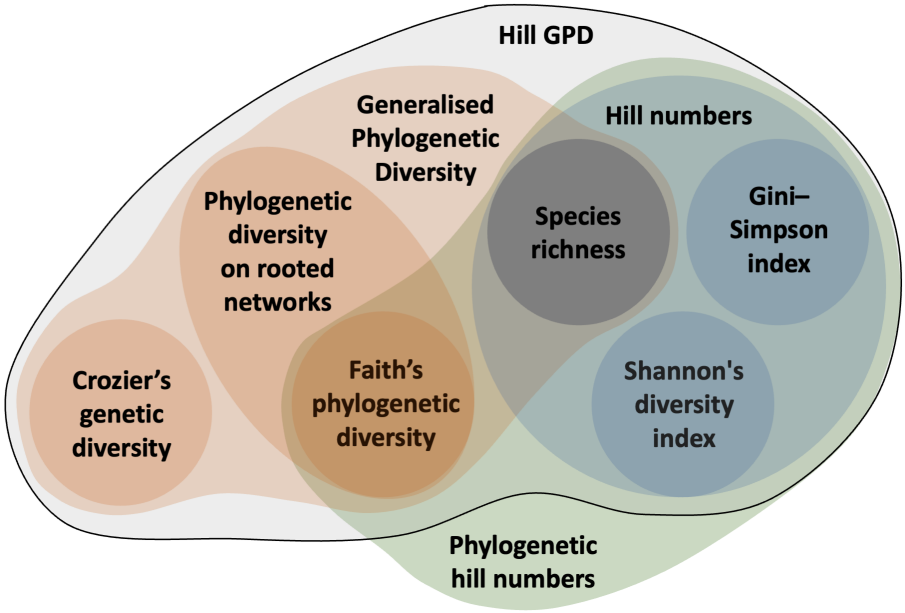

Fig. 4. Connections between overlapping concepts and generalisations. Blue corresponds to species abundance data, orange corresponds to phylogenetic data. Green corresponds to existing work connecting PD and hill numbers.

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 literature; Tilman et al., 1996, 2001), the authors ranked sixteen 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 an improvement on both species richness and a range of other functional diversity measures.

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

We sourced the Cedar Creek Biodiversity Experiment data from (Tilman, 2021) to provide both plot biomass and a matrix of species presence by plot. We sourced the phylogenetic tree directly from (Cadotte et al., 2009). All analyses were done 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. The tree was rate-smoothed (Sanderson, 2002) setting the total tree depth to 1 in order for comparability with Cadotte et al. (2009).

We developed an R function to iteratively calculate GPD 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 GPD for every one of the empirical plots. We repeated this process for a range of information loss 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 all the same diversity metrics for just the *Poaceae* and again for just the *Pentapetalae* as well as for the total (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, *Poaceae* only and *Pentapetalae* only. We discuss our findings in the Results section.

ii) Applying information loss and Partitioned GPD 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 '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 purely 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 for living fossils is needed to underpin research and resolve some of the controversy. They go 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 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) are emphatic that their approach does not invoke a ladder of progress. However, for living fossils to be defined partly based on having few evolutionary changes since divergence still feels reminiscent of the problematic characteristics. We will apply the idea of information loss and GPD to conceptualise living fossils and thereby show how the phenomenon of living fossils can be explained without the need to invoke any heterogeneity in rates of evolution or morphological change.

We propose that the taxa (say, extant vertices of a dated information graph) called living fossils do not necessarily possess more ancestral features *in total* than other extant vertices, but rather the features they do have are more *rare*, or maybe *unique* to the living fossil among extant vertices. To see this, consider information accumulation and loss occurring together at equal pace across every edge in a dated information graph. Now consider an ancient vertex with two descendent edges, one leading to a single extant vertex (a candidate living fossil), the other leading to a diverse group of extant vertices. Because information loss is constant everywhere, all extant vertices will inherit an equal total quantity 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 GPD concepts can be used to quantitatively identify living fossils from first principles on a dated information 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) and we use the term 'living-fossil-ness' first coined by Bennett et al. (2018). However, our approach beyond this is very different to 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 whilst 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. Information 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 information graph. The concept of an ancestral information graph gives us the ability to select 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 information graph is constructed from a dated information graph by artificially augmenting information accumulation (α) to focus on information generated at the time when fossils of interest appeared in the fossil record. For example, by setting information accumulation to zero along many edges and only retaining it on edges leading to ancestral nodes of interest. We quantitatively define 'living-fossil-ness' of a vertex v on an ancestral information graph as $\log_{10}(\text{Partitioned GPD}(v, A))$ where A is the set of all extant vertices and $v \in A$. Information loss will typically be an exponential process so we take the log of Partitioned GPD to make 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 source mammal trees as a posterior distribution from (Upham et al., 2019) with edge-length units in millions of years. To construct ancestral information 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 that K consists of just the most recent common ancestor vertex for all mammals. For the Cretaceous-Paleogene boundary case, K contains a collection 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 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 final results equally whilst enabling the scores to fall in a convenient range. Information loss was calculated under standard conditions with a single information form and $\beta_i(e)$ being given by Eqn. 5. For each of our ancestral information graphs, we calculated Partitioned GPD for every extant vertex using Monte Carlo methods. To see how, consider Eqn. 11, the first sum disappears due to there being only one information 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 information 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

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

In this equation S is the set of information graphs (trees) resulting from repeated random removals of edges from edge set E, and T is an example information graph (tree) from set S and $E(T)$ is the set of all edges remaining in T. Our ancestral information graph has $\alpha_i(e) = 0$ for most edges e and because information is only accumulated at a moment in time, no path through the information 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 done with a single traversal of edges, to get Partitioned GPD 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 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. So, 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 and Globally Endangered 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 done on relatively small clades such as mammals or birds in isolation (EDGE of Existence Programme, 2022). The opportunity to upscale 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 IUCN Red List of threatened species (IUCN, 2022). Whilst 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 to a GPD-based approach.

Our aim is to illustrate the difference between GPD 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 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 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 evolutionary distinctiveness as well as the terminal branch length component of ED (Isaac et al., 2007). Unique GPD and Partitioned GPD, unlike ED, depend in part on the structure of the whole tree, not only on the number of each descendants shared by each vertex. Our calculations of both Unique GPD and Partitioned GPD are thus approximations that can be made with only total numbers of descendants from each vertex. We do these approximations in two different ways making opposite assumptions about the structure of the unseen parts of the tree. In the first scenario we assume 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 assume 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 GPD and Unique GPD arising from the terminal edge. We estimated total Unique GPD by calculating, from first principles, the proportion of information 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 GPD, $P(v,x) \subset V$ is the set of vertices along a path between vertex v and vertex x 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 do a different calculation for each ancestral edge and sum the results, the α term gives the total information arising on an edge and the product of various β and $1-\beta$ terms gives the proportion of this information that survives along the path to the focal vertex whilst also being cut off in all other descendants of all other interior vertices.

We estimated Partitioned GPD via two different methods depending on whether we are assuming 'star' or 'stick' trees of other descendants from interior vertices. In all cases we need to sum over ancestral edges, calculate the information 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 information that is copied different numbers of times. In the case of a 'stick' tree we can approximate number of copies with with

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

Where $G(y)$ gives the descendent 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 information. As an approximation we just get the mean number of such copies based on the proportion $(1 - \beta(G(y)))$ surviving information loss along the stem of the 'stick tree' and the number of copies that would result $R(g(y))$. 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 apply 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 is not available for most species in our analysis. For comparison, we calculate EDGE scores for some species based on the formula $\ln(1 + \mathrm{ED}) + \mathrm{GE} \cdot \ln(2)$ from (Isaac et al., 2007) where $\mathrm{GE} \in \{0, 1, 2, 3, 4\}$ is chosen depending 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

The strongest correlation of plot biomass against plot GPD for all species was achieved for the smallest ρ, where GPD converges to PD. The correlation coefficients of PD and species richness do coincide exactly with those of GPD under standard conditions at each of its limits as expected. Considering only the species richness of the *Pentapetalae*, ignoring the contribution of any *Poaceae* achieves a better correlation with biomass than any GPD calculated for all species together. PD for *Pentapetalae* correlated better with total biomass than the species richness of *Pentapetalae*. The strongest correlation of all 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 measured, but unlike the other cases performed better as species richness (*Poaceae*) than it did as PD(*Poaceae*). GPD for *Poaceae* was optimised when $\rho = 10$ where it performed better than either species richness or PD.

ii) Applying information loss and Partitioned GPD 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 fifteen living fossils from the Triassic-Jurassic boundary are all Monotremes or South American Marsupials. The rankings did not appear sensitive to the value of ρ though the precise Living-fossil-ness numbers were (fig 6). 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 now top followed by the three species of *Solenodon*. *Ornithorhynchus anatinus* comes next,

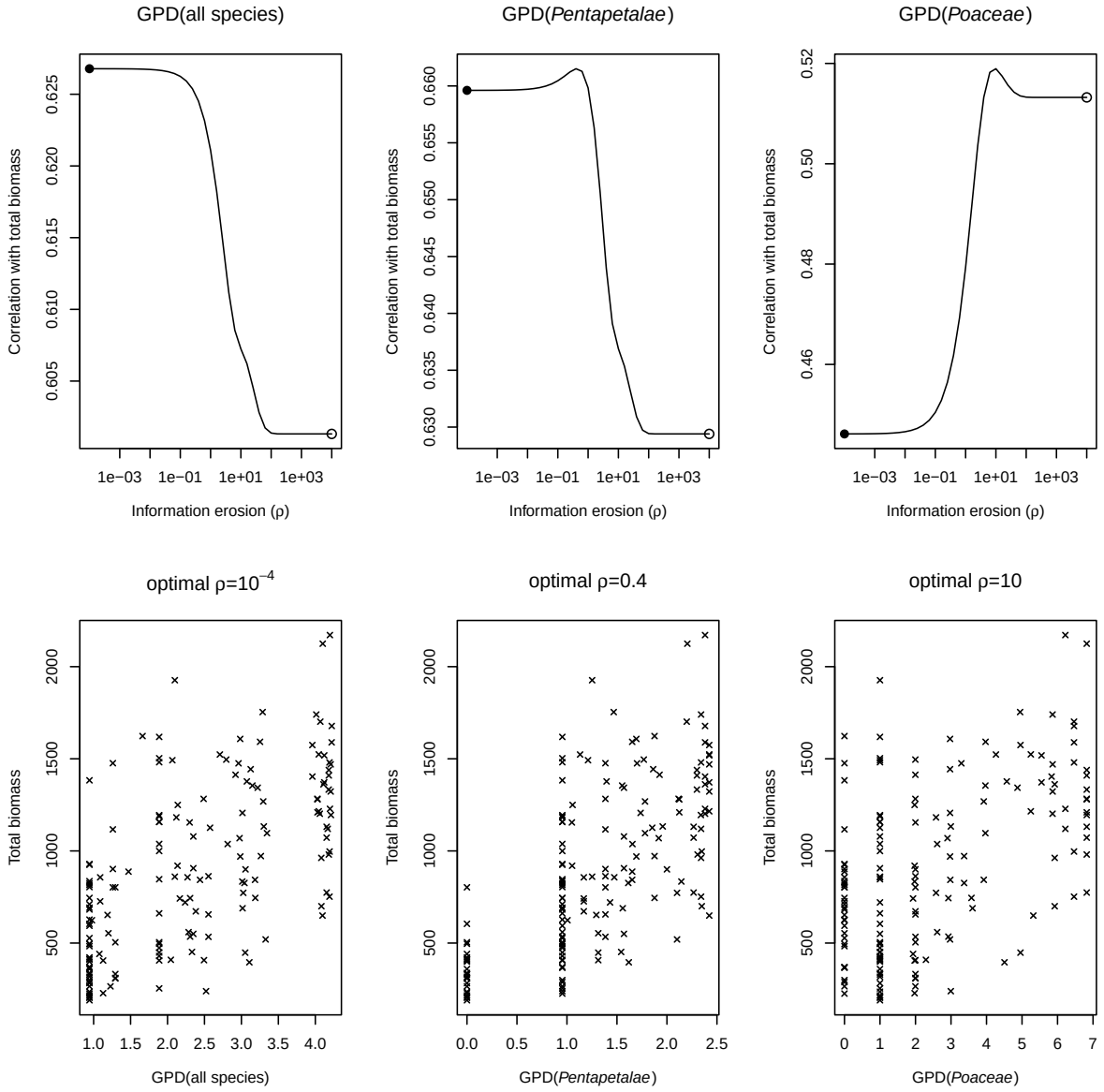

Fig. 5. Relationship between GPD under standard conditions and community productivity. The top row of panels show Pearson's correlation coefficient between GPD and total biomass in the Cedar Creek plots. The horizontal axis shows information loss ρ. The solid circle gives the same correlation against PD (corresponding to GPD under standard conditions with small ρ) and the open circle gives the correlation with species richness (corresponding to GPD under standard conditions with large ρ). The bottom row of panels show scatter plots of Biomass against GPD for the value of ρ that gave the strongest correlation. The three columns correspond to measurement of total GPD including all species (emphPentapetalae and *Poaceae* together) (left), measurement of GPD of *Pentapetalae* (centre), and GPD of *Poaceae* true grasses (right).

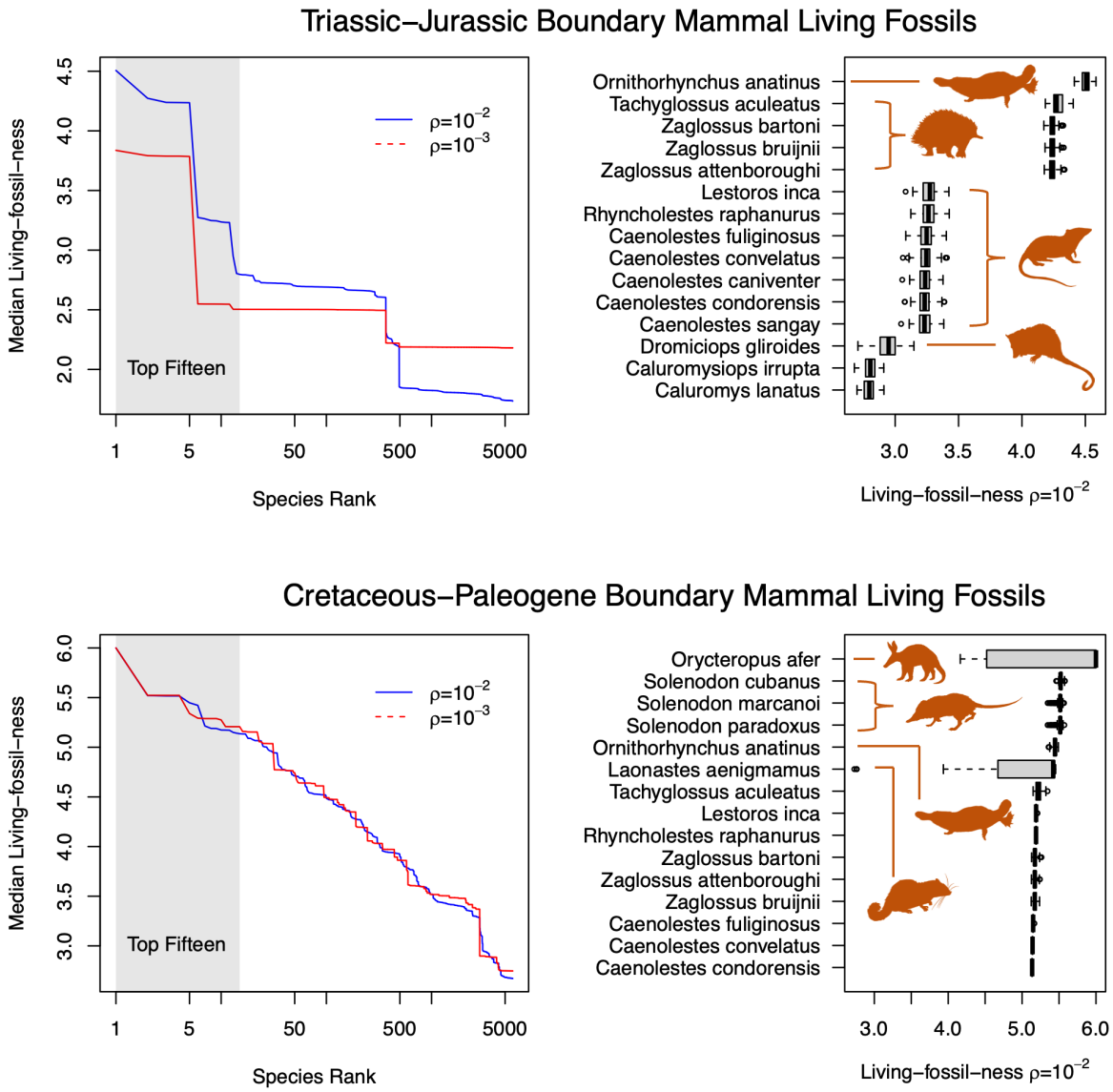

Fig. 6. Living-fossil-ness for mammal species calculated with reference to two different points in history t: the Triassic-Jurassic Boundary $t = 200$ Mya and the Cretaceous-Paleogene Boundary $t = 66$ Mya. The left panels show the living-fossil-ness with rank in log scale for two different values of information loss: $\rho = 10^{-2}$ (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 fifteen species only. 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 information loss using 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 public domain.

followed by another new entry *Laonastes aenigmamus*, a rodent. After this, the top species are all *Tachyglossidae* and *Paucituberculata* species. Uncertainty in 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 eleven 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 *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 and *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. 7). 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 and Unique GPD with $\rho = 0$ were, as expected, identical to ED and terminal edge length respectively (see supplementary material). The partitioned and Unique GPD with $\rho = 10^{-3}$ showed a somewhat similar patter to ED though the extremely high values for *Pygsuia biforma* and *Capsaspora owczarzaki* are somewhat dampened and almost all of their partitioned GPD score comes from unique GPD.

The Partitioned GPD with $\rho = 10^{-2}$ was substantively different from ED. The Partitioned GPD 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, a factor of about 3 difference, ED scores for the same species differed by a factor of about 23. *Xenoturbella churro* and *Latimeria chalumnae* all fell disproportionately far and so fell in the ranking of Partitioned GPD with $\rho = 10^{-2}$ compared to their ranking under ED - both had a comparatively small terminal branch length (Fig. 7). In the limit as $\rho \to \infty$, both Unique GPD and Partitioned GPD converge to unity for all species (see supplementary information).

Discussion

i) Evaluating ϕ_ρ as a predictor of community productivity

One motivation for formulating GPD was to provide a process-based measure of diversity that scales continuously between species richness and PD. Our results show that GPD does give the same result as PD and species richness at its two extremes (Fig. 5 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; Owen et al., 2019; Mazel et al., 2019) to a question about the extent of information loss. 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 cut off (Chao et al., 2010). GPD is different because its approach is mechanistic and based on the biological process of information loss. Information loss exists as a biological reality, occurring at a rate that is neither zero (PD) nor infinity (species richness); it should be somewhere in between.

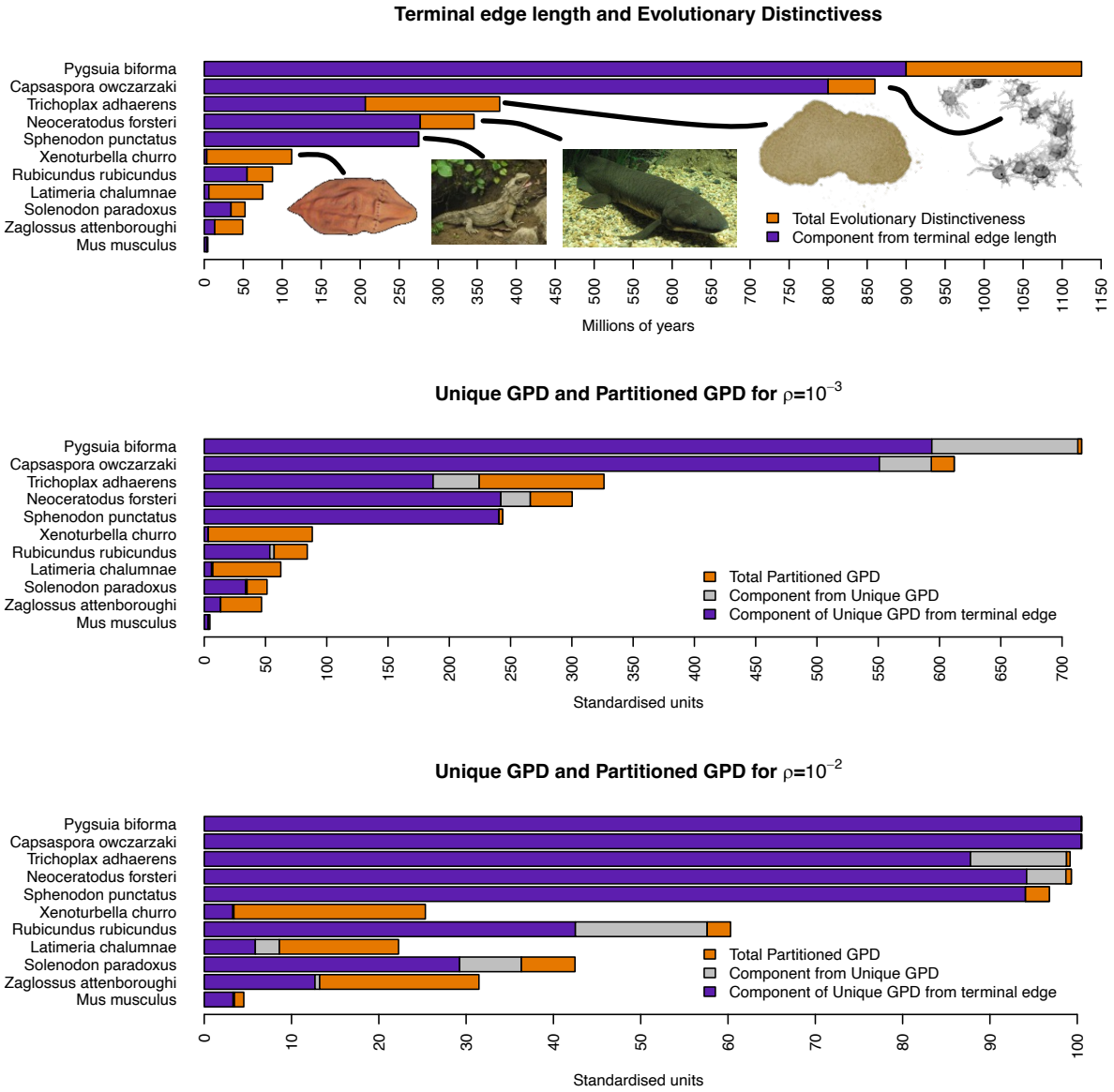

Fig. 7. Comparison of ED, terminal edge length, Unique GPD and Partitioned GPD for selected species ordered by ED. The stacked bars of the top panel show ED separated into two components: the terminal edge legth, and the remainder which is a contribution from interior edges. The middle and bottom panel show an equivalent for GPD with $\rho = 10^{-3}$ and $\rho = 10^{-2}$ respectively. The stacked bars of the GPD based graphs show Partitioned GPD separated out into three components: A contribution from Unique GPD arising on the terminal edge, a contribution from Unique GPD arising on interior edges but where the information 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 there was only 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, *Sphenodon punctatus* by Sid Mosdell from New Zealand

We found that GPD 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 GPD but 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 information loss, which can only reduce the influence of such branches, is optimal when set to zero. The phylogeny of 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 those two groups. This is confirmed with our finding that species richness of *Pentapetalae* alone performed better as a predictor of biomass than any measure of total diversity. These results are likely due to the presence 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 GPD of *Pentapetalae* with an optimal ρ of 0.4 (see Fig. 5). This outperformed both species richness and PD and thus demonstrates the potential of GPD as a general diversity measure. Whilst we find that the diversity of *Poaceae* is a poor predictor of community productivity, we highlight it as one example where species richness outperforms PD whilst GPD still outperforms both species richness and PD (Fig. 5).

The differences in 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 reasons why the best measure of overall diversity is not the same as the best predictor of community productivity; it can thus be misleading to compare PD to 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 GPD is likewise based more on the conceptual reasons supporting its use, rather than its encouraging performance in this particular case.

ii) Applying information loss and Partitioned GPD to quantify living fossils

To the best of our knowledge, our method for using Partitioned GPD on ancestral information graphs is the only attempt focused on quantifying living fossils, other than the work of Bennett et al. (2018). Whilst Bennett et al. (2018) provided inspiration for our quantification procedure, our approach is different in several important respects. First, we only calculate 'living-fossil-ness' for individual vertices of a data information graph whilst Bennett et al. (2018) primarily designate clades of species as living fossils. Second, we calculate 'living-fossil-ness' differently depending on which period in geological history the fossils of interest come from. Third, our calculations are process-based, incorporating the concept of information loss. 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). Whilst partitioned GPD converges to ED as a special case, our use of partitioned GPD is quite different because we only consider information created at a given moment in time. The lack of information loss 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 identify have a history of being referred to as living fossils in 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 find mention of South American

Marsupisals as living fossils in grey literature. Our measure of living-fossil-ness therefore demonstrates that the taxa historically identified as living fossils can be recovered from our GPD-based concept of living fossils that requires no heterogeneity in rates of evolution (Fig. 6). 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. So, 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. 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.

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 to 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 *Capsaspora owczarzaki* and *Pygsuia biforma* are not technically 'EDGE species' until they are considered under threat with 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 rearranges 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, and one 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 GPD because the deeper parts of the tree are subject to more information loss 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 information loss is a real force that cannot be ignored in phylogenetic analyses at such scales.

We saw that the greater the value of information loss, the greater the departure of Partitioned GPD from ED. For $\rho = 10^{-2}$, a relatively modest rate of information loss 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, remains at the low end of the distribution by a very substantial margin. Secondly, species with very short terminal edge lengths are disproportionately penalised by

Partitioned GPD. The information accumulated near the tips of the tree has greater value as there has not yet been so much opportunity for it to be lost in the first place, so the cost of partitioning this is high in terms of the final Partitioned GPD 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. Overall, Partitioned GPD provides a less extreme range of final values for downstream conservation prioritisation, continues to clearly identify the least ED species, and increases the penalty for species with short terminal edges based on a sound rationale.

Our example was based on ED values and corresponding EDGE scores (Isaac et al., 2007). More recent work extending the concept of EDGE has used mappings of IUCN Red List status onto 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 deeper edges of the tree have reduced influence because there is a small probability of all their descendants going 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 GPD. We need a coherent system for attributing value (GPD) 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 to 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 to 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 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' extinction of the Playtpus *Ornithorhynchus anatinus* with a terminal branch of 63 million years. Whilst 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 a large scale quantitative analyses of present PD gain or loss compared to background rates of PD gain or loss. When we recast in terms of GPD, however, we see that any generation of new diversity at the tips of the tree is already balanced against the natural process of information loss, just as speciation is balanced against a natural extinction (which we note runs at a very low rate compared to extinction from anthropogenic causes). Future work may thus be able to meaningfully compare the background rate of GPD loss to the current anthropogentic rate of GPD loss.

The future of information graphs and information loss

The foundations for GPD are based on a new concept of an information graph: a Directed Acyclic Graph where both processes of information accumulation and loss are defined along edges. This creates an intentional separation between the question of useful metrics on a given information graph, and the question of how to acquire an information graph in the first place. We hope that that future work will build information graphs

directly from genetic data by adapting tree building methods. 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 information graphs *de novo* is to incorporate information loss. We have shown with Theorem 2 that information accumulation and loss interact in a way that cannot be expressed with just a single dimension of data, such as an edge length. At least two dimensions are needed, one for information accumulation and the other for information loss. It may be that we use more then one information form, e.g. so that the information within an endosymbiont has its own accumulation and loss that may be distinct from that of the host. We speculate that information loss 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 because there is so much information loss, and information accumulation 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 edge lengths cannot possibly store the joint effects both information accumulation and information loss.

To build an information 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. The potential reward might be an object with less compromise and uncertainty, that enables more of the source data to inform downstream analyses. Beyond this, future information graphs need not only come 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 information is inherited from each parent. This illustrates how the information graph is more flexible, even 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 GPD calculus, we developed a model of exponential information decay that can be applied to a trees with branch lengths, given ρ, a rate of information loss. 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 is 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, for example 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 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 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 information loss 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 5), but these are

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 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 1 unit of edge 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.

Concluding remarks

In this manuscript, we began by pointing out a range of conceptual challenges with PD. First, features or future options for humanity cannot reasonably be directly proportionate to edge length for extremely long edges as PD would assert (e.g. *Pygsuia biforma* is about than 23 times that of the top EDGE mammal *Zaglossus attenboroughi*). Second, following PD, a single extant descendant of a deep edge is enough to guarantee 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. Third, 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. We have suggested that all these challenges are symptoms of one underlying cause: assuming features accumulate and are never lost. A key realisation is that the same concerns likely apply to any other metric that is underpinned by the sums of branch lengths in some way.

We introduced 'generalised phylogenetic diversity' (GPD), and an associated GPD calculus, which incorporates the process of 'information loss'. We have introduced Unique GPD, generalised phylogentic difference, Partitioned GPD and Hill GPD to illustrate GPD 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 information graph with extant vertex range maps, we would like to see a generalised phylogenetic endemism following (Rosauer et al., 2009) that spreads the value of information equally across the area in space where that information exists. Throughout the framing of GPD, we do not assume the underlying structure is a tree, GPD is therefore capabale 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 adoption of GPD easier as a straight swap for PD on any tree with edge lengths, given a value of ρ. Our three applications of GPD all rely on its information loss component and they cover ecology, evolution and conservation. GPD can outperform PD and species richness as a predictor of community productivity; metrics on an ancestral information graph can detect living fossils and resolve their controversy; and conservation prioritisation at the scale of the whole tree of life is distorted without information loss to re-balance 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 (Winter et al., 2013). 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, to do so requires them to have metrics with certain properties. It is easy to create metrics with the required properties from scratch and so they do so. The 'cost' of accidentally duplicating or reinventing something that already exists is very small for a metric compared to other research endeavours. The main message of this paper is thus not our individual metrics, but rather the conceptual importance of information loss and the problems associated with just summing branch lengths. Researchers should either account for information loss in the metrics they create, or stick to questions where the symptoms of

information loss are unlikely to manifest - for example 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 the number of years things have been evolving for, but it is hard to ignore information loss if we are making value judgements on the basis of something else that accumulates over those years, such as biological features, or future options for humanity. If we may summarise the original message of PD as 'species are not equal', our message is 'million-year-long sections of branch are not equal either'.

Acknowledgements

JR thanks Yan Wong, Ryan Chisholm, and Rosindell lab group members for useful discussions and suggestions. 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 supporting the ongoing development of this work.

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 1148 pages.

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.

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 .

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.

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

Forest, F., R. Grenyer, M. Rouget, and et. al. 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. 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 .

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. Anat. Rec. (Hoboken) 304:2080–2084.

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.

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.

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.

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.

Tilman, D. 2021. Plant aboveground biomass data: Biodiversity ii: Effects of plant biodiversity on population and ecosystem processes ver 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.

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.

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

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.